

Fundamental Theories of Physics 162

Luc Blanchet
Alessandro Spallicci
Bernard Whiting *Editors*

Mass and Motion in General Relativity



Springer

Fundamental Theories of Physics

Volume 162

Series Editors

Philippe Blanchard, Universität Bielefeld, Bielefeld, Germany

Paul Busch, University of York, Heslington, York, United Kingdom

Bob Coecke, Oxford University Computing Laboratory, Oxford, United Kingdom

Detlef Duerr, Mathematisches Institut, München, Germany

Roman Frigg, London School of Economics and Political Science, London, United Kingdom

Christopher A. Fuchs, Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada

Giancarlo Ghirardi, Università di Trieste, Trieste, Italy

Domenico Giulini, Leibnitz Universität Hannover, Hannover, Germany

Gregg Jaeger, Boston University CGS, Boston, USA

Claus Kiefer, Universität Köln, Köln, Germany

Klaas Landsman, Radboud Universiteit Nijmegen, Nijmegen, The Netherlands

Christian Maes, K.U. Leuven, Leuven, Belgium

Hermann Nicolai, Max-Planck-Institut für Gravitationsphysik, Golm, Germany

Vesselin Petkov, Concordia University, Montreal, Canada

Alwyn van der Merwe, University of Denver, Denver, USA

Rainer Verch, Universität Leipzig, Leipzig, Germany

Reinhard Werner, Leibnitz Universität Hannover, Hannover, Germany

Christian Wüthrich, University of California, San Diego, La Jolla, USA

For further volumes:

<http://www.springer.com/series/6001>

Luc Blanchet · Alessandro Spallicci
Bernard Whiting
Editors

Mass and Motion in General Relativity

 Springer

Editors

Luc Blanchet
Institut d'Astrophysique de Paris
98 bis Boulevard Arago
75014 Paris
France
blanchet@iap.fr

Alessandro Spallicci
Université d'Orléans
Observatoire des Sciences
de l'Univers en région Centre
LPC2E-CNRS, 3A Avenue
de la Recherche Scientifique
45071 Orléans
France
spallicci@cnsr-orleans.fr

Bernard Whiting
University of Florida
Department of Physics
P.O. Box 118440, Gainesville
FL 32611-8440, U.S.A.
bernard@phys.ufl.edu

ISBN 978-90-481-3014-6 e-ISBN 978-90-481-3015-3
DOI 10.1007/978-90-481-3015-3
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2010938712

© Springer Science+Business Media B.V. 2011

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Cover design: Spi Publisher Services

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Mass and Motion in General Relativity*

From the infinitesimal scale of particle physics to the cosmic scale of the universe, research is concerned with the nature of mass. While there have been spectacular advances in physics during the past century, mass still remains a mysterious entity at the forefront of current research. Particle accelerators in the quest for the Higgs boson responsible for the mass of particles, laser interferometers that are sensitive enough to respond to gravitational waves generated by the motion of astrophysical bodies, equivalence principle tests of the relationship between gravitational and inertial mass are among the most ambitious and expensive experiments that fundamental physics has ever envisaged.

Our current perspective on gravitation has arisen over millennia, through falling apples, lift thought experiments and stars spiraling into black holes. In this volume, the world's leading scientists offer a multifaceted approach to mass by giving a concise and introductory presentation into their particular research on gravity. The main theme is mass and its motion within general relativity and other theories of gravity, particularly for compact bodies. Within this framework, all articles are tied together coherently, covering post-Newtonian and related methods applied to in-spiraling compact binaries, as well as the self-force approach to the analysis of motion.

All contributions reflect the fundamental role of mass in physics, from issues related to Newton's laws, via the effect of self-force and radiation reaction within theories of gravitation, to the role of the Higgs boson in modern physics. Precision measurements are described in detail; modified theories of gravity reproducing experimental data are investigated as alternatives to dark matter and the fundamental problem of reconciling the theory of gravity with the physics of quantum fields is addressed.

Radiation and motion have been hotly debated within general relativity from the inception of the theory well beyond the theoretician's arena. Mass and motion are intimately intertwined as self-acceleration depends directly on the mass of the body

*Lectures from the School on Mass held at Orléans on 23–25 June 2008

Organised by the Observatoire des Sciences de l'Univers en région Centre OSUC, Université d'Orléans UO, Centre National de la Recherche Scientifique CNRS

experiencing it. Recent developments have shown that the computation of radiation reaction is unavoidable for determining the gravitational waveforms emitted not only by large bodies in binary formation but also from sources such as the capture of stellar size objects by super-massive black holes.

The main theme of this volume is indeed mass and its motion within general relativity (and other theories of gravity), particularly for compact bodies, to which many articles directly refer.

Within this framework, after a presentation of the mass and momentum in general relativity (Jaramillo and Gourgoulhon), there are chapters on post-Newtonian (Blanchet, Schäfer), effective one-body (Damour and Nagar) methods as well as on the self-force approach to the analysis of motion (Wald with Gralla, Detweiler, Poisson, Barack, Gal'tsov). post-Newtonian and self-force methods converge in their common domain of applicability (Blanchet, Detweiler, Le Tiec and Whiting). A snapshot on the state of the art of the self-force (Burko) and the historic development of the field including future perspectives for the classic free fall problem (Spallicci) conclude this central part.

Auxiliary chapters set the context for these theoretical contributions within a wider context. The space mission LISA (Jennrich) has been designed to detect the gravitational waves from EMRI captures. Motion in modern gravitation demands an account of the relation between vacuum fluctuations and inertia (Jaekel and Reynaud). A volume centred on the fundamental role of mass in physics should face issues related to the basic laws of mechanics proposed by Newton (Lämmerzahl) and precision measurements (Davis).

The role of the Higgs boson within physics is to give a mass to elementary particles (Djouadi), by interacting with all particles required to have a mass and thereby experiencing inertia.

Motion of stars and of galaxies are explicable according to most researchers by only evoking yet undetected matter and energy constituting around 95% of our universe. A proposed alternative to dark matter theories is due to the modified theories of gravity (Esposito-Farèse) such as MOND (MODified Newtonian Dynamics). Even if general relativity does not explain gravity, there still remains the fundamental problem of reconciling any theory of gravity with the physics of quantum fields (Noui), itself so well verified experimentally.

The book is based upon the lectures of the School on Mass held in Orléans, France, in June 2008. The school was funded by CNRS *Centre National de la Recherche Scientifique*, INSU *Institut National des Sciences de l'Univers*, UO *Université d'Orléans*, Région Centre, *Conseil Régional du Loiret*, *Observatoire de Paris* and was organised by OSUC *Observatoire des Sciences de l'Univers en région Centre* and its associated laboratory LPC2E *Laboratoire de Physique et Chimie de l'Environnement et de l'Espace*.

The editors wish to thank the OSUC director (Elisabeth Vergès) for continuous support and organisation of the school; the OSUC staff (S. Bouquet, T. Cantalupo, L. Catherine, N. Rolland) who dealt with all issues related to the practical organisation and running of two international events (the School followed up by the 11th Capra meeting on radiation reaction); the LPC2E director (M. Tagger)

for suggestions and hosting the Capra workshop; the local CNRS delegation (P. Letourneux) for assistance and support; M. Volkov (Univ. Tours) for suggestions and all members of the scientific and organisation committees, especially S. Cordier (MAPMO - Univ. Orléans).

Both events are shown on the OSUC web pages: <http://www.cnrs-orleans.fr/osuc/conf/>

The contributions to this book have been anonymously refereed and revised by the editors.

Luc Blanchet
Alessandro Spallicci
Bernard Whiting
Editors



Contents

Mass and Motion in General Relativity	v
The Higgs Mechanism and the Origin of Mass	1
Abdelhak Djouadi	
1 The Standard Model and the Generation of Particle Masses	1
1.1 The Elementary Particles and Their Interactions	2
1.2 The Standard Model of Particle Physics	3
1.3 The Higgs Mechanism for Mass Generation	5
2 The Profile of the Higgs Particle	8
2.1 Characteristics of the Higgs Boson	8
2.2 Constraints on the Higgs Boson Mass	9
2.3 The Higgs Decay Modes and Their Rates	11
3 Higgs Production at the LHC	13
3.1 The Large Hadron Collider	13
3.2 The Production of the Higgs Boson	14
3.3 Detection of the Higgs Boson	15
3.4 Determination of the Higgs Boson Properties	18
4 The Higgs Beyond the Standard Model	20
5 Conclusions	22
References	23
Testing Basic Laws of Gravitation – Are Our Postulates on Dynamics and Gravitation Supported by Experimental Evidence?	25
Claus Lämmerzahl	
1 Introduction – Why Gravity Is So Exceptional	25
2 Key Features of Gravity	27
3 Standard Tests of the Foundations of Special and General Relativity	28
3.1 Tests of Special Relativity	28
3.2 Tests of the Universality of Free Fall	30
3.3 Tests of the Universality of the Gravitational Redshift	31
3.4 The Consequence	32

4	Tests of Predictions	34
4.1	The Gravitational Redshift.....	35
4.2	Light Deflection	36
4.3	Perihelion/Periastron Shift	37
4.4	Gravitational Time Delay	38
4.5	Lense–Thirring Effect.....	40
4.6	Schiff Effect	41
4.7	The Strong Equivalence Principle	41
5	Why New Tests?	42
5.1	Dark Clouds – Problems with GR.....	42
5.2	The Search for Quantum Gravity	45
5.3	Possible New Effects	45
6	How to Search for “New Physics”	46
6.1	Better Accuracy and Sensitivity	46
6.2	Extreme Situations	47
6.3	Investigation of “Exotic” Issues	49
7	Testing “Exotic” but Fundamental Issues.....	50
7.1	Active and Passive Mass	50
7.2	Active and Passive Charge	51
7.3	Active and Passive Magnetic Moment	52
7.4	Charge Conservation	52
7.5	Small Accelerations	53
7.6	Test of the Inertial Law	54
7.7	Can Gravity Be Transformed Away?	57
8	Summary	60
	References	60
	Mass Metrology and the International System of Units (SI)	67
	Richard S. Davis	
1	Introduction	67
2	The SI.....	68
2.1	Base Units/Base Quantities	68
2.2	Gaussian Units.....	70
2.3	Planck Units, Natural Units, and Atomic Units.....	71
3	Practical Reasons for Redefining the Kilogram	71
3.1	Internal Evidence Among 1 kg Artifact Mass Standards.....	71
3.2	Fundamental Constants	73
3.3	Electrical Metrology	75
3.4	Relative Atomic Masses	77
4	Routes to a New Kilogram	78
5	Realizing a New Kilogram Definition in Practice	79
5.1	Watt Balances	80
5.2	Silicon X-Ray Crystal Density (XRCD)	81
5.3	Experimental Results.....	82

6 Proposals for a New SI 83
 6.1 Consensus Building and Formal Approval 83
 6.2 An SI Based on Defined Values of a Set of Constants 84
 7 Conclusion 84
 References 85

Mass and Angular Momentum in General Relativity 87

José Luis Jaramillo and Ericourgoulhon

1 Issues Around the Notion of Gravitational Energy in General Relativity ... 88
 1.1 Energy–Momentum Density for Matter Fields 88
 1.2 Problems when Defining a Gravitational Energy–Momentum 90
 1.3 Notation 92
 2 Spacetimes with Killing Vectors: Komar Quantities 94
 2.1 Komar Mass 94
 2.2 Komar Angular Momentum 95
 3 Total Mass of Isolated Systems in General Relativity 95
 3.1 Asymptotic Flatness Characterization of Isolated Systems 95
 3.2 Asymptotic Euclidean Slices 96
 3.3 ADM Quantities 97
 3.4 Bondi Energy and Linear Momentum 105
 4 Notions of Mass for Bounded Regions: Quasi-Local Masses 108
 4.1 Ingredients in the Quasi-Local Constructions 108
 4.2 Some Relevant Quasi-Local Masses 109
 4.3 Some Remarks on Quasi-Local Angular Momentum 114
 4.4 A Study Case: Quasi-Local Mass of Black Hole IHs 115
 5 Global and Quasi-Local Quantities in Black Hole Physics 118
 5.1 Penrose Inequality: a Claim for an Improved Mass
 Positivity Result for Black Holes 119
 5.2 Black Hole (Thermo-)dynamics 119
 5.3 Black Hole Extremality: a Mass–Angular Momentum Inequality ... 121
 6 Conclusions 121
 References 123

Post-Newtonian Theory and the Two-Body Problem 125

Luc Blanchet

1 Introduction 125
 2 Post-Newtonian Formalism 128
 2.1 Einstein Field Equations 128
 2.2 Post-Newtonian Iteration in the Near Zone 131
 2.3 Post-Newtonian Expansion Calculated by Matching 135
 2.4 Multipole Moments of a Post-Newtonian Source 139
 2.5 Radiation Field and Polarization Waveforms 143
 2.6 Radiative Moments Versus Source Moments 145

3 Inspiralling Compact Binaries147

3.1 Stress–Energy Tensor of Spinning Particles147

3.2 Hadamard Regularization150

3.3 Dimensional Regularization153

3.4 Energy and Flux of Compact Binaries156

3.5 Waveform of Compact Binaries160

3.6 Spin–Orbit Contributions in the Energy and Flux162

References164

Post-Newtonian Methods: Analytic Results on the Binary

Problem167

Gerhard Schäfer

1 Introduction167

2 Systems in Newtonian Gravity in Canonical Form.....169

3 Canonical General Relativity and PN Expansions171

3.1 Canonical Variables of the Gravitational Field173

3.2 Brill–Lindquist Initial-Value Solution for Binary Black Holes175

3.3 Skeleton Hamiltonian176

3.4 Functional Representation of Compact Objects179

3.5 PN Expansion of the Routh Functional185

3.6 Near-Zone Energy Loss Versus Far-Zone Energy Flux185

4 Binary Point Masses to Higher PN Order187

4.1 Conservative Hamiltonians187

4.2 Dynamical Invariants.....188

4.3 ISCO and the PN Framework.....190

4.4 PN Dissipative Binary Dynamics.....192

5 Toward Binary Spinning Black Holes192

5.1 Approximate Hamiltonians for Spinning Binaries.....196

6 Lorentz-Covariant Approach and PN Expansions.....200

6.1 PM and PN Expansions202

6.2 PN Expansion in the Near Zone203

6.3 PN Expansion in the Far Zone205

7 Energy Loss and Gravitational Wave Emission206

7.1 Orbital Decay to 4 PN Order206

7.2 Gravitational Waveform to 1.5 PN Order207

References209

The Effective One-Body Description of the Two-Body Problem.....211

Thibault Damour and Alessandro Nagar

1 Introduction211

2 Motion and Radiation of Binary Black Holes: PN Expanded Results.....213

3 Conservative Dynamics of Binary Black Holes: the EOB Approach215

4 Description of Radiation–Reaction Effects in the EOB Approach224

4.1 Resummation of \hat{F}^{Taylor} Using a One-Parameter Family of Padé Approximants: Tuning v_{pole} 227

4.2 Parameter-Free Resummation of Waveform and Energy Flux230

5 EOB Dynamics and Waveforms238
 5.1 Post–Post–Circular Initial Data238
 5.2 EOB Waveforms239
 5.3 EOB Dynamics241
 6 EOB and NR Waveforms243
 7 Conclusions248
 References249

Introduction to Gravitational Self-Force253

Robert M. Wald

1 Motion of Bodies in General Relativity253
 2 Point Particles in General Relativity254
 3 Point Particles in Linearized Gravity255
 4 Lorenz Gauge Relaxation256
 5 Hadamard Expansions256
 5.1 Hadamard Expansions for a Point Particle Source258
 6 Equations of Motion Including Self-Force259
 6.1 The MiSaTaQuWa Equations259
 6.2 The Detweiler–Whiting Reformulation260
 7 How Should Gravitational Self-Force Be Derived?261
 References262

Derivation of Gravitational Self-Force263

Samuel E. Gralla and Robert M. Wald

1 Difficulties with Usual Derivations263
 2 Rigorous Derivation Requirements264
 3 Limits of Spacetimes264
 4 Our Basic Assumptions265
 4.1 Additional Uniformity Requirement265
 5 Geodesic Motion266
 6 Corrections to Motion267
 6.1 Calculation of the Perturbed Motion268
 7 Interpretation of Results269
 8 Self-Consistent Equations269
 9 Summary270
 References270

Elementary Development of the Gravitational Self-Force271

Steven Detweiler

1 Introduction271
 1.1 Outline273
 1.2 Notation274
 2 Newtonian Examples of Self-Force and Gauge Issues275
 3 Classical Electromagnetic Self-Force277

4	A Toy Problem with Two Length Scales That Creates a Challenge for Numerical Analysis	278
4.1	An Approach Which Avoids the Small Length Scale	279
4.2	An Alternative That Resolves Boundary Condition Issues	281
5	Perturbation Theory	282
5.1	Standard Perturbation Theory in General Relativity	283
5.2	An Application of Perturbation Theory: Locally Inertial Coordinates	285
5.3	Metric Perturbations in the Neighborhood of a Point Mass	287
5.4	A Small Object Moving Through Spacetime	289
6	Self-Force from Gravitational Perturbation Theory	291
6.1	Dissipative and Conservative Parts	292
6.2	Gravitational Self-Force Implementations	293
7	Perturbative Gauge Transformations	295
8	Gauge Confusion and the Gravitational Self-Force	297
9	Steps in the Analysis of the Gravitational Self-Force	298
10	Applications	300
10.1	Gravitational Self-Force Effects on Circular Orbits of the Schwarzschild Geometry	300
10.2	Field Regularization Via the Effective Source	301
11	Concluding Remarks	304
	References	306
	Constructing the Self-Force	309
	Eric Poisson	
1	Introduction	309
2	Geometric Elements	311
3	Coordinate Systems	312
4	Field Equation and Particle Motion	316
5	Retarded Green's Function	316
6	Alternate Green's Function	318
7	Fields Near the World Line	319
8	Self-Force	321
9	Axiomatic Approach	322
10	Conclusion	324
	References	325
	Computational Methods for the Self-Force in Black Hole Spacetimes	327
	Leor Barack	
1	Introduction and Overview	327
1.1	The MiSaTaQuWa Formula	329
1.2	Gauge Dependence	330
1.3	Implementation Strategies	331

- 2 Mode-Sum Method335
 - 2.1 An Elementary Example336
 - 2.2 The Mode-Sum Formula338
 - 2.3 Derivation of the Regularization Parameters339
- 3 Numerical Implementation Strategies343
 - 3.1 Overcoming the Gauge Problem344
 - 3.2 Numerical Representation of the Point Particle347
- 4 An Example: Gravitational Self-Force in Schwarzschild Via
1+1D Evolution in Lorenz Gauge352
 - 4.1 Lorenz-Gauge Formulation352
 - 4.2 Numerical Implementation354
- 5 Toward Self-Force Calculations in Kerr: the Puncture Method
and m -Mode Regularization356
 - 5.1 Puncture Method in 2+1D356
 - 5.2 m -Mode Regularization358
- 6 Reflections and Prospects360
- References364

Radiation Reaction and Energy–Momentum Conservation367

Dmitri Gal'tsov

- 1 Introduction367
- 2 Energy–Momentum Balance Equation369
 - 2.1 Decomposition of the Stress Tensor371
 - 2.2 Bound Momentum374
 - 2.3 The Rest Frame (Nonrelativistic Limit)377
- 3 Flat Dimensions Other than Four378
- 4 Local Method for Curved Space-Time379
 - 4.1 Hadamard Expansion in Any Dimensions380
 - 4.2 Divergences381
 - 4.3 Four Dimensions384
 - 4.4 Self and Radiative Forces in Curved Space-Time386
- 5 Gravitational Radiation387
 - 5.1 Bianchi Identity387
 - 5.2 Vacuum Background389
 - 5.3 Gravitational Radiation for Non-Geodesic Motion390
- 6 Conclusions391
- References392

The State of Current Self-Force Research395

Lior M. Burko

- 1 Introduction395
- 2 The Teukolsky Equation397
 - 2.1 The Inhomogeneous Teukolsky Equation
with a Distributional Source397
 - 2.2 Adiabatic Waveforms398

2.3	Numerical Solution of the Teukolsky Equation	399
2.4	The Linearized Einstein Equations	400
3	Frequency-Domain Calculations of the Self-Force	401
3.1	Mode-Sum Regularization	401
3.2	The Detweiler–Whiting Regular Part of the Self-Force	402
4	Time-Domain Calculations of the Self-Force	403
4.1	1+1D Numerical Simulations	403
4.2	2+1D Numerical Simulations	404
5	Post-adiabatic Self-Force-Driven Orbital Evolution	408
5.1	The Importance of Second-Order Self-Forces	408
5.2	Conservative Self-Force Effects	412
	References	413

High-Accuracy Comparison Between the Post-Newtonian and Self-Force Dynamics of Black-Hole Binaries

	Luc Blanchet, Steven Detweiler, Alexandre Le Tiec, and Bernard F. Whiting	415
1	Introduction and Motivation	416
2	The Gauge-Invariant Redshift Observable	418
3	Regularization Issues in the SF and PN Formalisms	419
4	Circular Orbits in the Perturbed Schwarzschild Geometry	421
5	Overview of the 3PN Calculation	423
5.1	Iterative PN Computation of the Metric	423
5.2	The Example of the Zeroth-Order Iteration	426
6	Logarithmic Terms at 4PN and 5PN Orders	427
6.1	Physical Origin of Logarithmic Terms	427
6.2	Expression of the Near-Zone Metric	429
7	Post-Newtonian Results for the Redshift Observable	430
8	Numerical Evaluation of Post-Newtonian Coefficients	433
8.1	Overview	434
8.2	Framework for Evaluating PN Coefficients Numerically	435
8.3	Consistency Between Analytically and Numerically Determined PN Coefficients	437
8.4	Determining Higher Order PN Terms Numerically	438
8.5	Summary	439
	References	441

LISA and Capture Sources

	Oliver Jennrich	443
1	LISA – A Mission to Detect and Observe Gravitational Waves	443
1.1	Mission Concept	444
1.2	Sensitivity	445
1.3	Measurement Principle	446
2	Capture Sources	448
3	Science Return	449

4 Detection451
 4.1 Capture Rates451
 4.2 Signal Characteristics452
 4.3 Data Analysis454
 5 Summary and Conclusions456
 References457

Motion in Alternative Theories of Gravity461

Gilles Esposito-Farèse

1 Introduction461
 2 Modifying the Matter Action462
 3 Modified Motion in Metric Theories?464
 4 Scalar-Tensor Theories of Gravity468
 4.1 Weak-Field Predictions469
 4.2 Strong-Field Predictions471
 4.3 Binary-Pulsar Tests472
 4.4 Black Holes in Scalar-Tensor Gravity476
 5 Extended Bodies477
 6 Modified Newtonian Dynamics479
 6.1 Mass-Dependent Models?480
 6.2 Aquadratic Lagrangians or k-Essence481
 6.3 Difficulties482
 6.4 Nonminimal Couplings485
 7 Conclusions486
 References487

Mass, Inertia, and Gravitation491

Marc-Thierry Jaekel and Serge Reynaud

1 Introduction491
 2 Vacuum Fluctuations and Inertia494
 2.1 Linear Response Formalism494
 2.2 Response to Motions498
 2.3 Relativity of Motion502
 2.4 Inertia of Vacuum Fields504
 3 Mass as a Quantum Observable508
 3.1 Quantum Fluctuations of Mass508
 3.2 Mass and Conformal Symmetries510
 4 Metric Extensions of GR514
 4.1 Radiative Corrections515
 4.2 Anomalous Curvatures518
 4.3 Phenomenology in the Solar System520
 5 Conclusion526
 References527

Motion in Quantum Gravity531
Karim Noui

- 1 Introduction531
 - 1.1 The Problem of Defining Motion in Quantum Gravity531
 - 1.2 Quantum Gravity532
 - 1.3 Three-Dimensional Quantum Gravity Is a Fruitful Toy Model534
 - 1.4 Outline of the Article.....535
- 2 Casting an Eye Over Loop Quantum Gravity536
 - 2.1 The Classical Theory: Main Ingredients536
 - 2.2 The Route to the Quantization of Gravity538
 - 2.3 Spin-Networks Are States of Quantum Geometry539
 - 2.4 The Problem of the Hamiltonian Constraint541
- 3 Three-Dimensional Euclidean Quantum Gravity542
 - 3.1 Construction of the Noncommutative Space543
 - 3.2 Constructing the Quantum Dynamics550
 - 3.3 Particles Evolving in the Fuzzy Space552
 - 3.4 Reduction to One Dimension553
- 4 Discussion558
- References559

Free Fall and Self-Force: an Historical Perspective561
Alessandro Spallicci

- 1 Introduction562
- 2 The Historical Heritage563
- 3 Uniqueness of Acceleration and the Newtonian Back-Action565
- 4 The Controversy on the Repulsion and on the Particle Velocity at the Horizon569
- 5 Black Hole Perturbations574
- 6 Numerical Solution578
- 7 Relativistic Radial Fall Affected by the Falling Mass582
 - 7.1 The Self-Force582
 - 7.2 The Pragmatic Approach.....587
- 8 The State of the Art590
 - 8.1 Trajectory591
 - 8.2 Regularisation Parameters592
 - 8.3 Effect of Radiation Reaction on the Waveforms During Plunge592
- 9 Beyond the State of the Art: the Self-Consistent Prescription593
- 10 Conclusions595
- References597

Index605

The Higgs Mechanism and the Origin of Mass

Abdelhak Djouadi

Abstract The Higgs mechanism plays a key role in the physics of elementary particles: in the context of the Standard Model, the theory which, describes in a unified framework the electromagnetic, weak, and strong nuclear interactions, it allows for the generation of particle masses while preserving the fundamental symmetries of the theory. This mechanism predicts the existence of a new type of particle, the scalar Higgs boson, with unique characteristics. The detection of this particle and the study of its fundamental properties is a major goal of high-energy particle colliders, such as the CERN Large Hadron Collider or LHC.

1 The Standard Model and the Generation of Particle Masses

The end of the last millennium witnessed the triumph of the Standard Model of elementary particles, the quantum and relativistic theory which describes in a unified framework three of the four fundamental forces in Nature: the electromagnetic, weak, and strong nuclear interactions. In particular, major progress has been achieved in the last decade as, on the one hand, the discovery of the top quark has finally allowed to fully reconstruct the puzzle formed by matter elementary particles and, on the other hand, very high precision experiments have asserted the validity of the model for describing the three particle interactions with an unprecedentedly high degree of accuracy. Nevertheless, one cornerstone of the theory still remains to be tested: the mechanism by which the particles acquire mass while preserving the fundamental symmetries of the theory. This mechanism predicts the existence of a new type of particle, the scalar Higgs boson, which is expected to be produced and studied at the CERN Large Hadron Collider (LHC), which will soon start operation.

In this mini-review, I present a pedagogical introduction to the Standard Model and the Higgs mechanism for mass generation. I then briefly describe the basic

A. Djouadi (✉)

Laboratoire de Physique Théorique, CNRS et Université Paris-Sud, F-91405 Orsay, France
e-mail: Abdelhak.Djouadi@th.u-psud.fr

properties of the Higgs particle and discuss the prospects for producing it at the LHC and for studying its basic properties. The Higgs sectors of some scenarios for new physics beyond the Standard Model will be briefly commented upon.

1.1 The Elementary Particles and Their Interactions

Let us start by briefly summarizing the particle content of the Standard Model and the basic interactions to which it is subject [1], also sketched in Figs. 1 and 2.

Particles of: matter ($s=\frac{1}{2}$) force ($s=1$) mass ($s=0$)
3 families of fermions gauge bosons Higgs

$Q \rightarrow$ $m \rightarrow$	quark up u $+2/3$ $\sim 5 \text{ MeV}$	quark charm c $+2/3$ 1.6 GeV	quark top t $+2/3$ 172 GeV	gluon g 0 0	Higgs H 0 $\geq 114 \text{ GeV}$
	quark down d $-1/3$ $\sim 5 \text{ MeV}$	quark strange s $-1/3$ 0.2 GeV	quark bottom b $-1/3$ 4.9 GeV	photon γ 0 0	
	neutrino e ν_e 0 ~ 0	neutrino m ν_m 0 ~ 0	t neutrino ν_t 0 ~ 0	boson Z Z^0 0 91.2 GeV	
	electron e -1 0.5 MeV	muon m -1 0.1 GeV	tau t -1 1.7 GeV	bosons W W^\pm ± 1 80.4 GeV	

Fig. 1 The elementary particles of the Standard Model, their spin, electric charges, and their masses in Giga-Electron Volts (GeV) and in units where the speed of light c is equal to unity

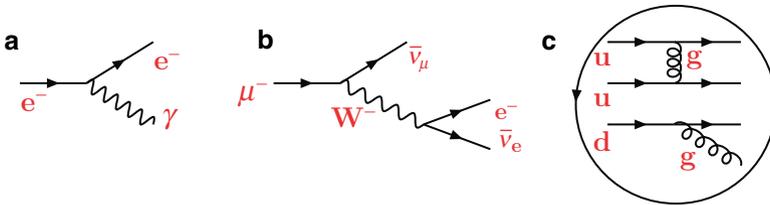


Fig. 2 Diagrams (called Feynman diagrams) illustrating the three fundamental interactions of the Standard Model. (a) The electromagnetic interaction, where an electron emits a photon, continuing with altered momentum; (b) the weak interaction responsible for the decay of a muon, via the exchange of a W boson, into an electron and muonic and electronic antineutrinos; (c) the strong interaction where the u,u,d quarks constituting the proton interact by exchanging or emitting gluons

The particles that constitute the building blocks of matter have intrinsic magnetic moment or spin equal to $s = 1/2$ and are called fermions as they obey to Fermi–Dirac statistics.¹ They appear in three families; see Fig. 1. The first family forms ordinary matter: it consists of the electron and its associated neutrino, which are called leptons, as well as the up and down quarks with fractional electric charges, and which form nuclear matter, that is, the protons and neutrons. The two other families are perfect replicas of the former: the leptons and quarks that constitute them have exactly the same quantum numbers but larger masses. They decay into the fermions e , ν_e , and u of the first family which, in contrast, are absolutely stable. Note that the top quark, discovered in 1995, is 330,000 times heavier than the electron, observed by Thomson a century earlier. The latter is far heavier than the neutrinos, which have very small masses that can be safely neglected in the present discussion.

To be complete, one should note that for each particle is associated an antiparticle that has the same properties but opposite electric charge; these are usually noted with a bar, \bar{f} for the antifermion of the fermion f .

Besides, one has the force particles that mediate the fundamental interactions between the various fermions. They have a spin equal to unity, $s = 1$, and are called bosons as they obey to Bose–Einstein statistics.² The photon, denoted γ , is the messenger of the electromagnetic interaction to which are subject charged particles, that is, all fermions except neutrinos. The W^+ , W^- , and Z^0 bosons mediate the weak nuclear interaction responsible for the radioactive decay of heavy particles and which, in principle, concerns all fermions. Finally, eight gluons are the messengers of the strong nuclear force that binds the atomic nuclei, and which concerns only quarks.

Note that there is a fourth fundamental force in Nature, the gravitational interaction for which the messenger is the hypothetical graviton of spin 2. It has a magnitude that is far too weak to play a role at the energies that are being probed in laboratory experiments. It is thus neglected, except in some cases discussed later.

1.2 *The Standard Model of Particle Physics*

The quantum and relativistic theory that describes in a unified framework the electromagnetic, weak, and strong forces of elementary particles is called the Standard Model [2, 3]. It is based on a very powerful principle, local or gauge symmetry: the fields corresponding to the particles,³ as well as the particle interactions, are invariant with respect to local transformations (i.e., for any space–time point) of a given internal symmetry group. The model is a generalization of Quantum

¹ The exclusion principle, put forward by Wolfgang Pauli in 1925, forbids to two fermions to be in the same quantum configuration.

² In contrast to fermions, several bosons can occupy the same quantum configuration and, thus, can aggregate.

³ In a quantum theory, to each particle is associated a field that has a given number of degrees of freedom. For instance, the fields associated to a fermion or to the (massless) photon have two degrees of freedom, while a real scalar field has a single degree of freedom.

Electro-Dynamics (QED) [4], the quantum and relativistic theory of electromagnetism which describes the interaction of electrically charged particles through the exchange of photons. The latter is invariant under local phase transformations⁴ described by a symmetry group noted $U(1)_Q$ and conserves the quantum number that is the electric charge Q .

The symmetry group of the Standard Model is slightly more complicated and is denoted by $SU(3)_C \times SU(2)_L \times U(1)_Y$.

- For the strong interaction [3], based on the symmetry group $SU(3)_C$, the quarks appear in three different states differentiated by a quantum number called color (which has nothing to do with the usual color) that they exchange via eight intermediate massless gluons.⁵
- The electromagnetic and weak interactions are combined to form the electroweak interaction [2], which is based on the symmetry group $SU(2)_L \times U(1)_Y$. The fermions appear in two quantum configurations called left- and right-handed chiralities corresponding, for massless fermions, to the two possibilities for the projection of spin onto the direction of motion ($s = \pm \frac{1}{2}$). The fermions with left-handed chiralities of each family are assembled in a doublet of weak isospin, while the fermions with a right-handed chirality are in singlets of weak isospin. In the case of first-family leptons, for instance, the left-handed electron and its associated neutrino always appear in the form of a doublet (ν_{eL} has isospin $+\frac{1}{2}$ while e_L has isospin $-\frac{1}{2}$), while the right-handed electron appears in a singlet e_R (with isospin equal to 0); there is no right-handed neutrino ν_{eR} . The same holds for quarks: the left-handed quarks form a doublet (u_L) and the right-handed ones u_R, d_R are singlets.
- For a given particle, the quantum number of hypercharge Y is given by the electric charge and the isospin, $Y = 2Q - 2I$.

The electroweak interaction is mediated by the exchange of the gauge bosons⁶ W^\pm, Z^0 and the photon γ . While the photon, the messenger of the long range

⁴ In QED, the Lagrangian density that describes the theory is invariant under phase transformations on the charged fermionic fields collectively denoted by ψ , $\psi(x_\mu) \rightarrow e^{iQ\theta} \psi(x_\mu)$, where $x_\mu = (\mathbf{x}, t)$ is the space-time four-vector and Q the electric charge of the fermion. These transformations are called gauge or local transformations since the parameter θ depends on the space-time four-vector. The photon field mediating the interaction and described by the four-vector $A_\mu = (\mathbf{A}, A_0)$, transforms as: $A_\mu(x_\mu) \rightarrow A_\mu(x_\mu) - \frac{1}{Q} \partial_\mu \theta(x_\mu)$, where ∂_μ is the derivative with respect to x_μ . In fact, the interaction of fermions via the exchange of photons can be induced in a minimal way in the Lagrangian density of the free fermion and photon systems, by substituting the usual derivative ∂_μ by what is called the covariant derivative: $D_\mu \equiv \partial_\mu - iQA_\mu$. The gauge transformation group is noted $U(1)_Q$ for the group of unitary matrices of dimension one.

⁵ The transformations of the $SU(3)_C$ symmetry group of the strong interaction, called Quantum Chromo-Dynamics or QCD, are generated by eight 3×3 unitary matrices with determinant equal to unity. The quarks are triplets of the group (they appear in three colors) while the gluons correspond to the eight generators of the group (there are $n^2 - 1$ generators for $SU(n)$) and are non-massive.

⁶ The three generators of the $SU(2)_L$ group [$n^2 - 1 = 3$ for $n = 2$], which can be identified with the three 2×2 Pauli matrices that generate spatial rotations, correspond to the three-vector bosons

electromagnetic force, has zero-mass, the W^\pm, Z^0 gauge bosons should be massive since they mediate the weak force that has a short range.

The Standard Model combines esthetics, since gauge invariance provides a symmetry and is related to a geometrical principle, economy as the number of gauge bosons is fixed and their interactions uniquely determined in a minimal way once the symmetry group is chosen, mathematical coherence and, thus, the possibility of predicting any phenomenon with infinite precision in principle. Last but not least, it had a blatant experimental success as some of its predictions have been confirmed at the permille level of accuracy [1, 5]. This makes the Standard Model one of the most successful and most precisely verified theories in Physics.

1.3 The Higgs Mechanism for Mass Generation

A cornerstone of the Standard Model is the mechanism that generates the particle masses while preserving the gauge invariance of the theory. Indeed, the direct introduction of masses for the fermions and for the gauge bosons that mediate the weak interaction violates the invariance with respect to the transformations of the electroweak symmetry group. In principle, gauge bosons should remain massless to preserve a local symmetry.⁷ This is for instance the case of the photon for which the zero-mass ensures the invariance of electromagnetism with respect to phase transformations. On the other hand, the fact that the left- and right-handed fermions do not have the same isospin quantum numbers prevents them from acquiring a mass in a gauge invariant way under isospin symmetry.⁸

It is the Higgs–Brout–Englert mechanism [6, 7], commonly called the Higgs mechanism, which allows the generation of particle masses while preserving the gauge symmetry of electroweak interactions.

The Higgs mechanism postulates the existence of a doublet (under isospin) of complex scalar fields,

$$\Phi = \begin{pmatrix} \text{Re}\Phi^+ + i\text{Im}\Phi^+ \\ \text{Re}\Phi^0 + i\text{Im}\Phi^0 \end{pmatrix}, \quad (1)$$

$W_\mu^1, W_\mu^2, W_\mu^3$, the messengers of the interaction. The gauge boson associated with the unique generator of the $U(1)_Y$ group is noted B_μ . The four gauge bosons of the electroweak group, $W_\mu^1, W_\mu^2, W_\mu^3$, and B_μ are not the physical ones; the latter are linear combinations of the former:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), Z_\mu^0 = \cos\theta_W W_\mu^3 + \sin\theta_W B_\mu, A_\mu = -\sin\theta_W W_\mu^3 + \cos\theta_W B_\mu,$$

where θ_W is the electroweak mixing angle.

⁷ A mass term for the photon and thus a term that is bilinear in the fields, $M_A^2 A_\mu A^\mu$ (with the notation $A_\mu A^\mu = \sum_\mu A_\mu A_\mu = A_0^2 - \mathbf{A} \cdot \mathbf{A}$), will violate the invariance with the transformations under the group $U(1)_Q$ since one would have: $M_A^2 A_\mu A^\mu \rightarrow M_A^2 (A_\mu - \frac{1}{Q} \partial_\mu \theta)(A^\mu - \frac{1}{Q} \partial^\mu \theta) \neq M_A^2 A_\mu A^\mu$.

⁸ In the case of the first family of leptons for instance, since $(\begin{smallmatrix} \nu_{eL} \\ e_{eL} \end{smallmatrix})$ forms an isospin doublet while e_R forms an isospin singlet, one cannot form a mass term for the electron (which is bilinear in the electron field), $m_e \bar{e}_L e_R$, as this term violates $SU(2)_L$ symmetry.

to which one associates a potential that is invariant under the transformations of the $SU(2)_L \times U(1)_Y$ electroweak symmetry group,

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2. \quad (2)$$

In this equation, μ^2 stands for the mass term of the field Φ and λ the (positive) coupling constant of its self-interaction. For positive values of μ^2 , the potential $V(\Phi)$ has the usual form of an inverted bell in which the minimum of the field Φ , corresponding to the state of vacuum which should be stable, has zero value. In this case, we simply have four additional scalar fields corresponding to four new degrees of freedom or scalar particles, which does not help much toward the solution of the mass generation problem. The situation becomes much more interesting if the mass squared term μ^2 is negative. In this case, the potential $V(\Phi)$ has the shape of a bottle of Champagne bottom as shown in Fig. 3. The minimum of the potential is not reached for a zero value of the field Φ (or, rather, for its neutral component Φ^0) as usual, but at the nonzero value $v = \sqrt{-\mu^2/\lambda}$ that is called the nonzero vacuum expectation value of the field Φ .

When interpreting the field content of the theory starting from this nonsymmetric (the small ball of Fig. 3 having chosen a given minimum) but physical vacuum, one realizes that three degrees of freedom, among the four degrees of freedom of the complex doublet field Φ , have disappeared from the spectrum: they have been “absorbed” by three gauge bosons of the electroweak interaction. These spin-1 fields, initially massless and with two components or degrees of freedom called transverse components, will acquire an additional degree of freedom corresponding to their longitudinal component, a characteristic signature of massive spin-1 fields.

The $SU(2)_L \times U(1)_Y$ symmetry is then still present but, since the vacuum is not symmetric, it is not apparent: it is said to be spontaneously broken. Thus, it is the spontaneous breaking of the electroweak symmetry that generates the masses of the vector bosons W^\pm and Z^0 in a gauge invariant way. The photon remains massless as it should be to explicitly preserve the gauge invariance of electromagnetism.⁹

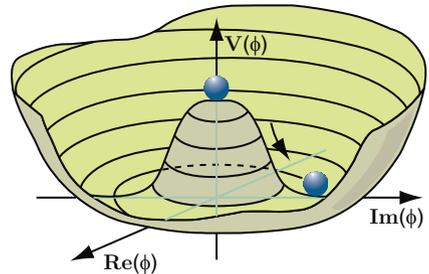


Fig. 3 The potential of the scalar field Φ with its minimum at the value v of the field

⁹ Some technical details of this mechanism are as follows. One first imposes to the Lagrangian density of the field $\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$ to be invariant under the local transformations of the $SU(2)_L \times U(1)_Y$ symmetry group. The most general form of the Lagrangian is given by:

Using the same scalar field Φ , the masses of the Standard Model fermions can also be generated in a gauge invariant manner by introducing, for each fermionic field with left- and right-handed chiralities (and thus, not for the neutrinos which have only left-handed chiralities) interaction terms with the scalar field. After spontaneous symmetry breaking, one identifies the magnitude of the various interaction terms with the experimentally measured values of the fermion masses.¹⁰

$$\mathcal{L}_\Phi = (D^\mu \Phi)^\dagger (D_\mu \Phi) - V(\Phi), \quad V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2$$

where μ^2 is the mass term and λ the self-coupling constant. The covariant derivative D_μ induces the interactions of the field Φ with the gauge boson fields W_μ^a, B_μ :

$$D_\mu = \partial_\mu - ig_2 \frac{\tau_a}{2} W_\mu^a - ig_1 \frac{Y}{2} B_\mu$$

with $\frac{1}{2}\tau^a$ and $\frac{Y}{2}$ the generators of the $SU(2)_L$ and $U(1)_Y$ groups with coupling constants g_2 and g_1 . For $\mu^2 < 0$, the minimum of the potential is at the (vacuum expectation) value $v = \sqrt{-\mu^2/\lambda}$. To obtain the physical fields, one must describe the Lagrangian \mathcal{L}_Φ with the true and stable vacuum and the various steps to follow are:

- Write the doublet field Φ in terms of four real fields $\theta_{1,2,3}(x)$ and $H(x)$ and use the expansion series of an exponential, which to first order gives:

$$\Phi(x) = \begin{pmatrix} \theta_2 + i\theta_1 \\ \frac{1}{\sqrt{2}}(v+H) - i\theta_3 \end{pmatrix} \simeq e^{i\theta_a(x)\tau^a(x)/v} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v+H(x) \end{pmatrix}$$

- Use the freedom to perform a gauge transformation on Φ to eliminate the three fields $\theta_{1,2,3}$

$$\Phi(x) \rightarrow e^{-i\theta_a(x)\tau^a(x)} \Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v+H(x) \end{pmatrix}$$

- Develop the kinetic term $|D_\mu \Phi|^2$ of \mathcal{L}_Φ which, gives

$$\frac{1}{2}(\partial_\mu H)^2 + \frac{g_2^2}{8}(v+H)^2 |W_\mu^1 + iW_\mu^2|^2 + \frac{1}{8}(v+H)^2 |g_2 W_\mu^3 - g_1 B_\mu|^2.$$

After having defined the new fields W_μ^\pm, Z_μ^0 , and A_μ , with $\sin \theta_W = g_1/\sqrt{g_1^2 + g_2^2}$, one identifies the terms that are bilinear in these fields with the masses of the associated particles. One realizes then that the $\theta_{1,2,3}$ degrees of freedom have been absorbed by the fields W^\pm and Z^0 to form their longitudinal components and thus their masses are given by:

$$M_W = \frac{1}{2}v g_2, \quad M_Z = \frac{1}{2}v \sqrt{g_2^2 + g_1^2}, \quad M_A = 0.$$

With the value v given by the Fermi constant of weak interaction, $v = 1/(\sqrt{2}G_F)^{1/2} = 246$ GeV, and the experimentally measured values of the coupling constant g_2 and g_1 , one recovers the correct masses for the W^\pm and Z^0 bosons. The photon, instead, remains massless, $M_A = 0$, as it should. For more details on the Higgs mechanism, see for instance the detailed review of Ref. [7].

¹⁰ To generate the fermion masses, one introduces a Lagrangian density describing the fermion–Higgs interactions in an $SU(2)_L \times U(1)_Y$ gauge invariant way. In the case of the electron and the neutrino for instance, and taking into account that the leptons with left-handed chirality are in $SU(2)_L$ doublets while the electron with right-handed chirality is in a singlet, one would have:

$$\mathcal{L}_f = f_e (\bar{e}, \bar{\nu})_L \Phi e_R \rightarrow f_e (\bar{\nu}_e, \bar{e}_L) \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v+H \end{pmatrix} e_R.$$

One then identifies the masses of the electron and the neutrino with the bilinear terms in the fields:

$$m_e = f_e v / \sqrt{2} \text{ and } m_\nu = 0.$$

Finally, among the four initial degrees of freedom of the field Φ and after three have been absorbed by the W^\pm and Z^0 gauge bosons to acquire their masses, one degree of freedom will be left over. This residual degree of freedom corresponds to a physical particle,¹¹ the Higgs boson H , the ‘‘Grail of particle physics.’’

The Standard Model, despite all of its brilliant successes, will only be complete and validated once this particle has been observed and its fundamental properties determined. This is the major goal of high-energy colliders and, in particular, of the CERN LHC which has recently started operation.

2 The Profile of the Higgs Particle

2.1 Characteristics of the Higgs Boson

The Higgs boson has remarkable characteristics, which means that it has a unique status in the table of elementary particles given in Fig. 1.

First of all, in contrast to matter particles with spin 1/2 and to gauge particles with spin 1, it has spin zero.¹² It is therefore a boson, as it has integer spin, but it does not mediate gauge interactions.

Another unique property of the Higgs particle is that it interacts with or couples to elementary particles proportionally to their masses: the more massive is the particle, the stronger is its interaction with the Higgs boson.¹³ Thus, the Higgs particle will couple more strongly to the messengers of the weak interactions, the W^\pm and Z^0 bosons, the masses of which are of the order of hundred GeV. It couples also more strongly to the top quark, the heaviest particle in the Standard Model, and, to a lesser extent, the bottom quark and the τ leptons, than to the fermions of the first and second generations which have much smaller masses. Furthermore, it does not couple to the neutrinos, which are considered as being massless.

The Higgs boson does not couple directly to photons and gluons as the latter have no mass (in the case of gluons, a direct coupling is also absent because

¹¹ The mass of the Higgs boson can be simply deduced from the scalar Higgs potential by isolating the terms that are bilinear in the H fields, $\frac{1}{2}M_H^2 H^\dagger H$, and one obtains

$$M_H = \sqrt{2\lambda v^2}.$$

¹² Scalar or spin-zero particles also exist in Nature, but not at the fundamental level: the π mesons, for instance, are spin-zero particles but they are bound states of spin 1/2 quarks.

¹³ The interactions of the Higgs boson with the other particles, that is, the terms in the density Lagrangian involving the fields H and two fermionic fields or gauge bosonic fields are described by the same terms giving the masses, since H always appears in the combination $H + v$. The interaction of the Higgs boson to the particles is thus proportional to their masses:

$$\mathcal{L}_{Hff} \propto m_f/v, \quad \mathcal{L}_{HW+W-} \propto g_2 M_W, \quad \mathcal{L}_{HZZ} \propto g_2 M_Z / \cos \theta_W.$$

the Higgs boson does not carry color quantum numbers). However, couplings can be induced in an indirect way through quantum fluctuations. Indeed, according to Heisenberg's uncertainty principle of Quantum Mechanics, the Higgs boson can emit pairs of very heavy particles (such as top quarks for instance) and immediately absorb them; but these virtual particles can, in the meantime, emit photons or gluons. Higgs–photon–photon and Higgs–gluon–gluon couplings are then generated. However, they are expected to be rather small, as they imply intermediate interactions of the virtual particles to photons and gluons, which have a small intensity.

Finally, the Higgs boson has also self-interactions, residual of those of the original scalar field Φ shown in the Higgs potential of Eq. 2; the magnitude of these triple and quartic self-interactions are also proportional to the Higgs boson mass (in fact, Higgs mass squared).¹⁴

2.2 Constraints on the Higgs Boson Mass

The Higgs boson mass M_H is the only free and unknown parameter of the Standard Model, since the coupling constants of the three fundamental interactions that it describes as well as the masses of the fermion and the gauge boson have been experimentally determined. Once this parameter is fixed, the entire profile of the Standard Model Higgs boson is uniquely determined. In particular, its couplings to the other particles, its production and decay rates can be calculated.

Nonetheless, M_H is not a completely free parameter as it is subject to some experimental and theoretical constraints [8] that we briefly summarize.

The experimental constraints come mainly from the ancestor of the LHC, the Large Electron Positron collider LEP, an electron–positron collider that has operated in the 1990s with an energy ranging approximately from 90 to 210 GeV. Important constraints come also from the Tevatron, the proton–antiproton collider of Fermilab near Chicago with an energy of 2 TeV. The LEP experiment has first allowed a comprehensive direct search for the Higgs boson and the absence of any signal at LEP led to the lower bound of 114 GeV on its mass. Current results from the Tevatron indicate that a Higgs particle with a mass comprised between 160 and 170 GeV is probably excluded. In addition, the high-precision measurement of some electroweak observables – of the order of one permille for some of them – led to indirect constraints on M_H . Indeed, even if it is too heavy to be produced directly in collider experiments, the Higgs boson appears in the small but measurable quantum fluctuations of, for instance, the masses of the Z^0 and W^\pm bosons, which have been

¹⁴ The self interactions between three or four Higgs bosons can be readily obtained from the scalar Higgs potential after electroweak symmetry breaking and read:

$$\mathcal{L}_{HHH} \propto 3M_H^2/v, \quad \mathcal{L}_{HHHH} \propto 3M_H^2/v^2.$$

The triple and quartic Higgs couplings are thus proportional to the Higgs mass squared.

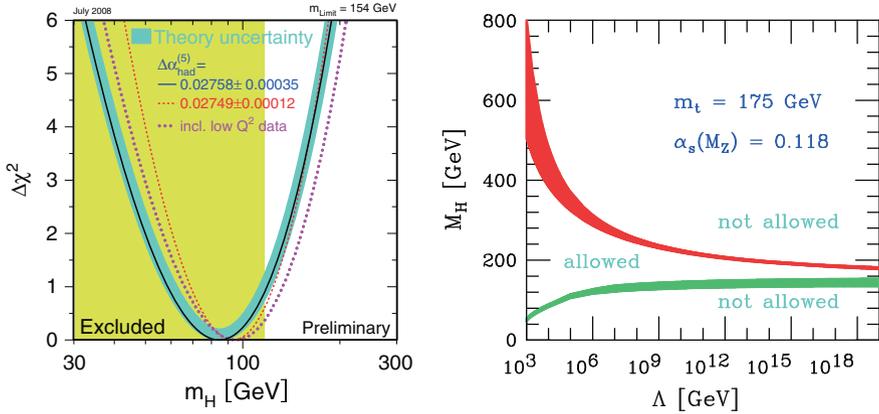


Fig. 4 *Left*: the preferred values of the Higgs boson mass in the Standard Model (the minima of the curves) after a global fit of all electroweak precision data. The full curve in black represents the 68% confidence level result which leads to $M_H = 84^{+36}_{-26}$ GeV, the blue band includes the theoretical uncertainties and the dotted curves are for the results when some experimental inputs are slightly changed. The domain in yellow represents the excluded region, $M_H \leq 114$ GeV, from direct Higgs searches at LEP; from Ref. [5]. *Right*: the trivality bound from the finiteness of the Higgs self-coupling (upper curved in red) and the vacuum stability bound from the requirement of the positivity of the self-coupling (lower curved in green) on the Higgs boson mass as a function of the new physics or cut-off scale Λ ; the allowed region lies between the bands and the colored/shaded bands illustrate the impact of various uncertainties; from Ref. [9]

determined with high accuracy. A global analysis of all electroweak high-precision data available today [5] allows the imposition of an upper bound, $M_H \lesssim 180$ GeV, with a 95% confidence level or probability;¹⁵ see the left-hand side of Fig. 4.

Theoretical constraints on the Higgs boson mass can also be derived from considerations on the energy scale for which the Standard Model is valid before some new physics beyond the model manifests itself. A first constraint is obtained from the requirement that the theory remains unitary, an important constraint that, in Quantum Mechanics, is related to the conservation of probabilities. For a too heavy Higgs particle, some processes such as the scattering of W^\pm and Z^0 bosons (in which a Higgs particle can be exchanged) would have amplitudes that increase with energy and would eventually violate unitarity at energies above 1 TeV. To preserve unitarity in the context of the Standard Model, a Higgs particle with a mass below approximately 1 TeV is required.

Another constraint emerges from the fact that the self-coupling of the Higgs boson, which is proportional to M_H^2 , evolves with energy by virtue of quantum fluctuations (virtual fermions, gauge, and Higgs bosons are exchanged in the coupling

¹⁵ These constraints, being indirect, are however valid only in the context of the Standard Model and could be less severe in some of its extensions in which new phenomena might also contribute to the observables via quantum fluctuations; see later.

among three or four Higgs particles). This evolution is rather strong and at some stage, the coupling becomes extremely large and the theory completely loses its predictability.¹⁶ If the energy scale up to which the coupling $\lambda \propto M_H^2$ remains small enough, and the Standard Model effectively valid, is of the order of the Higgs mass itself, M_H should be less than approximately 1 TeV.¹⁷ On the other hand, for small values of the self-coupling, and hence of the Higgs boson mass, the quantum fluctuations tend to drive the coupling to negative values and, thus, completely destabilize the scalar Higgs potential to the point where the minimum is not stable anymore (the scalar potential of Fig. 3 is inverted and the minimum is reached when the field is at $-\infty$). Requiring that the self-coupling stays positive and the minimum stable up to energies of about 1 TeV implies that the Higgs boson should have a mass above approximately 70 GeV. However, if the Standard Model is to be extended to ultimate scales, such as for instance the Planck scale $M_P \sim 10^{18}$ GeV, these requirements on the self-coupling from finiteness and positivity become much more constraining and the Higgs mass should lie in the range $120 \text{ GeV} \lesssim M_H \lesssim 180 \text{ GeV}$. This is a rather narrow margin that is close to the one obtained from the direct and indirect experimental constraints.

2.3 *The Higgs Decay Modes and Their Rates*

Since the Higgs boson couples to particles proportionally to their masses, it will have the tendency to decay into the heaviest particle allowed by kinematics (of course, one needs that the sum of the masses of the final particles does not exceed the mass of the decaying Higgs). For a mass of the order of 100 GeV, the Higgs boson will prefer to decay into a pair of bottom quarks and, to a lesser extent, a pair of charm quarks or τ leptons, which have smaller masses. The hierarchy of the decay rates is given by the mass squared of these particles. The probability for these decays to occur, or the branching ratios, are shown in Fig. 5 (center) and as can be seen, for Higgs masses below 130 GeV, the decays into bottom quark final states are by far dominant with a probability of the order of 80%, while the probability for decays into charm quarks and τ lepton pairs is of the order of a few percent.

¹⁶ In the Standard Model, and in particle physics in general, we do not know how to solve exactly the equations of motion of the fields. The procedure is then to deal with a free field (noninteracting) theory and to treat the interactions as small perturbations; this is possible as the coupling constants of the electromagnetic and weak interactions, as well as the strong coupling constant at high energy, are small enough. The system is thus expanded in series of the coupling constant and solved order by order in the (perturbative) series. This approach fails if a coupling constant is too strong so that the series is not convergent.

¹⁷ The exact value is $M_H \lesssim 650 \text{ GeV}$ when only the first terms of the perturbation series are included. This value is remarkably close to the one obtained from numerical simulations in lattice gauge theory where the theory can be solved exactly.

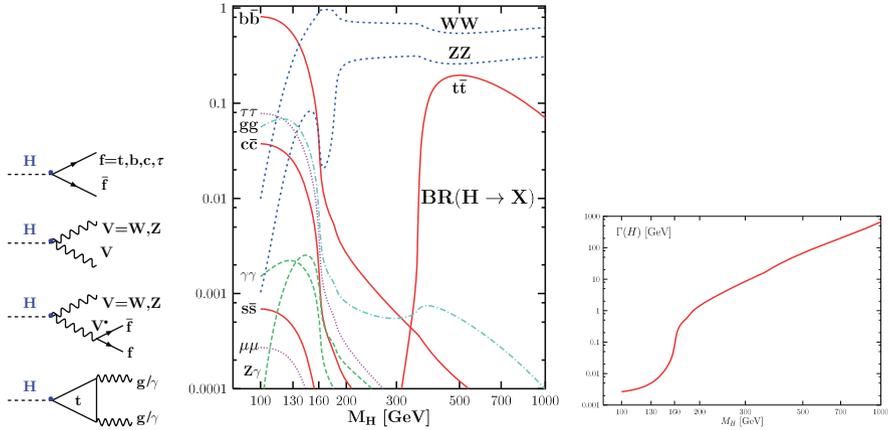


Fig. 5 The dominant decay processes of the Higgs boson including the loop and three-body decays (*left*) and the rates or branching ratios of these Higgs decays as a function of M_H (*center*), together with the total Higgs decay width as a function of M_H (*right*); from Ref. [10]

Nevertheless, some decay channels that should normally not appear can be induced by quantum fluctuations. This is for instance the case of Higgs decays into two gluons, two photons, or a photon plus a Z^0 boson. In particular, the decay rate into two gluons, induced by a loop involving a virtual top quark which couples strongly to the Higgs boson, can be comparable to the decay rates into charm quarks and τ leptons. Instead, the decay mode into two photons and into a photon plus a Z^0 boson (which are induced mainly by top quark and W boson loops) are very rare, a consequence of the fact that the electromagnetic coupling is much smaller than the strong interaction coupling. For Higgs masses below 130 GeV, the probability for these two decay modes to occur is at the few permille level.

In addition, the Higgs boson can decay into two rather massive particles, with the sum of their masses larger than M_H , but one of which is virtual and decays into two real particles with smaller masses. This is the case, for example, of the Higgs decay into two W^\pm (or Z^0) bosons for masses below $2M_W$ ($2M_Z$) and thus, one of the W^\pm (Z^0) bosons must be virtual and decays into a pair of rather light fermions. In fact, for values of the Higgs mass above 130 GeV (and below $2M_W$), the rate for the three-body Higgs decay into a real and a virtual W^\pm boson becomes comparable and even larger than the otherwise dominating two-body decay into a pair of bottom–antibottom quarks. This is due to the fact that the virtuality of the gauge boson is partially compensated by the stronger coupling of the Higgs boson to the W^\pm bosons compared to the Higgs coupling to bottom quarks.

For Higgs boson masses of the order of 180 GeV and beyond, the Higgs decays into two pairs of real W^+W^- and Z^0Z^0 bosons largely dominate with branching fractions of two to one in favor of the former channel. Even for a Higgs boson with a mass larger than 350 GeV, for which the decay channel into pairs of top quarks becomes kinematically open, these two channels remain dominant thanks to the longitudinal components of the W^\pm , Z^0 bosons which significantly enhance the rates.

Finally, one should note that the total decay width of the Higgs boson, the inverse of its lifetime, is of the order of only a few MeV for Higgs masses close to 100 GeV but it considerably increases with the Higgs mass to reach the GeV range for $M_H \propto 180$ GeV and becomes of the same order of M_H when the latter approaches 1 TeV; see the right-hand side of Fig. 5. Thus, the Higgs boson is a very narrow resonance for small masses but the resonance becomes very wide for a very heavy Higgs particle.

3 Higgs Production at the LHC

3.1 The Large Hadron Collider

The LHC, located at CERN near Geneva, forms a circular ring of length 27 km buried 100 meters underground; see Fig. 6. It is the largest scientific instrument ever built and its construction represented a major technological challenge. The LHC is a proton–proton collider operating at an energy of 14 TeV in the center of mass. Since the protons are formed by three quarks, the effective energy, that is, the energy in the quark center of mass, is of the order of 5 TeV. This energy is largely sufficient to probe in depth the TeV scale.

An important characteristic is the luminosity delivered by the machine, corresponding to its ability to produce particle collisions. Integrated over time, it has the inverse unit of a cross section, that is the probability of an interaction during the collision; the product of the two quantities gives the expected number of events. The usual unit of a cross section is cm^2 , but since the events are extremely rare the commonly used unit is the picobarn (pb), $1 \text{ pb} = 10^{-36} \text{ cm}^2$, or femtobarn, $1 \text{ fb} = 10^3 \text{ pb}$. The luminosity expected at the LHC is of the order of 10 fb^{-1} per year in the early operating stage and should increase to 100 fb^{-1} per year in the following years.

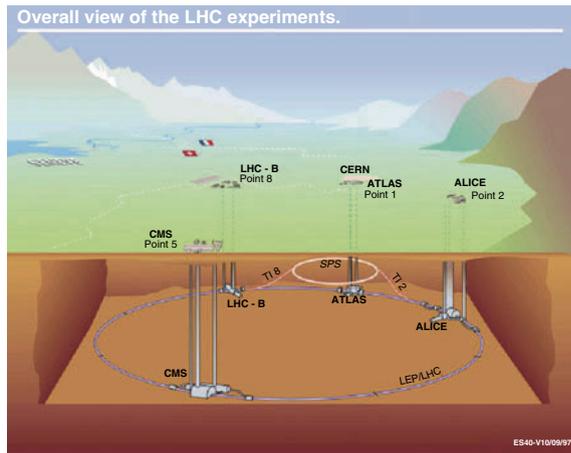


Fig. 6 The LHC tunnel and the various associated experiments including the multipurpose experiments ATLAS and CMS (courtesy from CERN)

The two multipurpose detectors ATLAS and CMS [11] (there are also other detectors dedicated to, for instance, the physics of the bottom quark and that of heavy ions) have been devised to deliver the maximal amount of information on the interactions that occur in their inner part and to cover a large spectrum of possible signatures from known and new physical phenomena. Their potential has particularly been optimized to detect the Higgs particle for masses comprised between 100 GeV and 1 TeV, in the main production modes and in the most important decay channels such as decays into charged or neutral leptons, photons, and heavy quarks.

3.2 The Production of the Higgs Boson

At the LHC, the Higgs particle could in principle be simply produced in the annihilation of an up or down type quark that lies inside the initial protons and its corresponding antiquark;¹⁸ however, since the masses of these first family quarks are very small, the Higgs production rates turn out to be completely negligible. The Higgs particle should therefore be produced via a radiation from a rather heavy particle such as the massive gauge bosons W^\pm , Z^0 or the top quark, exploiting the large Higgs couplings to these particles. Four production processes are then at our disposal for Higgs production at the LHC [7]; see the left-hand side of Fig. 7.

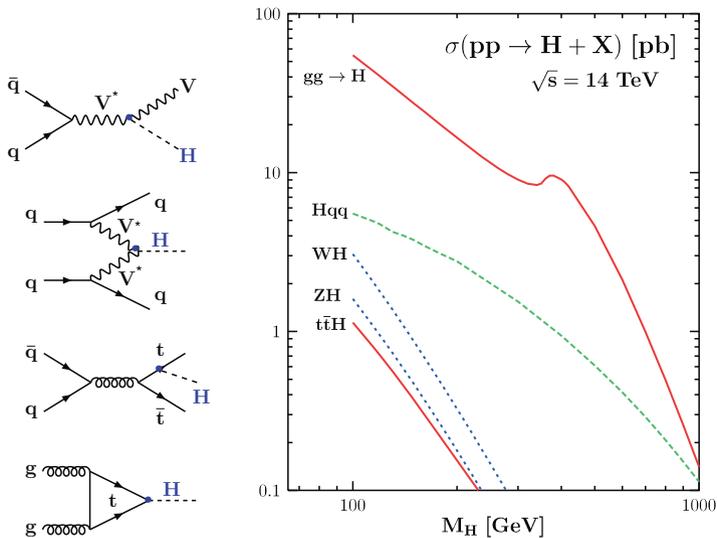


Fig. 7 *Left:* Feynman diagrams for the dominant Higgs production mechanisms at the LHC. *Right:* the Higgs production cross sections of these processes at the LHC (in pb) as a function of the Higgs mass (the higher order corrections have been included); from Ref. [7]

¹⁸ While the quarks naturally appear inside the protons, antiquarks as well as gluons appear via quantum fluctuations with a rather high probability.

There is first Higgs production in association with a massive $V = W^\pm$ or Z^0 boson, a quark and its antiquark partner from the initial protons annihilate to produce a virtual gauge boson, which then immediately decays into a real gauge boson and a Higgs particle. There is also the process in which the quarks inside the protons radiate (virtual) massive gauge bosons which then annihilate to produce a Higgs particle; the final state would then consist of a Higgs boson and two quarks with a very characteristic kinematics. A third production process exploits the very large Higgs coupling to top quarks: a pair of top and antitop quarks is created in the annihilation of the quarks and antiquarks (as well as the gluons) of the initial protons and a Higgs boson is then emitted from one of the heavy final state particles.

Finally, there is the process in which two gluons from the protons (which again appear through the quantum fluctuations of the u and d quarks forming the protons) annihilate and, via a loop of virtual heavy top quarks, produce a Higgs particle. This process is the inverse of the one that allows to the Higgs boson to decay into two gluons discussed above. It turns out that this gluon fusion process is by far the dominant Higgs production mechanism at the LHC. Indeed, the smallness of the Higgs coupling to gluons, that is generated through tiny quantum fluctuations, is compensated by the large Higgs coupling to the top quark, the favorable kinematics since only one heavy particle is produced in the final state and, also, by the high probability of finding a gluon in the proton at the very high energies that come into play at the LHC.

The cross sections or production rates for these various Higgs production processes are shown in the right-hand side of Fig. 7, with a unit that is the pb, which would correspond to 10, 000 events for the standard luminosity expected at the LHC. For relatively small Higgs masses, say below 200 GeV, the gluon fusion mechanism has a cross section of the order of several tens of pb while the other processes have cross sections that are one or several order of magnitude below. More than one million events involving a Higgs particle can be thus expected after several years of collider running. The production rates rapidly decrease with the Higgs mass and when M_H approaches 1 TeV, the cross sections of the gluon fusion and W^\pm, Z^0 vector boson fusion processes become comparable, of the order of a fraction of a pb, corresponding to a thousand events containing a Higgs particle.

One should note that in the previous Higgs production processes, it is very important to take into account the quantum corrections of (at least) the strong interaction, that is, when additional gluons are exchanged in loops or emitted in the final states. For instance, in the case of the gluon fusion process, these higher order corrections lead to an increase of the Higgs production cross section by approximately a factor two.

3.3 Detection of the Higgs Boson

Producing the Standard Model Higgs particle at the LHC is thus relatively easy, thanks to the high energy of the collider and its expected luminosity. However,

detecting the Higgs particle in a very complex environment is another story. Indeed, as the protons are not elementary objects and since the colliding beams contain a very large number of protons, in general, there is not one single collision in any event, but several. This is shown in the left-hand side of Fig. 8 where a Higgs event, as seen by one of the LHC detectors, has been simulated and where one can observe that several tracks (corresponding to particles) appear with only a few of them corresponding to the decay products of the Higgs boson.

In addition, the production rates of all other known Standard Model particles, which are viewed as uninteresting background events that one should get rid of, are simply gigantic. This is also illustrated in the left-hand side of Fig. 8, where the cross sections for several Standard Model processes are compared to that of Higgs production in the gluon–gluon fusion process for $M_H = 150$ GeV. For instance, the total cross section for the production of hadrons, that is, light quarks and gluons which are subject to strong interactions, is ten orders of magnitude larger than that of a Higgs boson with a relatively low mass. Even the cross sections for the production of W^\pm and Z^0 bosons are three to four orders of magnitude larger. Detecting the

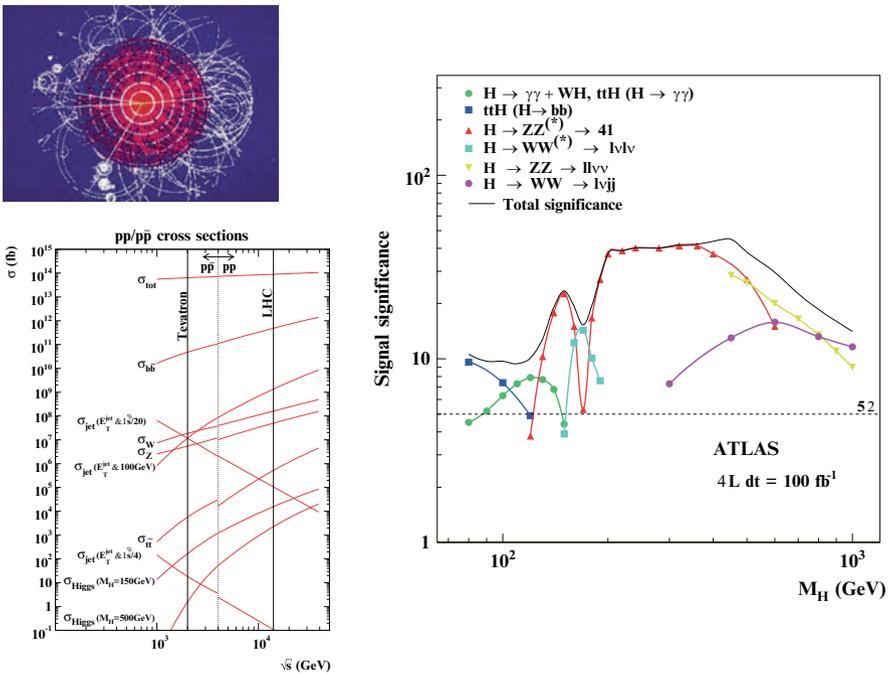


Fig. 8 *Left*: simulation of a Higgs event in a proton–proton collision at the LHC and the cross sections for various production mechanisms (including a Higgs production mechanism for $M_H = 150$ GeV) as a function of the center of mass energy [12]. *Right*: the statistical significance of a signal for a Higgs particle at the ATLAS experiment for various final state Higgs decays [11]

Higgs particle in this hostile environment resembles finding a needle in a haystack; the challenges to be met are simply enormous.

To be able to detect a Higgs particle, one should take advantage in an optimal manner of the kinematical characteristics of the signal events that are, in general, quite different from those of the background events. In addition, one should focus on the decay modes of the Higgs particles (and those of the particles that are produced in association with it such as W^\pm , Z^0 bosons or top quarks) that are easier to extract from the background events. Pure hadronic modes such as Higgs decays into quarks or gluons have to be discarded although much more frequent in most cases.

For a Higgs boson with a mass close to 100 GeV, an interesting signature would be the decay into two very energetic photons, a configuration that is rarely mimicked by the background events. Although this Higgs decay mode has a very small probability to occur, at most a few permille for a light Higgs boson as shown in Fig. 5, the production rates are large enough to compensate and to allow for a significant number of signal events which can be disentangled from the backgrounds.¹⁹

Another interesting signal configuration, valid for Higgs bosons with masses larger than approximately 180 GeV, would be the decay into two Z^0 bosons which then decay into electron–positron or muon–antimuon pairs. This final state with four charged leptons is a rather clean signature (often called the Higgs golden mode) with little background,²⁰ allowing for a relatively easy detection of the Higgs particle up to rather large masses. At higher Higgs masses, when the production cross sections become smaller, this four-charged lepton signature can be supplemented by final states in which one of the Z^0 bosons decays either into neutrinos or quark–antiquark pairs, which occur more frequently and increase the statistics. In addition, the signature involving Higgs decays into W^+W^- pairs with the W^\pm bosons decaying into charged lepton and neutrino pairs (and, for higher Higgs mass, the more frequent one with one of the W^\pm bosons decaying into two quarks) could be used.²¹

A gigantic effort has been made by the experimental collaborations [11], with the precious help of theorists in particle phenomenology, to determine with the highest accuracy the Higgs discovery potential at the LHC in the most important production channels and the experimentally interesting decay modes. Taken into account were all the backgrounds from Standard Model processes and the expected experimental

¹⁹ A characteristic of the signal is that the square of the sum of the two photon four-momenta (called invariant mass) should correspond (as a result of energy–momentum conservation) to the four-momentum squared of the Higgs boson, which is equal to M_H^2 . Therefore, the Higgs signal events “peak at an invariant mass M_H ,” while the background events should have a continuous invariant mass spectrum with no particular peak.

²⁰ Here again, one expects the two Z^0 bosons (reconstructed from their leptonic decays) from the Higgs decays to, “peak at an invariant mass M_H ,” while the background, from direct Z^0 boson pair production, for instance, should have a continuous invariant mass spectrum.

²¹ Here, there is no invariant mass peak as the neutrinos from W^\pm boson decays escape detection and only appear indirectly as missing energy momentum (however, some kinematical distributions have a striking behavior that can be observed). In this case, the signal is a significant excess of events compared to the background; both should therefore be determined with a high confidence.

environment; the performance of the machine and the characteristics of the ATLAS and CMS detectors have been simulated in the most precise way.

The end result is that with the integrated total luminosity expected at the LHC and adding all the production and decay channels, the Higgs particle cannot escape detection in the entire mass range that is allowed theoretically. The right-hand side of Fig. 8 shows the result obtained from a simulation of the ATLAS experiment²² with a luminosity of 100 pb^{-1} for various Higgs decay signatures. As one can see, the statistical significance of the Higgs signal events with respect to the sum of all (irreducible) background events, $\sigma = N_{\text{signal}}/\sqrt{N_{\text{background}}}$, is for all values of M_H , much larger than $\sigma = 5$, the value beyond which one can claim discovery with a very high degree of confidence. The most difficult regions are the low Higgs mass region $M_H \approx 120 \text{ GeV}$, where the main detection channel is the rare Higgs decays into two photons, and the very high mass region, $M_H \approx 1 \text{ TeV}$, where the production cross sections are small; the significance is, however, large enough for discovery.

3.4 Determination of the Higgs Boson Properties

Another goal of the LHC, which is as important as the Higgs discovery itself, would be to determine the fundamental properties of the Higgs particle once it is observed. This would allow a check that the Higgs mechanism is indeed responsible for the spontaneous breaking of the electroweak symmetry and, hence, of the generation of the weak gauge boson and fermion masses.

At the LHC, the Higgs boson mass could be measured in a very accurate way (below the percent level) by exploiting the two very clean decay channels discussed above: the two photon decay in the low mass range and the decay into four charged leptons via Z^0 bosons in the intermediate and high Higgs mass ranges. The latter channel would also allow measurement of the total Higgs decay width for $M_H \gtrsim 200 \text{ GeV}$ when it starts to be experimentally resolvable (for smaller Higgs masses, the total Higgs decay width is too small to be resolved experimentally).

An indication on the spin of the Higgs boson would be provided by the observation of the decay into two photons: because of angular momentum conservation, a spin-1 particle cannot decay into two spin-1 particles. However, this leaves the possibility for a higher (even and integer) spin such as, for instance, a spin-2 particle as is the case for the graviton, the hypothetical messenger of the gravitational interaction. A more unambiguous way to determine the Higgs spin would be to take advantage of the four charged lepton Higgs decay mode and to observe some correlations between the angles formed by two of the final state leptons, which exhibit a characteristic signature of an initially decaying spin-zero particle.

²² Similar results are obtained by the CMS experiment and in practice, one should combine the results of the two experiments to increase the statistics and thus the significance of the signal.

The determination of the Higgs couplings to gauge bosons and fermions is possible at the LHC through the measurement of the cross sections times the branching ratios, given by the event rate in the various search channels. However, the accuracy in this determination is rather limited because of the small statistics that one obtains after applying the cuts that suppress the large backgrounds which are often plagued with uncertainties, and the various systematical errors such as the common uncertainty in the absolute luminosity. In addition, when one attempts to interpret the measurements, theoretical uncertainties from the limited precision on the quark/gluon densities in the proton and from the higher-order corrections should be taken into account. Furthermore, the couplings that can be measured will critically depend on the Higgs boson mass. For instance, in the mass range above $M_H \sim 2M_W$, only the couplings to gauge bosons can be accessed directly and the $H\bar{t}t$ coupling can be probed indirectly in the gluon fusion mechanism.

Nevertheless, and as shown in the left-hand side of Fig. 9, a statistical precision of the order of 10 to 30% can be achieved for some ratios of partial widths, which are proportional to the Higgs couplings squared. Under some theoretical assumptions, these measurements can be translated into absolute Higgs partial widths in the various decay channels and hence, into the square of the Higgs couplings to gauge bosons and fermions (in fact, mainly top quarks), as shown in the right-hand side of Fig. 9. The accuracies deteriorate when the systematical errors are added.

A precise measurement of the trilinear Higgs self-coupling, which is the first nontrivial probe of the Higgs potential and, probably, the most decisive test of the Higgs mechanism, is unfortunately not possible at the LHC. Indeed, to probe this coupling, one needs to consider processes in which two Higgs particles are produced, the leading one being double Higgs production in the gluon–gluon fusion

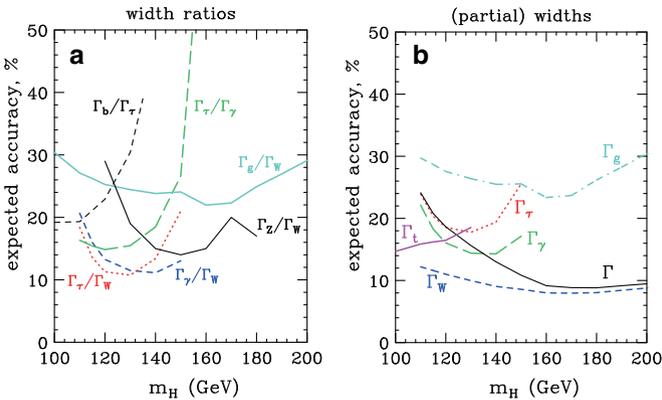


Fig. 9 *Left*: relative accuracy expected at the LHC with a luminosity of 200 fb^{-1} for the measurement of various ratios of Higgs boson partial widths that are proportional to the square of the Higgs couplings to the particles. *Right*: the relative accuracy expected in the indirect determination of the partial and total widths Γ_i and Γ with some theoretical assumptions. Only the statistical errors have been included and no detector simulation has been made. From Ref. [13]

mechanism: a virtual Higgs particle is produced in the usual gluon–gluon to Higgs mechanism and then splits into two real Higgs particles. The cross sections for such a process are rather tiny even for low values of M_H and the corresponding backgrounds are very large. Furthermore, the contribution of the diagram with the Higgs self-coupling is diluted by another possibility for the same final state: the radiation of both Higgs particles from the internal heavy top quarks.

Thus, the measurement of some Higgs couplings (such as the couplings to W^\pm , Z^0 and eventually top quarks) can be performed at the LHC only at the ten percent level at most, while some couplings are not accessible (such as the couplings to bottom and charm quarks, τ leptons and muons, photons, as well as the Higgs self-coupling that is essential to reconstruct the scalar potential responsible for the breaking of electroweak symmetry) and will have to await for the successor of the LHC. The latter will be, ideally, an electron–positron collider with a center of mass energy ranging from 300 GeV to 1 TeV, a very high integrated luminosity and a clear environment which would allow for high precision tests of the Higgs properties. An international project for such a machine, the International Linear Collider, is under way and involves the major laboratories in high-energy physics [14].

4 The Higgs Beyond the Standard Model

Despite of its success in describing all data available today, the Standard Model is far from being considered to be perfect in many respects. Indeed, it does not explain the proliferation of fermions (why three fermion families?) and the large hierarchy in their mass spectra (in particular, it does not say much about the observed small masses for the neutrinos which are assumed to be massless) and does not unify in a satisfactory way the electromagnetic, weak, and strong forces (as one has three different symmetry groups with three different coupling constants which, in addition, almost fail to meet at a common value during their evolution with the energy scale) and ignores the fourth force, the gravitational interaction. Furthermore, it does not contain a massive, electrically neutral, weakly interacting, and absolutely stable particle that would account for dark matter which is expected to represent 25% of the energy content of the Universe and fails to explain the baryon asymmetry in the Universe: why there are (by far) more particles than antiparticles.

However, the main problem that makes particle physicists believe that the Standard Model is simply an effective theory, valid only at the energy scales that have been explored so far, that is, much below 1 TeV, and should be replaced by a more fundamental theory at the TeV scale, is related to the particular status of the Higgs boson. Indeed, contrary to fermions and gauge bosons, the Higgs particle has a mass that cannot be protected against quantum corrections (i.e., when the Higgs boson emits and reabsorbs virtual particles). These corrections tend to drive the Higgs mass to very large values, of the order of the scale of the underlying new physics (which serves as a cutoff) or the Planck scale (the ultimate scale), while we need it to be close to the 100 GeV range. Thus, the Standard Model cannot be extrapolated

up to energies higher than the TeV scale where some new physics should emerge. This is the main reason which makes that particle physicists expect that something new, in addition to the Higgs particle, should manifest itself at the LHC.

Among the many possibilities for this new physics beyond the Standard Model, the option that emerges in the most natural way is Supersymmetry. Supersymmetry combines internal gauge symmetries with space–time symmetries and relates fermions and bosons: to each particle, it predicts the existence of a super-partner (and thus, at least doubles the Standard Model particle spectrum) which should have the same properties but with a spin different by a unit $1/2$ and also a different mass as Supersymmetry must be broken in Nature. The lightest of these new particles is the ideal candidate for dark matter in the Universe. Supersymmetry protects the Higgs mass from acquiring large values as the dominant quantum corrections from standard particles are exactly compensated by the contributions of their supersymmetric partners.²³ These new particles should not be heavier than 1 TeV so as not to spoil this compensation and, thus, they should be produced at the LHC.

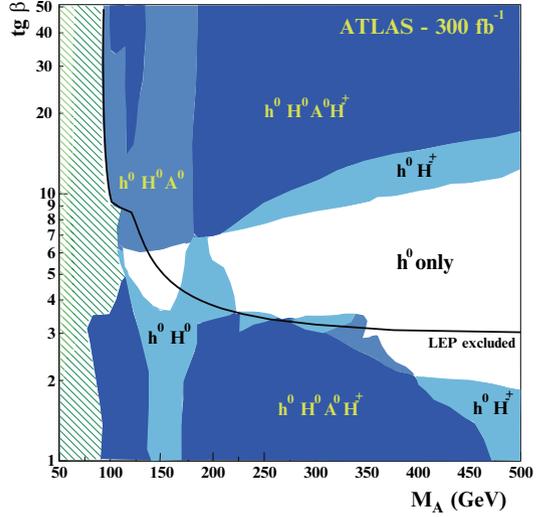
In the minimal supersymmetric extension of the Standard Model, two doublets of complex scalar fields are necessary to break the electroweak symmetry and to give masses to gauge bosons and (separately) to isospin up and down fermions. This leads to an extended Higgs sector compared to the Standard Model: rather than one, one would have five Higgs particles, three neutral ones (noted h^0 , H^0 , and A^0), and two charged ones (noted H^\pm); for a detailed review, see Ref. [15]. The lightest neutral Higgs particle h^0 has, in general, the same properties as the standard Higgs boson but, by virtue of Supersymmetry, a mass that is below 140 GeV. At least this particle should be produced at the LHC; the other Higgs bosons, could also be detected if they are not too heavy and their couplings to fermions not too tiny. This is illustrated in Fig. 10, in which is shown the number of Higgs particles of this minimal supersymmetric model, which can be observed by the ATLAS experiment at the LHC (with a luminosity of 300 fb^{-1}) in the plane formed by the two free parameters (to first approximation) of the model: M_A , the mass of one of the neutral Higgs bosons and $\tan\beta$, the ratio of the nonzero vacuum expectation values of the two Higgs fields.

Other extensions of the Standard Model, such as non-minimal supersymmetric theories for instance, predict an even richer Higgs spectrum.²⁴ In contrast, some

²³ In addition, the contribution of the supersymmetric particles to the energy evolution of the gauge coupling constants means that the latter can indeed meet at a single point at a scale slightly below the Planck scale; thus, the three interactions can be unified into one single interaction with one coupling constant and hence, one symmetry group. Note also that Supersymmetry has many other theoretical virtues: it is the first nontrivial extension of the Poincaré group in quantum field theory, which, when made local, necessarily includes Einstein's theory of gravity, and it appears naturally in Superstring theories in which the elementary particles we observe are the excitation modes of elementary strings with Planck length, $\approx 10^{-33}$ cm. These features may help to reach the ultimate goal of particle physics: the unification of all fundamental forces including gravity.

²⁴ For consistency reasons and to cope with experimental data, only singlets and (an even number of) Higgs doublets can be added. In the most general supersymmetric extension with an arbitrary Higgs content, the lightest Higgs should have a mass below 200 GeV and be observed at the LHC.

Fig. 10 The number of Higgs bosons of the minimal supersymmetric model that can be produced at the LHC in the ATLAS experiment with a luminosity of 300 fb^{-1} in the $(M_A, \tan\beta)$ plane



new physics models such as theories with extra space–time dimensions²⁵ or models inspired from the strong interactions but at the TeV scale, do not incorporate any Higgs particle in their spectrum. However, to preserve the unitarity of the theory, a new ingredient should appear in the scattering of the massive W^\pm and Z^0 bosons, and its effects should be measurable at the LHC. Thus, even if no Higgs particle is detected at the LHC some new phenomenon should be observed.

5 Conclusions

Thus, several scenarios for the generation of the elementary particle masses are possible. In addition to the one of the Standard Model with only one Higgs particle (and which has been discussed in some detail), there are scenarios with an extended and richer Higgs sector as in supersymmetric theories and scenarios with no Higgs boson at all as in some versions of extra space–time dimensional models. To the question: which option Nature has chosen? the LHC will soon provide an answer.

²⁵ For instance, if there is an extra-dimensional space where only gravitons can propagate, the weakness of the gravity interaction can be attributed to the existence of large extra space dimensions. In this scenario, the four-dimensional Planck mass is a fictitious mass scale, and the fundamental gravity mass scale in the higher dimension could be close to the TeV scale which then technically solves the hierarchy problem of the Standard Model and brings gravity into the game. In some models, the symmetry breaking is triggered by specific boundary conditions for the gauge fields in the compactification of the extra space dimensions, giving rise to Higgsless models.

References

1. For a compilation of the latest experimental results, see: Particle Data Group, C. Amsler et al., Phys. Lett. B **667**, 1 (2008)
2. The original references for the Standard Model of the electroweak interactions are: S. Glashow, Nucl. Phys. **22**, 579 (1961); S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967); A. Salam, in *Elementary Particle Theory*, ed. by N. Svartholm (Almqvist and Wiksells, Stockholm, 1969), p. 367; G. 't Hooft, M. Veltman, Nucl. Phys. B **44**, 189 (1972)
3. The original references for the strong interaction part of the Standard Model are: M. Gell-Mann, Phys. Lett. **8**, 214 (1964); G. Zweig, CERN-Report 8182/TH401 (1964); H. Fritzsch, M. Gell-Mann, H. Leutwyler, Phys. Lett. B **47**, 365 (1973); D. Gross, F. Wilczek, Phys. Rev. Lett. **30**, 1343 (1973); H.D. Politzer, Phys. Rev. Lett. **30**, 1346 (1973)
4. P. Dirac, Proc. R. Soc. Lond. A **114**, 243 (1927); P. Jordan, W. Pauli, Z. Phys. **47**, 151 (1928); W. Heisenberg, W. Pauli, Z. Phys. **56**, 1 (1929); S. Tomonaga, Progr. Theor. Phys. **1**, 27 (1946); J. Schwinger, Phys. Rev. **73**, 416 (1948); R. Feynman, Phys. Rev. **76**, 749 (1949)
5. The LEP Collaborations ALEPH, DELPHI, L3, OPAL, and the LEP Electroweak Working Group, arXiv:hep-ex/0612034v2 (2007), <http://lepewwg.web.cern.ch/LEPEWWG/>
6. The original references for the Higgs mechanism are: P.W. Higgs, Phys. Rev. Lett. **12**, 132 (1964); F. Englert, R. Brout, Phys. Rev. Lett. **13**, 321 (1964); G. Guralnik, C. Hagen, T. Kibble, Phys. Rev. Lett. **13**, 585 (1964)
7. For a detailed review, see: A. Djouadi, Phys. Rep. **457**, 1 (2008)
8. ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, the LEP Working Group for Higgs Boson Searches, Phys. Lett. B **565**, 61 (2003); Tevatron New Phenomena, Higgs working group, for the CDF and DZero collaborations, arXiv:0903.4001v1 [hep-ex] (2009)
9. T. Hambye, K. Riessellmann, Phys. Rev. D **55**, 7255 (1997)
10. A. Djouadi, J. Kalinowski, M. Spira, Comput. Phys. Commun. **108**, 56 (1998)
11. Detailed discussions can be found in: the Technical Design Reports of the CMS Collaboration (CERN/LHCC/2006-021, June 2006), ATLAS Collaboration, arXiv:0901.0512v4 [hep-ex] (2009). For a summary of experimental Higgs searches, see for instance: J.G. Branson, D. Denegri, I. Hinchliffe, F. Gianotti, F.E. Paige, P. Spiccas (CMS and ATLAS Coll.), Eur. Phys. J. **Direct C4**, N1 (2002)
12. The LHC/ILC Study Group; G. Weiglein et al., Phys. Rep. **426**, 47 (2006)
13. M. Dührssen, S. Heinemeyer, H. Logan, D. Rainwater, G. Weiglein, D. Zeppenfeld, Phys. Rev. D **70**, 113009 (2004)
14. A. Djouadi, J. Lykken, K. Mönig, Y. Okada, M. Oreglia, S. Yamashita, the ILC Collaboration, *Physics at the ILC*, vol. 2 of the *International Linear Collider Reference Design Report*, arXiv:0709.1893v1 [hep-ph] (2009)
15. A. Djouadi, Phys. Rep. **459**, 1

Testing Basic Laws of Gravitation – Are Our Postulates on Dynamics and Gravitation Supported by Experimental Evidence?

Claus Lämmerzahl

Abstract Gravity is the most fundamental interaction; it not only describes a particular interaction between matter, but also encompasses issues such as the notion of space and time, the role of the observer, and the relativistic measurement process. Gravity is geometry and, in consequence, allows the existence of horizons and black holes, nontrivial topologies, a cosmological big bang, time-travel, warp drive, and other phenomena unknown in nonrelativistic physics. Here we present the experimental basis of General Relativity, addressing its foundations encoded in the Einstein Equivalence Principle and its predictions in the weak and strong gravity regimes. We discuss several approaches in the search to reveal an influence of the much sought-after quantum theory of gravity. We emphasize assumptions underlying the dynamics – for example, Newton’s axioms and conservation laws – and the current extent to which they are supported by experiment. We discuss conditions under which gravity can be transformed away locally, and examine higher order time derivatives in the equations of motion.

1 Introduction – Why Gravity Is So Exceptional

Gravity is the most fundamental interaction in physics: it is not only a very particular interaction between particles, but also it is related to the notion of space and time, the description of the observer, and the relativistic measurement process. Thus, *any issue related to gravity is also of concern for the description of all other interactions.*

Even by itself, General Relativity (GR), the relativistic theory of gravity, is highly interesting. Since GR is related to the space–time geometry, the gravitational interaction modifies the structure of space–time and leads to surprising phenomena, such as black holes. It is remarkable that we have a theory capable of predicting that a

C. Lämmerzahl (✉)

ZARM, University of Bremen, Am Fallturm, 28359 Bremen, Germany

e-mail: laemmerzahl@zarm.uni-bremen.de

region of space–time can “disappear” and no longer be accessible to the observer. Other unexpected effects, like lensing or cosmological implications such as the big bang, have had a big impact on science, and even on the philosophy of science; in particular, they have attracted very much the attention of the general public.

It is fascinating to follow the present observational exploration of black holes, for example, in the center of our Milky Way [146]. In parallel, there are mathematical studies of known black hole solutions of GR, and the search for new solutions of the Einstein field equations, such as the solution for a disk of dust [127]. There are also numerical studies of the merging of binary black holes which, when spinning, may exhibit an unexpected acceleration [56].

GR in general, and solutions with black holes in particular, have lead to very beautiful, highly interesting, and exceedingly stimulating mathematics studies. In particular, these studies include questions about the geometry and the topology of black holes and our universe. These issues have stimulated a veritable laboratory for *gedanken* experiments, which have lead to consideration of the information paradox [68], time travel [123] (for a recent discussion see, e.g., [88]), warp drive [5] (for a more recent discussion see, e.g., [89]), etc.

In recent years, increasing effort has been spent on developing a quantum theory of gravity. A large number of people have attempted to develop a unification of quantum theory along the lines of string theory [111], loop quantum gravity [81, 140] or noncommutative geometry [126] (see also references therein). While string theory lays emphasis on the particle content of our physical world and neglects somewhat the geometrical nature of gravity, loop quantum gravity starts from gravity as space–time geometry and neglects the particle content. Within string theory, higher dimensional theories are experiencing a renaissance and, for example, black holes display even more unusual features than are known from four dimensions [51, 82].

Gravity is one area in physics where something new is expected, which will undoubtedly lead to another revolution in the physics paradigm. Very unusual effects are expected to arise in quantum gravity and there are both theoretical and experimental efforts under way in the search for the new phenomena it should entail. Until now all experiments are in agreement with standard GR. However, substantial efforts are being made to find experimental signatures of quantum gravity. Any experimental result in this direction will guide the development of the theory itself. New experiments have been designed and new technologies have been developed to improve available accuracy in the search for possible quantum gravity effects. It is speculated that perhaps the LHC has the potential to see related phenomena.

Since gravity is such a fundamental interaction – it covers the notion of space–time, the space–time geometry, the observer, the measurement process, etc. – it is clear that thinking about gravity and questioning its underlying principles can open up many unusual possibilities that should be tested by experiment. These range from questioning Newton’s axioms, conservation laws, the time dependence of constants, etc. One may also speculate whether under extreme situations, like extremely weak gravity, small accelerations, large accelerations, highest energies, ultralow temperatures etc., some of the principles underlying today’s physics lose their meaning.

Similarly exciting is quantum theory. The experimental realization of the strange behavior of quantum systems is always truly astonishing, as Bohr said: “If quantum mechanics hasn’t profoundly shocked you, you haven’t understood it yet.” However, since quantum theory is based on a scheme that is not directly related to experiments, that is, there is no real operational approach to quantum theory, it is much more difficult to systematically question various assumptions underlying quantum theory. For a survey of experiments testing quantum theory see [102].

In this chapter, we first describe the remarkable features of GR and then present its experimental basis. This basis consists in the principles underlying the fact that today gravity is described by a metric tensor representing the space–time geometry. This metric theory then predicts certain effects which, for Einstein’s GR, acquire particular values. Then we give reasons why it is important to improve these experiments and to perform new ones, and we also present a strategy for such new tests, where emphasis is placed on tests of gravity and relativity in extreme situations. Finally we focus on unusual questions related to possible effects rarely discussed in the literature, like tests of Newton’s axioms and of conservation laws, etc. In fact, all tests of gravity can be regarded as searches for “new physics”. This is a considerably enlarged version of an earlier article [92].

2 Key Features of Gravity

Gravity is singled out and characterized by a set of universality principles that are shared by no other interaction.

1. Universal presence of gravity
 - Gravity is *everywhere*
 - Gravity *always can be transformed away* locally
2. Universal action on masses
 - Gravity acts on *all bodies*
 - Gravity acts on all bodies *in the same way*
3. Universal action on clocks
 - Gravity acts on *all clocks*
 - Gravity acts on all clocks *in the same way*
4. Universal creation of gravitational field
 - *Each mass* creates a gravitational field
 - Each mass creates a gravitational field *in the same way*

The last of these features means that all, say, spherically symmetric masses of the *same weight* create the same gravitational field. That means that a measurement of a gravitational field of a spherically symmetric body only gives the mass of the gravitating body and not its composition.

3 Standard Tests of the Foundations of Special and General Relativity

The basic structure of GR, and of all other physics, is encoded in the Einstein Equivalence Principle (EEP). This principle states that (i) if all nongravitational interactions are switched off, all pointlike particles move in a gravitational field in the same way, (ii) all nongravitational clocks¹ are influenced by the gravitational field in the same way, and (iii) locally, Special Relativity is valid, in that all physical laws are Lorentz covariant.

These principles are so important because they imply the following:

- The gravitational interaction is described by means of a metrical tensor. The mathematical frame for that is a Riemannian geometry.
- The equations of motion for a point particle, for a spin- $\frac{1}{2}$ -particle, of the electromagnetic field, etc. have to be the geodesic equation, the Dirac equation, the Maxwell equations in Riemannian space-times with a certain space-time metric.
- All these Riemannian metrics have to be the same.

Owing to their importance it is clear that these principles have to be confirmed with the highest possible accuracy. We describe appropriate experiments below.

3.1 Tests of Special Relativity

Lorentz invariance, the symmetry of SR which also holds locally in GR, is based on the constancy of the speed of light and the relativity principle. For recent reviews see, e.g., [9, 116].

3.1.1 The Constancy of the Speed of Light

The constancy of the speed of light has many aspects:

1. The speed of light should not depend on the velocity of the source. Otherwise, it would be possible to measure in one space-time event in one direction two light rays with different velocities. This independence from the velocity of the source has been confirmed in various experiments in the laboratory as well as by astrophysical observations. If the velocity of light depends on the velocity of the source, then this can be written as $c' = c + \kappa v$, where v is the velocity of the source (in some frame) and κ some parameter. Within this model, it is possible that the light of a star in a binary system may overtake light that was emitted earlier. Such a reversal of the chronological order has never been observed, allowing the estimate $\kappa \leq 10^{-11}$ [27]. Laboratory experiments performed at CERN used protons hitting a Beryllium target to create π^0 mesons with a velocity of

¹ Pendula and hourglasses are not allowed.

$v = 0.99975 c$. These moving mesons decay into photons whose velocity has been measured and compared with the velocity of photons emitted from a source at rest. No difference in the speed of the photons was found giving [7] $\kappa \leq 10^{-6}$ though, from a nonrelativistic point of view, one would expect almost $2c$. The constancy of c for photons appears to hold for all velocities of the source.

2. The speed of light does not depend on frequency or polarization. The best results for this are from astrophysics. From radiation at frequencies $7.1 \cdot 10^{18}$ Hz and $4.8 \cdot 10^{19}$ Hz of Gamma Ray Burst GRB930229 one obtains $\Delta c/c \leq 6.3 \cdot 10^{-21}$ [143]. In a theoretical model of a hypothetical photon rest mass the best restriction is $m_\gamma \leq 10^{-47}$ kg from radiation from GRB980703 [143]. Analysis of the polarization of light from distant galaxies yields an estimate of $\Delta c/c \leq 10^{-32}$ [84].
3. The speed of light is universal. This means that the velocity of all other massless particles, as well as the limiting maximum velocity of all massive particles, coincides with c . The maximum speed of electrons, neutrinos, and muons in vacuum has been shown in various laboratory experiments to coincide with the velocity of light at a level $|c - c_{\text{particle}}|/c \leq 10^{-6}$ [6,29,58,80]. Astrophysical observations of radiation from the supernova SN1987A yield an estimate for the comparison of photons and neutrinos, which is two orders of magnitude better [109,157].
4. The speed of light does not depend on the velocity of the laboratory. This can be tested, for example, by comparing the frequency of an optical resonator that depends on the speed of light and the frequencies of an atomic clock, in a modern version of the corresponding Kennedy–Thorndike experiment. The best estimate today yields $\Delta c/c \leq 10^{-16}$ [71].
5. The speed of light does not depend on the direction of propagation. This isotropy of the speed of light has been confirmed, by modern Michelson–Morley experiments using optical resonators, to a relative accuracy of $\Delta c/c \leq 10^{-17}$ [71].

These results mean that the speed of light is universal, so it can be interpreted as part of the space–time geometry. The implied causal structure is an essential part of the operational description of space–time proposed by Ehlers, Pirani, and Schild [50].

3.1.2 The Relativity Principle

The relativity principle states that the outcome of all experiments when performed identically within a laboratory, that is, without reference to the external world, is independent of the orientation and the velocity of the laboratory. This applies to the photon sector as well as to the matter sector. For the photon sector this can be tested with the Michelson–Morley and Kennedy–Thorndike type experiments already discussed above.

Regarding the matter sector, the corresponding tests are Hughes–Drever type experiments. In general, these are nuclear or electronic spectroscopy experiments. Such effects can be modeled by an anomalous inertial mass tensor [67] of the corresponding particle. For nuclei, one then gets estimates of the order

$\delta m/m \leq 10^{-30}$ [35, 103, 135]. Modeling with an anisotropic speed of light, as in the $TH\epsilon\mu$ -formalism [168], yields $\Delta c/c \leq 10^{-21}$. In addition to the possibility of an anisotropic mass tensor, there is also the possibility of an anomalous coupling of the spin to some given cosmological vector or tensor fields, which would destroy Lorentz invariance. Recent tests have given no evidence for any anomalous spin coupling either to the neutron [19, 20], to the proton [74], or to the electron [69, 72]. All anomalous spin couplings are absent to the order of 10^{-31} GeV (see also [165] for a review). Similarly, higher order derivatives in the Dirac and Maxwell equations generally lead to anisotropy effects [110].

A further consideration is that there could be intrinsic anisotropies in the Coulomb or Newtonian potentials [83, 84]. Anisotropies in the Coulomb potential should affect the lengths of optical cavities which, in turn, might influence the frequency of light in the cavity. It has been shown that the influence of the anisotropies of the Coulomb potential are smaller than the corresponding anisotropies in the velocity of light [124]. Anisotropies in the Newtonian potential of the Earth have recently been searched for using atomic interferometry [125], which has constrained the anisotropies at the 10^{-8} level.

Future spectroscopy of anti-hydrogen may yield further information about the validity of the PCT symmetry.

3.1.3 The Consequence

The consequence of the above experiments is that within the accuracy given by these experiments, vectors transform with the Lorentz-transformations. The best adapted mathematical framework thus introduces a four-dimensional space-time, which, locally, is equipped with a Minkowski metric $\eta_{ab} = \text{diag}(+1, -1, -1, -1)$. More can be found in standard textbooks; for example, see [139, 149].

3.2 Tests of the Universality of Free Fall

The Universality of Free Fall (UFF) states that all neutral point-like particles move in a gravitational field in the same way, that is, that the path of these bodies is independent of the composition of the body. The corresponding tests are described in terms of the acceleration of these particles in the reference frame of the gravitating body: the Eötvös factor compares the normalized accelerations of two bodies $\eta = (a_2 - a_1)/[\frac{1}{2}(a_2 + a_1)]$ in the same gravitational field. In the frame of Newton's theory this can be expressed as $\eta = (\mu_2 - \mu_1)/[\frac{1}{2}(\mu_2 + \mu_1)]$, where $\mu = m_g/m_i$ is the ratio of the gravitational to inertial mass. Though there are no point particles, it is possible experimentally to manufacture macroscopic bodies such that their higher gravitational multipoles are either very small or very well controlled. In other cases, a numerical integration yields the effective gravitational force on the extended body. Both these methods are used in the various tests of the UFF.

There are two principal schemes in which to perform tests of the UFF. The first scheme uses the free fall of bodies. In this case the full gravitational attraction toward the Earth can be exploited. However, these experiments suffer from the fact that the time-of-flight is limited to roughly 1 s and that a repetition needs new adjustment. The other scheme uses a restricted motion confined to one dimension only, namely a pendulum or a torsion balance. The big advantage is the periodicity of the motion, which by far beats the disadvantage that only a fraction of the gravitational attraction is used. In fact, the best test today of the UFF uses a torsion pendulum and confirms it at the level of $2 \cdot 10^{-13}$ [145]. Newly proposed tests in space, the approved mission MICROSCOPE [160], and the proposal STEP [108] will combine the full advantages of free fall and periodicity.

Quantum gravity inspired scenarios hint that the UFF might be violated below the 10^{-13} level [39, 40]. From cosmology with a dynamical vacuum energy (quintessence), a violation at the 10^{-14} level can also be derived [167]. If the validity of the UFF holds, we can impose bounds on the time variability of various constants, such as the fine structure constant and the electron-to-proton mass ratio [42].

According to GR, spinning particles couple to the space–time curvature [15, 70] and, thus, violate the UFF. However, the effect is far beyond any current experimental reach. Testing the UFF for spinning matter amounts to a search for an anomalous coupling of spin to gravity. Motivation for anomalous spin couplings came from the search for the axion, a candidate for the dark matter in the universe initially introduced to resolve the strong PC puzzle in QCD [122]. In these models, spin may couple to the gradient of the gravitational potential or to gravitational fields generated by the spin of the gravitating body. Tests of the first case by weighing polarized bodies show that, for polarized matter, the UFF is valid at a level of order 10^{-8} [73].

Charged particles, too, must couple to space–time curvature [44], again at a level that is too small to be detectable. It is possible to introduce a charge-dependent violation of the UFF by proposing a charge-dependent anomalous inertial and/or gravitational mass. It is also possible to choose the model such that, for a neutral atom, the UFF is fulfilled exactly while it is violated for isolated charges [45]. It has been suggested that a corresponding experiment be carried out in space [45].

3.3 Tests of the Universality of the Gravitational Redshift

A test of the universal influence of the gravitational field on clocks based on different physical principles requires clock comparison during their common transport through different gravitational potentials. There is a large variety of clocks that can be compared:

1. Light clocks (optical resonators)
2. Atomic clocks based on
 - (a) Hyperfine transitions
 - (b) Fine structure transitions
 - (c) Principal transitions

3. Molecular clocks based on
 - (a) Rotational transitions
 - (b) Vibrational transitions
4. Gravitational clocks based on revolution of planets or binary systems
5. The rotation of the Earth
6. Pulsar clocks based on the spin of stars
7. Clocks based on particle decay

At a phenomenological level, the comparison of two collocated clocks is given by

$$\frac{\nu_{\text{clock 1}}(x_1)}{\nu_{\text{clock 2}}(x_1)} = \left(1 - (\alpha_{\text{clock 2}} - \alpha_{\text{clock 1}}) \frac{U(x_1) - U(x_0)}{c^2} \right) \frac{\nu_{\text{clock 1}}(x_0)}{\nu_{\text{clock 2}}(x_0)} \quad (1)$$

where $\alpha_{\text{clock } i}$ are phenomenologically given clock-dependent parameters, U is the Newtonian potential, and x_0 and x_1 are two positions. If this frequency ratio does not depend on the gravitational potential then the gravitational redshift is universal. This is a null-test of the quantity $\alpha_{\text{clock 2}} - \alpha_{\text{clock 1}}$. It is obviously preferable to employ a large difference in the gravitational potential, which clearly shows the need for space experiments. In experiments today, the variation of the gravitational field is induced by the motion of the Earth around the Sun and thus requires that the clocks used have very good long-term stability.

The best test to date has been performed by comparing the frequency ratio of the 282 nm $^{199}\text{Hg}^+$ optical clock transition to the ground state hyperfine splitting in ^{133}Cs over 6 years. The result is $|\alpha_{\text{Hg}} - \alpha_{\text{Cs}}| \leq 5 \cdot 10^{-6}$ [14, 54]. Other tests compare Cs clocks with the hydrogen maser, and Cs or electronic transitions in I_2 with optical resonators. We are looking forward to ultrastable clocks on the ISS and on satellites in Earth orbit, or even in deep space as proposed by SPACETIME, OPTIS, and SAGAS [94, 113, 169], which should considerably improve the quality of the scientific results.

So far there are no tests using anti-clocks, that is, clocks made of antimatter. However, since the production of anti-hydrogen is a well established technique today, attempts to perform high-precision spectroscopy of anti-hydrogen have been proposed. These measurements should first test special relativistic CPT invariance but, as a long-term measurement, could also be used to test the Universality of the Gravitational Redshift for a clock based on anti-hydrogen.

3.4 The Consequence

A consequence of the validity of the EEP is that gravity can be described by a Riemannian metric, $g_{\mu\nu}$, a symmetric second rank tensor defined on a differentiable manifold that is identified as the collection of all possible physical events. The purpose of this metric is twofold: First, it governs the rate of clocks, that is,

$$s = \int ds, \quad ds = \sqrt{g_{\mu\nu} dx^\mu dx^\nu} \quad (2)$$

is the time shown by clocks where the integration is along the world-line of those clocks. Second, the metric gives the equation of motion for massive point particles as well as for light rays,

$$0 = D_\nu v \quad \Leftrightarrow \quad 0 = \frac{d^2 x^\mu}{ds^2} + \left\{ \begin{matrix} \mu \\ \rho\sigma \end{matrix} \right\} \frac{dx^\rho}{ds} \frac{dx^\sigma}{ds} \quad (3)$$

where D_ν is the covariant derivative along ν and

$$\left\{ \begin{matrix} \mu \\ \rho\sigma \end{matrix} \right\} = \frac{1}{2} g^{\mu\nu} (\partial_\rho g_{\nu\sigma} + \partial_\sigma g_{\nu\rho} - \partial_\nu g_{\rho\sigma})$$

is the Christoffel symbol. Here $x = x(s)$ is the world-line of the particle parametrized by its proper time and $v = dx/ds$ the tangent vector along this world-line. While $g(v, v) = 1$ for particles, we have $g(v, v) = 0$ for light, so that we must use some affine parameter to parametrize the world-line of a light ray. More on that can be found in many textbooks on gravity; see, for example, [66, 121, 166]. It can be shown that this notion also describes the propagation of, for example, the spin vector, $D_\nu S = 0$, where S is a particle spin. (This is valid at first order in the spin vector; in the case of spin–spin interactions as they appear for spinning binary systems, terms of $\mathcal{O}(S^2)$ have to be added, see, e.g., [53].) In generalized theories of gravity there might be additional terms in the equations of motion for v and for S .

For a general, static, spherically symmetric space–time metric, which we take to have the form:

$$ds^2 = g_{tt} dt^2 - g_{rr} dr^2 - r^2 (d\vartheta^2 + \sin^2 \vartheta d\varphi^2), \quad (4)$$

we obtain an effective equation of motion

$$\frac{1}{2} \left(\frac{dr}{ds} \right)^2 = \frac{1}{2} \left(\frac{E^2}{g_{tt} g_{rr}} - \frac{1}{g_{rr}} \left(1 + \frac{L^2}{r^2} \right) \right), \quad (5)$$

where E and L are the conserved (specific) energy and angular momentum, respectively. In the case of asymptotic flatness it is possible to uniquely define an effective potential [79]

$$U_{\text{eff}} = \frac{1}{2} \left(E^2 - 1 - \frac{E^2}{g_{tt} g_{rr}} + \frac{1}{g_{rr}} \left(1 + \frac{L^2}{r^2} \right) \right), \quad (6)$$

which completely governs the motion of the particle.

In order to solve the equations of motion one has to know the metric. The metric is given by independent field equations

$$G_{\mu\nu} = \kappa T_{\mu\nu}, \quad (7)$$

where G is a prescribed differential operator acting on the metric and T is the energy–momentum tensor of the matter creating the gravitational field.

4 Tests of Predictions

Gravity can be explored only through its action on test particles (or test fields). Accordingly, the gravitational interaction has been studied through the motion of stars, planets, satellites, and light. There are only very few experiments that demonstrate the effects of gravity on quantum fields.

Any metric theory of gravity leads to effects like the gravitational redshift, the deflection of light, the perihelion precession, the Lense–Thirring effect, the Schiff effect, etc. GR is singled out through certain values for these effects. In the case of weak gravitational fields, such as occur in the Solar system, and of asymptotic flatness, any deviation of a gravitational theory from GR can be parametrized by a few constants, namely the PPN parameters [168]. Many astrophysical observations and space experiments that probe fundamental physics are designed to make precise measurement of these effects and, thus, to better ascertain the PPN parameters.

For Einsteins GR we have, in the left hand side of the field Eq. 7,

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}, \quad (8)$$

where $R_{\mu\nu}$ and R are the Ricci tensor and Ricci scalar, respectively. For a spherically symmetric gravitating body we obtain the Schwarzschild metric

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = \left(1 - \frac{2M}{r}\right) dt^2 - \frac{1}{1 - \frac{2M}{r}} dr^2 - r^2 d\vartheta^2 - r^2 \sin^2 \vartheta d\varphi^2. \quad (9)$$

Use of this metric in the equation of motion yields an ordinary differential equation

$$\left(\frac{dr}{d\varphi}\right)^2 = \frac{r^4}{L^2} \left(E^2 - \varepsilon + \varepsilon \frac{2M}{r} - \frac{L^2}{r^2} + 2\frac{ML^2}{r^3}\right), \quad (10)$$

($\varepsilon = 1$ for massive particle, $\varepsilon = 0$ for light), which can be solved in terms of the Weierstrass \wp -function [65]

$$r(\varphi) = \frac{2M}{\wp(\frac{1}{2}\varphi; g_2, g_3) + \frac{1}{3}}, \quad (11)$$

where the Weierstrass invariants given by

$$g_2 = 4 \left(\frac{1}{3} - \varepsilon \left(\frac{2M}{L} \right)^2 \right) \quad (12)$$

$$g_3 = 4 \left(\frac{2}{27} + \frac{2}{3}\varepsilon \left(\frac{2M}{L} \right)^2 - E^2 \left(\frac{2M}{L} \right)^2 \right) \quad (13)$$

depend on M , E , and L . This solution can be used to calculate most of the Solar system effects.

The Kerr metric is a vacuum solution of Einstein’s field equation that describes a rotating black hole. This metric contains the product of $d\varphi dt$, which appears also in the metric of a rotating observer in Minkowski space–time. The gravitational field of a rotating star is not given by the Kerr solution but, for weak fields, the Kerr solution is a very good approximation to the solution for a rotating star (for which no exact solution exists) so one can, in practice, use the Kerr solution when describing effects related to the addition of rotation. In a weak field limit, the rotation of a star adds to the Schwarzschild metric (9) a term proportional to $J_i dt dx^i$, where J_i is the angular momentum of the rotating star. The solutions of the geodesic equation in the Kerr solution are quite complicated but are still given by elliptic integrals [34].

The situation in space–times with cosmological constant is much more complicated. A spherically symmetric mass in a universe with a cosmological constant is described by the Schwarzschild-de Sitter solution (see, e.g., [139]), and the corresponding geodesic equation can be solved by means of hyperelliptic integrals [62, 63]. Also in Kerr–de Sitter space–times the geodesic equation can now be solved analytically [61] (see also [64]), and even more generally in all Plebański–Demiański space–times without acceleration [60].

4.1 The Gravitational Redshift

The gravitational redshift compares the frequencies of a light ray measured by two different observers. The general situation is shown in Fig. 1. A light ray intersects the world-lines O_1 and O_2 of two observers at the space–time events x_1 and x_2 . The measured frequency is given by $\omega = k(u) = k_\mu u^\mu$, where k is the 4-wave

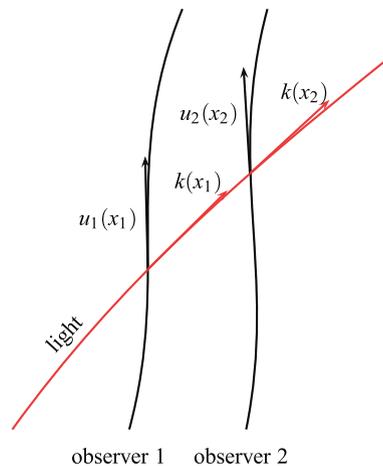


Fig. 1 The geometry of the gravitational redshift: a light ray crosses the world-lines of two observers that both measure the frequency of the light ray

vector of the light ray and u the 4-velocity of an observer. Accordingly, the gravitational redshift is given by the ratio

$$\frac{\nu_2}{\nu_1} = \frac{k(u_2)}{k(u_1)}, \quad (14)$$

($\omega = 2\pi\nu$). This relation gives the total redshift, consisting of the gravitational redshift and the Doppler effect.

In a stationary gravitational field this ratio can be presented in a very simple form. For a stationary gravitational field there exists a timelike Killing vector ξ , so that $k(\xi) = \omega_0 = \text{const}$ along the light ray. It then follows that

$$\frac{\nu_2}{\nu_1} = \sqrt{\frac{g_{tt}(x_1)}{g_{tt}(x_2)}} \approx 1 - \frac{GM}{c^2} \left(\frac{1}{r_1} - \frac{1}{r_2} \right), \quad (15)$$

where r_1 and r_2 are the radial positions of the two observers. The right part of the equation follows if we assume the validity of the Einstein theory of gravity.

This effect was observed first by Pound and Rebka [134] who confirmed the predictions to within 1%. Later, in a space experiment where the time of a hydrogen maser in a rocket was compared with the time of an identical hydrogen maser on Earth, the confirmation has been improved to 1 part in 10^4 [164]. The gravitational redshift also plays an important role in satellite navigation and positioning systems. In the passage of one day the redshift will account for a distance of several km on Earth.

A further aspect of the gravitational redshift is the coupling of gravity to the Maxwell field. Assuming a stationary situation, that is, assuming a Killing vector field ξ and an electromagnetic field strength F that is stationary, $\mathcal{L}_\xi F = 0$, it can be shown [78] that there exists a generalized scalar electrostatic potential ϕ so that $i_\xi F = d\phi$ (i being the inner product). With the observer's 4-velocity given by $u = e^{-\varphi}\xi$, where φ is a gravitational potential (in a Newtonian approximation it is mgz), we then have $d\phi = e^\varphi i_u F = e^\varphi E$ where E is the electric field measured by the observer u . Since ϕ is constant along the paths of charged particles, we have $\text{const.} = \Delta\phi \approx E(1 + \varphi)$. As a consequence, the voltage between two identical batteries depends on their position in the gravitational field. This has been experimentally verified at the percent level [77]. This also confirms the universality of the coupling of gravity to all forms of matter.

4.2 Light Deflection

The deflection of light was the first prediction of Einstein's GR to be confirmed by observation, which occurred only four years after the complete formulation of the theory. With the exact solution of the geodesic equation for light given in Eq. 11,

the deflection angle is defined as the difference between the angles φ_1 and φ_2 for which $\wp(\frac{\varphi}{2}; g_2, g_3) + \frac{1}{3} = 0$. Explicitly,

$$\delta\varphi = \frac{4}{\sqrt{e_1 - e_2}} F(\alpha, k), \quad \sin \alpha = \sqrt{-\frac{e_3 + \frac{1}{3}}{e_2 - e_3}}, \quad k^2 = \frac{e_2 - e_3}{e_1 - e_3} \quad (16)$$

where

$$F(\alpha, k) = \int_0^\alpha \frac{dx}{1 - k^2 \sin^2 x} \quad (17)$$

is the elliptic integral of the first kind [2] and $e_1 > e_2 > e_3$ are the three real zeros of the polynomial $4x^3 - g_2x - g_3$ (in our light deflection scenario $e_3 < -\frac{1}{3}$). Here, $e_2 = \frac{2M}{r_2} - \frac{1}{3}$ where r_2 is the radial coordinate of closest approach of the deflected light ray. In an approximation for weak gravitational fields or small mass M this is $\delta\varphi = M/b$, where b is the impact parameter. In the frame of the PPN formalism we obtain $\Delta\varphi = \frac{1}{2}(1 + \gamma)M/b$.

Today's observations use Very Long Baseline Interferometry (VLBI); this has lead to a confirmation of Einstein's theory at the 10^{-4} level [151].

4.3 Perihelion/Periastron Shift

The exact value of the perihelion shift is

$$\delta\varphi = \frac{2}{\sqrt{e_1 - e_3}} F\left(\frac{\pi}{2}, k\right) - 2\pi, \quad (18)$$

where again $k^2 = \frac{e_2 - e_3}{e_1 - e_3}$ and $e_1 > e_2 > e_3$ are the real zeros of the corresponding polynomial (the values of k , e_1 , e_2 , and e_3 are here different from the corresponding values in the previous subsection). Here $e_2 = \frac{2M}{r_2} - \frac{1}{3}$ and $e_3 = \frac{2M}{r_3} - \frac{1}{3}$ so that we can relate e_2 and e_3 to the minimum and maximum radial distances, r_2 and r_3 , of the orbit. In a post-Newtonian approximation one obtains $\delta\varphi = \frac{6\pi M}{a(1-e^2)}$, where a is the semimajor axis and e the eccentricity of the orbit. In the PPN framework this has to be multiplied with $(2 + 2\gamma - \beta)/3$.

It was first observed by Le Verrier in the nineteenth century that the perihelion shift of Mercury was larger than that calculated on a Newtonian basis from the influence of other planets. Today this post-Newtonian perihelion shift has been determined as $42''98$ per century, with an error of the order 10^{-3} [133].

Recently, a huge periastron shift of a candidate binary black hole system in the quasar OJ287 has been observed, where one black hole is small compared to the other [161]. The observed perihelion shift is approximately 39° per revolution, which takes 12 years.

4.4 Gravitational Time Delay

In the vicinity of masses, electromagnetic signals move slower than in empty space. This effect is referred to as the gravitational time delay, see Fig. 2, which has been confirmed by observations and experiments. There are two ways to detect this effect: (i) direct observation, that is, by comparing the time of flight of light signals in two situations for fixed sender and receiver, and (ii) by observing the change in the frequency induced by this gravitational time delay.

4.4.1 Direct Measurement

The gravitational time delay for signals that pass in the vicinity of a body of mass M is given by [168]

$$\delta t = 2(1 + \gamma) \frac{GM}{c^3} \ln \frac{4x_{\text{Sat}}x_{\text{Earth}}}{b^2}, \quad (19)$$

where x_{Sat} and x_{Earth} are the distances of the satellite and the Earth, respectively, from the gravitating mass and b is the closest distance of the signal to the gravitating mass. If the gravitating body is the Sun and if we take b to be the radius of the Sun, then the effect would be of the order 10^{-4} s, which is clearly measurable. Reflection of radar signals from the surface of Venus has confirmed this effect [150]. An improved result is obtained by using Mars ranging data from the Viking Mars mission [136]. GR, characterized by $\gamma = 1$, has thus been confirmed by $|\gamma - 1| \leq 10^{-4}$.

4.4.2 Measurement of Frequency Change

Though the time delay is comparatively small, the induced modification of the received frequency can indeed be measured with higher precision, the reason being that clocks are very precise and can thus resolve frequencies very precisely.

The corresponding change in the frequency is easily derived. The emission time of the first wave crest is t_{s1} . This first wave crest will be received at

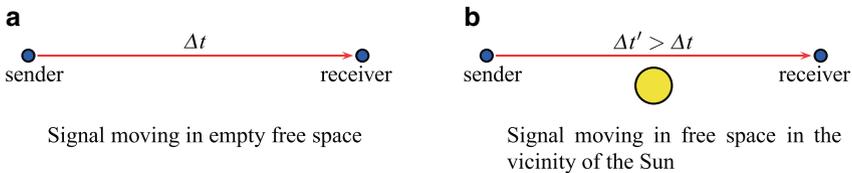


Fig. 2 Gravitational time delay. A signal from the sender to the receiver passing the Sun (b) needs a longer time than a signal in empty free space (a)

$t_{r1} = t_{s1} + \Delta t(t_{s1})$. Now, the second wave crest will be emitted at $t_{s2} = t_{s1} + \frac{1}{\nu_0}$ and received at $t_{r2} = t_{s1} + \Delta t(t_{s2})$. The measured frequency then is given by

$$\nu = \frac{1}{t_{r2} - t_{r1}}. \tag{20}$$

With the result (19) one can easily derive the relative frequency shift

$$y(t) = \frac{\nu - \nu_0}{\nu_0} = 2(\alpha + \gamma) \frac{GM}{c^3} \frac{1}{b(t)} \frac{db(t)}{dt}, \tag{21}$$

where ν_0 is the emitted frequency. It should be noted that, in this formula, it is the time dependence of the impact parameter that is responsible for the effect, which has been measured by the Cassini mission. The associated mission scenario is shown in Fig. 3. The calculated time dilation and frequency shifts are shown in Fig. 4. One

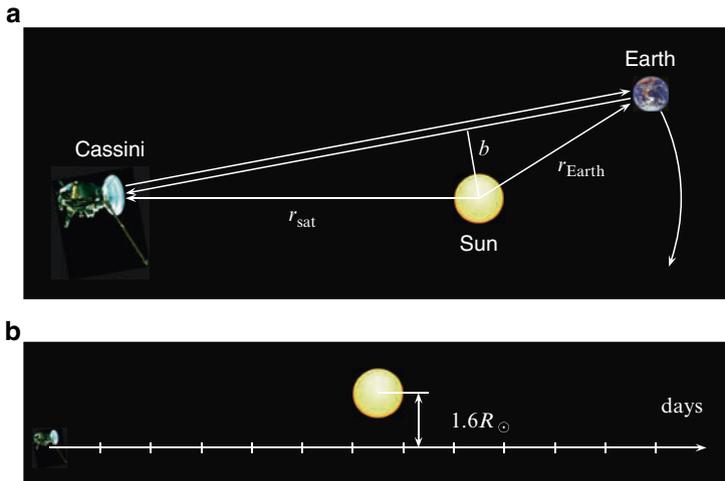


Fig. 3 Cassini mission scenario: (a) top view, (b) sight-of-line view from Earth

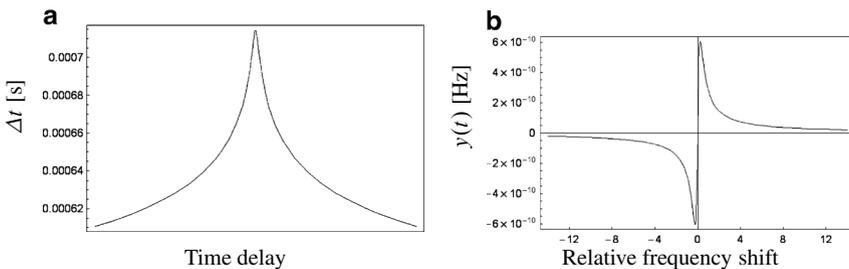


Fig. 4 (a) Calculated time delay, (b) relative frequency shift

important feature of the actual measurement was that three different wavelengths for the signals were used, which made it possible to eliminate dispersion effects near the Sun and to verify this time delay with an accuracy of 10^{-5} [23].

4.4.3 Remarks

The theoretical description of the gravitational time delay requires some additional remarks. In the above treatment – and this is the standard description of this effect – we compared a measurement in the presence of a gravitational field with a measurement without a gravitational field. However, within an exact framework for gravitational effects there is no definition for the unique identification of points with and without a gravitational field. Therefore, there is no definition of a gravitational time delay; there is no situation that can be taken as reference with respect to which the signal can be delayed.

Within an exact treatment there is only a combined effect due to the gravitational time delay, redshift, kinematical time delay (Doppler effect), and light bending. There is no way to isolate a gravitational time delay; this is only possible asymptotically, in the weak field approximation.

4.5 Lense–Thirring Effect

The metric component $J_i dt dx^i$ that reflects the rotation of a gravitating body can be regarded as representing a gravitomagnetic vector potential, the curl of which gives a Lorentz type gravitational force acting on bodies. The influence of this field on the trajectory of satellites results in a motion of the nodes (mathematically this is related to a period of the analytical solution of the geodesic equation), which has been measured by observing the LAGEOS satellites via laser ranging. Together with new data of the Earth's gravitational field obtained from the CHAMP and GRACE satellites, the confirmation recently reached the 10% level [36].

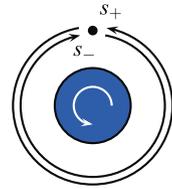
The gravitomagnetic field also influences the rate of clocks. It is easily shown that the geodesic equation for *circular* orbits in the equatorial plane reduces to

$$\frac{d\varphi}{dt} = \pm\Omega_0 + \Omega_{\text{Lense–Thirring}}, \quad (22)$$

where $\Omega_{\text{Lense–Thirring}}$ is the frame-dragging angular velocity that is proportional to the angular momentum of the gravitating source. The \pm is related to the two different directions of the circular orbit. From this we obtain the difference of the proper time of two counterpropagating clocks, see Fig. 5,

$$s_+ - s_- = 4\pi \frac{J}{M}. \quad (23)$$

Fig. 5 Clocks tick differently when orbiting a rotating mass in opposite directions along the same orbit



It should be remarked that this quantity does not depend on G and r . In principle, this effect can be calculated for arbitrary orbits. It decreases with increasing inclination and vanishes for polar orbits. For clocks in satellites orbiting the Earth, this effect can be as large as 10^{-7} s [115].

4.6 Schiff Effect

The gravitational field of a rotating gravitating body also influences the rotation of gyroscopes. This effect is currently being considered by the data analysis group of the GP-B mission that flew in 2004. Analysis is expected to be complete in 2010. Though the mission met all design requirements, a huge technological success, it turned out after the mission that contrary to all expectations and requirements the gyroscopes lost more energy than anticipated [57]. For updates of the data analysis one may contact GP-B's Web site [57]. Full analysis of the experiment requires the determination of further constants characterizing this spinning down effect, which affects the overall accuracy of the measurement of the Schiff effect that was expected to be of the order of 0.5%. Nevertheless, recent reports of the GP-B data analysis group give at the moment an error of about 10% [52, 57].

It should be noted that although both effects within GR are related to the gravitomagnetic field of a rotating gravitational source, the Lense–Thirring effect and the Schiff effect differ conceptually, even measuring different quantities, so they may be regarded as independent tests of GR. In a generalized theory of gravity, spinning objects may couple to different gravitational fields (like torsion) than the trajectory of orbiting satellites. Moreover, the Lense–Thirring effect is a global effect related to the whole orbit while the Schiff effect observes the Fermi-propagation of the spin of a gyroscope.

4.7 The Strong Equivalence Principle

The gravitational field of a body contains energy that adds to the rest mass of the gravitating body. The strong equivalence principle now states that EEP is also valid for self gravitating systems, that is, that the UFF is valid for the gravitational energy, too. This has been confirmed by Lunar Laser Ranging with an accuracy of 10^{-3}

[168] where the validity of the UFF had to be assumed. However, the latter has been tested separately for bodies of the same composition as the Earth and Moon and confirmed with an accuracy of $1.4 \cdot 10^{-13}$ [16].

5 Why New Tests?

It is evident that the number of high precision tests relating to gravity has increased considerably in the last decade. This is certainly not due to some impact from the official Einstein year 2005, but is the consequence of (i) improved technology, (ii) the quest for a quantum theory of gravity, and (iii) problems in the understanding of observational data within standard GR.

5.1 *Dark Clouds – Problems with GR*

Despite all the confirmation catalogued above, some serious problems with GR may exist. In most cases there is no doubt concerning the data. The main problem is the interpretation of the observations and measurements. Each phenomenon that cannot be explained within standard GR is, inevitably, motivation to propose new theories. One should, nevertheless, spend considerable effort in searching for conventional explanations. Below, besides the “standard” interpretation of the phenomena we also mention activities regarding more conventional explanations.

5.1.1 **Dark Matter**

It was first observed by Zwicky in 1933 that in the Coma cluster of about 1,000 galaxies, the galaxies move with a velocity that is much higher than what is expected from the standard laws of gravity. This feature has since been confirmed for many other galaxy clusters, and even for stars within galaxies; it has also been confirmed with gravitational lensing. The apparent gravitational field is too strong. In order to keep the Einstein equations one introduces dark matter that accounts for the observed strength of gravity [158]. Structure formation also appears to need this dark matter. However, so far there is no single observational hint at which particles might constitute this dark matter. Consequently, there are alternative attempts to describe the same effects by a modification [141] of the gravitational field equations, for example, by a term of Yukawa form, or by a modification of the dynamics of particles, as in the MOND ansatz [120, 142], which has recently been formulated in a relativistic framework [21]. With the current lack of direct detection of Dark Matter particles, all these attempts remain on an equal footing.

Another attempt to solve the dark matter problem involves taking into account the full nonlinear Einstein equation. There are suggestions that many of the observations that are usually “explained” by dark matter could be explained by a stronger gravitational field which emerges from more fully taking the Einstein equations into account [17, 37].

5.1.2 Dark Energy

Observations of type Ia supernovae indicate an accelerating expansion of the universe and that 75% of the total energy density consists of a dark energy component with negative pressure [131, 137]. Furthermore, WMAP measurements of the cosmic microwave background [152], the galaxy power spectrum, and the Lyman-alpha forest data lines [129, 159, 162] all support the existence of Dark Energy, rather than a modification of the basic laws of gravitation [130]. However, in this case too, there are attempts to give an explanation in terms of modified field equations; see, for example, [128]. Recently it has been claimed that dark energy or, equivalently, the observed acceleration of the universe can be explained by inhomogeneous cosmological models, such as the spherically-symmetric Lemaitre–Tolman–Bondi model, see, for example, [13, 33, 163].

Buchert and Ehlers [31] have shown, first in a Newtonian framework, that with a spatial averaging of matter and the gravitational field, rotation, and shear of matter can influence the properties of the averaged gravitational field as would be described in effective Friedman equations. Their observation also holds in the relativistic case [30]. Therefore, it is still an open question whether or not the need for dark energy is just the result of an incorrect averaging procedure. An influence of the averaging has certainly been found in the interpretation of existing data [106, 107].

5.1.3 Pioneer Anomaly

The Pioneer anomaly is an anomalous, unexplained acceleration of the Pioneer 10 and 11 spacecraft of

$$a_{\text{Pioneer}} = (8.74 \pm 1.33) \cdot 10^{-10} \text{ m/s}^2 \quad (24)$$

toward the Sun [11, 12]. This acceleration seems to have been turned on after the last flyby of Jupiter and Saturn, and has stayed constant within a 3% range. Until now, no convincing explanation has been found. An anisotropy of the thermal radiation might explain the acceleration. In particular, while the power provided by the plutonium decays exponentially with a half life of 87.5 y (which would mean a decrease of more than 10% during 10 years), the acceleration has stayed constant within a margin of 3%. Presently, much further work is being done on a good thermal modeling of the spacecraft [138], and a reanalysis of the early tracking data is still underway. Improvements in ephemerides are also helping to eliminate various proposed explanations and theories [154].

5.1.4 Flyby Anomaly

It has been observed on several occasions that satellites after an Earth swing-by possess a significant unexplained velocity increase of a few mm/s. This unexpected and unexplained velocity increase is called the *flyby anomaly*. For a summary of recent analyses, see [100]. In a recent article [10] a heuristic formula has been found, which describes all flybys

$$\Delta v = v \frac{\omega R}{c^2} (\cos \delta_{\text{in}} - \cos \delta_{\text{out}}) \quad (25)$$

where R and ω are the radius and the angular velocity, respectively, of the Earth, and δ_{in} and δ_{out} are the inclinations of the incoming and outgoing trajectory.

Although no explanation has been found so far, it is expected that the effect is either (i) a mismodeling of the thermal influence of the Earth's and the Sun's radiation on the satellite, (ii) a mismodeling of reference systems (this is supported by the fact that all the flybys can be modeled by Eq. 25 containing geometrical terms only), or (iii) a mismodeling of the satellite's body by a point mass. There are also more hypothetical considerations: in [118, 119] a model was introduced in which the inertial mass experiences a modification that depends on the Hubble scale and the acceleration of a body. Within this model, the additional term accounts for the Pioneer anomaly and also gives a modification of the velocities of spacecraft during a flyby. Another proposal [32] relates the flyby anomaly to an anisotropic speed of light, which, however, only resorts to a non-understood early measurement reported by D.C. Miller 75 years ago and neglects all new confirmations of the isotropy of light at the level of 10^{-17} . In [3], S. Adler discusses the possibility that the flyby anomaly may be related to dark matter around the Earth. This proposal would lead to severe restrictions on the dark matter model (e.g., a two component dark matter model around the Earth is needed), which are unlikely to be consistent with other observations. In [132] a modification of Special Relativity, based again on a violation of the relativity principle, has been used in a scheme for obtaining a modified velocity. Within a certain five-dimensional theory of gravity [55] an additional acceleration occurs, which may account for the flyby as well as the Pioneer anomaly. An attempt to understand the flyby anomaly on a conventional level has been carried forward by J.P. Mbelek [117], who claims that the observation was due to a mismodeling of Special Relativity in the orbit determination.

5.1.5 Increase of Astronomical Unit

The analysis of radiometric distances measured between the Earth and the major planets, and observations from Martian orbiters and landers from 1961 to 2003, both lead to reports of a secular increase of the Astronomical Unit of approximately 10 m/cy [87] (see also the article [153] and the discussion therein). This increase cannot be explained by a time-dependent gravitational constant G because the \dot{G}/G needed is larger than the restrictions obtained from LLR. Such an increase might be mimicked, though, by a long-term increase in the density of the solar plasma.

5.1.6 Quadrupole and Octupole Anomaly

Recently, an anomalous behavior of the low- l contributions to the cosmic microwave background has been reported. It has been shown that (i) there exists an alignment between the quadrupole and octupole with $>99.87\%$ C.L. [43], and (ii) that the quadrupole and octupole are aligned to the Solar system ecliptic to $>99\%$ C.L. [148]. No correlation with the galactic plane has been found.

The reason for this anomaly is totally unclear. One may speculate that an unknown gravitational field within the Solar system slightly redirects the incoming cosmic microwave radiation (in a similar way that motion with a certain velocity with respect to the rest frame of the cosmological background redirects the cosmic background radiation and leads to modifications of the dipole and quadrupole parts). Such a redirection should be more pronounced for low- l components of the radiation. It should be possible to calculate the gravitational field needed for such a redirection and then to compare that with the observational data of the Solar system and the other observed anomalies.

5.2 The Search for Quantum Gravity

There are many experiments proving that matter must be quantized and, indeed, all experiments in the quantum domain are in full agreement with quantum theory, with all its seemingly strange postulates and consequences. Consistency of the theory also requires that the fields to which quantized matter couples also have to be quantized. Therefore, the gravitational interaction has to be quantized too. However, though gravity is an interaction between particles, it also deforms the underlying geometry. This double role of gravity seems to prevent all quantization schemes from being successful in the gravitational domain.

The incompatibility of quantum mechanics and GR is not only due to the fact that it is not possible to quantize gravity according to known schemes, but also because time plays a different role in quantum mechanics and in GR. Moreover, it is expected that a quantum theory of gravity will solve the problem of the singularities appearing within GR. It is also hoped that such a theory would lead to a true unification of all interactions and, thus, to a better understanding of the physical world.

Any theory is characterized by its own set of constants. It is believed that the Planck energy $E_{\text{Pl}} \approx 10^{28}$ eV sets the scale for quantum gravity effects. All expected effects scale with this energy or the corresponding Planck length, Planck time, etc.

5.3 Possible New Effects

The low energy limit of string theory, as well as some semiclassical limit of loop quantum gravity and results from noncommutative geometry, suggest that many of the standard laws of physics will suffer some modifications. At a basic level

these modifications show up in the equations of the standard model (Dirac equation, Maxwell equations, etc.) and in Einstein's field equations. These modifications then result in the following (see, e.g., [9, 38, 116]):

- Violation of Lorentz invariance
 - Different limiting velocities of different particles
 - Modified dispersion relations leading to birefringence in vacuum
 - Modified dispersion relations leading to frequency-dependent velocity of light in vacuum
 - Orientation- and velocity-dependent effects
- Time and position dependence of constants (varying α , G , etc.)
- Modified Newtonian law at short and large distances

In recent years there have been increased efforts to search for these possible effects, so far without success.

Besides these effects expected to result from quantum gravity, there are some more “exotic” issues that are usually taken for granted but are also worth testing experimentally. Such issues include:

- Violations of Newton's inertial law $\mathbf{F} = m\ddot{\mathbf{x}}$.
- Violation of *actio = reactio*.
- Violation of charge conservation.
- Violation of mass or energy conservation.
- Questioning that gravity can be transformed away even if UFF is fulfilled.

In most cases there is no basic theory from which these effects can be derived, due in part to the fact that equations of motions cannot normally be derived without an action principle. Nevertheless, since these issues are at the very basis of our description of physical dynamics, they should be tested to the highest accuracy possible.

6 How to Search for “New Physics”

If one looks for “new physics” then one has to measure effects that have never previously been measured. Strategies by which it might be possible to find new things include (i) using more precise devices, (ii) exploring new parameter regions, and (iii) testing “exotic” ideas.

6.1 Better Accuracy and Sensitivity

It is clear that in searching for tiny effects, better accuracy is always a good strategy. It is amazing how the accuracy for testing Lorentz invariance, for example, has increased over the years. It took more than 20 years to improve the results of the

Brillet and Hall experiment of 1979 [28]; within another few years the accuracy improved by two orders of magnitude and it is still improving further.

It would be of interest to find examples where present-day technologies have, at least in principle, sensitivity to quantum gravity effects. One such example arises with gravitational wave interferometers [8], which currently have a strain sensitivity of 10^{-21} . With Advanced LIGO the sensitivity will become 10^{-24} . Thus, for a continuous gravitational wave with a frequency in the maximum sensitivity range between 10 and 1,000 Hz a continuous observation over one year would reach a sensitivity of slightly less than 10^{-28} . This is the sensitivity needed for observing Planck scale effects (10^{28} eV) by optical laboratory devices (which have an energy scale of ~ 1 eV). It is, thus, the level of sensitivity required to detect Planck-scale modifications in the dispersion relation for photons [8].

6.2 *Extreme Situations*

Often, “new physics” is discovered when new situations are explored. We discuss various scenarios of this kind.

6.2.1 **Extreme High Energy**

One possibility for exploring new physics is to probe physical processes at very high energies. With the LHC, where energies of the order 10^{13} eV should be achievable, it is hoped that signals of the Higgs particle and of supersymmetry will be found. This energy range is still far away from the quantum gravity scale. The best that one can do is to observe high energy cosmic rays that have energies of up to 10^{21} eV. It has, in fact, been speculated that the observations of high energy cosmic rays – which according to standard theories are forbidden owing to the GZK-cutoff – could indicate a modified dispersion relation [9, 116].

6.2.2 **Extreme Low Energy**

The other extreme, very low temperatures, might also provide a tool for investigating possible signals of quantum gravity. One may speculate that the influence of expected space–time fluctuations on the dynamics of quantum systems is more pronounced at very low temperatures. One may even speculate that such space–time fluctuations may give rise to a temperature threshold above absolute zero.

Very low temperatures may be achievable in BECs for which a long period of free evolution is possible. Recently a free evolution time of more than 1 s has been sustained at the Bremen drop tower where a BEC has been created during a period of 4.7 s of free fall [171]. These BECs may be used for novel investigations, including a search for deviations from standard physics predictions.

6.2.3 Large Distances

The unexplained phenomena, dark matter, dark energy, and the Pioneer anomaly are related to large distances. It is questionable whether the ordinary laws of gravity should be modified at large distances. Recently, some suggestions have been made:

- It has been examined whether a Yukawa modification of the Newtonian potential may account for galactic rotation curves [141].
- In the context of higher dimensional braneworld theories, deviations from Newton's potential arise [48]. At large distances the potential behaves like $1/r^2$, as one would expect from the Poisson equation in five dimensions. A comparison with cosmological and astrophysical observations has been reviewed in [112].
- From considering a running coupling constant, it has been suggested that the spatial parts of the space–time metric possess a part that grows linearly with distance [75]. This approach is in agreement with present solar system tests and also describes the Pioneer anomaly [76].

6.2.4 Small Accelerations

An acceleration, a , being of physical dimension m s^{-2} can be related to a length scale $l_0 = c^2/a$. Now, the largest length scale in our universe is the Hubble length $L_H = c/H$, where H is the Hubble constant. The corresponding acceleration is cH , at an order of magnitude that remarkably coincides with the Pioneer acceleration and the MOND acceleration scales. As a consequence, it really seems mandatory to perform experiments that explore physics for such small accelerations (see below).

6.2.5 Large Accelerations

Analogously, since the smallest length scale is the Planck length l_{Pl} , the corresponding acceleration is $a = 2 \cdot 10^{51} \text{ m s}^{-2}$, which, however, is far outside any experimental reach. For the smaller accelerations that might be reached by electrons in the fields of strong lasers, one might be able to detect Unruh radiation or to probe the physics near black holes [144, 147].

6.2.6 Strong Gravitational Fields

Most observations and tests of gravity are being performed in weak fields: the solar system, galaxies, galaxy clusters. Recently, it became possible to observe phenomena in strong gravitational fields: in binary systems and in the vicinity of black holes.

The observation of stars in the vicinity of black holes [146] may, in one or two decades, give improved measurements of the perihelion shift and of the Lense–Thirring effect. Binary systems present an even better laboratory for observing strong field effects.

The inspiral of binary systems, which has been observed with very high precision, can be completely explained by the loss of energy through the radiation of gravitational waves as calculated within GR [24]. The various data from such systems can be used to constrain hypothetical deviations from GR. As an example, such data can be used for a test of the strong equivalence principle [41] and of preferred frame effects and conservation laws [22] in the strong field regime.

Double pulsars have recently been detected and studied. These binary systems offer possibilities for analyzing spin effects, thus, opening up an entirely new domain for exploration of gravity in the strong field regime [85, 86]. Accordingly, the dynamics of spinning binary objects has been intensively analyzed [25, 53, 156].

6.3 Investigation of “Exotic” Issues

We describe several “unusual” questions which are rarely posed but that are worth investigating both experimentally and theoretically. A class of these peculiarities addresses Newton’s axioms, particularly their dynamical part related to forces:

1. Test of *actio = reactio*. Tests of this axiom can be encoded in a difference between active and passive charges (electric charge, masses, magnetic moments, etc., generally, any quantity that creates a corresponding field).
2. Test of the inertial law $m\ddot{x} = F$ where F is the force acting on a body. What is being measured here? The measured acceleration together with the knowledge of the mass (which can be determined, e.g., through elastic scattering) leads to the exploration of the force. This can be illustrated with the Lorentz force. If one sends charged particles through a condenser, their trajectory will be deflected in response to the voltage applied to the condenser. The deflection gives the force and the force defines the electric field E .

Therefore, the question of testing the inertial law may have at least two meanings:

- (a) Why are there no higher time derivatives in the inertial law? (In fact, owing to back reaction all equations of motion are of higher than second order. For charged particles, for example, we have the third order Abraham–Lorentz equation. This back reaction force can be calculated from the basic equations of motion which are of second order only. Therefore, the question is why the underlying basic equations of motion are of second order.)
 - (b) Does the inertial law hold for all forces, no matter how large or small? (in our example, do we have $m\ddot{x} = qE$ even if E becomes extremely large or small?)
3. Test of the superposition of forces.

7 Testing “Exotic” but Fundamental Issues

7.1 Active and Passive Mass

The notion of active and passive masses and their possible non-equality was first introduced and discussed by Bondi [26]. The *active mass* m_a is the source of the gravitational field (here we restrict to the Newtonian case with the gravitational potential U) $\Delta U = 4\pi m_a \delta(x)$, whereas the *passive mass* m_p reacts to it

$$m_i \ddot{x} = m_p \nabla U(x). \quad (26)$$

Here, m_i is the inertial mass and x the position of the particle. The equations of motion for a gravitationally bound two-body system then are

$$m_{1i} \ddot{x}_1 = G m_{1p} m_{2a} \frac{x_2 - x_1}{|x_2 - x_1|^3}, \quad m_{2i} \ddot{x}_2 = G m_{2p} m_{1a} \frac{x_1 - x_2}{|x_1 - x_2|^3}, \quad (27)$$

where 1, 2 refer to the two particles and G is the gravitational constant.

For the equation of motion of the center of mass, $X = (m_{1i} x_1 + m_{2i} x_2) / M_i$, we find

$$\ddot{X} = \frac{m_{1p} m_{2p}}{M_i} C_{21} \frac{x}{|x|^3} \quad \text{with} \quad C_{21} = \frac{m_{2a}}{m_{2p}} - \frac{m_{1a}}{m_{1p}} \quad (28)$$

where $M_i = m_{1i} + m_{2i}$ and x is the relative coordinate. Thus, if $C_{21} \neq 0$ then active and passive masses are different and the center of mass shows a self-acceleration along the direction of x . This is a violation of Newton’s *actio* equals *reactio*. A limit has been derived by Lunar Laser Ranging (LLR): no self-acceleration of the moon has been observed yielding a limit of $|C_{Al-Fe}| \leq 7 \cdot 10^{-13}$ [18].

The dynamics of the relative coordinate

$$\ddot{x} = -G \frac{m_{1p} m_{2p}}{m_{1i} m_{2i}} \left(m_{1i} \frac{m_{1a}}{m_{1p}} + m_{2i} \frac{m_{2a}}{m_{2p}} \right) \frac{x}{|x|^3}. \quad (29)$$

have been probed in a laboratory experiment by Kreuzer [90] with the result $|C_{21}| \leq 5 \cdot 10^{-5}$.

The issue of the equality of active and passive gravitational mass is of the same quality as the issue of the equality of inertial and passive gravitational mass. While the UFF is an equivalence of all bodies *reacting* to the gravitational field, here we have an equivalence of all masses *creating* a gravitational field: all (spherically symmetric) masses of the same weight create the same gravitational field, independent of their internal composition. The equality of active and passive masses constitutes a universality principle that we may call the *Universality of the Gravitational Field*.

It is interesting to note that there is no Lagrange function from which the equations of motion (27) can be directly derived. As a consequence there is no Hamiltonian, which means that there is no quantum version of this system. Only the equation of motion for the relative distance can be quantized.

7.2 Active and Passive Charge

Similarly, one can think of active and passive charges, which have been discussed recently [98]. Though electric charges have no direct link to gravity, a discussion of the similarities and differences to the gravitational case will underline the universality of this question and can lead to a better understanding of the gravitational case. As an example, we will see that on the one hand the weakness of the gravitational interaction helps in a search for a difference of active and passive masses, while on the other hand the fact that negative charges are possible may help in circumventing the short timescales present in the electromagnetic interaction, which at first sight are a big obstacle in searching for a difference in active and passive electric charges. Furthermore, since in the weak field approximation there are many similarities between gravity and electromagnetism, a different active and passive charge would give a strong indication of a possible difference of active and passive masses. Moreover, as charged bodies also gravitate, a difference in active and passive charges would probably lead to a modified behavior for interacting charged black holes. This realization has not yet been fully developed.

The resulting equations of an electrically bound system with different active and passive charges are similar to the equations for a gravitationally bound system with different active and passive masses. The only difficulty that arises here is that the self acceleration of the center of mass cannot be observed, since within atoms the timescale is too short so that, as a result, this effect averages out.

However, there is one substantial difference between this and the massive case: there are positive and negative charges. This opens up the possibility of defining active as well as passive neutrality. In order to exploit this possibility one has to consider a bound system in an external electric field E

$$m_{1i}\ddot{x}_1 = q_{1p}q_{2a} \frac{x_2 - x_1}{|x_2 - x_1|^3} + q_{1p}E(x_1), \quad m_{2i}\ddot{x}_2 = q_{2p}q_{1a} \frac{x_1 - x_2}{|x_1 - x_2|^3} + q_{2p}E(x_2), \quad (30)$$

where q_{1p} , q_{1a} , q_{2p} , and q_{2a} are the passive and active charges. The equations of motion of the center of mass and the relative coordinate are

$$\ddot{X} = \frac{q_{1p}q_{2p}}{M_i} \bar{C}_{21} \frac{x}{|x|^3} + \frac{1}{M_i} (q_{1p} + q_{2p}) E, \quad \ddot{x} = -\frac{1}{m_{\text{red}}} q_{1p}q_{2p} \bar{D}_{21} \frac{x}{|x|^3}, \quad (31)$$

where

$$\bar{C}_{21} = \frac{q_{2a}}{q_{2p}} - \frac{q_{1a}}{q_{1p}}, \quad \bar{D}_{21} = \frac{m_{1i}}{M_i} \frac{q_{1a}}{q_{1p}} + \frac{m_{2i}}{M_i} \frac{q_{2a}}{q_{2p}}. \quad (32)$$

Thus, if active and passive charges are different, the center of mass shows a self-acceleration along the direction of x , in addition to the acceleration caused by the external field E . Due to fast internal motion the self-acceleration of the center of mass is not observable.

However, it is now possible to define active neutrality through $0 = q_{a1} + q_{a2}$ as well as passive neutrality $0 = q_{p1} + q_{p2}$. We may now prepare an actively neutral

system by the condition that it creates no electric field (which may be explored by other test charges). This actively neutral system might be passively non-neutral and may react on an external electric field. Also, a passively neutral field may actively create an electric field. If actively neutral systems are also passively neutral, then the active and passive charge are proportional. These procedures can be carried out with high precision resulting in $\tilde{C}_{12} \leq 10^{-21}$ [98]. Atomic spectra represent a cleaner test but yield only an estimate of the order $\tilde{C}_{12} \leq 10^{-9}$ [98].

7.3 Active and Passive Magnetic Moment

A similar analysis can be carried out for magnetic fields created by magnetic moments. If active and passive magnetic moments are different, then again we would observe a self-acceleration of the center of mass. In this case atomic spectroscopy is more useful and yields an (unsurpassed) estimate $\tilde{C}_{12} \leq 10^{-5}$ [98].

7.4 Charge Conservation

Charge conservation is a very important feature of the ordinary Maxwell theory:

- It is basic for an interpretation of Maxwell-theory as a $U(1)$ gauge theory.
- It is necessary for the compatibility with standard quantum theory insofar as it relates to the conservation of probability.

Recently, some models that allow for a violation of charge conservation have been discussed. Within higher dimensional brane theories it has been argued that charge may escape into other dimensions [46,47], leading to charge nonconservation in four-dimensional space–time. Charge nonconservation may also occur in connection with variable-speed-of-light theories [104]. A very important aspect of charge nonconservation is its relation to the EEP, which is at the basis of GR [105]. Charge nonconservation necessarily appears if, phenomenologically, one introduces into the Maxwell equations, in a gauge-independent way, a mass for the photon [95,97].

The more important a particular feature of physics is, the more firmly this feature should be based on experimental facts. There seem to be only three classes of experiments related to charge conservation:

1. *Electron disappearing*: Charge is not conserved if electrons spontaneously disappear through $e \rightarrow \nu_e + \gamma$ or, more generally, through $e \rightarrow$ any neutral particles. Decays of this kind have been searched for using high-energy storage rings but they have not been observed [4, 155]. For the general process, the probability for such a process has been estimated to be $2 \cdot 10^{-22} \text{ year}^{-1}$ [155]; for two specific processes the probability is as low as $3 \cdot 10^{-26} \text{ year}^{-1}$ [4]. Even for a strict non-disappearance of electrons, the charge of an electron may vary in time and

thus may give rise to charge nonconservation. Thus, while charge-conservation implies the non-disappearance of electrons, electron non-disappearance does not imply charge conservation.

2. *Equality of electron and proton charge:* Another aspect of charge conservation is the equality of the absolute value of the charge of elementary particles like electrons and protons. Tests of the equality of q_e and q_p through the neutrality of atoms [49] yield very precise estimates because a macroscopic number of atoms can be observed. The result is $|(q_e - q_p)/q_e| \leq 10^{-19}$.
3. *Time-variation of α :* The most direct test of charge conservation is implied by the search for a time-dependence of the fine structure constant $\alpha = q_e q_p / \hbar c$. Since different hyperfine transitions depend in a different way on the fine structure constant, a comparison of various transitions is sensitive to a variation of α . Recent comparisons of different hyperfine transitions [114] lead to $|\dot{\alpha}/\alpha| \leq 7.2 \cdot 10^{-16} \text{ s}^{-1}$. This may be translated into an estimate for charge conservation $|\dot{q}_e/q_e| \leq 3.6 \cdot 10^{-16} \text{ s}^{-1}$, provided \hbar and c are constant and $q_p = q_e$. However, this direct translation does not hold within the framework of varying c theories. An estimate that is more than one order of magnitude better comes from an analysis of the natural OKLO reactor [38], but it requires some additional assumptions on the α -dependence of various nuclear quantities.

Apparently, we have *no dedicated direct experiment to test charge conservation*.

7.5 Small Accelerations

Since the effect of gravity is observed by its influence on orbits of satellites and stars, a modification of Newton's first law, $F = ma$, will dramatically change the interpretation of the orbits and, therefore, the relation between the observation and the deduced gravitational field. This is, for example, the basis of the MOND (MODified Newtonian Dynamics) ansatz proposed by Milgrom [120] and put into a relativistic formulation by Bekenstein [21].

The MOND ansatz replaces $m\ddot{x} = F$ by

$$m\ddot{x}\mu(|\ddot{x}|/a_0) = F, \quad (33)$$

where $\mu(x)$ is a function that behaves as

$$\mu(x) = \begin{cases} 1 & \text{for } |x| \gg 1 \\ x & \text{for } |x| \ll 1. \end{cases} \quad (34)$$

For Newtonian gravity this means that from the equation $F = m\nabla U$ we obtain the special cases

- For large accelerations: $\ddot{x} = \nabla U$.
- For small accelerations: $\ddot{x}|\ddot{x}| = a_0 \nabla U \rightarrow |\ddot{x}| = \sqrt{a_0 |\nabla U|}$.

This result for small accelerations, such as are present in the outer regions of galaxies, describes many galactic rotation curves very well, and may also reproduce dynamics of galactic clusters. The acceleration scale a_0 is of the order $10^{-10} \text{ m s}^{-2}$.

A recent laboratory experiment using a torsion balance tests the relation between the force acting on a body and the resulting acceleration [59]. No deviation from Newton's inertial law has been found for accelerations down to $5 \cdot 10^{-14} \text{ m s}^{-2}$. However, this does not mean that the MOND hypothesis is ruled out. Within MOND it is required that the full acceleration should be smaller than approximately $10^{-10} \text{ m s}^{-2}$, while in the above experiment only two components of the acceleration were small while the acceleration due to the Earth's attraction was still present. This means that better tests must be performed in space. An earlier test [1] went down to accelerations of $3 \cdot 10^{-11} \text{ m s}^{-2}$, though the applied force was nongravitational. It might be questioned whether the MOND ansatz applies to all forces or to the gravitational force only. There exists a short time and space window (of the order 1 s and 10 cm) for performing tests capable of such a distinction on Earth [170].

It has also been questioned whether the MOND ansatz can describe the Pioneer anomaly [12, 120] but positive confirmation has not been convincingly demonstrated. In any case, it is a very remarkable coincidence that the Pioneer acceleration, the MOND characteristic acceleration, and the cosmological acceleration are all of the same order of magnitude, $a_{\text{Pioneer}} \approx a_0 \approx cH$, where H is the Hubble constant.

What is the principal meaning of such tests? When we are testing $m\ddot{x} = F$ for small F , this at first sight means nothing. The only measured quantity in this equation is x as function of time from which we can derive \ddot{x} . Such measurements of \ddot{x} are used to *define* the force F and to explore the charge-to-mass ratio. Therefore, this kind of measurement does not provide any kind of test.

The only way to give these experiments a meaning is if one has a model for the force. If the force is given by, for example, a gravitating mass, $F = m\nabla U$ with $U = G \int \rho(x')/|x-x'|dV'$, then one may ask whether the acceleration decreases linearly with decreasing gravitating mass. If the gravitating mass is spherically symmetric, $U = GM/r$, then the question is whether $\ddot{x} \rightarrow \alpha\ddot{x}$ for $M \rightarrow \alpha M$, particularly in the case of small M . This is an operationally well-defined question.

Since all components of the acceleration should be extremely small, it is necessary to perform such tests in space. It has been suggested that such a test should be carried out in a satellite located at a Lagrange point of the Earth–Sun system.

7.6 Test of the Inertial Law

The question we ask here is how one can test experimentally whether equations of motion possess second or higher order time derivatives. If the equation of motion is of n th order, then the solution for the path depends on n initial conditions. To enable a theoretical description of such tests we set up equations of motion of higher order where the higher order terms are characterized by some parameters which vanish in the standard equations of motion. This means that, besides their mass, particles are characterized by further parameters related to the additional higher order time

derivatives. We solve these equations of motion and try to exploit already completed experiments, or propose new ones in order to obtain estimates on the extra parameters. So as not to be too general, we use the Lagrange formalism, which, for our purposes, is of higher order with a Lagrangian depending on higher derivatives. A complete description of a particle's dynamics requires the introduction of an interaction with, for example, the electromagnetic field. The structure of this coupling may differ from what we know in a more familiar, first order Lagrangian.

7.6.1 Higher Order Equation of Motion for Classical Particles

In order to get a feeling of what might happen we take for simplicity a (nonrelativistic) second order Lagrangian $L = L(t, x, \dot{x}, \ddot{x})$, see [101] for more details. The Euler–Lagrange equations read

$$0 = \frac{\partial L}{\partial x^i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}^i} + \frac{d^2}{dt^2} \frac{\partial L}{\partial \ddot{x}^i}. \quad (35)$$

It can be shown that these equations of motion remain the same if we add to the Lagrangian a total time derivative of a function $f(t, x, \dot{x})$,

$$\frac{d}{dt} f(t, x, \dot{x}) = \partial_t f(t, x, \dot{x}) + \dot{x}^i \frac{\partial}{\partial x^i} f(t, x, \dot{x}) + \ddot{x}^i \frac{\partial}{\partial \dot{x}^i} f(t, x, \dot{x}). \quad (36)$$

According to the gauge principle, one should replace the derivatives $\partial_t f(t, x, \dot{x})$, $\nabla f(t, x, \dot{x})$, and $\nabla_{\dot{x}} f(t, x, \dot{x})$ by gauge fields, which then yield gauge field strengths. However, it makes no sense to have velocity-dependent gauge fields. Therefore we assume that f is a polynomial in the velocities, $f(t, x, \dot{x}) = \sum_{k=0}^N f_{i_1, \dots, i_k}(x) \dot{x}^{i_1} \dots \dot{x}^{i_k}$.

In the simplest case, $N = 0$ and $L = \frac{1}{2} \varepsilon \ddot{x}^2 + \frac{m}{2} \dot{x}^2$. In this case the gauged Lagrange function reads $L = \frac{1}{2} \varepsilon \ddot{x}^2 + \frac{m}{2} \dot{x}^2 + q\phi + q\dot{x}^i A_i$ that yields as an equation of motion

$$\varepsilon \ddot{\ddot{x}} + m\ddot{x} = qE(x) + q\dot{x} \times B(x) = F(x), \quad (37)$$

where E and B are the electric and magnetic field derived as usual from the scalar and vector potentials ϕ and A . More general cases are discussed in [101].

This equation of motion may be solved in a first approximation by using, to begin with, the substitution $x = \varepsilon \bar{x} + x_0$ where x_0 is assumed to solve the equation of motion without the fourth order term. If we assume that the force is very smooth and that the deviation $\varepsilon \bar{x}$ is very small, that is, if $\bar{x} \cdot \nabla F(x_0) \ll m\ddot{\bar{x}}$ and can be neglected, then we obtain

$$\ddot{\bar{x}}_0 + \varepsilon \ddot{\bar{\ddot{x}}} + m\ddot{\bar{x}} = 0. \quad (38)$$

This equation can be integrated twice

$$\ddot{\bar{x}}_0 + \varepsilon \ddot{\bar{x}} + m\bar{x} = at + b, \quad (39)$$

where a and b are two integration constants. Inserting the equation for \ddot{x}_0 yields

$$\ddot{\bar{x}} + \frac{m}{\varepsilon}\bar{x} = -\frac{1}{m\varepsilon}F(x_0) + \frac{1}{\varepsilon}at + \frac{1}{\varepsilon}b. \quad (40)$$

With a new variable $\hat{x} = \bar{x} - \frac{1}{m}at + \frac{1}{m}b$ we have

$$\ddot{\hat{x}} + \frac{m}{\varepsilon}\hat{x} = -\frac{1}{m\varepsilon}F(x_0). \quad (41)$$

If ε is small (and m large),² then m/ε becomes large. Then the term $\frac{m}{\varepsilon}\hat{x}$ is dominant compared with the term on the right-hand side. If, furthermore, we take ε to be positive, then \hat{x} is a fast oscillating term (for negative ε we have runaway solutions). The total solution then is

$$x(t) = x_0(t) + \varepsilon \left(\hat{x}(t) + \frac{1}{m}at - \frac{1}{m}b \right). \quad (42)$$

This solution consists of the standard solution $x_0(t)$, which is the main motion, a small displacement, a small linearly growing term, and a small fast oscillating term, a kind of *zitterbewegung*. From ordinary observations, a and b should be very small. Neglecting these particular contributions, the standard solution of the standard second order equation of motion seems to be rather robust against the addition of a higher order term.

The question now is how to search for the deviations from the standard solution. One way might be to look for the linearly growing term, which, however, requires a long observation time. Another way might be to search for a fundamental variation in the final position resulting from well-defined initial conditions. Some corresponding proposals have been worked out in [101].

7.6.2 Higher Order Equation of Motion for Quantum Particles

It is easier to consider the question of the order of the time derivative at the quantum level. If one adds, for example, a second time derivative to the Schrödinger equation, then this will change the spacing between the energy levels. A comparison with measurements yields an estimate on the strength of such a term [93]. A higher order time derivative in the Maxwell equations would, for example, modify the dispersion relation by adding cubic or higher order energy terms. Such additional terms could, in principle, be observed in high energy cosmic radiation or in experiments with gravitational wave interferometers, as described above in Section 6.1.

² We assume that ε is independent of m .

7.7 Can Gravity Be Transformed Away?

It might be thought that, with the validity of the UFF, it would be possible to eliminate gravity from the equations of motion of a neutral point particle. This is not the case. The UFF merely implies that the equation of motion should have the general form $\ddot{x}^\mu + \Gamma^\mu(x, \dot{x}) = 0$, where it is essential that no particle parameters enter this equation. If gravity can be transformed away (Einstein elevator), then the second term has to be bilinear in the velocity $\Gamma^\mu(x, \dot{x}) = \Gamma_{\rho\sigma}^\mu \dot{x}^\rho \dot{x}^\sigma$. This is not the case, for example, in Finsler geometries or in the model presented in [100]. These are examples where the UFF is valid but Einstein's elevator fails to hold; they constitute a gravity-induced violation of Lorentz invariance.

7.7.1 Finsler Geometry

An indefinite Finslerian geometry is given by

$$ds^2 = F(x, dx) \quad \text{with} \quad F(x, \lambda dx) = \lambda^2 F(x, dx), \quad (43)$$

so that

$$ds^2 = g_{\mu\nu}(x, dx) dx^\mu dx^\nu \quad \text{with} \quad g_{\mu\nu}(x, y) = \frac{1}{2} \frac{\partial^2 F(x, y)}{\partial y^\mu \partial y^\nu}, \quad (44)$$

where $g_{\mu\nu}(x, dx)$ is a kind of metric, which, however, depends on the vector it is acting on. The motion of light rays and point particles is to be described by the action principle $0 = \delta \int ds^2$.

There are two main consequences of such a Finslerian framework. (i) Since the Christoffel connection depends on the 4-velocity, it cannot be transformed away, so the equation of motion will not reduce to $\ddot{x}^\mu = 0$ for all possible particle 4-velocities. Therefore, gravity cannot be transformed away in the whole tangent space as it can be in GR. (ii) There is no coordinate transformation by which the Finslerian metric could acquire a Minkowskian form. Therefore, a Finslerian metric violates Lorentz invariance.

A very simple example of a Finslerian metric is given by

$$ds^2 = F(dx^\mu) = dt^2 - D(dx^i), \quad D(\lambda dx^i) = \lambda^2 D(dx^i), \quad (45)$$

with

$$(D(dx^i))^r = D_{i_1 \dots i_{2r}} dx^{i_1} \dots dx^{i_{2r}} = (\delta_{ij} dx^i dx^j)^r + \phi_{i_1 \dots i_{2r}} dx^{i_1} \dots dx^{i_{2r}}, \quad (46)$$

where $i, j, \dots = 1, 2, 3$. The anisotropy is encoded in the tensor field $\phi_{i_1 \dots i_{2r}}$, which, by comparison with many experiments, can be assumed to be very small: $\phi_{i_1 \dots i_{2r}} \ll 1$.

7.7.2 Testing Finslerian Anisotropy in Tangent Space

In [96] this ansatz was used for describing tests of Finslerian models, in the photon sector given by $ds^2 = 0$, using Michelson–Morley experiments. From a comparison with the best available optical data, see page 29 in Section 3.1.1, one deduces that $\phi_{i_1 \dots i_{2r}} \leq 10^{-16}$.

In the matter sector, within the nonrelativistic realm, one may start with a Hamiltonian of the form

$$H = H(p) \quad \text{with} \quad H(\lambda p) = \lambda^2 H(p), \quad (47)$$

where $p_i = -i\hbar\partial_i$. For a “power-law” ansatz we have

$$H = \frac{1}{2m} \left(g^{i_1 \dots i_{2r}} p_{i_1} \dots p_{i_{2r}} \right)^{\frac{1}{r}}. \quad (48)$$

The deviation from the standard case may again be parametrized as

$$H = \frac{1}{2m} \left(\Delta^p + \phi^{i_1 \dots i_{2r}} p_{i_1} \dots p_{i_{2r}} \right)^{\frac{1}{r}} \approx \frac{1}{2m} p^2 \left(1 + \frac{1}{r} \frac{\phi^{i_1 \dots i_{2r}} p_{i_1} \dots p_{i_{2r}}}{p^{2r}} \right). \quad (49)$$

The second term is a nonlocal operator that has influence on, for example,

- The degeneracy of Zeeman levels given by $H_{\text{tot}} = H + \sigma \cdot B$. If H_0 deviates from p^2 then the Zeeman levels split, as can be explored in Hughes–Drever type experiments, which lead to estimates $\phi^{i_1 \dots i_{2r}} \leq 10^{-30}$, see Section 3.1.2.
- On the phase shift in atomic interferometry. The atom–photon interaction leads to a phase shift

$$\delta\phi \sim H(p+k) - H(p) \approx \frac{k^2}{2m} + \frac{1}{m} \left(\delta^{il} + \frac{1}{r} \frac{\phi^{i l i_3 \dots i_{2r}} p_{i_3} \dots p_{i_{2r}}}{p^{2(r-1)}} \right) p_i k_l, \quad (50)$$

where we have used $k \ll p$. This is a modified Doppler term: while rotating the whole apparatus we get different Doppler terms.

7.7.3 Finslerian Geodesic Equation

In Finslerian space–time gravity cannot, in general, be transformed away. In [99] we discuss a Finslerian model of gravity by appropriately modifying the ansatz (45) for a Finslerian metric function

$$ds^2 = h_{00} dt^2 - \left(h_{i_1 i_2} \dots h_{i_{2r-1} i_{2r}} + \phi_{i_1 \dots i_{2r}} \right) dx^{i_1} \dots dx^{i_{2r}}, \quad (51)$$

which reduces to a Riemannian space–time for $\phi_{i_1 \dots i_{2r}} = 0$. For the case of a spherically symmetric Finsler space–time, it is possible to calculate the geodesic equation to first order in the Finslerian deviation $\phi_{i_1 \dots i_{2r}}$. We assumed for $h_{\mu\nu}$ the Schwarzschild form and found, for circular orbits, a modified Kepler law

$$\frac{r^3}{T^2} = \left(1 - \frac{A(r)}{r^4}\right) \frac{GM}{4\pi^2}, \quad (52)$$

where $A(r)$ is an arbitrary function, related to one component of the spherically symmetric tensor $\phi_{i_1 \dots i_{2r}}$.

For a radial free fall we obtain

$$\frac{d^2 r}{d\tau^2} = - \left(1 - B(r) \left(1 - \frac{2GM}{r}\right)^2\right) \frac{GM}{r^2}, \quad (53)$$

where τ is the proper time and $B(r)$ another function related to another component of the spherically symmetric tensor $\phi_{i_1 \dots i_{2r}}$. In the Newtonian approximation this gives

$$\frac{d^2 r}{dt^2} = - (1 - B(r)) \frac{GM}{r^2}. \quad (54)$$

Comparison of (52) with (54) reveals that radial motion and circular motion “feel” different gravitational constants, which, in general, may depend on the radial distance [99],

$$\frac{r^3}{T^2} = \frac{G_1 M}{4\pi^2}, \quad \frac{d^2 r}{dt^2} = - \frac{G_2 M}{r^2}. \quad (55)$$

The geodesic equation in Finsler space–time thus implies that the gravitational attraction of a body falling vertically towards the center of the Earth is different from the gravitational attraction that keeps a satellite on its bound orbit, see Fig. 6. From the orbit of the Earth around the Sun one can determine GM of the Sun with a relative accuracy of approximately 10^{-9} . This mass can be taken to determine the gravitational field of the Sun and the acceleration that bodies experience within standard theory. The acceleration of a satellite on a radial escape orbit can be measured with an accuracy of the order 10^{-10} m/s², which would allow a determination of GM of the Sun with an accuracy of the order 10^{-8} (at a distance of approximately 1 AU). As for the Earth, the gravitational acceleration of a body falling on Earth can be measured with an accuracy of 10^{-8} m/s² [91] leading to a relative accuracy of the determination of GM of the Earth of the order 10^{-9} . So, if all observations and measurements are compatible within standard theory, then the equality of the acceleration of horizontally moving satellites and planets and vertically falling

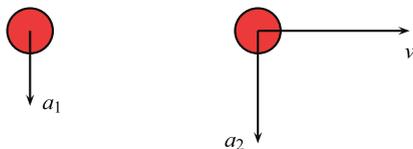


Fig. 6 A body falling toward the center of the Earth may feel a gravitation acceleration toward the center of the Earth different from that of a body moving horizontally

bodies is confirmed to within the order of 10^{-8} . As a consequence, the functions G_1 and G_2 , or A/r^4 and B , should differ by less than 10^{-8} .

It is clear from the given formulae that Finsler geometry offers the possibility of having different properties for escape and bound orbits (the gravitational attraction depends on the orbit) and, thus, is in the position to describe effects like the Pioneer anomaly; for example, a very simple choice in this case might be $A = 0$ and $B = B_0 r^2$ (assuming that the observed anomalous acceleration is of gravitational origin and not a systematic error). Further studies on experimental and observational consequences of Finsler gravity are in progress [99].

8 Summary

In this chapter, we have described the underlying principles of GR encoded in the EEP, and their corresponding experimental verification. We have also described observations relating to the predictions of GR, ranging from the weak field Solar system to strong field effects in compact binary systems. Besides the standard principles, we also focussed some attention on assumptions that are usually taken for granted, even though their experimental basis is sometimes not strong, or the interpretation of related experiments is not unique. These assumptions include charge conservation, equality of active and passive mass, charge, and magnetic moment, the order of the time derivative in classical and quantum equations of motion, and the issue of whether gravity can be transformed away locally.

Acknowledgements I would like to thank H. Dittus, V. Kagramanova, J. Kunz, D. Lorek, P. Rademaker, and V. Perlick for discussions and the German Aerospace Center DLR as well as the German Research Foundation and the Centre for Quantum Engineering and Space-Time Research QUEST for financial support.

References

1. A. Abramovici, Z. Vager, Phys. Rev. D **34**, 3240 (1986)
2. M. Abramowitz, I.A. Stegun (eds.), *Handbook of Mathematical Functions* (Dover Publications, New York, 1968)
3. S.L. Adler, Phys. Rev. D **79**, 023505 (2009)
4. Y. Aharonov, F.T. Avignone III, R.L. Brodzinski, J.I. Collar, E. Garcia, H.S. Miley, A. Morales, J. Morales, S. Nussinov, A. Ortiz de Solórzano, J. Puimedon, J.H. Reeves, C. Sáenz, A. Salinas, M.L. Sarsa, J.A. Villar, Phys. Lett. B **353**, 168 (1995)
5. M. Alcubierre, Class. Q. Grav. **11**, L73 (1994)
6. J. Alspector, G.R. Kalbfleisch, N. Baggett, E.C. Fowler, B.C. Barish, A. Bodek, D. Buchholz, F.J. Sciulli, E.J. Siskind, L. Stutte, H.E. Fisk, G. Krafczyk, D.L. Nease, O.D. Fackler, Phys. Rev. Lett. **36**, 837 (1976)
7. T. Alväger, F.J.M. Farley, J. Kjellmann, I. Wallin, Phys. Lett. **12**, 260 (1964)
8. G. Amelino-Camelia, C. Lämmerzahl, Class. Q. Grav. **21**, 899 (2004)
9. G. Amelino-Camelia, C. Lämmerzahl, A. Macias, H. Müller, in *Gravitation and Cosmology*, ed. by A. Macias, C. Lämmerzahl, D. Nuñez, AIP Conf. Proc. **758** (AIP, Melville, 2005) p. 30

10. J.D. Anderson, J.K. Campbell, J.E. Ekelund, J. Ellis, J.F. Jordan, *Phys. Rev. Lett.* **100**, 091102 (2008)
11. J.D. Anderson, P.A. Laing, E.L. Lau, A.S. Liu, M.M. Nieto, S.G. Turyshev, *Phys. Rev. Lett.* **81**, 2858 (1998)
12. J.D. Anderson, P.A. Laing, E.L. Lau, A.S. Liu, M.M. Nieto, S.G. Turyshev, *Phys. Rev. D* **65**, 082004 (2002)
13. P.S. Apostolopoulos, N. Broudzakis, N. Tetradis, E. Tsavara, *J. Cosm. Astrop. Phys.* **0606**, 009 (2006)
14. N. Ashby, T.P. Heavner, T.E. Parker, A.G. Radnaev, Y.O. Dudin, *Phys. Rev. Lett.* **98**, 070802 (2007)
15. J. Audretsch, *J. Phys. A* **14**, 411 (1981)
16. S. Baessler, B.R. Heckel, E.G. Adelberger, J.H. Gundlach, U. Schmidt, H.E. Swanson, *Phys. Rev. Lett.* **83**, 3585 (1999)
17. H. Balasin, D. Grumiller, *Int. J. Mod. Phys. D* **17**, 475 (2008)
18. D.F. Bartlett, D. van Buren, *Phys. Rev. Lett.* **57**, 21 (1986)
19. D. Bear, R.E. Stoner, R.L. Walsworth, V.A. Kostelecký, C.D. Lane, *Phys. Rev. Lett.* **85**, 5038 (2000)
20. D. Bear, R.E. Stoner, R.L. Walsworth, V.A. Kostelecký, C.D. Lane, *Phys. Rev. Lett.* **89**, 209902(E) (2002)
21. J.D. Bekenstein, *Phys. Rev. D* **70**, 083509 (2004)
22. J.F. Bell, T. Damour, *Class. Q. Grav.* **13**, 3121 (1996)
23. B. Bertotti, L. Iess, P. Tortora, *Nature* **425**, 374 (2003)
24. L. Blanchet, *Living Rev. Rel.* **9**, URL: <http://www.livingreviews.org/lrr-2006-4>
25. L. Blanchet, A. Buonanno, G. Faye, *Phys. Rev. D* **74**, 104034 (2006); Erratum, *ibidem* **75**, 049903(E) (2007); Erratum, *ibidem* **81**, 089901(E) (2010)
26. H. Bondi, *Rev. Mod. Phys.* **29**, 423 (1957)
27. K. Brecher, *Phys. Rev. Lett.* **39**, 1051 (1977)
28. A. Brilliet, J.L. Hall, *Phys. Rev. Lett.* **42**, 549 (1979)
29. B.C. Brown, G.E. Masek, T. Maung, E.S. Miller, H. Ruderman, W. Vernon, *Phys. Rev.* **30**, 763 (1973)
30. T. Buchert, *Gen. Rel. Grav.* **40**, 467 (2008)
31. T. Buchert, J. Ehlers, *Astron. Astrophys.* **320**, 1 (1997)
32. R.T. Cahill, *Prog. Phys.* **3**, 9 (2008)
33. M.N. Celerier, *Astron. Astrophys.* **353**, 63 (2000)
34. S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Oxford University Press, Oxford, 1983)
35. T.E. Chupp, R.J. Hoara, R.A. Loveman, E.R. Oteiza, J.M. Richardson, M.E. Wagshul, *Phys. Rev. Lett.* **63**, 1541 (1989)
36. I. Ciufolini, *Gen. Rel. Grav.* **36**, 2257 (2004)
37. F.I. Cooperstock, S. Tieu, arXiv:astro-ph/0507619v1 (2005)
38. T. Damour, M. Lilley, arXiv:0802.4169v1 [hep-th] (2008)
39. T. Damour, F. Piazza, G. Veneziano, *Phys. Rev. D* **66**, 046007 (2002)
40. T. Damour, F. Piazza, G. Veneziano, *Phys. Rev. Lett.* **89**, 081601 (2002)
41. T. Damour, G. Schäfer, *Phys. Rev. Lett.* **66**, 2549 (1991)
42. T. Dent, in *SUSY06: 14th Int. Conf. Supersymmetry and the Unification of Fundamental Interactions*, ed. by J.L. Feng, AIP Conf. Proc. **903** (AIP, Melville, 2007) p. 665
43. A. de Oliveira-Costa, M. Tegmark, M.J. Devlin, L. Page, A.D. Miller, C.B. Netterfield, Y. Xu, *Phys. Rev. D* **71**, 043004 (2005)
44. B.S. DeWitt, R.W. Brehme, *Ann. Phys. (N.Y.)* **9**, 220 (1960)
45. H. Dittus, C. Lämmerzahl, H. Selig, *Gen. Rel. Grav.* **36**, 571 (2004)
46. S.L. Dubovsky, V.A. Rubakov, arXiv:hep-th/0204205v1 (2002)
47. S.L. Dubovsky, V.A. Rubakov, P.G. Tinyakov, *J. High Energy Phys.* **08**, 041 (2000)
48. D. Dvali, G. Gabadadze, M. Porrati, *Phys. Lett. B* **485**, 208 (2000)
49. H.F. Dylla, J.G. King, *Phys. Rev. A* **7**, 1224 (1973)

50. J. Ehlers, F.A.E. Pirani, A. Schild, in *General Relativity, Papers in Honour of J.L. Synge*, ed. by L. O’Raifeartaigh (Clarendon, Oxford, 1972), p. 63
51. R. Emparan, H.S. Reall, *Living Rev. Rel.* **11**, URL: <http://www.livingreviews.org/lrr-2008-6>
52. C.W.F. Everitt, M. Adams, W. Bencze, S. Buchman, B. Clarke, J.W. Conklin, D.B. DeBra, M. Dolphin, M. Heifetz, D. Hipkins, T. Holmes, G.M. Kaiser, J. Kolodziejczak, J. Li, J. Lipa, J.M. Lockhart, J.C. Mester, B. Muhlfelder, Y. Oshima, B.W. Parkinson, M. Salomon, A. Silbergleit, V. Solmonik, K. Stahl, M. Taber, J.P. Turneure, S. Wang, P.W. Worden, in *Probing the Nature of Gravity: Confronting Theory and Experiments in Space*, ed. by C.W.F. Everitt, M.C.E. Huber, R. Kallenbach, G. Schäfer, B.F. Schutz, R.A. Treumann. Space Science Series of ISSI (Springer, New York, 2010), p. 53
53. G. Faye, L. Blanchet, A. Buonanno, *Phys. Rev. D* **74**, 104033 (2006)
54. T.M. Fortier, N. Ashby, J.C. Bergquist, M.J. Delaney, S.A. Diddams, T.P. Heavner, L. Hollberg, W.M. Itano, S.R. Jefferts, K. Kim, F. Levi, L. Lorini, W.H. Oskay, T.E. Parker, J. Shirley, J.E. Stalnaker, *Phys. Rev. Lett.* **98**, 070801 (2007)
55. M.B. Gerrard, T.J. Sumner, arXiv:0807.3158v3 [gr-qc] (2010)
56. J.A. Gonzalez, M.D. Hannam, U. Sperhake, B. Brügmann, S. Husa, *Phys. Rev. Lett.* **98**, 231101 (2007)
57. Gravity Probe B website, <http://einstein.stanford.edu/>
58. Z.G.T. Guiragossian, G.B. Rothbart, M.R. Yearian, R.A. Gearhart, J.J. Murray, *Phys. Rev. Lett.* **34**, 335 (1975)
59. J.H. Gundlach, S. Schlamminger, C.D. Spitzer, K.-Y. Choi, B.A. Woodahl, J.J. Coy, E. Fischbach, *Phys. Rev. Lett.* **98**, 150801 (2007)
60. E. Hackmann, V. Kagramanova, J. Kunz, C. Lämmerzahl, *Europhys. Lett.* **88**, 30008 (2009)
61. E. Hackmann, V. Kagramanova, J. Kunz, C. Lämmerzahl, *Phys. Rev. D* **81**, 044020 (2009)
62. E. Hackmann, C. Lämmerzahl, *Phys. Rev. D* **78**, 024035 (2008)
63. E. Hackmann, C. Lämmerzahl, *Phys. Rev. Lett.* **100**, 171101 (2008)
64. E. Hackmann, C. Lämmerzahl, A. Macias, in *New Trends in Statistical Physics: Festschrift in Honour of Leopoldo Garcia-Colin’s 80 Birthday* (World Scientific, Singapore, 2010), p. 301
65. Y. Hagihara, *Japan. J. Astron. Geophys.* **8**, 67 (1931)
66. J.B. Hartle, *Gravitation, an Introduction to Einstein’s General Relativity* (Addison Wesley, San Francisco, 2003)
67. M.P. Haugan, *Ann. Phys. (N.Y.)* **118**, 156 (1979)
68. S.W. Hawking, *Phys. Rev. D* **14**, 2460 (1976)
69. B.R. Heckel, in *CPT and Lorentz Symmetry III*, ed. by V.A. Kostelecký (World Scientific, Singapore, 2004), p. 133
70. F.W. Hehl, *Phys. Lett. A* **36**, 225 (1971)
71. S. Herrmann, A. Senger, K. Möhle, M. Nagel, E.V. Kovalchuk, A. Peters, *Phys. Rev. D* **80**, 105011 (2009)
72. L.-S. Hou, W.-T. Ni, Y.-C. Li, *Phys. Rev. Lett.* **90**, 201101 (2003)
73. C.-H. Hsieh, P.-Y. Jen, K.-L. Ko, K.-Y. Li, W.-T. Ni, S.-S. Pan, Y.-H. Shih, R.-J. Tyan, *Mod. Phys. Lett. A* **4**, 1597 (1989)
74. M.A. Humphrey, D.F. Phillips, E.M. Mattison, R.L. Walsworth, *Phys. Rev. A* **68**, 063807 (2003)
75. M.-T. Jaekel, S. Reynaud, *Mod. Phys. Lett. A* **20**, 1047 (2005)
76. M.-T. Jaekel, S. Reynaud, in *Lasers, Clocks and Drag-Free Control: Exploration of Relativistic Gravity in Space*, ed. by H. Dittus, C. Lämmerzahl, S. Turyshev (Springer, Berlin, 2007), p. 193
77. A.K. Jain, J.E. Lukens, J.-S. Tsai, *Phys. Rev. Lett.* **58**, 1165 (1987)
78. V. Kagramanova, J. Kunz, C. Lämmerzahl, *Class. Q. Grav.* **40**, 1249 (2008)
79. V. Kagramanova, J. Kunz, C. Lämmerzahl, *Gen. Rel. Grav.* **40**, 1249 (2008)
80. G.R. Kalbfleisch, N. Baggett, E.C. Fowler, J. Alsppector, *Phys. Rev. Lett.* **43**, 1361 (1979)
81. C. Kiefer, *Quantum Gravity* (Oxford University Press, Oxford, 2004)
82. B. Kleihaus, J. Kunz, F. Navarro-Lerida, in *Recent Developments in Gravitation and Cosmology*, ed. by A. Macias, C. Lämmerzahl, A. Camacho, *AIP Conf. Proc.* **977** (AIP, Melville, 2008) p. 94

83. V.A. Kostelecký, Phys. Rev. D **69**, 105009 (2004)
84. V.A. Kostelecký, M. Mewes, Phys. Rev. D **66**, 056005 (2002)
85. M. Kramer, I.H. Stairs, R.N. Manchester, M.A. MacLaughlin, A.G. Lyre, R.D. Ferdman, M. Burgag, D.R. Lorimer, A. Possenti, N. D'Amico, J. Sarkission, G.B. Hobbs, J.E. Reynolds, P.C.C. Freire, F. Camilo, Science **314**, 97 (2006)
86. M. Kramer, I.H. Stairs, R.N. Manchester, M.A. MacLaughlin, A.G. Lyre, R.D. Ferdman, M. Burgag, D.R. Lorimer, A. Possenti, N. D'Amico, J. Sarkission, B.C. Joshi, P.C.C. Freire, F. Camilo, Ann. Phys. (Leipzig) **15**, 34 (2006)
87. G.A. Krasinsky, V.A. Brumberg, Celest. Mech. Dyn. Astron. **90**, 267 (2004)
88. S. Krasnikov, Phys. Rev. D **65**, 064013 (2002)
89. S. Krasnikov, Phys. Rev. D **67**, 104013 (2003)
90. L.B. Kreuzer, Phys. Rev. **169**, 1007 (1968)
91. K. Kuroda, N. Mio, Phys. Rev. **42**, 3903 (1990)
92. C. Lämmerzahl, Eur. J. Phys. Special Topics **163**, 255 (2008)
93. C. Lämmerzahl, Ch.J. Bordé, in *Gyros, Clocks, and Interferometers: Testing Relativistic Gravity in Space*, ed. by C. Lämmerzahl, C.W.F. Everitt, F.W. Hehl (Springer, Berlin, 2001), p. 464
94. C. Lämmerzahl, C. Ciufolini, H. Dittus, L. Iorio, H. Müller, A. Peters, E. Samain, S. Scheithauer, S. Schiller, Gen. Rel. Grav. **36**, 2373 (2004)
95. C. Lämmerzahl, M.P. Haugan, Phys. Lett. A **282**, 223 (2001)
96. C. Lämmerzahl, D. Lorek, H. Dittus, Gen. Rel. Grav. **41**, 1345 (2008)
97. C. Lämmerzahl, A. Macias, H. Müller, Phys. Rev. D **71**, 025007 (2005)
98. C. Lämmerzahl, A. Macias, H. Müller, Phys. Rev. A **75**, 052104 (2007)
99. C. Lämmerzahl, V. Perlick, in preparation
100. C. Lämmerzahl, O. Preuss, H. Dittus, in *Lasers, Clocks and Drag-Free Control: Exploration of Relativistic Gravity in Space*, ed. by H. Dittus, C. Lämmerzahl, S.G. Turyshev, Lect. Notes Phys. **562** (Springer, Berlin, 2007), p. 75
101. C. Lämmerzahl, P. Rademaker, arXiv:0904.4779v1 [gr-qc] (2009)
102. S.K. Lamoreaux, Int. J. Mod. Phys. **7**, 6691 (1992)
103. S.K. Lamoreaux, J.P. Jacobs, B.R. Heckel, F.J. Raab, E.N. Fortson, Phys. Rev. Lett. **57**, 3125 (1986)
104. S. Landau, P.D. Sisterna, H. Vucetich, arXiv:gr-qc/0105025v1 (2001)
105. S. Landau, P.D. Sisterna, H. Vucetich, Phys. Rev. D **63**, 081303 (2001)
106. N. Li, D.J. Schwarz, EAS Publications Series **36**, 83 (2009)
107. N. Li, M. Seikel, D.J. Schwarz, Fortschr. Phys. **56**, 65 (2008)
108. N. Lockerbie, J.C. Mester, R. Torii, S. Vitale, P.W. Worden, in *Gyros, Clocks, Interferometers...: Testing Relativistic Gravity in Space*, ed. by C. Lämmerzahl, C.W.F. Everitt, F.W. Hehl, Lect. Notes Phys. **562** (Springer, Berlin, 2001), p. 213
109. M.J. Longo, Phys. Rev. D **36**, 3276 (1987)
110. D. Lorek, C. Lämmerzahl, in *Proc. 11th Marcel Grossmann Mtg*, Berlin 23–29 July 2006, ed. by H. Kleinert, R.T. Jantzen, R. Ruffini (World Scientific, Singapore, 2008), p. 2618
111. J. Louis, T. Mohaupt, S. Theisen, in *Approaches to Fundamental Physics*, ed. by I.-O. Stamatescu, E. Seiler, Lect. Notes Phys. **721** (Springer, Berlin, 2007), p. 289
112. A. Lue, Phys. Rep. **423**, 1 (2005)
113. L. Maleki, JPL report (2001)
114. H. Marion, F. Pereira Dos Santos, M. Abgrall, S. Zhang, Y. Sortas, S. Bize, I. Maksimovic, D. Calonico, J. Grünert, C. Mandache, P. Lemonde, G. Santarelli, Ph. Laurent, A. Clairon, C. Salomon, Phys. Rev. Lett. **90**, 150801 (2003)
115. B. Mashhoon, F. Gronwald, H.I.M. Lichtenegger, in *Gyros, Clock, Interferometers...: Testing Relativistic Gravity in Space*, ed. by C. Lämmerzahl, C.W.F. Everitt, F.W. Hehl, Lect. Notes Phys. **562** (Springer, Berlin, 2001), p. 83
116. D. Mattingly, Living Rev. Rel. **8**, URL (cited on 4 April 2006): <http://www.livingreviews.org/lrr-2005-5>
117. J.P. Mbelek, arXiv:0809.1888v3 [gr-qc] (2009)

118. M.E. McCulloch, J. Brit. Interpl. Soc. **61**, 373 (2008)
119. M.E. McCulloch, Mon. Not. R. Astron. Soc. **389**, L57 (2008)
120. M. Milgrom, New Astr. Rev. **46**, 741 (2002)
121. C.W. Misner, K. Thorne, J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973)
122. J.E. Moody, F. Wilczek, Phys. Rev. D **30**, 130 (1984)
123. M.S. Morris, K.S. Thorne, U. Yurtsever, Phys. Rev. Lett. **61**, 1446 (1988)
124. H. Müller, C. Braxmaier, S. Herrmann, A. Peters, C. Lämmerzahl, Phys. Rev. D **67**, 056006 (2003)
125. H. Müller, S.-W. Chiow, S. Herrmann, S. Chu, K.-Y. Chung, Phys. Rev. Lett. **100**, 031101 (2008)
126. F. Müller-Hoissen, in *Recent Developments in Gravitation and Cosmology*, ed. by A. Macias, C. Lämmerzahl, A. Camacho, AIP Conf. Proc. **977** (AIP, Melville, 2008), p. 12
127. G. Neugebauer, R. Meinel, Phys. Rev. Lett. **75**, 3046 (1995)
128. S. Nojiri, S.D. Odintsov, Int. J. Geom. Mod. Phys. **4**, 115 (2007)
129. J.O. Overduin, W. Priester, Naturwiss. **88**, 229 (2001)
130. P.J.E. Peebles, B. Ratra, Rev. Mod. Phys. **75**, 559 (2003)
131. S. Perlmutter, G. Aldering, G. Goldhaber, R.A. Knop, P. Nugent, P.G. Castro, S. Deustua, S. Fabbro, A. Goobar, D.E. Groom, I.M. Hook, A.G. Kim, M.Y. Kim, J.C. Lee, N.J. Nunes, R. Pain, C.R. Pennypacker, R. Quimby, C. Lidman, R.S. Ellis, M. Irwin, R.G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B.J. Boyle, A.V. Filippenko, T. Matheson, A.S. Fruchter, N. Panagia, H.J.M. Newberg, W.J. Couch, The Supernova Cosmology Project, Astrophys. J. **517**, 565 (1999)
132. W. Petry, arXiv:0806.0334v1 [physics.gen-ph] (2008)
133. E.V. Pitjeva, Astron. Lett. **31**, 340 (2005)
134. R.V. Pound, G.A. Rebka, Phys. Rev. Lett. **4**, 337 (1960)
135. J.D. Prestage, J.J. Bollinger, W.M. Itano, D.J. Wineland, Phys. Rev. Lett. **54**, 2387 (1985)
136. R.D. Reasenberg, I.I. Shapiro, P.E. MacNeil, R.B. Goldstein, J.C. Breidenthal, J.P. Brenkle, D.L. Cain, T.M. Kaufman, T.A. Komarek, A.I. Zygierbaum, Astrophys. J. Lett. **234**, L219 (1979)
137. A.G. Riess, A.V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P.M. Garnavich, R.L. Gilliland, C.J. Hogan, S. Jha, R.P. Kirshner, B. Leibundgut, M.M. Phillips, D. Reiss, B.P. Schmidt, R.A. Schommer, R.C. Smith, J. Spyromilio, C. Stubbs, N.B. Suntzeff, J. Tonry, Astron. J. **116**, 1009 (1998)
138. B. Rievers, C. Lämmerzahl, S. Bremer, M. List, H. Dittus, New J. Phys. **11**, 113032 (2009)
139. W. Rindler, *Relativity* (Oxford University Press, Oxford, 2001)
140. C. Rovelli, *Quantum Gravity, Cambridge Monographs on Mathematical Physics* (Cambridge University Press, Cambridge, 2004)
141. R.H. Sanders, Astron. Astrophys. **136**, L21 (1984)
142. R.H. Sanders, S.S. McGough, Ann. Rev. Astron. Astrophys. **40**, 263 (2002)
143. B.E. Schaefer, Phys. Rev. Lett. **82**, 4964 (1999)
144. G. Schäfer, R. Sauerbrey, arXiv:astro-ph/9805106v1 (1998)
145. S. Schlamminger, K.-Y. Choi, T.A. Wagner, H. Gundlach, E.A. Adelberger, Phys. Rev. Lett. **100**, 041101 (2008)
146. R. Schoedel, A. Eckart, T. Alexander, D. Merritt, R. Genzel, A. Sternberg, L. Meyer, F. Kul, J. Moulta, T. Ott, C. Straubmeier, Astron. Astrophys. **469**, 125 (2007)
147. R. Schützhold, G. Schaller, D. Habs, Phys. Rev. Lett. **97**, 121302 (2006)
148. D.J. Schwarz, G.D. Starkman, D. Huterer, C.J. Copi, Phys. Rev. Lett. **93**, 221301 (2004)
149. R. Sexl, H. Urbantke, *Relativity, Groups, Particles* (Springer, Wien, 2001)
150. I.I. Shapiro, Phys. Rev. Lett. **13**, 789 (1964)
151. S.S. Shapiro, J.L. Davis, D.E. Lebach, J.S. Gregory, Phys. Rev. Lett. **92**, 121101 (2004)
152. D.N. Spergel, R. Bean, O. Doré, M.R. Nolta, C.L. Bennett, J. Dunkley, G. Hinshaw, N. Jarosik, E. Komatsu, L. Page, H.V. Peiris, L. Verde, M. Halpern, R.S. Hill, A. Kogut, M. Limon, S.S. Meyer, N. Odegard, G.S. Tucker, J.L. Weiland, E. Wollack, E.L. Wright, Astrophys. J. **170**, 377 (2007)

153. E.M. Standish, in *Transits of Venus: New Views of the Solar System and Galaxy, Proceedings IAU Colloquium No. 196*, ed. by D.W. Kurtz (Cambridge University Press, Cambridge, 2005), p. 163
154. E.M. Standish, in *Recent Developments in Gravitation and Cosmology*, ed. by A. Macias, C. Lämmerzahl, A. Camacho, AIP Conf. Proc. **977** (AIP, Melville, 2008), p. 254
155. R.I. Steinberg, K. Kwiatkowski, W. Maenhaut, N.S. Wall, Phys. Rev. D **12**, 2582 (1975)
156. J. Steinhoff, G. Schäfer, S. Hergt, Phys. Rev. D **77**, 104018 (2008)
157. L. Stodolsky, Phys. Lett. B **201**, 353 (1988)
158. T. Sumner, Living Rev. Rel. **5**, URL (cited on 20 November 2005): <http://www.livingreviews.org/lrr-2002-4>
159. M. Tegmark, M.A. Strauss, M.R. Blanton, K. Abazajian, S. Dodelson, H. Sandvik, X. Wang, D.H. Weinberg, I. Zehavi, N.A. Bahcall, F. Hoyle, D. Schlegel, R. Scoccimarro, M.S. Vogele, A. Berlind, T. Budavari, A. Connolly, D.J. Eisenstein, D. Finkbeiner, J.A. Frieman, J.E. Gunn, L. Hui, B. Jain, D. Johnston, S. Kent, H. Lin, R. Nakajima, R.C. Nichol, J.P. Ostriker, A. Pope, R. Scranton, U. Seljak, R.K. Sheth, A. Stebbins, A.S. Szalay, I. Szapudi, Y. Xu, J. Annis, J. Brinkmann, S. Burles, F.J. Castander, I. Csabai, J. Loveday, M. Doi, M. Fukugita, B. Gillespie, G. Hennessy, D.W. Hogg, Z. Ivezic, G.R. Knapp, D.Q. Lamb, B.C. Lee, R.H. Lupton, T.A. McKay, P. Kunszt, J.A. Munn, L. O'Connell, J. Peoples, J.R. Pier, M. Richmond, C. Rockosi, D.P. Schneider, C. Stoughton, D.L. Tucker, D.E. van den Berk, B. Yanny, D.G. York, Phys. Rev. D **69**, 103501 (2004)
160. P. Touboul, in *Gyros, Clocks, Interferometers...: Testing Relativistic Gravity in Space*, ed. by C. Lämmerzahl, C.W.F. Everitt, F.W. Hehl, Lect. Notes Phys. **562** (Springer, Berlin, 2001), p. 274
161. M.J. Valtonen, H.J. Lehto, K. Nilsson, J. Heidt, L.O. Takalo, A. Sillanpää, C. Villforth, M. Kidger, G. Poyner, T. Pursimo, S. Zola, J.-H. Wu, X. Zhou, K. Sadakane, M. Drozd, D. Koziel, D. Marchev, W. Ogloza, C. Porowski, M. Siwak, G. Stachowski, M. Winiarski, V.-P. Hentunen, M. Nissinen, A. Liakos, S. Dogru, Nature **452**, 851 (2008)
162. C. van de Bruck, W. Priester, in *Dark Matter in Astrophysics and Particle Physics*, Proc. 2nd Int. Conf. on Dark Matter in Astrophysics and Particle Physics, Heidelberg 20–25 July, 1998, ed. by H.V. Klapdor-Kleingrothaus, L. Baudis (Institute of Physics, Philadelphia, 1999), p. 181
163. R.A. Vanderveld, E.E. Flanagan, I. Wasserman, Phys. Rev. D **74**, 023506 (2006)
164. R.F.C. Vessot, M.W. Levine, E.M. Mattison, E.L. Blomberg, T.E. Hoffmann, G.U. Nystrom, B.F. Farrel, R. Decher, P.B. Eby, C.R. Baugher, J.W. Watts, D.L. Teuber, F.D. Wills, Phys. Rev. Lett. **45**, 2081 (1980)
165. R.L. Walsworth, in *Special Relativity*, ed. by J. Ehlers, C. Lämmerzahl, Lect. Notes Phys. **702** (Springer, Berlin, 2006), p. 493
166. S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972)
167. C. Wetterich, Phys. Lett. B **561**, 10 (2003)
168. C.M. Will, *Theory and Experiment in Gravitational Physics*, revised edition (Cambridge University Press, Cambridge, 1993)
169. P. Wolf, Ch.J. Bordé, A. Clairon, L. Duchayne, A. Landragin, P. Lemonde, G. Santarelli, W. Ertmer, E. Rasel, F.S. Cataliotti, M. Inguscio, G.M. Tino, P. Gill, H. Klein, S. Reynaud, C. Salomon, E. Peik, O. Bertolami, P. Gil, J. Páramos, C. Jentsch, U. Johann, A. Rathke, P. Bouyer, L. Cacciapuoti, D. Izzo, P. De Natale, B. Christophe, P. Touboul, S.G. Turyshev, J. Anderson, M.E. Tobar, F. Schmidt-Kaler, J. Vigué, A.A. Madej, L. Marmet, M.-C. Angonin, P. Delva, P. Tourrenc, G. Metris, H. Müller, R. Walsworth, Z.H. Lu, L.J. Wang, K. Bongs, A. Toncelli, M. Tonelli, H. Dittus, C. Lämmerzahl, G. Galzerano, P. Laporta, J. Laskar, A. Fienga, F. Roques, K. Sengstock, Exp. Astron. **23**, 651 (2009)
170. A. Yu. Ignatiev, Phys. Rev. Lett. **98**, 101101 (2007)
171. T. van Zoest, N. Gaaloul, Y. Singh, H. Ahlers, W. Herr, S.T. Seidel, W. Ertmer, E. Rasel, M. Eckart, E. Kajari, S. Arnold, G. Nandi, W.P. Schleich, R. Walser, A. Vogel, K. Sengstock, K. Bongs, W. Lewoczko-Adamczyk, M. Schiemangk, T. Schuldt, A. Peters, T. Königmann, H. Müntinga, C. Lämmerzahl, H. Dittus, T. Steinmetz, T.W. Hänsch, J. Reichel, Science **328**, 1540 (2010)

Mass Metrology and the International System of Units (SI)

Richard S. Davis

Abstract The International System of Units (SI) is widely used in science, industry, and commerce because it caters simultaneously to the needs of all. In the early twenty-first century, this means defining the units of time, length, mass, and electricity in terms of the fundamental constants of physics, and then “realizing” these definitions to sufficient accuracy on the human scale of the second, meter, kilogram, and ampere. This program has already been successful except for the kilogram, which is still defined in terms of an artifact constructed in the late nineteenth century. Although quantum-based electrical standards are widely used, the SI voltages or resistances produced by these standards depend on the values of constants that are at present based on experimental values derived from the artifact kilogram. This chapter presents the current state of affairs, which is unsatisfactory, and proceeds to describe work that will lead to a redefinition of the kilogram, probably in terms of a fixed value for the Planck constant.

1 Introduction

What does it mean to measure a mass? In this chapter we present two possibilities that will be discussed in detail. The first has been used since antiquity: choose one object as the standard, S , and then measure the ratio of the mass, $m(X)$, to $m(S)$ for any X . The mass of X is then said to be $m(X)/m(S)$ in units of $m(S)$. An object known as the international prototype of the kilogram serves to define the unit “kilogram” in the SI. Secondary mass standards used throughout the world are ultimately traceable to the international prototype, which is conserved and used at the International Bureau of Weights and Measures (BIPM). This system works to a few parts in 10^8 at the 1 kg level. Experimental uncertainties generally increase if mass determinations very different from 1 kg are required. It would be preferable if

R.S. Davis (✉)

Bureau International des Poids et Mesures, Pavillon de Breteuil, 92312 Sèvres cedex, France

e-mail: rdavis@bipm.org

$m(S)$ were an invariant, or “fundamental,” quantity such as the mass of the electron, m_e . However, there is a mismatch of 30 orders of magnitude between 1 kg and m_e and this presents practical challenges.

A second way to measure the mass of X is through its energy equivalent: $m(X) = E/c_0^2$. We might then make use of the relation $E = hf$ to devise a nondestructive measurement of the mass of X in terms of its Compton frequency:

$$m(X) = (h/c_0^2) f. \quad (1)$$

A variant of 1 has been tested experimentally to about 500 parts in 10^9 [37], where $m(X)$ was the mass equivalent of a nuclear binding energy. However, if X is of the order of 1 kg then $f \approx 10^{50}$ Hz, which seems to be experimentally inaccessible.

The following describes how the world of metrology is meeting the challenge to redefine the kilogram in terms of fundamental constants.

2 The SI

2.1 Base Units/Base Quantities

The SI has seven base units, each of which associated with a base quantity [3]. These are the second, s (time); the meter, m (length); the kilogram, kg (mass); the ampere, A (electrical current); the kelvin, K (temperature); the mole, mol (amount of substance); and the candela, cd (luminous intensity). The kelvin and the candela will not be discussed further in this chapter. The SI has evolved from the meter, kilogram, second, ampere (MKSA) system [3], and has a formal mechanism for evolving as our knowledge grows. This has important practical consequences as will be discussed below in Section 5.

It is obviously not the goal of the SI to define a minimal set of base units, for clearly this has not been achieved [15]. Rather the SI attempts to be useful to the greatest number of communities so that, for instance, merchants can measure bolts of cloth and industries can source precision parts in the same units used by physicists to measure the Bohr radius, and all can achieve an accuracy that is not limited by the definition of the meter. Let us now look briefly at the current definitions of the five base units of interest here. Each definition fixes the value of some physical property or constant.

The present definition of the *second*, which dates from 1968, specifies an exact frequency, $\nu_{\text{hfs}}(\text{Cs})$, for the hyperfine splitting in the ground state of a caesium-133 atom (it is understood that the atom is at rest at a temperature of 0 K). The definition anticipates that the second will be “realized” using an atomic clock. The value of the hyperfine splitting is indeed a constant of nature, although the theoretical model for alkali atoms and ions is less well developed than for atomic hydrogen [33, 35]. Note that this definition is now more than 40 years old and technology has advanced greatly. The hfs of Cs-133 is perhaps no longer the most judicious choice on which to base the second, but such speculation is beyond the scope of this article.

The present definition of the *meter* essentially defines a fixed value for c_0 , the speed of light in vacuum. The definition states that the meter is the distance traveled by light in a vacuum during a duration of $(1/\{c_0\})[s]$. Here we use the curly brackets of quantity calculus to indicate the numerical value of c_0 , devoid of units; the unit itself is given in square brackets [12]. The definition anticipates that high-accuracy length measurements will be performed using masers or lasers. The present meter definition dates from 1983 and it is instructive to understand why 15 years elapsed from the redefinition of the second to the redefinition of the meter. It was necessary to demonstrate a practical way to “realize” the new definition and to show that this realization leads to improved length measurements. Essentially this meant measuring a laser frequency in units provided by an atomic clock. Results from the key experiment, which reported an improved measurement of c_0 , were published about 35 years ago [17]. The authors stated that “The main limitation [to the determination of c_0] is asymmetry of the krypton...line defining the meter.” The meter had been defined in 1960 to be a certain multiple of a particular krypton emission line. Even though this definition relied on the property of an atom, there was a technical limitation to the experimental determination of a much more fundamental physical quantity. This was the principal motivation for the redefinition, which, for administrative reasons of the type outlined in Section 6.1, came some years later. The long history of ever-improving measurements of c_0 was effectively brought to an end in 1983 by the present definition of the meter.

The definition of the *kilogram* has not changed since the foundations of the SI were laid in 1889. Since the SI unit “kilogram” and its related quantity “mass” are the main subject of this chapter, the kilogram definition will be given in full [3]:

The kilogram is the unit of mass. It is equal to the mass of the international prototype of the kilogram.

This is a quintessential artifact definition. In the 1880s a number of similar objects were fashioned from a particular binary alloy (Pt90%/Ir10%) and one of these objects was selected to represent 1 kg, while maintaining historical continuity with the previous artifact representation of the kilogram [10]. Whereas the definitions of the second and the meter rely on fixed values for physical constants, an artifact definition cannot do this. All the definition of the kilogram can accomplish is to assign a mass of exactly 1 kg to the mass of a particular object. If the mass of that artifact changes (e.g., relative to the electron rest mass), its SI value nevertheless remains exactly 1 kg. The mass of any object X is the ratio of its mass to the mass, m_0 , of the international kilogram artifact:

$$m_X = \left\{ \frac{m_X}{m_0} \right\} [\text{kg}]. \quad (2)$$

The latter is assigned a value of exactly 1 kg in the SI.

Two more SI base units, the ampere and the mole, also depend on the kilogram. All definitions are quoted from [3]:

The ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 m apart in vacuum, would produce between these conductors a force equal to 2×10^{-7} newton per metre of length.

The newton is the SI unit of force, derived from the kilogram, meter and second. One may note that the ampere definition implicitly fixes the value of the magnetic constant, μ_0 , to be exactly $4\pi \times 10^{-7} \text{N/A}^2$. Recall that

$$c_0^2 \epsilon_0 \mu_0 = 1, \quad (3)$$

where ϵ_0 is the electric constant that, by inspection, also has a fixed value in the SI. The ampere definition essentially describes a *Gedanken* experiment, impossible to realize in practice. Nevertheless, laboratory experiments can exploit well-known physical principles to realize the ampere definition with a practical geometry. The ampere definition ensures coherence between electrical and mechanical units by assigning a fixed value to a unit with dimensions N/A^2 .

Finally the *mole* also relies on the kilogram:

The mole is the amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kg of carbon 12.

It is understood that the definition refers to an atom that is free and at rest in its ground state. The mole is principally of interest to chemists but it is indeed a base unit of the SI and its present definition depends on the kilogram. The number of elementary entities per mole referred to in the definition is of course the Avogadro constant, N_A :

$$N_A = \frac{0.012 \text{ kg/mol}}{m(^{12}\text{C})}. \quad (4)$$

Thus the SI value of the Avogadro constant has the same relative uncertainty as the mass of an atom of ^{12}C . We will present much more on this issue in Section 2. A useful history of the mole and its utility to chemistry can be found in [26].

2.2 Gaussian Units

The SI defines units only. Their associated quantities are defined through standard physics. Nevertheless, the equations of physics – especially those of electrodynamics – take different forms in different unit systems. An analysis of the differences between the SI and other commonly used systems of units, most notably the cgs-Gaussian units, is given by J.D. Jackson in his deservedly well-known textbook [20]. Jackson has chosen to use SI in all but the final chapters of his latest edition, and his reasons for this choice are instructive. He concludes that “the reality is that scientists must be conversant in many languages....”

2.3 Planck Units, Natural Units, and Atomic Units

The Planck units are perhaps the best known of the natural unit systems, and for this reason provide a useful contrast to the SI. In usual formulations, the Planck units correspond to the quantities mass, length, time, and temperature. The sizes of the Planck units, relative to their respective SI counterparts, are thought by some to be insightful. Best values of the Planck units are periodically tabulated by the CODATA Task Group on Fundamental Constants [31, 32]. In addition, Planck units provide a useful context for viewing proposals to update the SI. We recall that Planck units are constructed from the following physical constants: Newtonian constant of gravitation (G), Planck constant (h) divided by 2π , the Boltzmann constant (k), and c_0 . If μ_0 is added to this list, one can also derive a Planck unit of electrical charge. We will see that the proposals for a new SI go a long way to defining units in terms of fixed values for certain fundamental constants, some of them already in the Planck set. This will be discussed further in Section 6.2.

Other unit systems based on fundamental constants are also used, in particular the so-called natural units and the atomic units [31]. In both of these systems, the electron mass, m_e , serves to define the unit of mass.

3 Practical Reasons for Redefining the Kilogram

As discussed above, there are excellent reasons in principle for basing a unit system on physical constants rather than artifacts. In this section we discuss practical reasons for redefining the kilogram as soon as possible. Such motivation comes from three major areas: internal evidence from mass comparisons among 1 kg standards, possible confusion in the analysis and use of fundamental constants, and adoption of conventional units for electrical metrology (which is now based on quantum standards) and chemistry (which relies heavily on the ^{12}C atom as a unit of mass). We now examine each of these in detail.

3.1 Internal Evidence Among 1 kg Artifact Mass Standards

When the international prototype was put into service in 1889, some 40 additional copies had also been manufactured, each of the same material, shape, surface finish, and mass (to within a manufacturing tolerance of ± 1 mg). Later, additional copies were manufactured and the number is now approaching 100. Six of these are official copies that are stored at the BIPM along with the international prototype itself. Most of the others have become national 1-kg prototypes of the Member States of the BIPM. In the SI, all mass values are traceable to the international prototype. In practice, the set of the oldest prototypes has been calibrated just three times: the first during the years just prior to adoption of the kilogram definition in 1889,

the second over several years roughly centered around 1950, and the third from 1989 to 1991. The latter two campaigns are known as the second and third verifications. The history of this work is reviewed in [10]. Additional prototypes maintained at the BIPM are used to maintain traceability to the international prototype between verifications.

Comparisons to the international prototype can be made directly or indirectly by means of sensitive balances known as mass comparators. These already existed in the 1880s although their standard deviations and convenience of use are now much improved. Nevertheless, 100 years ago metrologists were prepared to make hundreds of painstaking measurements in order to reduce the random uncertainty of the final result to about $5\mu\text{g}$, or 5 parts in 10^9 (5 ppb) of 1 kg. Essentially, a comparator can be used to determine mass ratios. The smallest uncertainties are obtained when the physical properties of the artifacts being compared are as close to identical as possible. Rather than using Eq. 2, traceability to the international prototype is obtained through a chain that is more or less long depending on the requirements of the end user:

$$\{m_X\}[\text{kg}] = \left\{ \frac{m_X}{m_n} \right\} \cdot \left\{ \frac{m_n}{m_{n-1}} \right\} \cdots \left\{ \frac{m_2}{m_1} \right\} \cdot \left\{ \frac{m_1}{m_0} \right\} [\text{kg}]. \quad (5)$$

Simply put, mass metrology consists of minimizing the uncertainty of the ratio measurements shown on the right-hand side of Eq. 5 and ensuring that the masses of the artifacts that appear in both numerator and denominator (known as transfer standards) are stable. The uncertainties to which national metrology institutes adhering to the CIPM Mutual Recognition Arrangement can disseminate the unit of mass can be found on the BIPM Web site [4]. Secondary laboratories can be traceable to the SI through these laboratories. A system of laboratory accreditation is available, for example, through the International Laboratory Accreditation Cooperation [21].

Although m_0 is by definition 1 kg, we know that its mass could change with respect to a fundamental constant such as the electron mass. However, we have as yet no experimental evidence for a change in mass of m_0 with respect to a fundamental constant [11], which is another way of saying that it is difficult to determine with sufficient accuracy the ratio of m_0 to the mass of a fundamental constant.

Nevertheless, we have ample internal evidence that the masses of the set of prototypes are slightly unstable among themselves. This was suspected from the results of the second verification and confirmed by the third verification. Results for the oldest national prototypes having a complete calibration history over 100 years are shown in Fig. 1. The x -axis represents the mass of the international prototype and the straight lines represent the changes in calibration of each prototype from the time of its initial calibration through the third verification. The three experimental points are connected by straight lines to help visibility. Thus each of the three calibrations of a prototype X is plotted in Fig. 1 as

$$\left(\left(\frac{m_X}{m_0} \right)_t - \left(\frac{m_X}{m_0} \right)_{t=0} \right) / 10^{-9}, \quad (6)$$

where t is the number of years since 1889. All prototypes were cleaned and washed prior to calibration. One sees a general trend for the masses of the national

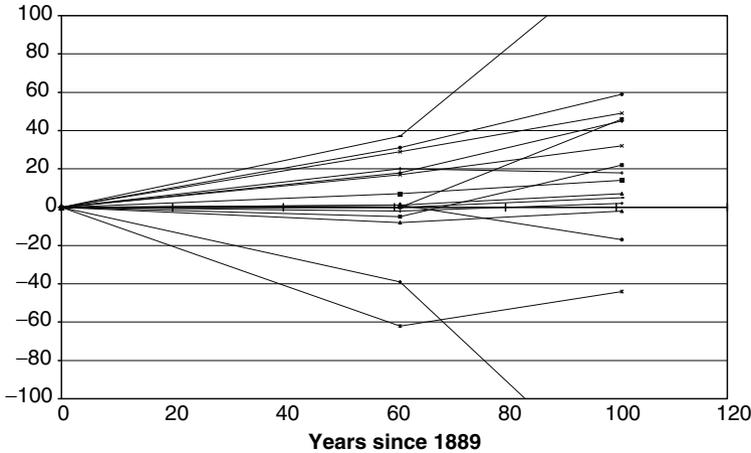


Fig. 1 Changes in mass since their initial calibration of the oldest national prototypes with respect to the international prototype. The y-axis is in ppb ($\mu\text{g}/\text{kg}$)

prototypes to increase by about 0.5 ppb/year with respect to the international prototype. Does this mean that the international prototype is losing mass (as one can read in the mass media)? Perhaps the national prototypes are generally gaining mass; or perhaps none of these objects are stable with respect to a fundamental constant of mass. By 1991, after the results of the third verification were known, it was suggested that experiments to link the mass of the international prototype to an atomic mass or a physical constant should be pursued with a goal of achieving an uncertainty of 20 ppb in order to monitor the stability of the international prototype [36]. This suggestion was then endorsed by the relevant international bodies, see Section 5. However, advances in quantum physics have more recently led to a call from other areas of metrology for a major overhaul of the SI, including – and especially – a redefinition of the kilogram.

3.2 Fundamental Constants

In presenting the Planck units, natural units, and atomic units, it has been assumed that there are certain physical quantities that are fundamental and, therefore, ideal standards from which to create a SI. An often-quoted statement by James Clerk Maxwell makes a prescient and eloquent argument for basing units on the physical constants:

The dimensions of our earth and its time of rotation, though relative to our present means of comparison very permanent, are not so by physical necessity. The earth might contract by cooling, or it might be enlarged by a layer of meteorites falling on it, or its rate of revolution might slowly slacken, and yet it would continue to be as much a planet as before.

But a molecule, say of hydrogen, if either its mass or its time of vibration were to be altered in the least, would no longer be a molecule of hydrogen.

If, then, we wish to obtain standards of length, time and mass which shall be absolutely permanent, we must seek them not in the dimensions, or the motion, or the mass of our planet, but in the wavelength, the period of vibration, and the absolute mass of these imperishable and unalterable and perfectly similar molecules [25].

Maxwell was, of course, criticizing the original metric system, which based the meter on a specified fraction of the earth's circumference and the kilogram on the mass of a cubic decimeter of water at its temperature of maximum density. The ghost of this system is seen in our present kilogram, which agrees with this earlier definition to within about 3 parts in 10^5 . However, "water" is not a well-defined substance and the original definition was replaced by one based on a single, solid artifact: first the *kilogramme des Archives*, and then the present international prototype [10]. In some sense the density maximum of a well-characterized sample of water is a physical constant but we no longer consider it to be "fundamental." This is because its properties are influenced by a number of effects that are difficult to model to arbitrarily high accuracy: isotopic abundances, dissolved gases and other impurities, thermal expansion, compressibility, etc. The maximum density of a particular isotopic mix of water has been determined to a relative uncertainty of about 10^{-6} and this limit is not due to shortcomings of the present definition of the kilogram.

Our understanding of which quantities in nature are fundamental constants evolves with our knowledge. For instance, the fine-structure constant, α , which today is determined by an experiment whose results are analyzed using QED perturbation theory [18], may one day be calculable from first principles. A possible analogy to the value of π , which in antiquity was determined by measurement, is sometimes cited. We may one day find that α , or other constants, are time - dependent (and thus not really constant) [24, 35]. It may be that string theories will lead us to revise our notions of "fundamental" constants. This paper will not enter the debate over which constants are the most fundamental. In the following, we will assume that the fundamental constants at our disposal are those that are listed in the CODATA 2006 recommendation [31].

In any case, the values of all fundamental constants containing the kilogram in their dimension must be traceable to the international prototype. How this situation is best remedied will be the subject of this chapter. Of course at some level of precision the mass of the international prototype must be less stable than the values of the constants that are traceable to it. However, this phenomenon has not yet been observed. A more practical concern is the experimental uncertainty in determining such constants with respect to the present definition of the kilogram. Thus, for instance, every time there is an improved experimental determination of the Planck constant, its SI value changes (within the previously accepted uncertainty, one hopes) and the uncertainty of the new value is improved. We will see below that the recommended relative uncertainty of h is approximately 5×10^{-8} . The relative uncertainty of the electron rest mass, m_e , is also about 5×10^{-8} but the

Table 1 The variances, covariances, and correlation coefficients of the values of a selected group of constants based on the 2006 CODATA adjustment. The numbers in bold above the main diagonal are 10^{16} times the numerical values of the relative covariances; the numbers in bold on the main diagonal are 10^{16} times the numerical values of the relative variances; and the numbers in italics below the main diagonal are the correlation coefficients^a (Table L of [31], used with permission)

	α	h	e	m_e	N_A	m_e/m_μ	F
α	0.0047	0.0002	0.0024	-0.0092	0.0092	-0.0092	0.0116
h	<i>0.0005</i>	24.8614	12.4308	24.8611	-24.8610	-0.0003	-12.4302
e	<i>0.0142</i>	<i>0.9999</i>	6.2166	12.4259	-12.4259	-0.0048	-6.2093
m_e	<i>-0.0269</i>	<i>0.9996</i>	<i>0.9992</i>	24.8795	-24.8794	0.0180	-12.4535
N_A	<i>0.0269</i>	<i>-0.9996</i>	<i>-0.9991</i>	<i>-1.0000</i>	24.8811	-0.0180	12.4552
m_e/m_μ	<i>-0.0528</i>	<i>0.0000</i>	<i>-0.0008</i>	<i>0.0014</i>	<i>-0.0014</i>	6.4296	-0.0227
F	<i>0.0679</i>	<i>-0.9975</i>	<i>-0.9965</i>	<i>-0.9990</i>	<i>0.9991</i>	<i>-0.0036</i>	6.2459

^aThe relative covariance is $u_r(x_i, x_j) = u(x_i, x_j)/(x_i x_j)$, where $u(x_i, x_j)$ is the covariance of x_i and x_j ; the relative variance is $u_r^2(x_i) = u_r(x_i, x_i)$; and the correlation coefficient is $r(x_i, x_j) = u(x_i, x_j)/[u(x_i)u(x_j)]$.

relative uncertainty of h/m_e is orders of magnitude lower than either of these. Why? Dimensionally, h/m_e is independent of the unit of mass and thus the definition of the kilogram is irrelevant to this ratio.

In addition to a list of fundamental constants with recommended values and uncertainties, the authors of CODATA 2006 also provide a covariance matrix to handle the uncertainties of combinations of constants whose values are correlated – chiefly by their traceability to a macroscopic kilogram artifact. The relevant table in CODATA 2006 is reproduced below as Table 1. The correlation coefficients that are nearly +1 or -1 are due to traceability to the artifact kilogram. Mills et al. [28] point out that redefining the kilogram in terms of fixed values for h , c_0 , and $\nu_{\text{hfs}}(\text{Cs})$ would help create a quantum SI, in accord with currently accepted physics and would thus lead to dramatic changes in Table 1. The same authors also propose defining the ampere in terms of fixed values for e and $\nu_{\text{hfs}}(\text{Cs})$, the mole in terms of a fixed number of entities with no reference to the kilogram, and the kelvin in terms of a fixed value of the Boltzmann constant. Mohr et al. [31] conclude that such redefinitions would represent “a significant advance in our knowledge of the values of the constants.” Although the meaning of knowledge in this context might be debated by epistemologists, the ideas of Mills et al. have undoubtedly been a motor for change.

3.3 Electrical Metrology

In the second half of the last century, the precise measurement of electrical quantities was revolutionized by two quantum mechanical effects. The first of these was discovered by Brian Josephson, who in due course received the Nobel Prize. Voltage standards based on the Josephson effect are known as Josephson junctions or Josephson arrays [19]. In essence, such devices are frequency-to-voltage transducers:

$$V(n) = \frac{nf}{K_J}, \quad (7)$$

where f is a microwave frequency and K_J is the ‘‘Josephson constant.’’ The voltage characteristic of the device is quantized in discrete steps. In Eq. 7 n is the integer value of the n th step. A remarkable feature of these devices is that K_J depends only on fundamental constants:

$$K_J = \frac{2e}{h}. \quad (8)$$

Over the years since Josephson’s theoretical prediction, there has been neither theory nor experiment to suggest that Eq. 8 is an approximation or that there is some missing, device-dependent correction. On the contrary, the relative difference in voltage between two Josephson devices illuminated by the same frequency and biased to the same step has been measured to be less than 10^{-18} and this constitutes an excellent experimental test of the strong equivalence principle as applied to charged particles in a gravitational field [22].

In a second development, Klaus von Klitzing discovered the quantized Hall effect (QHE) in semiconductors at very low cryogenic temperature and very high magnetic induction. In the conventional Hall effect, an electric current, I , passes along a sample in the presence of an external magnetic induction perpendicular to the current. A voltage, V_H (the Hall voltage), appears perpendicular to both the current and the induction.

The classical Hall probe is used as a transducer that converts magnetic induction to resistance V_H/I , with a proportionality constant that is material dependent.

The QHE is quite different [19, 41]. Without going into detail, it is sufficient to state that the quantum Hall (QH) resistance, R_H , for any QH device is given by

$$R_H(i) = \frac{R_K}{i}, \quad (9)$$

where i is a small integer and R_K is known as the von Klitzing constant:

$$R_K = \frac{h}{e^2}. \quad (10)$$

There are two remarkable points to be made about Eq. 10 and von Klitzing made them both in his Nobel lecture. The first is that R_K is device independent: ‘‘...one may come to the conclusion that such a complicated system like a semiconductor is not useful for very fundamental discoveries. Indeed, most of the experimental data in solid state physics are analyzed on the basis of simplified theories...’’ [40]. In subsequent years, different QH devices made of different semiconductors have been found to agree within relative uncertainties of some parts in 10^9 [13].

A second remarkable point, which is a manifestation of the fundamental nature of the QHE, is the relation of R_K to α :

$$\alpha = \frac{\mu_0 c_0}{2R_K}. \quad (11)$$

The quantity $\mu_0 c_0$ is sometimes referred to as the impedance of vacuum, Z_0 , which in the present SI has a fixed value of approximately 377Ω . Therefore an independent measurement of R_K provides a value for the fine-structure constant. Such a measurement is possible by comparing the quantized Hall resistance of a device to the impedance of a “calculable capacitor,” which is determined by measuring a single length. Within the uncertainty of these measurements, the resulting value of α is consistent with much more precise QED determinations [31]. As an exercise, one can use Table 1 to show that the variance of R_K as defined in Eq. 10 equals the variance of α .

Electrical metrology has thus been in a situation where the most precise measurements of voltage and resistance are derived from quantum standards based on fundamental constants but SI values of these constants are currently traceable to the artifact definition of the kilogram.

Quantum standards for voltage and resistance can, of course, be combined using Ohm’s law to produce a quantum standard for electrical current. However, an independent quantum standard based on single-electron tunnelling (SET) would close the so-called quantum metrology triangle and thus provide robustness to the existing system [19]. Keller has reviewed progress in this area [23].

In 1990, electrical metrologists adopted “conventional” values for K_J and R_K , known as K_{J-90} and R_{K-90} . The conventional values have, by definition, no uncertainty. Thus they are not SI values although they are based on the CODATA 1989 recommended SI values.

The community of electrical metrologists, represented by the Consultative Committee for Electricity and Magnetism (CCEM) of the International Committee for Weights and Measures (CIPM), recommend that SI values of both h and e be fixed as soon as possible and these values should be based on the best available SI values of h and e (and not on the conventional values adopted in 1990) [6]. Consequences of this recommendation are:

- Fixing a value for either h or e redefines the kilogram, see Table 1 and [11].
- Supposing that fixing a value for h is used to redefine the kilogram, then fixing a value for e redefines the ampere.
- Fixing values for both h and e overdetermines the new unit system, one manifestation of which would be a value of α defined by a committee. In order for α to remain an experimentally determined quantity that is independent of unit systems, the proposal is that the value of μ_0 would no longer be exactly $4\pi \times 10^{-7} \text{N/A}^2$, see Eqs. 10 and 11. Instead the magnetic constant would acquire the same relative uncertainty as α . Within that uncertainty, the value of μ_0 could change with improved measures of α .

3.4 Relative Atomic Masses

The field of chemistry has long ago solved its kilogram problem by listing atomic masses relative to $m(^{12}\text{C})$, the mass of an atom of carbon-12. This is accomplished by defining the atomic mass unit, u (also referred to as the dalton, Da), as

$$u = \frac{1}{12}m(^{12}\text{C}) = \frac{M_u}{N_A}, \quad (12)$$

where M_u is the molar mass constant, currently defined as exactly 10^{-3}kg/mol . Except for the notation given to the molar mass constant, we have already seen this relation in Eq. 4. The mass in kilogram of any atomic or subatomic particle X is given by

$$m(\text{X}) = A_r(\text{X})u, \quad (13)$$

and, obviously, $A_r(^{12}\text{C}) = 12$ (exactly).

$A_r(\text{X})$ is known as the relative atomic mass of X. Returning to Table 1, we see that the correlation coefficient between m_e and N_A is -1.0000 ; but it is not *exactly* -1 (the correlation coefficient between $m(^{12}\text{C})$ and N_A is exactly -1). In fact, elsewhere in CODATA 2006 we learn that the uncertainty of $A_r(\text{electron})$ is two orders of magnitude smaller than the uncertainty of m_e .

One might think that the chemical community, as represented by the Consultative Committee for Amount of Substance: Metrology in Chemistry (CCQM), would wish to fix a value for N_A – yet another way to redefine the kilogram. However, their current thinking is more nuanced: redefine the mole based on a fixed value for the Avogadro constant, with no reference to the kilogram [8] (in broad agreement with [28], but with additional details attended to). See [29] for further discussion.

4 Routes to a New Kilogram

There are several relations among the fundamental constants that are needed to discuss the redefinition of the kilogram in sufficient detail. The first is the Compton frequency, f_m , of a mass, m :

$$f_m = \frac{m c_0^2}{h}. \quad (14)$$

This is a well-known relationship in particle physics. Thus the kilogram could formally be redefined by fixing a value for h and then specifying the defining frequency f_{m_0} , which would, however, be enormous – of order 10^{50}Hz :

$$f_{m_0} = \frac{m_0 c_0^2}{h}. \quad (15)$$

In the present SI, m_0 is the mass of the international prototype (1 kg exactly), c_0 has a fixed value, and so an experimental determination of f_{m_0} , assuming this were possible, yields the SI value of h . We leave to the next section a description of how f_{m_0} is determined using devices known as watt balances. Obviously, if h were to have a fixed value in a new SI, the same measurement of f_{m_0} would determine the mass of the artifact kilogram. The experimental uncertainty could no longer be assigned to h and would instead be associated with m_0 .

In this new SI, a practical realization of 1 kg could also be made via the electron mass, m_e . We can rewrite the SI definition of the Rydberg constant, R_∞ [31], as follows:

$$\frac{m_e c_0^2}{h} = \frac{2R_\infty c_0}{\alpha^2}. \quad (16)$$

The value of R_∞ is known to a relative uncertainty of about 7×10^{-12} from measurements of the atomic spectra of hydrogen and deuterium. The relative uncertainty of α^2 is about 1.4×10^{-9} , according to CODATA, but the recommended value of α may change by about seven times this uncertainty due to the correction of previously published QED calculations of the anomalous magnetic moment of the electron [1]. With this caveat in mind, the dominant uncertainty in m_e still comes from the experimental determination of h . This can be verified from Eq. 16 and the correlations shown in Table 1. Conversely, an experimental value for h would be available through an accurate determination of m_e . The *relative* atomic mass of the electron, $A_r(e)$, is known from experiment to a relative uncertainty of about 5×10^{-10} .

By definition, the mass in kg of a ^{12}C atom is the ratio of its mass to the international prototype, as shown in Eq. 2. Therefore, making use of Eqs. 12, 13, and 16:

$$f_{m_0} = \left\{ \frac{m_0}{m(^{12}\text{C})} \right\} \frac{2}{\alpha^2} \left(\frac{12R_\infty c_0}{A_r(e)} \right) = \frac{N_A m_0}{M_u} \frac{2}{\alpha^2} \left(\frac{R_\infty c_0}{A_r(e)} \right). \quad (17)$$

Thus an accurate measurement of the mass of an atom of carbon-12 or, equivalently, a determination of the Avogadro constant, will serve as a check on measurements of f_{m_0} obtained directly from Eq. 15.

5 Realizing a New Kilogram Definition in Practice

If the kilogram is redefined as a consequence of fixing a value for h , then the present uncertainty assigned to the Planck constant, currently about 50×10^{-9} , would inevitably be attached to the value of the international prototype of the kilogram. The community of mass metrologists, represented by the Consultative Committee for Mass and Related Quantities (CCM) of the CIPM have recommended that a number of conditions be met [7] prior to redefinition. These involve reducing the present uncertainty of experimental determinations of the Planck constant, maintaining apparatus and expertise that will allow measurements of 1-kg artifacts in terms of h into the future, and accounting for present discrepancies among experimental determinations of h (e.g., [11]).

Below, we examine how the Planck constant is currently determined with respect to the kilogram using devices known as watt balances. The experiment is usually analyzed as a variant of Eq. 14. This interpretation is made especially well in [5]. We also discuss below how the Planck constant may be determined through a measurement of the Avogadro constant by means of the method of silicon X-ray crystal density (XRCD).

5.1 Watt Balances

Watt balances designed to operate at an accuracy level of parts in 10^8 are complicated devices and several detailed papers discuss their design features in much greater detail than can be done here; see, for example [16, 39]. This type of measurement was first proposed in 1975 by Bryan Kibble of the National Physical Laboratory (UK), prior to the discovery of the QHE. Now that voltage and resistance measurements can both be made traceable to quantum standards, the watt balance is generally presented as a determination of h in SI units. The experiment consists of two parts. In the first part, the gravitational force, mg , on a test mass, m , is balanced by an equal and opposite force produced by a linear electromagnetic motor. The required force, F , developed by the motor is proportional to an electrical current, I , such that

$$F = IK = mg, \quad (18)$$

where K is an instrumental constant that depends on the induction of the stator magnet as well as on complicated geometrical terms.

In the second part of the experiment, the same motor is configured as voltage generator. The moving element of the motor, the “coil,” is forced to travel vertically at a velocity, v . This induces a voltage U across the ends of the coil. Under ideal conditions,

$$U = vK. \quad (19)$$

The constant, K , is common to both modes and can therefore be eliminated:

$$mgv = IU, \quad (20)$$

hence the name “watt balance.” It was Kibble’s insight that although Eq. 20 equates mechanical and electrical power, there is no parasitic loss due to power dissipated in the experiment. Equation 20 equates virtual power. Assuming that current is measured as U'/R and that voltages and resistance are measured using Josephson and quantized Hall devices, Eqs. 7–9, 10, then the final equation (somewhat simplified) becomes:

$$\frac{mc_0^2}{h} = \frac{nf'n'f'c_0^2}{4igv}, \quad (21)$$

where v and g are ultimately measured in terms of Doppler shifts, clock times, and phase changes [5]. The Josephson effect is used in both the determination of the voltage and current (the latter via Ohm’s law) and this accounts for the primed quantities in Eq. 21. Formally, the right-hand side of Eq. 21 is the experimental determination of the de Broglie–Compton frequency of a macroscopic body of mass m , where the value of m is traceable to the mass of the international prototype shown schematically in Eq. 5. As a practical matter, results of watt balance experiments are reported as measured SI values of h .

5.2 Silicon X-Ray Crystal Density (XRCD)

Based on Eq. 17, it will be sufficient to show how the ratio $m_0/m(^{12}\text{C})$ is determined experimentally. The experiment is a worldwide collaborative effort, which is described in several excellent review papers; see, for example [2]. Here we will describe the experiment schematically.

The XRCD is fundamentally a counting experiment, designed to link the atomic and macroscopic domains. Note, for instance, that the ratio $m_0/m(^{12}\text{C})$, which appears in Eq. 17, is the number of ^{12}C atoms in 1 kg – a very large number.

The “trick” behind silicon XRCD is the following: a 1 kg perfect crystal of ^{28}Si is manufactured in the shape of a sphere. Silicon is chosen to take advantage of the prodigious amount of research and development already available from the semiconductor industry. The volume V of the sphere can be measured to high accuracy by means of optical interferometry. Since the sphere is subject to thermal expansion and compressibility, V is determined at a specified reference temperature (22.5°C) and pressure (0 Pa). The interatomic spacing of crystal samples that were adjacent to the sphere prior to its fabrication is determined by X-ray interferometry, under the same reference conditions. One may then deduce the volume a_0^3 of a unit cell of the crystal. A perfect silicon crystal has eight atoms per unit cell, and thus the number of atoms, n , within the sphere is known:

$$n = 8 \frac{V}{a_0^3}. \quad (22)$$

The same number can also be determined as the ratio of the mass of the sphere, $m(\text{sph})$, to the mass of a single atom of ^{28}Si :

$$n = \frac{m(\text{sph})}{m(^{28}\text{Si})}, \quad (23)$$

which we now rewrite to express experimental operations in terms of dimensionless ratios:

$$n = \frac{m(\text{sph})}{m_0} \frac{m(^{12}\text{C})}{m(^{28}\text{Si})} \frac{m_0}{m(^{12}\text{C})}. \quad (24)$$

The first ratio on the right-hand side of Eq. 24 represents a determination of the mass of the sphere (nominally 1 kg) with respect to the mass of the international prototype [34]. The second ratio represents a comparison of atomic masses, which has already been determined to very high accuracy [14]. The last ratio, the one actually needed for Eq. 17, can now be found by eliminating n from Eqs. 22 and 24:

$$\frac{m_0}{m(^{12}\text{C})} = \frac{m_0}{m(\text{sph})} \frac{m(^{28}\text{Si})}{m(^{12}\text{C})} \frac{8V}{a_0^3}. \quad (25)$$

Equation 25 shows the explicit role of m_0 in this determination. However, the relation is more frequently written in terms of the Avogadro constant (thought to be

more fundamental than the mass of a carbon-12 atom) and the density, $\rho(^{28}\text{Si})$, of a macroscopic sample of silicon-28 at the reference conditions (a quantity which will be the same for all perfect crystals):

$$N_A = \frac{A_r(^{28}\text{Si})M_u}{\rho(^{28}\text{Si})} \frac{8}{a_0^3}, \quad (26)$$

where the relation

$$m(\text{sph}) = \left\{ \frac{m(\text{sph})}{m_0} \right\} [\text{kg}] \quad (27)$$

is implicit.

In concluding this section, we emphasize that corrections to the simple view of XRCD presented here should not be overlooked. For example, crystal dislocations must be controlled, chemical purity must be known, and relative abundances of the three naturally occurring silicon isotopes must be known. (In [2], plans are described to produce crystals of highly enriched ^{28}Si . This has already been accomplished and measurements with the enriched material are progressing well.) Many corrections to Eqs. 25 and 26 are thus required in order to describe real materials of the highest quality. A more fundamental correction, for the mass equivalent of the crystal binding energy, is still negligible compared to present experimental uncertainties.

5.3 *Experimental Results*

CODATA 2006 has collected all relevant experimental results up to publication and has discussed them as determinations of h . As always, the CODATA Task Group on Fundamental Constants has been scrupulous in taking correlations among the inputs into account. At the time of publication, the recommended value for h had a relative uncertainty of about 50×10^{-9} , based on a weighted mean of all available input data. The value of h is largely determined by the experimental result from NIST (USA) watt balance, which has by far the lowest uncertainty (all uncertainties are first calculated *a priori*). Even with its larger uncertainty, the XRCD result is inconsistent with the NIST result, and so the CODATA Task Group decided to enlarge the *a priori* uncertainties of these input data *a posteriori* by a factor of 1.5, which does not change the value of the weighted mean but does improve statistical consistency based on a χ -squared test [31].

Subsequent to the cutoff date for data considered for the CODATA 2006, a new value of h became available from researchers at the NPL [38]. This result, referred to as “preliminary” by the authors, is statistically different from the CODATA 2006 recommendation. The recent data are summarized in Table 2, which also shows previously published results from both the NIST and the NPL. Except for the last line, all information in Table 2 can be found in Tables XXXV and XLVII of [31].

Thus the most recent watt balance result is 300(83) ppb above the CODATA 2006 recommended value for h . One must always search carefully to make sure that, for example, differences between results from watt balances and XRCD are not due to new physics. This point is examined in [31].

Table 2 Results in chronological order since 1990. The NPL-07 result [38] was published after the deadline for CODATA 2006. All other results are discussed in [31]

Type of experiment	Identification	h /(J s)	Relative standard uncertainty
Watt Balance	NPL-90	$6.6260682(13) \times 10^{-34}$	200×10^{-9}
Watt Balance	NIST-98	$6.62606891(58) \times 10^{-34}$	87×10^{-9}
Silicon XRCD	2005	$6.6260745(19) \times 10^{-34}$	290×10^{-9}
Watt Balance	NIST-07	$6.62606891(24) \times 10^{-34}$	36×10^{-9}
CODATA 2006	Combined	$6.62606896(33) \times 10^{-34}$	50×10^{-9}
Watt Balance	NPL-07	$6.62607095(44) \times 10^{-34}$	66×10^{-9}

Recall from Fig. 1 that the set of prototype kilograms seems to be diverging in mass from the international prototype by only about 0.5 ppb/year and one can understand why some mass metrologists view the immediate redefinition of the kilogram in terms of the Planck constant as premature. The chosen value of h will, of course, ensure that the mass of m_0 and all mass values derived from m_0 will be unchanged on the day the redefinition takes effect. Thus challenges for the mass community will be: (1) to deal with the significant additional uncertainty, which will be common to all macroscopic mass standards (and thus will not affect current uncertainties for differences between mass standards); and (2) to keep the ensemble of artifact mass standards and macroscopic quantities such as force and pressure, which are derived from such mass standards, closely linked to the SI.

6 Proposals for a New SI

6.1 Consensus Building and Formal Approval

The Consultative Committee for Units (CCU) of the CIPM is responsible for editing the SI brochure [3], the definitive SI documentation. The CCU also serves as a forum for proposals to redefine SI units. It receives such proposals principally from other Consultative Committees (CCs) and from scientific unions. The CIPM may place a formal resolution before the General Conference of Weights and Measures (CGPM), which takes place every four years. Ultimately the CGPM must approve significant changes to the SI on behalf of the 54 member states of the BIPM. The 23rd meeting of the CGPM took place in November 2007, and the next meeting, the 24th, will be held in October 2011.

At the 2007 CGPM meeting, the President of the CCU presented a progress report that summarized a broad consensus of all interested parties. This presentation is publicly available on the Internet [27].

For its part, the 2007 meeting of the CGPM passed Resolution 12, *On the possible redefinition of certain base units of the SI* [9]. Resolution 12 notes the desirability that new definitions of the base units “should be easily understood” and requests the

CIPM “to take whatever preparations are necessary so that, if the results of experiments are found to be satisfactory and the needs of users met, formal proposals for changes in the definitions of the kilogram, ampere, kelvin and mole can be put to the 24th General Conference.”

Given the pace of research, it is possible that the 2011 deadline may be met. The chief obstacle is a satisfactory realization of 1 kg by means of the experiments represented schematically by Eqs. 21 and 25.

A final concern is how to word a definition of the kilogram based on a fixed value for h in a way that is “easily understood.” A definition that specifies the de Broglie–Compton frequency, f_{m_0} , of 1 kg is problematic for two reasons: relatively few members of the public will understand this definition and those who do may be concerned that $1/f_{m_0}$ is seven orders of magnitude smaller than the Planck time, $(\hbar G/c_0^5)^{1/2}$. Regarding the issue of public comprehension, we may note that the definition of the second already refers to “hyperfine levels of the ground state of the caesium 133 atom” [3], and this illustrates that public comprehension, although important, is not an overriding issue.

6.2 *An SI Based on Defined Values of a Set of Constants*

Mohr [30] has recently demonstrated how a new SI could be defined by choosing a basis set of constants (above we have been concerned with ν_{hfs} , c_0 , h , e) the values of which would then be chosen to adjust the sizes of the base units seamlessly to their traditional values. There is no need to specify which of the constants defines the unit of mass, just as there is no necessity to say which of the Planck set defines the Planck mass. The choice is unique. Mohr’s publication and its references should be consulted by anyone with a serious interest in unit systems.

7 Conclusion

The SI is moving to a system entirely based on fixed values for a selected set of physical constants that appear in QED. The present artifact definition of the kilogram will be replaced as part of this program. The remaining issues are how to determine the masses of macroscopic artifacts, which will still be used in conventional mass metrology, in terms of the new definition. At present, the most promising experimental methods are based on watt balance and silicon XRCD technologies. Unfortunately, the results of such measurements are not yet as consistent as one would wish. However, work is progressing relatively quickly and technical problems may be resolved (and some expectations may be reassessed) in time for formal approval of a new kilogram definition by the next meeting of the CGPM in late 2011, or the following in 2015.

References

1. T. Aoyama, M. Hayakawa, T. Kinoshita, M. Nio, N. Watanabe, *Phys. Rev. D* **77**, 053012 (2008)
2. P. Becker, *Metrologia* **40**, 366 (2003)
3. BIPM (Bureau International des Poids et Mesures), *The International System of Units*, 8th edn. (BIPM, Sèvres, 2006), http://www.bipm.org/en/si/si_brochure/. Accessed 4 January 2009
4. BIPM KCDB (Bureau International des Poids et Mesures, Key Comparison Database), *Calibration and Measurement Capabilities, Mass standards*, http://kcdb.bipm.org/AppendixC/country_list.asp?Service=M/Mass.1.1. Accessed 4 January 2009
5. C.J. Bordé, *Phil. Trans. R. Soc. A* **363**, 2177 (2005)
6. CCEM (Consultative Committee for Electricity and Magnetism), *Recommendation E1 (2007): Proposed Changes to the International System of Units (SI)*. Document CCEM/2007-44, <http://www.bipm.org/cc/CCEM/Allowed/25/CCEM2007-44.pdf> (2007). Accessed 4 January 2009
7. CCM (Consultative Committee for Mass and Related Quantities), *Report of the CCM to the CIPM, 15 May 2007 (Document CCU/07-14)*. (BIPM, Sèvres, 2007), http://www.bipm.org/wg/CCM/CCM-WGSI-kg/Allowed/General/CCM_document_to_CCU-CIPM_May07.pdf. Accessed 4 January 2009
8. CCQM (Consultative Committee for Amount of Substance: Metrology in Chemistry), *Report of the 14th Meeting (3–4 April 2008) to the International Committee for Weights and Measures*. (BIPM, Sèvres, 2008), <http://www.bipm.org/utills/common/pdf/CCQM14.pdf>. Accessed 4 January 2009
9. CGPM (General Conference for Weights and Measure), *Resolution 12 of the 23rd Meeting of the CGPM (2007)*, <http://www.bipm.org/en/CGPM/db/23/12/> Accessed 4 January 2009
10. R.S. Davis, *Metrologia* **40**, 299 (2003)
11. R.S. Davis, *Phil. Trans. R. Soc. A* **363**, 2249 (2005)
12. J. de Boer, *Metrologia* **31**, 401 (1995)
13. F. Delahaye, T.J. Witt, R.E. Elmquist, R.F. Dziuba, *Metrologia* **37**, 173 (2000)
14. F. DiFilippo, V. Natarajan, K.R. Boyce, D.E. Pritchard, *Phys. Rev. Lett.* **73**, 1481 (1994)
15. M.J. Duff, L.B. Okun, G. Veneziano, *J. High Energy Phys.* **023**, 3 (2002)
16. A. Eichenberger, B. Jeckelmann, Ph. Richard, *Metrologia* **40**, 356 (2003)
17. K.M. Evenson, J.S. Wells, F.R. Petersen, B.L. Danielson, G.W. Day, R.L. Berger, J.L. Hall, *Phys. Rev. Lett.* **29**, 1346 (1972)
18. G. Gabrielse, D. Hameke, T. Kinoshita, M. Nio, B. Odom, *Phys. Rev. Lett.* **97**, 030802 (2006); Erratum, *ibidem* **99**, 039902(E) (2007)
19. J. Gallop, *Phil. Trans. R. Soc. A* **363**, 2221 (2005)
20. J.D. Jackson, *Classical Electrodynamics*, 3rd edn. (Wiley, New York, 1998)
21. ILAC (International Laboratory Accreditation Cooperation), <http://www.ilac.org/>
22. A.K. Jain, J.E. Lukens, J.S. Tsai, *Phys. Rev. Lett.* **58**, 1165 (1987)
23. M. Keller, *Metrologia* **45**, 102 (2008)
24. S.N. Lea, *Rep. Prog. Phys.* **70**, 1473 (2007)
25. J.C. Maxwell, in *The Scientific Papers of J.C. Maxwell*, ed. by W.D. Niven, vol. 2 (Cambridge University, Cambridge, 1890), p. 225
26. M.L. McGlashan, *Metrologia* **31** 447 (1995)
27. I.M. Mills, *Report of the CCU to the 23rd meeting of the CGPM, November 2007*. Powerpoint presentation, <http://www.bipm.org/en/committees/cc/ccu/> Accessed 4 January 2009
28. I.M. Mills, P.J. Mohr, T.J. Quinn, B.N. Taylor, E.R. Williams, *Metrologia* **43**, 227 (2006)
29. J.T.M. Milton, J.M. Williams, S.J. Bennett, *Metrologia* **44**, 356 (2007)
30. P.J. Mohr, *Metrologia* **45**, 129 (2008)
31. P.J. Mohr, B.N. Taylor, D.B. Newell, *J. Phys. Chem. Ref. Data* **37**, 1187 (2008), <http://physics.nist.gov/cuu/Constants/JPCRDMTN2008.pdf> Accessed 4 January 2009
32. NIST (National Institute for Standards and Technology, USA), in *Latest (2006) Values of the Constants*, <http://physics.nist.gov/cuu/Constants/index.html>. Accessed 4 January 2009
33. B.W. Petley, *Metrologia* **44**, 69 (2007)
34. A. Picard, N. Bignell, M. Borys, S. Downes, S. Mizushima, *Metrologia* **46**, 1 (2009)

35. J.D. Prestage, R.L. Tjoelker, L. Maleki, *Phys. Rev. Lett.* **74**, 3511 (1995)
36. T.J. Quinn, *IEEE Trans. Instrum. Meas.* **40**, 81 (1991)
37. S. Rainville, J.K. Thompson, E.G. Myers, J.M. Brown, M.S. Dewey, E.G. Kessler Jr, R.D. Deslattes, H.G. Börner, M. Jentschel, P. Mutti, D.E. Pritchard, *Nature* **438**, 1096 (2005)
38. I.A. Robinson, B.P. Kibble, *Metrologia* **44**, 427 (2007)
39. R.L. Steiner, E.R. Williams, R. Liu, D.B. Newell, *IEEE Trans. Instrum. Meas.* **56**, 592 (2007)
40. K. von Klitzing, The quantized Hall effect, Nobel Lecture 9 December 1985. Nobel Prize Web, http://nobelprize.org/nobel_prizes/physics/laureates/1985/klitzing-lecture.pdf. Accessed 4 January 2009
41. K. von Klitzing, *Phil. Trans. R. Soc. A* **363**, 2203 (2005)

Mass and Angular Momentum in General Relativity

José Luis Jaramillo and Ericourgoulhon

Abstract We present an introduction to mass and angular momentum in General Relativity. After briefly reviewing energy–momentum for matter fields, first in the flat Minkowski case (Special Relativity) and then in curved spacetimes with or without symmetries, we focus on the discussion of energy–momentum for the gravitational field. We illustrate the difficulties rooted in the Equivalence Principle for defining a local energy–momentum density for the gravitational field. This leads to the understanding of gravitational energy–momentum and angular momentum as nonlocal observables that make sense, at best, for extended domains of spacetime. After introducing Komar quantities associated with spacetime symmetries, it is shown how total energy–momentum can be unambiguously defined for isolated systems, providing fundamental tests for the internal consistency of General Relativity as well as setting the conceptual basis for the understanding of energy loss by gravitational radiation. Finally, several attempts to formulate quasi-local notions of mass and angular momentum associated with extended but finite spacetime domains are presented, together with some illustrations of the relations between total and quasi-local quantities in the particular context of black hole spacetimes. This article is not intended to be a rigorous and exhaustive review of the subject, but rather an invitation to the topic for nonexperts.

J.L. Jaramillo (✉)

Instituto de Astrofísica de Andalucía, CSIC, Apartado Postal 3004, Granada 18080, Spain and

Laboratoire Univers et Théories, Observatoire de Paris, CNRS, Université Paris Diderot, 5 place Jules Janssen, F-92190 Meudon, France

e-mail: jarama@iaa.es

E.ourgoulhon (✉)

Laboratoire Univers et Théories, Observatoire de Paris, CNRS, Université Paris Diderot, 5 place Jules Janssen, F-92190 Meudon, France

e-mail: eric.ourgoulhon@obspm.fr

1 Issues Around the Notion of Gravitational Energy in General Relativity

1.1 Energy–Momentum Density for Matter Fields

Let us first consider mass and angular momentum associated with matter in the absence of gravity, in a flat Minkowski spacetime. The density of energy and linear momentum associated with a distribution of matter are encoded in the energy–momentum tensor $T_{\mu\nu}$, corresponding to the Noether current conserved under infinitesimal spacetime translations in a Lagrangian framework. This general conservation property, namely $\partial_\mu T^{\mu\nu} = 0$ in inertial Minkowski coordinates, plays a key role in our discussion. Indeed, together with the presence of symmetries, it permits the introduction of conserved quantities or *charges*. Given a space-like hypersurface Σ and considering the unit time-like vector n^μ normal to it, we can define the conserved quantity associated with the symmetry k^μ and the domain D ($\subset \Sigma$) as

$$Q_D[k^\mu] = \int_D k^\rho T_{\nu\rho} n^\nu \sqrt{\gamma} d^3x, \quad (1)$$

where $\sqrt{\gamma} d^3x$ denotes the induced volume element in D . The conservation of $T_{\mu\nu}$ and the characterisation of k^μ as a symmetry imply the conservation of the vector $T^\mu{}_\nu k^\nu$, that is, $\partial_\mu (T^\mu{}_\nu k^\nu) = 0$. Applying then the Stokes theorem, it follows the equality between the change in time of $Q_D[k^\mu]$ and the flux of $\gamma^\mu{}_\rho T^{\rho\nu} k_\nu$ through the boundary of D (where $\gamma^\mu{}_\nu$ is the projector on D). Minkowski spacetime symmetries are given by Poincaré transformations. Therefore, we can associate conserved quantities with the infinitesimal generators corresponding to translations T_a^ν , rotations J_i^μ , and boosts K_i^μ (here the label a for translation generators runs in $\{0, 1, 2, 3\}$, whereas i is a space-like index in $\{1, 2, 3\}$). In this manner, a 4-momentum $P_a[D]$ and an angular momentum $J_i[D]$ associated with the distribution of matter in $D \subset \Sigma$ can be defined as

$$P_a[D] = \int_D T_{\mu\nu} T_a^\nu n^\mu \sqrt{\gamma} d^3x, \quad J_i[D] = \int_D T_{\mu\nu} J_i^\nu n^\mu \sqrt{\gamma} d^3x. \quad (2)$$

More generally, we can combine together the rotation and boost generators J_i^μ and K_i^μ into a vector-field-valued antisymmetric matrix $M_{[ab]}^\mu$ (where $J_i^\mu = {}^3\epsilon_i{}^{jk} M_{[jk]}^\mu$ and $K_i^\mu = M_{[0i]}^\mu$) and write the conserved quantities

$$J_{[ab]}[D] = \int_D T_{\mu\nu} M_{[ab]}^\nu n^\mu \sqrt{\gamma} d^3x. \quad (3)$$

The mass and (Pauli–Lubanski) spin are constructed as

$$m^2[D] := -\eta^{ab} P_a[D] P_b[D], \quad S^a[D] := \frac{1}{2} \epsilon^{abcd} P_b[D] J_{[cd]}[D], \quad (4)$$

in terms of which Poincaré Casimirs (invariant under Poincaré transformations) can be expressed.

In the non-flat case, (matter) energy–momentum tensor acts as the source of gravity through the Einstein equation and, consistently with Bianchi identities, satisfies the divergence-free condition analogous to the flat conservation law:

$$G_{\mu\nu} := {}^4R_{\mu\nu} - \frac{1}{2}{}^4R g_{\mu\nu} = 8\pi T_{\mu\nu}, \quad \nabla_\nu T^{\nu\mu} = 0. \quad (5)$$

The same strategy employed in the flat case for defining physical quantities associated with matter, that is, using conserved currents corresponding to some symmetry, can be followed in non-flat spacetimes $(\mathcal{M}, g_{\mu\nu})$ presenting Killing vectors k^μ . The vector $T^\mu{}_\nu k^\nu$ is conserved, that is, $\nabla_\mu (T^\mu{}_\nu k^\nu) = 0$, and provides a current density for the conserved quantity $Q_D[k^\mu]$ defined by expression (1). The physical interpretation of $Q_D[k^\mu]$ depends of course on the nature of the Killing vector k^μ . Actually, $Q_D[k^\mu]$ does not actually depend on the slice Σ in the sense that its value is the same in the domain of dependence of D (this precisely corresponds to the conserved nature of this charge).

In a general spacetime with no symmetries the previous strategy ceases to work, and ambiguities in the definition of mass and angular momentum enter into scene. One can still calculate the flux of $T^\mu{}_\nu \xi^\nu$ for a given vector ξ^ν , and define the associated quantity $Q_D[\xi^\mu]$. However, the latter will now depend on the slice Σ and, in addition, its explicit dependence on ξ^μ introduces some degree of arbitrariness in the discussion. In this context, given a space-like 3 + 1 foliation $\{\Sigma_t\}$ of the spacetime with time-like normal vector n^μ , the current $P^\mu := -T^{\mu\nu} n_\nu$ can be interpreted as the energy–momentum density associated with (Eulerian) observers *at rest* with respect to Σ_t . That is, $E := T^{\mu\nu} n_\mu n_\nu$ stands as the matter energy density and $p^\mu := -\gamma^\mu{}_\rho T^{\rho\nu} n_\nu$ as the momentum density, where $\gamma_{\mu\nu}$ is the induced metric on Σ_t (see Eq. 12 below for the complete 3 + 1 decomposition of $T_{\mu\nu}$). In particular, we can calculate the (matter) energy associated with observers n^μ over the spatial region D by direct integration

$$E[D] = \int_D E \sqrt{\gamma} d^3x = \int_D T^{\mu\nu} n_\mu n_\nu \sqrt{\gamma} d^3x. \quad (6)$$

By imposing the dominant energy condition on the matter energy–momentum tensor (see Section 3.3), the vector $-T^{\mu\nu} n_\nu$ is future directed and non-space-like. Its Lorentzian norm is therefore non-positive and an associated (matter) mass density m can be given as $m^2 := -P^\mu P_\mu = -(-T^{\mu\rho} n_\rho)(-T^{\nu\sigma} n_\sigma) g_{\mu\nu} = E^2 - p^i p_i \geq 0$. The corresponding mass $M[D]$ in the extended region D would be

$$M[D] := \int_D \sqrt{E^2 - p^i p_i} \sqrt{\gamma} d^3x. \quad (7)$$

Note the difference between the construction of $M[D]$ and that of $m[D]$ in the Minkowskian case: for the latter one first integrates to obtain the charges and then calculates a Minkowskian norm, whereas for constructing $M[D]$ that order is reversed; in addition, different metrics are employed in each case (cf. Section 2.2 in [85]).

1.2 Problems when Defining a Gravitational Energy–Momentum

In the characterization of the physical properties of the gravitational field, in particular its energy–momentum and angular momentum, we could try to follow a similar strategy to that employed for the matter fields. This would amount to identify appropriate local densities that would then be integrated over finite spacetime regions. However, such an approach rapidly meets important conceptual difficulties.

A local (point-like) density of energy associated with the gravitational field cannot be defined in General Relativity. Reasons for this can be tracked to the Equivalence Principle. Illustrated in a heuristic manner, this principle can be used to get rid of the gravitational field on a given point of spacetime. Namely, a free falling point-like particle does not *feel* any gravitational field so that, in particular, no gravitational energy density can be identified at spacetime points.

In a Lagrangian setting, these basic conceptual difficulties are reflected in the attempts to construct a gravitational energy–momentum tensor, when mimicking the methodological steps followed in the matter field case. We can write generically the gravitational-matter action as

$$S = S_{\text{EH}} + S_{\text{m}} = \frac{1}{16\pi} \int_{\mathcal{M}} {}^4R \sqrt{-g} d^4x + \int_{\mathcal{M}} L_{\text{m}}(g_{\mu\nu}, \Phi_i, \nabla_{\mu} \Phi_i, \dots) \sqrt{-g} d^4x. \quad (8)$$

where S_{EH} denotes the Einstein–Hilbert action and Φ_i in the matter Lagrangian L_{m} account for the matter fields. The symmetric energy–momentum for matter is obtained from the variation of the matter action S_{m} with respect to the metric

$$T_{\mu\nu} := \frac{-2}{\sqrt{-g}} \frac{\delta S_{\text{m}}}{\delta g^{\mu\nu}}, \quad (9)$$

whereas the field equations for the matter fields follow from the variation with respect to the matter fields Φ_i . On the contrary, the gravitational action S_{EH} only depends on the gravitational field, since any further background structure would be precluded by diffeomorphism invariance (a feature closely tied to the physical Equivalence Principle). Einstein equation for the gravitational field follows from the variation of the total action with respect to the metric field $g_{\mu\nu}$, with no gravitational analogue of the symmetric matter energy–momentum tensor $T_{\mu\nu}$. Attempts to construct a symmetric energy–momentum tensor for the gravitational field either recover the Einstein tensor $G_{\mu\nu}$ or can only be related to higher-order gravitational energy–momentum objects, such as the Bel–Robinson tensor (see, e.g., [84]). Again, the absence of a tensorial (i.e., point-like geometric) quantity representing energy–momentum for the gravitational field is consistent with, and actually a consequence of, the Equivalence Principle.

The natural interpretation of the symmetric matter energy–momentum tensor $T_{\mu\nu}$ as introduced in Eq. 9 is that of the *current source* for the gravitational field, obtained as a conserved current associated with spacetime translations. Alternative, in terms of the Noether theorem [75] it is natural to introduce a (nonsymmetric)

canonical energy–momentum tensor for matter from which a symmetric one can be constructed through the Belinfante–Rosenfeld procedure [15, 16, 47]. The application of this construction to the gravitational field naturally leads to the discussion of gravitational energy–momentum *pseudo-tensors* [85]. The underlying idea consists in decomposing the Einstein tensor $G_{\mu\nu}$ into a part that can be identified with the energy–momentum and a second piece that can be expressed in terms of a pseudo-potential. That is [25]

$$G_{\mu}{}^{\nu} := -8\pi t_{\mu}{}^{\nu} + \frac{1}{2\sqrt{-g}}\partial_{\lambda}(H_{\mu}{}^{\nu\lambda}), \quad (10)$$

where $t_{\mu\nu}$ is the gravitational energy–momentum pseudo-tensor and $H_{\mu\nu}{}^{\lambda}$ is the superpotential. Einstein’s equation is then written as

$$\partial_{\lambda}(H_{\mu}{}^{\nu\lambda}) = 16\pi\sqrt{-g}(t_{\mu}{}^{\nu} + T_{\mu}{}^{\nu}) =: 16\pi\mathcal{T}_{\mu}{}^{\nu}. \quad (11)$$

Objects $t_{\mu}{}^{\nu}$ and $H_{\mu}{}^{\nu\lambda}$ are not tensorial quantities. This means that their value at a given spacetime point is not a well-defined notion. Moreover, their very definition needs the introduction of some additional background structure and some choice of preferred coordinates is naturally involved. Different pseudo-tensors exist in the literature, for example, those introduced by Einstein, Papapetrou, Bergmann, Landau and Lifshitz, Møller, or Weinberg (e.g. see references in [25]).

As an alternative to the pseudo-tensor approach, there also exist attempts in the literature aiming at constructing truly tensorial energy–momentum quantities. However they also involve the introduction of some additional structure, either in the form of a background object or by fixing a gauge in some given formulation of General Relativity (cf. comments on the tetrad formalism approach in [85]).

1.2.1 Nonlocal Character of the Gravitational Energy

As illustrated above, crucial conceptual and practical caveats are involved in the association of energy and angular momentum with the gravitational field. For these reasons, one might legitimately consider gravitational energy and angular momentum in General Relativity as intrinsically meaningless notions in *generic* situations, in such a way that the effort to derive explicit general local expressions actually represents an ill-defined problem (cf. remarks in [73] referring to the quest for a local expression of energy in General Relativity). Having said this and after accepting the nonexistence of a *local* (point-like) notion of energy density for the gravitational field, one may also consider gravitational energy–momentum and angular momentum as notions intrinsically associated with extended domains of the spacetime and then look for restricted settings or appropriate limits where they can be properly defined.

In fact, making a sense of the energy and angular momentum for the gravitational field in given regions of spacetime is extremely important in different

contexts of gravitational physics, as it can be illustrated with examples coming from mathematical relativity, black hole physics, lines of research in Quantum Gravity, or relativistic astrophysics. From a structural point of view, having a well-defined *mass positivity* result is crucial for the internal consistency of the theory, as well as for the discussion of the solutions stability. Moreover, the possibility of introducing appropriate positive-definite (*energy*) quantities is often a key step in different developments in mathematical relativity, in particular when using variational principles. In the study of the physical picture of black holes, appropriate notions of mass and angular momentum are employed. In particular, they play a key role in the formulation of black hole thermodynamics (e.g., [88]), a cornerstone in different approaches to Quantum Gravity. In the context of relativistic astrophysics and numerical relativity, the study of relativistic binary mergers, gravitational collapse, and the associated generation/propagation of gravitational radiation also requires appropriate notions of energy and angular momentum (see e.g., [64] for a further discussion on the intersection between numerical and mathematical relativity).

Once the nonlocal nature of the gravitational energy–momentum and angular momentum is realized, the conceptual challenge is translated into the manner of determining the appropriate physical parameters associated with the gravitational field in an extended region of spacetime. An unambiguous answer has been given in the case of the total mass of an isolated system. However, the situation is much less clear in the case of extended but finite spacetime domains. In a broad sense, existing attempts either enforce some additional structure that restricts the study to an appropriate subset of the solution space of General Relativity, or alternatively they look for a genuinely geometric characterization aiming at fulfilling some expected *physical* requirements. In this article we present an overview of some of the relevant existing attempts and illustrate the kind of additional structures they involve. We do not aim here at an exhaustive review of the subject, but rather we intend to provide an introduction to the topic for nonexperts. In this sense we follow essentially the expositions in [43, 79, 85, 87] and refer the reader interested in further developments to the existing literature, in particular to the excellent and comprehensive review by Szabados [85].

1.3 Notation

Before proceeding further, we set the notation, some of whose elements have already been anticipated above. The signature of spacetime $(\mathcal{M}, g_{\mu\nu})$ is chosen to be $\text{diag}[-1, 1, 1, 1]$ and Greek letters are used for spacetime indices in $\{0, 1, 2, 3\}$. We denote the Levi–Civita connection by ∇_μ and the volume element by ${}^4\epsilon = \sqrt{-g} dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3$. We make $G = c = 1$ throughout.

1.3.1 3 + 1 Decompositions

In our presentation of the subject, 3+1 foliations of spacetime $(\mathcal{M}, g_{\mu\nu})$ by space-like 3-slices $\{\Sigma_t\}$ will play an important role. Given a height-function t , the

time-like unit normal to Σ_t will be denoted by n^μ and the 3+1 decomposition of the evolution vector field by $t^\mu = Nn^\mu + \beta^\mu$, where N is the lapse function and β^μ is the shift vector. The induced metric on the space-like 3-slice Σ_t is expressed as $\gamma_{\mu\nu} = g_{\mu\nu} + n_\mu n_\nu$, with D_μ the associated Levi–Civita connection and volume element ${}^3\epsilon = \sqrt{\gamma} dx^1 \wedge dx^2 \wedge dx^3$, so that ${}^3\epsilon_{\mu\nu\rho} = n^{\sigma 4} \epsilon_{\sigma\mu\nu\rho}$. The extrinsic curvature of $(\Sigma_t, \gamma_{\mu\nu})$ in $(\mathcal{M}, g_{\mu\nu})$ is defined as $K_{\mu\nu} := -\frac{1}{2} \mathcal{L}_n \gamma_{\mu\nu} = -\gamma_\mu^\rho \nabla_\rho n_\nu$. The 3 + 1 decomposition of the (matter) stress–energy tensor, in terms of an Eulerian observer n^μ in rest with respect to the foliation $\{\Sigma_t\}$, is

$$T_{\mu\nu} = E n_\mu n_\nu + p_{(\mu} n_{\nu)} + S_{\mu\nu}, \quad (12)$$

where the matter energy and momentum densities are given by $E := T_{\mu\nu} n^\mu n^\nu$ and $p_\mu := -T_{\nu\rho} n^\nu \gamma^\rho_\mu$, respectively, whereas the matter stress tensor is $S_{\mu\nu} := T_{\rho\sigma} \gamma^\rho_\mu \gamma^\sigma_\nu$. Latin indices running in $\{1, 2, 3\}$ will be employed in expressions only involving objects intrinsic to space-like Σ_t slices.

1.3.2 Closed 2-Surfaces

Closed 2-surfaces \mathcal{S} , namely topological spheres in our discussion, will also be relevant in the following. The normal bundle $T^\perp \mathcal{S}$ can be spanned by a time-like unit vector field n^μ and a space-like unit vector field s^μ , which we choose to satisfy the orthogonality condition $n^\mu s_\mu = 0$. When considering \mathcal{S} as embedded in a space-like 3-surface Σ , n^μ can be identified with the time-like normal to Σ and s^μ with the normal to \mathcal{S} tangent to Σ . In the generic case, n^μ and s^μ can be defined up to a boost transformation: $n'^\mu = \cosh(\eta)n^\mu + \sinh(\eta)s^\mu$ and $s'^\mu = \sinh(\eta)n^\mu + \cosh(\eta)s^\mu$, with η a real parameter. Alternatively, one can span $T^\perp \mathcal{S}$ at $p \in \mathcal{S}$ in terms of the null normals defined by the intersection between the normal plane to \mathcal{S} and the light-cone at the spacetime point p . The directions defined by the *outgoing* ℓ^μ and the *ingoing* k^μ null normals (satisfying $k^\mu \ell_\mu = -1$) are uniquely determined, though it remains a boost-normalization freedom: $\ell'^\mu = f \cdot \ell^\mu$, $k'^\mu = \frac{1}{f} \cdot k^\mu$. The induced metric on \mathcal{S} is given by: $q_{\mu\nu} = g_{\mu\nu} + k_\mu \ell_\nu + \ell_\mu k_\nu = g_{\mu\nu} + n_\mu n_\nu - s_\mu s_\nu = \gamma_{\mu\nu} - s_\mu s_\nu$, the latter expression applying when \mathcal{S} is embedded in $(\Sigma, \gamma_{\mu\nu})$. The Levi–Civita connection associated with $q_{\mu\nu}$ will be denoted by ${}^2D_\mu$ and the volume element by ${}^2\epsilon = \sqrt{q} dx^1 \wedge dx^2$, that is, ${}^2\epsilon_{\mu\nu} = n^\rho s^\sigma \epsilon_{\rho\sigma\mu\nu}$. When integrating tensors on \mathcal{S} with components normal to the sphere, it is convenient to express the volume element as $dS_{\mu\nu} = (s_\mu n_\nu - n_\mu s_\nu) \sqrt{q} d^2x$ (this is just a convenient manner of reexpressing ${}^4\epsilon_{\mu\nu\rho\sigma}$ for integrating over \mathcal{S} after a contraction with the appropriate tensor; cf., e.g., Eq. 13).

The *second fundamental tensor* of $(\mathcal{S}, q_{\mu\nu})$ in $(\mathcal{M}, g_{\mu\nu})$ is defined as $\mathcal{K}^\alpha_{\mu\nu} := q^\rho_\mu q^\sigma_\nu \nabla_\rho q^\alpha_\sigma$, that can be expressed as $\mathcal{K}^\alpha_{\mu\nu} = n^\alpha \Theta_{\mu\nu}^{(n)} + s^\alpha \Theta_{\mu\nu}^{(s)} = k^\alpha \Theta_{\mu\nu}^{(\ell)} + \ell^\alpha \Theta_{\mu\nu}^{(k)}$, where the *deformation tensor* $\Theta_{\mu\nu}^{(v)}$ associated with a vector v^μ normal to \mathcal{S} is defined as $\Theta_{\mu\nu}^{(v)} = q^\rho_\mu q^\sigma_\nu \nabla_\rho v_\sigma$. We set a specific notation for the cases corresponding to s^μ and n^μ , namely, $H_{\mu\nu} := \Theta_{\mu\nu}^{(s)}$, the extrinsic curvature of $(\mathcal{S}, q_{\mu\nu})$ inside a 3-slice $(\Sigma, \gamma_{\mu\nu})$, and $L_{\mu\nu} := -\Theta_{\mu\nu}^{(n)}$.

Information about the extrinsic curvature of $(S, q_{\mu\nu})$ in $(\mathcal{M}, g_{\mu\nu})$ is completed by the *normal fundamental forms* associated with normal vectors ν^μ . In particular, we define the 1-form $\Omega_\mu^{(\ell)} := k^\rho q^\sigma{}_\mu \nabla_\sigma \ell_\rho$. This form is not invariant under a boost transformation, and transforms as $\Omega_\mu^{(\ell')} = \Omega_\mu^{(\ell)} + {}^2D_\mu \ln f$ in the notation above. Other normal fundamental forms can be defined in terms of normals k^μ , n^μ , and s^μ , but they are all related up to total derivatives.

2 Spacetimes with Killing Vectors: Komar Quantities

As commented above, some additional structure is needed to introduce meaningful notions of gravitational energy and angular momentum. Let us first consider spacetimes admitting isometries. This represents the most straightforward generalization of the definition of physical parameters as conserved quantities under existing symmetries. Requiring the presence of Killing vectors represents our first example of the enforcement of an additional structure on the considered spacetime.

Given a Killing vector field k^μ in the spacetime $(\mathcal{M}, g_{\mu\nu})$ and \mathcal{S} a space-like closed 2-surface, let us define the Komar quantity [67] k_K as

$$k_K := -\frac{1}{8\pi} \oint_{\mathcal{S}} \nabla^\mu k^\nu dS_{\mu\nu} \quad (13)$$

(see previous section for the notation $dS_{\mu\nu}$ for the volume element on \mathcal{S}). Let us consider \mathcal{S} as embedded in a space-like 3-slice Σ and let us take a second closed 2-surface \mathcal{S}' such that either \mathcal{S}' is completely contained in \mathcal{S} or vice versa, and let us denote by V the region in Σ contained between \mathcal{S} and \mathcal{S}' . The previously defined Komar quantity k_K is then *conserved* in the sense that its value does not depend on the chosen 2-surface as long as no matter is present in the intermediate region V

$$k_K^{\mathcal{S}} = 2 \int_V \left(T_{\mu\nu} - \frac{1}{2} T g_{\mu\nu} \right) n^\mu k^\nu \sqrt{\gamma} d^3x + k_K^{\mathcal{S}'}, \quad (14)$$

where $T = T_{\mu\nu} g^{\mu\nu}$.

Remark 1. Two important points must be stressed: (a) the definition of k_K is geometric and therefore coordinate independent, and (b) k_K is associated with a closed 2-surface with no need to refer to any particular embedding in a 3-slice Σ (in the discussion above the latter has been only introduced for pedagogical reasons).

2.1 Komar Mass

Stationary spacetimes admit a time-like Killing vector field k^μ . The associated conserved Komar quantity is known as the Komar mass

$$M_K := -\frac{1}{8\pi} \oint_{\mathcal{S}} \nabla^\mu k^\nu dS_{\mu\nu}. \quad (15)$$

This represents our first notion of mass in General Relativity. It is instructive to write the Komar mass in terms of 3 + 1 quantities. Given a 3-slicing $\{\Sigma_t\}$ and choosing the evolution vector $t^\mu = Nn^\mu + \beta^\mu$ to coincide with the time-like Killing symmetry, we find

$$M_K = \frac{1}{4\pi} \oint_S (s^i D_i N - K_{ij} s^i \beta^j) \sqrt{q} d^2x. \quad (16)$$

2.2 Komar Angular Momentum

Let us consider now an axisymmetric spacetime, where the axial Killing vector is denoted by ϕ^μ . That is, ϕ^μ is a space-like Killing vector whose action on \mathcal{M} has compact orbits, two stationary points (the poles), and is normalized so that its natural affine parameter takes values in $[0, 2\pi)$. The Komar angular momentum is defined as

$$J_K := \frac{1}{16\pi} \oint_{S_t} \nabla^\mu \phi^\nu dS_{\mu\nu}. \quad (17)$$

Note (apart from the sign choice) the factor 1/2 with respect to the Komar quantity ϕ_K , known as the *Komar anomalous factor* (it can be explained in the context of a bimetric formalism by writing the conserved quantities in terms of an *Einstein energy-momentum flux* density that can be expressed as the sum of half the Komar contribution plus a second term: in the angular momentum case this second piece vanishes, whereas for the mass case it equals half the Komar term; cf. [65]). Adopting a 3-slicing adapted to axisymmetry, that is, $n^\mu \phi_\mu = 0$, we have:

$$J_K = \frac{1}{8\pi} \oint_S K_{ij} s^i \phi^j \sqrt{q} d^2x = \frac{1}{8\pi} \oint_S \Omega_\mu^{(\ell)} \phi^\mu \sqrt{q} d^2x. \quad (18)$$

3 Total Mass of Isolated Systems in General Relativity

3.1 Asymptotic Flatness Characterization of Isolated Systems

The characterization of an isolated system in General Relativity aims at capturing the idea that spacetime becomes flat when we move *sufficiently far* from the system, so that spacetime approaches that of Minkowski. However, the very notion of *far away* becomes problematic due to the absence of an a priori background spacetime. In addition, we must consider different *kinds of infinities*, since we can move away from the system in space-like and also in null directions. Different strategies exist in the literature for the formalization of this asymptotic flatness idea, and not all

of them are mathematically equivalent. Traditional approaches attempt to specify the adequate falloff conditions of the curvature in appropriate coordinate systems at *infinity*. These approaches have the advantage of embodying the weakest versions of asymptotic flatness. We will illustrate their use in the discussion of spatial infinity in Section 3.2. However, the use of coordinate expressions in this strategy also introduces the need of verifying the intrinsic nature of the obtained results, something that it is not always straightforward. For this reason, a geometric manner of describing asymptotic flatness is also desirable, without relying on specific coordinates. This has led to the conformal compactification picture, where infinity is *brought to a finite distance* by an appropriate spacetime conformal transformation. More concretely, one works with an unphysical spacetime $(\tilde{\mathcal{M}}, \tilde{g}_{\mu\nu})$ with boundary, such that the physical spacetime $(\mathcal{M}, g_{\mu\nu})$ is conformally equivalent to the interior of $(\tilde{\mathcal{M}}, \tilde{g}_{\mu\nu})$, that is, $\tilde{g}_{\mu\nu} = \Omega^2 g_{\mu\nu}$. *Infinity* is captured by the boundary $\partial\tilde{\mathcal{M}}$ and is characterized by the vanishing of the conformal factor, $\Omega = 0$. The whole picture is inspired in the structure of the conformal compactification of Minkowski spacetime. The conformal boundary is the union of different pieces, which are classified according to the metric type of the geodesics reaching their points. This defines (past and future) null infinity \mathcal{I}^\pm , spatial infinity i^0 , and (past and future) time-like infinity i^\pm , that is, $\partial\tilde{\mathcal{M}} = \mathcal{I}^\pm \cup i^0 \cup i^\pm$. The conformal spacetime is represented in the so-called Carter–Penrose diagram. Falloff conditions for the characterization of asymptotic flatness are substituted by differentiability conditions on the fields at null and spatial infinity (isolated systems do not require flatness conditions on time-like infinity). Null infinity was introduced in the conformal picture by Penrose [77, 78], the discussion of asymptotic flatness at spatial infinity was developed by Geroch [39], and a unified treatment was presented in [5, 8] (see also [11, 56]). We will briefly illustrate the different approaches to asymptotic flatness in the following sections, but we refer the reader to the existing bibliography (e.g., [37, 87]) for further details.

3.2 Asymptotic Euclidean Slices

The following two sections are devoted to the discussion of conserved quantities at spatial infinity, but they also illustrate the coordinate-based approach to asymptotic flatness. A slice Σ endowed with a space-like 3-metric γ_{ij} is *asymptotically Euclidean* (flat), if there exists a Riemannian background metric f_{ij} such that:

- (i) f_{ij} is flat, except possibly on a compact domain D of Σ .
- (ii) There exists a coordinate system $(x^i) = (x, y, z)$ such that outside D , $f_{ij} = \text{diag}(1, 1, 1)$ (*Cartesian-type coordinates*) and the variable $r := \sqrt{x^2 + y^2 + z^2}$ can take arbitrarily large values on Σ .
- (iii) When $r \rightarrow +\infty$

$$\begin{aligned} \gamma_{ij} &= f_{ij} + O(r^{-1}), & \frac{\partial\gamma_{ij}}{\partial x^k} &= O(r^{-2}), \\ K_{ij} &= O(r^{-2}), & \frac{\partial K_{ij}}{\partial x^k} &= O(r^{-3}). \end{aligned} \quad (19)$$

Given an asymptotically flat spacetime foliated by asymptotically Euclidean slices $\{\Sigma_t\}$, *spatial infinity* is defined by $r \rightarrow +\infty$ and denoted as i^0 .

3.2.1 Asymptotic Symmetries at Spatial Infinity

As commented in the discussion of the Komar quantities, the existence of symmetries provides a natural manner of defining physical parameters as conserved quantities. In the context of spatial infinity, the spacetime diffeomorphisms preserving the asymptotic Euclidean structure (19) are referred to as asymptotic symmetries. Asymptotic symmetries close a Lie group. Since the spacetime is asymptotically flat, one would expect this group to be isomorphic to the Poincaré group. However, the set of diffeomorphisms $(x^\mu) = (t, x^i) \rightarrow (x'^\mu) = (t', x'^i)$ preserving conditions (19) is given by

$$x'^\mu = \Lambda^\mu_\nu x^\nu + c^\mu(\theta, \varphi) + O(r^{-1}), \quad (20)$$

where Λ^μ_ν is a Lorentz matrix and the c^μ 's are four functions of the angles (θ, φ) related to coordinates $(x^i) = (x, y, z)$ by the standard spherical formulas: $x = r \sin \theta \cos \varphi$, $y = r \sin \theta \sin \varphi$, $z = r \cos \theta$. This group indeed contains the Poincaré symmetry, but it is actually much larger due to the presence of *angle-dependent* translations. The latter are known as *supertranslations* and are defined by $c^\mu(\theta, \varphi) \neq \text{const}$ and $\Lambda^\mu_\nu = \delta^\mu_\nu$ in the group representation (20). The corresponding abstract infinite-dimensional symmetry preserving the structure of spatial infinity (*Spi*) is referred to as the *Spi* group [5,8]. The existence of this (infinite-dimensional) Lie structure of asymptotic symmetries has implications in the definition of a global physical mass, linear, and angular momentum at spatial infinity (see below).

3.3 ADM Quantities

Hamiltonian techniques are particularly powerful for the systematic study of physical parameters, considered as conserved quantities under symmetries acting as canonical transformations in the solution (phase) space of a theory. In this sense, the Hamiltonian formulation of General Relativity provides a natural framework for the discussion of global quantities at spatial infinity. This was the original approach adopted by Arnowitt, Deser, and Misner in [4] and we outline here the basic steps.

First, a variational problem for the class of spacetimes we are considering must be set. For a correct formulation we need to specify: (a) the dynamical fields we are varying, (b) the domain \mathcal{V} over which these fields are varied together with the prescribed value of their variations at the boundary $\partial\mathcal{V}$, and (c) the action functional S compatible with the field equations. As integration domain \mathcal{V} we consider the region bounded by two space-like 3-slices Σ_{t_1} and Σ_{t_2} and an outer time-like tube \mathcal{B} . Σ_{t_1} and Σ_{t_2} can be seen as part of a 3-slicing $\{\Sigma_t\}$ with metric and extrinsic curvature given by (γ_{ij}, K^{ij}) , whereas \mathcal{B} has $(\chi_{\mu\nu}, P^{\mu\nu})$ as induced metric and

extrinsic curvature. That is, $\chi_{\mu\nu} = g_{\mu\nu} - u_\mu u_\nu$ and $P_{\mu\nu} = -\gamma_\mu^\rho \nabla_\rho u_\nu$, where u^μ is the unit space-like normal to \mathcal{B} . The dynamical field whose variation we consider is the spacetime metric $g_{\mu\nu}$, under boundary conditions $\delta g_{\mu\nu}|_{\partial\mathcal{V}} = 0$ (note that we impose nothing on variations of the derivatives of $g_{\mu\nu}$). The appropriate gravitational Einstein–Hilbert action then reads (cf., e.g., [79]; the discussion has a straightforward extension to incorporate matter)

$$S = \frac{1}{16\pi} \int_{\mathcal{V}} {}^4R \sqrt{-g} d^4x + \frac{1}{8\pi} \left\{ - \int_{\Sigma_{t_2}} (K - K_0) \sqrt{\gamma} d^3x + \int_{\Sigma_{t_1}} (K - K_0) \sqrt{\gamma} d^3x + \int_{\mathcal{B}} (P - P_0) \sqrt{-\chi} d^3x \right\}, \quad (21)$$

where K and P are the traces of the extrinsic curvatures of the hypersurfaces Σ_{t_i} and \mathcal{B} , respectively, as embedded in $(\mathcal{M}, g_{\mu\nu})$. The subindex 0 corresponds to their extrinsic curvatures as embedded in $(\mathcal{M}, \eta_{\mu\nu})$. The boundary term guarantees the well-posedness of the variational principle, that is, the functional differentiability of the action and the recovery of the correct Einstein field equation, under the assumed boundary conditions for the dynamical fields.

Making use of the 3 + 1 fields decompositions, and considering the intersections $\mathcal{S}_t := \mathcal{B} \cap \Sigma_t$ between space-like 3-slices Σ_t and the time-like hypersurface \mathcal{B} , we can express the action (21) as

$$S = \frac{1}{16\pi} \int_{t_1}^{t_2} \left\{ \int_{\Sigma_t} N ({}^3R + K_{ij} K^{ij} - K^2) \sqrt{\gamma} d^3x + 2 \oint_{\mathcal{S}_t} (H - H_0) \sqrt{q} d^2x \right\} dt \quad (22)$$

where H and H_0 denote the trace of the extrinsic curvature of the 2-surface \mathcal{S}_t as embedded in (Σ_t, γ_{ij}) and (Σ_t, f_{ij}) , respectively. The Lagrangian density L can be read from the form of the action (22). The 3-metric γ_{ij} plays the role of the dynamical variable and the dependence of L on $\dot{\gamma}_{ij}$ follows from the explicit expression of the extrinsic curvature K_{ij} in terms of the lapse and the shift, that is,

$$K_{ij} = \frac{1}{2N} \left(\gamma_{ik} D_j \beta^k + \gamma_{jk} D_i \beta^k - \dot{\gamma}_{ij} \right). \quad (23)$$

In particular no derivatives of N and β^i appear in (22), indicating that the lapse function and the shift vector are not dynamical variables. The Hamiltonian description is obtained by performing a Legendre transformation from variables $(\gamma_{ij}, \dot{\gamma}_{ij})$ to canonical ones (γ_{ij}, π^{ij}) , where

$$\pi^{ij} := \frac{\delta L}{\delta \dot{\gamma}_{ij}} = \frac{1}{16\pi} \sqrt{\gamma} (K \gamma^{ij} - K^{ij}). \quad (24)$$

The Hamiltonian density \mathcal{H} is then given by

$$\mathcal{H} = \pi^{ij} \dot{\gamma}_{ij} - L, \quad (25)$$

and the Hamiltonian follows from an integration over a 3-slice, resulting in (cf. [43, 79] for details)

$$H = \frac{1}{16\pi} \left\{ - \int_{\Sigma_t} (NC_0 + 2\beta^i C_i) \sqrt{\gamma} d^3x - 2 \oint_{S_t} [N(H - H_0) - \beta^i (K_{ij} - K\gamma_{ij})s^j] \sqrt{q} d^2x \right\}, \quad (26)$$

where

$$\begin{aligned} C_0 &:= {}^3R + K^2 - K_{ij}K^{ij}, \\ C_i &:= D_j K_i^j - D_i K. \end{aligned} \quad (27)$$

Functionals C_0 and C_i vanish on solutions of the Einstein equation (in vacuum). More specifically, equations $C_0 = 0$ and $C_i = 0$, respectively, represent the Hamiltonian and momentum constraints of General Relativity, corresponding to the contraction of the Einstein equation (5) with n^μ . From a geometric point of view, they are referred to as the Gauss–Codazzi relations and represent conditions for the embedding of (Σ_t, γ_{ij}) as a submanifold of a spacetime $(\mathcal{M}, g_{\mu\nu})$ with vanishing $n^\mu G_{\mu\nu}$. The evaluation of the gravitational Hamiltonian (26) on solutions to the Einstein equation yields

$$H_{\text{solution}} = -\frac{1}{8\pi} \oint_{S_t} [N(H - H_0) - \beta^i (K_{ij} - K\gamma_{ij})s^j] \sqrt{q} d^2x. \quad (28)$$

Remark 2. Note that in the absence of boundaries the gravitational Hamiltonian vanishes on physical solutions. This is a feature of diffeomorphism invariant theories [58] and reflects the fact that the Hamiltonian, considered as the generator of a canonical transformation, does not move points in the solution space of the theory. In other words, it is a generator of gauge transformations, something consistent with the interpretation of the Hamiltonian as the generator of diffeomorphisms. Note also that the situation changes in the presence of boundaries, where diffeomorphisms not preserving boundary conditions do not correspond to gauge transformations, indicating the presence of residual degrees of freedom (this is of relevance, for instance, in certain aspects of the quantum theory).

3.3.1 ADM Energy

We focus on solutions corresponding to isolated systems and consider 3-slices Σ_t that are asymptotically Euclidean in the sense of conditions (19) (we refer the reader to [1] for a discussion of the total energy in cosmological asymptotically Anti-de

Sitter spacetimes and to [34] for its discussion in higher curvature gravity theories). We choose the lapse and the shift so that the evolution vector t^μ is associated with some asymptotically inertial observer for which $N = 1$ and $\beta^i = 0$ at spatial infinity. In particular, this flow vector t^μ generates asymptotic time translations that, in this asymptotically flat context, constitute actual (asymptotic) symmetries. Conserved quantities under time translations have the physical meaning of an energy. In the present case, the conserved quantity is referred to as the ADM energy. The latter is obtained from expression (28) by making $N = 1$ and $\beta^i = 0$ and taking the limit to spatial infinity, namely $r \rightarrow \infty$ in the well-defined sense of Section 3.2. That is,

$$E_{\text{ADM}} := -\frac{1}{8\pi} \lim_{\mathcal{S}_{(t,r \rightarrow \infty)}} \oint_{\mathcal{S}_t} (H - H_0) \sqrt{q} d^2x. \quad (29)$$

This ADM energy represents the total energy contained in the slice Σ_t . Using the explicit expression of the extrinsic curvature in terms of metric components, the ADM energy can be written as

$$E_{\text{ADM}} = \frac{1}{16\pi} \lim_{\mathcal{S}_{(t,r \rightarrow \infty)}} \oint_{\mathcal{S}_t} \left[\mathcal{D}^j \gamma_{ij} - \mathcal{D}_i (f^{kl} \gamma_{kl}) \right] s^i \sqrt{q} d^2x, \quad (30)$$

where \mathcal{D}_i stands for the connection associated with the metric f_{ij} and, consistently with notation in Section 1.3, s^i corresponds to the unit normal to \mathcal{S}_t tangent to Σ_t and oriented toward the exterior of \mathcal{S}_t (note that when $r \rightarrow \infty$ the normalization with respect to γ_{ij} and f_{ij} are equivalent). In particular, if we use the Cartesian-like coordinates employed in (19) we recover the standard form (see, e.g., [87])

$$E_{\text{ADM}} = \frac{1}{16\pi} \lim_{\mathcal{S}_{(t,r \rightarrow \infty)}} \oint_{\mathcal{S}_t} \left(\frac{\partial \gamma_{ij}}{\partial x^j} - \frac{\partial \gamma_{jj}}{\partial x^i} \right) s^i \sqrt{q} d^2x. \quad (31)$$

Remark 3. We note that asymptotic flatness conditions (19) guarantee the finite value of the integral since the $O(r^2)$ part of the measure $\sqrt{q} d^2x$ is compensated by the $O(r^{-2})$ parts of $\partial \gamma_{ij} / \partial x^j$ and $\partial \gamma_{jj} / \partial x^i$. It is very important to point out that finiteness of the ADM energy relies on the subtraction of the *reference* value H_0 in Eq. 29.

Conformal Decomposition Expression of the ADM Energy

A useful expression for the ADM energy in certain formulations of the Einstein equation is given in terms of a conformal decomposition of the 3-metric

$$\gamma_{ij} = \Psi^4 \tilde{\gamma}_{ij}. \quad (32)$$

Choosing the representative $\tilde{\gamma}_{ij}$ of the conformal class by the unimodular condition $\det(\tilde{\gamma}_{ij}) = \det(f_{ij}) = 1$, conditions (19) translate into

$$\begin{aligned} \Psi &= 1 + O(r^{-1}), & \frac{\partial \Psi}{\partial x^k} &= O(r^{-2}), \\ \tilde{\gamma}_{ij} &= f_{ij} + O(r^{-1}), & \frac{\partial \tilde{\gamma}_{ij}}{\partial x^k} &= O(r^{-2}), \end{aligned} \quad (33)$$

for the conformal factor and the conformal metric. Then it follows [43]

$$E_{\text{ADM}} = -\frac{1}{2\pi} \lim_{S_{(t,r \rightarrow \infty)}} \oint_{S_t} s^i \left(\mathcal{D}_i \Psi - \frac{1}{8} \mathcal{D}^j \tilde{\gamma}_{ij} \right) \sqrt{q} \, d^2x. \quad (34)$$

Note that whereas in the time-symmetric ($K_{\mu\nu} = 0$) conformally flat case the Komar mass is given in terms of the monopolar term in the asymptotic expansion of the (adapted) lapse N [cf. Eq. 16], the ADM energy is given by the monopolar term in ψ (the latter holds more generally under a vanishing *Dirac-like* gauge condition on $\mathcal{D}^j \tilde{\gamma}_{ij}$).

Example 1 (Newtonian Limit). As an application of expression (34) we check that the ADM energy recovers the standard result in the Newtonian limit. For this we assume that the gravitational field is weak and static. In this setting it is always possible to find a coordinate system $(x^\mu) = (x^0 = ct, x^i)$ such that the metric components take the form

$$-d\tau^2 = g_{\mu\nu} dx^\mu dx^\nu = -(1 + 2\Phi) dt^2 + (1 - 2\Phi) f_{ij} dx^i dx^j, \quad (35)$$

where again f_{ij} is the flat Euclidean metric in the 3-dimensional slice and Φ is the Newtonian gravitational potential, solution of the Poisson equation $\Delta\Phi = 4\pi\rho$ where ρ is the mass density (we recall that we use units in which the Newton's gravitational constant G and the light velocity c are unity). Then, using $\Psi = (1 - 2\Phi)^{1/4} \approx 1 - \frac{1}{2}\Phi$, Eq. 34 translates into

$$E_{\text{ADM}} = \frac{1}{4\pi} \lim_{S_{(t,r \rightarrow \infty)}} \oint_{S_t} s^i \mathcal{D}_i \Phi \sqrt{q} \, d^2x = \frac{1}{4\pi} \int_{\Sigma_t} \Delta\Phi \sqrt{f} \, d^3x \quad (36)$$

where in the second step we have assumed that Σ_t has the topology of \mathbb{R}^3 and have applied the Gauss–Ostrogradsky theorem (with $\Delta = \mathcal{D}_i \mathcal{D}^i$). Using now that Φ is a solution of the Poisson equation, we can write

$$E_{\text{ADM}} = \int_{\Sigma_t} \rho \sqrt{f} \, d^3x, \quad (37)$$

and we recover the standard expression for the total mass of the system at the Newtonian limit (as it will be seen in next section, in a non-boosted slice like this, mass is directly given by the energy expression).

3.3.2 ADM 4-Momentum and ADM Mass

ADM Linear Momentum

Linear momentum corresponds to the conserved quantity associated with an invariance under spatial translations. In the asymptotically flat case, the ADM momentum is associated with space translations preserving the falloff conditions (19) expressed in terms of the Cartesian-type coordinates (x^i). Given one of such coordinate systems, the three vectors $(\partial_i)_{i \in \{1,2,3\}}$ represent asymptotic symmetries generating asymptotic spatial translations that correspond to a choice $N=0$ and $\beta^i_{(\partial_j)} = \delta^i_j$ in the evolution vector t^μ . Substituting these values for the lapse and shift in the Hamiltonian expression evaluated on solutions (28), we obtain the conserved quantity under the infinitesimal translation ∂_i :

$$P_i := \frac{1}{8\pi} \lim_{S_{(t,r \rightarrow \infty)}} \oint_{S_t} (K_{ik} - K\gamma_{ik}) s^k \sqrt{q} d^2x. \quad (38)$$

Remark 4. Asymptotic falloff conditions (19) guarantee the finiteness of expression (38) for P_i .

The *ADM momentum* associated with the hypersurface Σ_t is defined as the linear form $(P_i) = (P_1, P_2, P_3)$. Its components actually transform as those of a linear form under changes of Cartesian coordinates $(x^i) \rightarrow (x'^i)$, which asymptotically correspond to a rotation and/or a translation. For discussing transformations under the full Poincaré group, we must introduce the ADM 4-momentum defined as

$$(P_\mu^{\text{ADM}}) := (-E_{\text{ADM}}, P_1, P_2, P_3). \quad (39)$$

Under a coordinate change $(x^\mu) = (t, x^i) \rightarrow (x'^\mu) = (t', x'^i)$ which preserves the asymptotic conditions (19), that is, any coordinate change of the form (20), components P_μ^{ADM} transform under the vector linear representation of the Lorentz group

$$P'^{\text{ADM}}_\mu = (\Lambda^{-1})^\nu_\mu P_\nu^{\text{ADM}}, \quad (40)$$

as first shown by Arnowitt, Deser, and Misner in [4]. Therefore (P_μ^{ADM}) can be seen as a linear form acting on vectors at spatial infinity i^0 and is called the *ADM 4-momentum*.

ADM Mass

Having introduced the ADM 4-momentum, its Minkowskian length provides a notion of mass. The ADM mass is therefore defined as

$$M_{\text{ADM}}^2 := -P_\mu^{\text{ADM}} P_\mu^{\text{ADM}}, \quad M_{\text{ADM}} = \sqrt{E_{\text{ADM}}^2 - P_i P^i}. \quad (41)$$

Remark 5. In the literature, references are found where the term *ADM mass* actually refers to this length of the ADM 4-momentum and other references where it refers to its time component that we have named here as the ADM energy. These differences somehow reflect traditional usages in Special Relativity where the term *mass* is sometimes reserved to refer to the Poincaré invariant (rest-mass) quantity, and in other occasions is used to denote the boost-dependent time component of the energy–momentum.

The ADM mass is a time independent quantity. Time evolution is generated by the Hamiltonian in expression (26). The time variation of a given quantity F defined on the phase space is expressed as the sum of its Poisson bracket with the Hamiltonian (accounting for the implicit time dependence through the time variation of the phase space variables) and the partial derivative of F with respect to time. Since in expression (26) there is no explicit time dependence, constancy of the ADM mass follows:

$$\frac{d}{dt}M_{\text{ADM}} = 0. \quad (42)$$

As a consequence of this, the ADM mass is a property of the whole (asymptotically flat) spacetime.

Remark 6 (Relation Between ADM and Komar Masses). Komar mass is defined only in the presence of a time-like Killing vector k^μ (more generally, cf. [71] for an early critical account of its physical properties). However, in the asymptotically flat case we can discuss the relation between the ADM mass and the Komar mass associated with an asymptotic inertial observer. Though the relation is not straightforward from explicit expressions (18) and (30), it can be shown [11, 14] that for any foliation $\{\Sigma_t\}$ such that the associated unit normal n^μ coincides with the time-like Killing vector k^μ at infinity (i.e., $N \rightarrow 1$ and $\beta^i \rightarrow 0$) we have

$$M_K = M_{\text{ADM}}. \quad (43)$$

As a practical application, this relation has been used as a quasi-equilibrium condition in the construction of initial data for compact objects in quasi-circular orbits (e.g., [46]).

Positivity of the ADM Mass

One of the most important results in General Relativity is the proof of the positivity of the ADM mass under appropriate energy conditions for the matter energy–momentum tensor. This is important first on conceptual grounds, since it represents a crucial test of the internal consistency of the theory. A violation of this result would evidence an essential instability of the solutions of the theory. It is also relevant on a practical level, since this theorem (and/or related results) pervades the everyday practice of (mathematical) relativists.

The theorem states that, under the *dominant energy condition*, the ADM mass cannot be negative, that is, $M_{\text{ADM}} \geq 0$. Moreover, $M_{\text{ADM}} = 0$ if and only if the spacetime is Minkowski. This result was first obtained by Schoen and Yau [81, 82] and then recovered using spinorial techniques by Witten [92] (see in this sense [33] for a previous related mass positivity result in supergravity).

The dominant energy condition essentially states that the local energy measured by a given observer is always positive, and that the flow of energy associated with this observer cannot travel faster than light. More precisely, given a future-directed time-like vector v^μ , this condition states that the vector $-T^\mu{}_\nu v^\nu$ is a future-oriented causal vector. Vector $-T^\mu{}_\nu v^\nu$ represents the energy–momentum 4-current density as seen by the observer associated with v^μ , in an analogous decomposition to that in Eq. 12. From the dominant energy condition it follows $E := T_{\mu\nu} v^\mu v^\nu \geq 0$, that is, the local density cannot be negative (*weak energy condition*) and, more generally, $E \geq \sqrt{P^i P_i}$.

3.3.3 ADM Angular Momentum

Pushing forward the strategy followed for defining the ADM mass and linear momentum, one would attempt to introduce total angular momentum as the conserved quantity associated with rotations at spatial infinity. More specifically, in the Cartesian-type coordinates used for characterizing asymptotically Euclidean slices (19), infinitesimal generators $(\phi_i)_{i \in \{1,2,3\}}$ for rotations around the three spatial axes are

$$\phi_x = -z\partial_y + y\partial_z, \quad \phi_y = -x\partial_z + z\partial_x, \quad \phi_z = -y\partial_x + x\partial_y, \quad (44)$$

which constitute Killing symmetries of the asymptotically flat metric. When using the associated *lapse* functions and *shift* vectors in the Hamiltonian expression (28), namely, $N = 0$ and $\beta_{(\phi_j)}^i = (\phi_j)^i$, the following three quantities result

$$J_i := \frac{1}{8\pi} \lim_{S_{(t,r \rightarrow \infty)}} \oint_{S_t} (K_{jk} - K\gamma_{jk}) (\phi_i)^j s^k \sqrt{q} d^2x, \quad i \in \{1, 2, 3\}. \quad (45)$$

However, the interpretation of J_i as the components of an angular momentum faces two problems:

1. First, asymptotic falloff conditions (19) are not sufficient to guarantee the finiteness of expressions (45).
2. Second, in contrast with the linear momentum case, the quantity $(J_i) = (J_1, J_2, J_3)$ does not transform appropriately under transformations (20) preserving (19). This can be tracked to the existence of supertranslations. In particular, the so-defined angular-momentum vector (J_i) depends non-covariantly on the particular coordinates we have chosen.

For this reason, it is not appropriate to refer to an ADM angular momentum in the same sense that we use the ADM term for mass and linear momentum quantities. A manner of removing the above-commented ambiguities consists in identifying an appropriate subclass of Cartesian-type coordinates where, first, the J_i components are finite and, second, they transform as the components of a linear form. Among the different strategies proposed in the literature, we comment here on the one proposed by York [94] in terms of further conditions on the conformal metric $\tilde{\gamma}_{ij}$ introduced in Eq. 32 and the trace of the extrinsic curvature K . Namely,

$$\frac{\partial \tilde{\gamma}_{ij}}{\partial x^j} = O(r^{-3}), \quad K = O(r^{-3}), \quad (46)$$

representing *asymptotic gauge conditions*. That is, they actually impose restrictions on the choice of coordinates but not on the geometric properties of space-time at spatial infinity. First condition in Eq. 46 is known as the *quasi-isotropic gauge*, whereas the second one is referred to as the *asymptotic maximal gauge*.

Remark 7. Note that, in contrast with the total angular momentum defined at spatial infinity, no ambiguity shows up in the definition of the Komar angular momentum in Eq. 17.

3.4 Bondi Energy and Linear Momentum

We could introduce Bondi (or Trautman–Bondi–Sachs) energy at null infinity following the same approach we have employed for the ADM energy, that is, by taking the appropriate limit of Eq. 28 with $N = 1$ and $\beta^i = 0$. In the present case, instead of keeping t constant and making $r \rightarrow \infty$ as we did in Eq. 29, we should introduce retarded and advanced time coordinates (respectively, $u = t - r$ and $v = t + r$) and consider the limit

$$E_{\text{BS}} := -\frac{1}{8\pi} \lim_{S(u,v \rightarrow \infty)} \oint_{S_u} (H - H_0) \sqrt{q} d^2x. \quad (47)$$

The full discussion of this limit would require the introduction of the appropriate falloff conditions for the metric components in a special class of coordinate system adapted to null infinity (Bondi coordinates). This is in the spirit of the original discussion on the energy flux of gravitational radiation from an isolated system by Bondi, Van der Burg, and Metzner [18], and Sachs [80]. However, aiming at providing some flavor of the geometric approach to asymptotic flatness, we rather outline here a discussion in the setting of the conformal compactification approach.

3.4.1 Null Infinity

A smooth spacetime (\mathcal{M}, g) is *asymptotically simple* [76] (see e.g., also [37]) if there exists another (unphysical) smooth Lorentz manifold $(\tilde{\mathcal{M}}, \tilde{g})$ such that

- (i) \mathcal{M} is an open submanifold of $\tilde{\mathcal{M}}$ with (smooth) boundary $\partial\tilde{\mathcal{M}}$.
- (ii) There is a smooth scalar field Ω on $\tilde{\mathcal{M}}$, such that $\Omega > 0$, $\tilde{g}_{\mu\nu} = \Omega^2 g_{\mu\nu}$ on \mathcal{M} , and $\Omega = 0$, $\partial_\mu \Omega \neq 0$ on $\partial\tilde{\mathcal{M}}$.
- (iii) Every null geodesic in \mathcal{M} begins and ends on $\partial\tilde{\mathcal{M}}$.

An asymptotically simple spacetime is *asymptotically flat* (at null infinity) if, in addition, the Einstein vacuum equation is satisfied in a neighbourhood of $\partial\tilde{\mathcal{M}}$ (or the energy–momentum decreases sufficiently fast in the matter case). In this case the boundary $\partial\tilde{\mathcal{M}}$ consists, at least, of a null hypersurface with two connected components $\mathcal{I} = \mathcal{I}^- \cup \mathcal{I}^+$, each one with topology $S^2 \times \mathbb{R}$ (note that in Minkowski $\partial\tilde{\mathcal{M}}$ also contains the points i^0, i^\pm). Boundaries \mathcal{I}^- and \mathcal{I}^+ represent past and future null infinity, respectively.

3.4.2 Symmetries at Null Infinity

In order to characterize a vector ξ^μ in \mathcal{M} as an infinitesimal asymptotic symmetry at (future) null infinity \mathcal{I}^+ , we must assess the vanishing of $\mathcal{L}_\xi g_{\mu\nu}$ as one gets to \mathcal{I}^+ . For this, we require first that ξ^μ , considered as a vector field in the unphysical spacetime (i.e., under the immersion of \mathcal{M} into $\tilde{\mathcal{M}}$), can be smoothly extended to \mathcal{I}^+ . Then ξ^μ is characterized as an asymptotic symmetry by demanding that $\Omega^2 \mathcal{L}_\xi g_{\mu\nu}$ can also be smoothly extended to \mathcal{I}^+ and vanishes there, that is,

$$\left(\tilde{\nabla}_\mu \xi_\nu + \tilde{\nabla}_\nu \xi_\mu - 2\Omega^{-1} \xi^\rho \tilde{\nabla}_\rho \Omega \tilde{g}_{\mu\nu} \right) \Big|_{\mathcal{I}^+} = 0. \quad (48)$$

Two vector fields ξ^μ and ξ'^μ are considered to generate the same infinitesimal asymptotic symmetry if their extensions to \mathcal{I}^+ coincide. The equivalence class of such vector fields, which we will still denote by ξ^μ , generates the asymptotic symmetry group at \mathcal{I}^+ . This is known as the Bondi–Metzner–Sachs (BMS) group and is universal in the sense that it is same for every asymptotically flat spacetime. The BMS group is infinite-dimensional, as it was the case of the Spi group at spatial infinity. It does not only contain the Poincaré group, but actually is a semi-direct product of the Lorentz group and the infinite-dimensional group of *angle dependent* supertranslations (see details in, e.g., [87]). The key point for the present discussion is that it possesses a unique *canonical* set of asymptotic 4-translations characterized as the only 4-parameter subgroup of the supertranslations that is a *normal* subgroup of the BMS group. This leads us to the Bondi–Sachs 4-momentum.

3.4.3 Bondi–Sachs 4-Momentum

As mentioned above, the original introduction of the Bondi energy was based in the identification of certain expansion coefficients in the line element of radiative spacetimes in adapted (Bondi) coordinates [18]. A Hamiltonian analysis, counterpart of the approach adopted in Section 3.3 for introducing the ADM mass, can be found in [89]. Here we rather follow a construction based on the Komar mass expression. Though Eq. 13 only defines a conserved quantity for a Killing vector k^μ , the vector fields ξ_a^μ ($a \in \{0, 1, 2, 3\}$) corresponding to the 4-translations at \mathcal{I}^+ get closer to an infinitesimal symmetry as one approaches \mathcal{I}^+ . Therefore, one can expect that a Komar-like expression makes sense for a given cross section S_u of \mathcal{I}^+ . This is indeed the case and the evaluation of the integral does not depend on how we get to S_u . However, the integral does depend on the representative ξ^μ in the class of vectors corresponding to the asymptotic symmetry. This is cured by imposing a divergence-free condition on ξ^μ [41]. Bondi–Sachs 4-momentum at $S_u \subset \mathcal{I}^+$ is then defined as

$$P_a^{\text{BS}} := -\frac{1}{8\pi} \lim_{(S \rightarrow S_u)} \oint_S \nabla^\mu \xi_a^\nu dS_{\mu\nu}, \quad \nabla_\mu \xi_a^\mu = 0. \quad (49)$$

Alternatively, ambiguities in the Komar integral can be solved by dropping the condition on the divergence and adding a term $\alpha \nabla_\mu \xi_a^\mu$ to the surface integral. When $\alpha = 1$ the resulting integral is called the *linkage* [91]. The discussion of Bondi–Sachs angular momentum is more delicate. We refer the reader to the discussion in Section 3.2.4 of [85].

Bondi Energy and Positivity of Gravitational Radiation Energy

Bondi energy E_{BS} (the zero component of the Bondi–Sachs 4-momentum) is a decreasing function of the retarded time. More concretely, Bondi energy satisfies a *loss equation*

$$\frac{dE_{\text{BS}}}{du} = - \int_{S_u} F \sqrt{q} d^2x, \quad (50)$$

where $F \geq 0$ can be expressed in terms of the squares of the so-called *news functions*. In [11] it is shown that if the *news tensor* satisfies the appropriate conditions, then Bondi mass coincides initially with the ADM mass (see also [56]). Bondi energy is interpreted as the remaining of the ADM energy in the process of energy extraction by gravitational radiation. As for the ADM mass, a positivity result holds for the Bondi mass [61, 83]. These properties constitute the underlying conceptual/structural justification of our understanding of energy radiation by gravitational waves: gravitational radiation carries positive energy away from isolated radiating systems, and the total radiated energy cannot be bigger than the original total ADM energy.

4 Notions of Mass for Bounded Regions: Quasi-Local Masses

As commented in Section 1.2, the convenience of associating energy–momentum with the gravitational field in given regions of the spacetime is manifest in very different contexts of gravity physics. More specifically, mathematical and numerical General Relativity or approaches to Quantum Gravity provide examples where we need to associate such an energy–momentum with a *finite* region of spacetime. This can be either motivated by the need to define appropriate physical/astrophysical quantities, or by the convenience of finding quasi-local quantities with certain desirable mathematical properties (e.g., positivity, monotonicity, etc.) in the study of a specific problem.

There exist many different approaches for introducing quasi-local prescriptions for the mass and angular momentum. Some of them can be seen as *quasi-localizations* of successful notions for the physical parameters of the total system, such as the ADM mass, whereas other attempts constitute genuine ab initio methodological constructions, mainly based on Lagrangian or Hamiltonian approaches. An important drawback of most of them in the context of the present article is that, typically, they involve *constructions* that are difficult to capture in short mathematical definitions without losing the underlying physical/geometrical insights. An excellent and comprehensive review is reference [85] by Szabados.

4.1 Ingredients in the Quasi-Local Constructions

First, the relevant bounded spacetime domain must be identified. Typically, these are compact space-like domains D with a boundary given by a closed 2-surface \mathcal{S} . Explicit expressions, such as relevant associated integrals, are formulated in terms of either the (3-dimensional) domain D itself or on its boundary \mathcal{S} . In particular, conserved-current strategies permit to pass from the 3-volume integral to a conserved-charge-like 2-surface integral. In other cases, 2-surface integrals are a consequence of the need of including boundary terms for having a correct variational formulation (as it was the case in the Hamiltonian formulation of Section 3.3).

We have already presented an example of quasi-local quantity in Section 2, namely the Komar quantities. Since symmetries will be absent in the generic case, an important ingredient in most quasi-local constructions is the prescription of some vector field that plays the role that infinitesimal symmetries had played in case of being present. In connection with this, one usually needs to introduce some background structure that can be interpreted as a kind of gauge choice.

Finally, different *plausibility* criteria for the assessment of the proposed quasi-local expressions (e.g., positivity, monotonicity, recovery of known limits, etc.), need to be considered (see [85]).

4.2 Some Relevant Quasi-Local Masses

4.2.1 Round Spheres: Misner–Sharp Energy

In some special situations, as it is the case of isolated systems above and some exact solutions, there is agreement on the form of the gravitational field energy–momentum. Another interesting case is that of spherically symmetric spacetimes, where the rotation group $SO(3)$ acts transitively as an isometry. Orbits under this rotation group are *round* spheres \mathcal{S} . Then, using the areal radius r_A as a coordinate ($4\pi r_A^2 = A$), an appropriate notion of mass/energy was given by Misner and Sharp [72]

$$E(\mathcal{S}) := \frac{1}{8} r_A^3 R_{\mu\nu\rho\sigma} {}^2\epsilon^{\mu\nu} {}^2\epsilon^{\rho\sigma}, \quad (51)$$

where ${}^2\epsilon_{\mu\nu} = n^\rho s^\sigma {}^4\epsilon_{\rho\sigma\mu\nu}$ (cf. Section 1.3) is the volume element on \mathcal{S} . This expression is related to the so-called Kodama vector K^μ , which can be defined in spherically symmetric spacetimes and such that $\nabla_\mu (G^{\mu\nu} K_\nu) = 0$. The current $S^\mu = G^{\mu\nu} K_\nu$ is thus conserved and, taking D as a solid ball of radius r_A , the flux of S^μ through the round boundary ∂D actually equals the change in time of the mass expression (51). Misner–Sharp proposal is considered as the *standard form* of quasi-local mass for round spheres.

4.2.2 Brown–York Energy

The rationale of the approach in Ref. [23] to quasi-local energy strongly relies on the well-posedness of a variational problem for the gravitational action. The adopted variational formulation is essentially the one outlined in Section 3.3 (where the discussion was in fact based in the treatment in [79] adapted from [23]). However, if the main interest is placed in the expressions of quasi-local parameters and not in the details of the symplectic geometry of the system phase space, a full Hamiltonian analysis does not need to be undertaken and one can rather follow a Hamilton–Jacobi one. The latter starts from action (21) defined on the spacetime domain \mathcal{V} . We recall that the boundary $\partial\mathcal{V}$ is given by two space-like hypersurfaces Σ_1 and Σ_2 and a time-like tube \mathcal{B} , such that the 2-spheres \mathcal{S}_i are the intersections between Σ_i and \mathcal{B} . The metric and extrinsic curvatures on Σ_i are given by $\gamma_{\mu\nu}$ and $K^{\mu\nu}$, whereas those on \mathcal{B} are denoted by $\chi_{\mu\nu}$ and $P^{\mu\nu}$. A Hamilton–Jacobi principal function can then be introduced by evaluating the action S on classical trajectories. An arbitrary function S^0 of the data on the boundaries can be added to S [it is the responsible of the *reference* terms with subindex 0 in expression (21)]. The principal function is given by $S_{\text{Cl}} := (S - S^0)|_{\text{Cl}}$ and Hamilton–Jacobi equations are obtained from its variation with respect to the data at the final slice Σ_2 . One of the Hamilton–Jacobi equations leads to the definition of a *surface stress–energy–momentum* tensor as

$$\tau^{\mu\nu} := \frac{-2}{\sqrt{-\chi}} \frac{\delta S_{\text{Cl}}}{\delta \chi_{\mu\nu}} = \frac{1}{8\pi} \{ (P \chi^{\mu\nu} - P^{\mu\nu}) - (P_0 \chi^{\mu\nu} - P_0^{\mu\nu}) \}. \quad (52)$$

This tensor satisfies a conservation-like equation with a source given in terms of the matter energy–momentum tensor $T^{\mu\nu}$. This motivates the definition of the *charge* $Q_S(\xi^\mu)$ associated with a vector ξ^μ as

$$Q_S(\xi^\mu) := \oint_S \xi_\rho \tau^{\rho\nu} n_\nu \sqrt{q} d^2x, \quad (53)$$

whose change along the tube \mathcal{B} is given by a matter flux. This expression is analogous to Eq. 1 in the matter case (here $\mathcal{S} \subset \mathcal{B}$ and n^μ is the time-like unit normal to \mathcal{S} and tangent to \mathcal{B}).

Using the 2 + 1 decomposition induced by a 3 + 1 space-like slicing $\{\Sigma_t\}$, we can decompose the tensor $\tau^{\mu\nu}$ as we did for the matter energy–momentum tensor $T^{\mu\nu}$ in Eq. 12. Writing explicitly the time-like components, it results

$$\begin{aligned} \varepsilon &:= n_\mu n_\nu \tau^{\mu\nu} = -\frac{1}{8\pi}(H - H^0), \\ j_\mu &:= -q_{\mu\nu} n_\rho \tau^{\nu\rho} = \frac{1}{8\pi} q_{\mu\nu} s_\rho (K\gamma^{\nu\rho} - K^{\nu\rho})|_0^{\text{Cl}}. \end{aligned} \quad (54)$$

Expressing the vector ξ^μ in the 3+1 decomposition $\xi^\mu = \xi n^\mu + \xi_\perp^\mu$ and considering a 2-surface \mathcal{S} lying in a slice of $\{\Sigma_t\}$, we have

$$Q_S(\xi^\mu) = \oint_S \xi_\rho \tau^{\rho\nu} n_\nu \sqrt{q} d^2x = \oint_S (\xi\varepsilon - \xi_\perp^\rho j_\rho) \sqrt{q} d^2x. \quad (55)$$

The Brown–York energy is then [cf. with the ADM mass expression (41)]

$$E_{\text{BY}}(\mathcal{S}, n^\mu) := Q_S(n^\mu) = -\frac{1}{8\pi} \oint_S (H - H_0) \sqrt{q} d^2x. \quad (56)$$

Note that this expression explicitly depends on the manner in which \mathcal{S} is inserted in some space-like 3-slice. In this sense, it corresponds to an energy (depending on a boost) rather than a mass.

Kijowski, Epp, Liu-Yau, and Kijowski-Liu-Yau Expressions

We briefly comment on some expressions that can be related to the Brown–York energy. Studying more general boundary conditions than the ones in [23], Kijowski proposed the following quasi-local expression for the mass [66]

$$E_{\text{Kij}} := \frac{1}{16\pi} \oint_S \frac{(H_0)^2 - (H^2 - L^2)}{H_0} \sqrt{q} d^2x, \quad (57)$$

where $H = H_{\mu\nu} q^{\mu\nu}$ and $L = L_{\mu\nu} q^{\mu\nu}$ are the traces of the extrinsic curvatures of \mathcal{S} with respect to unit orthogonal space-like s^μ and time-like n^μ vectors, that is, $n^\mu s_\mu = 0$ (cf. notation in Section 1.3). Apart from the choice of the background terms H_0 , this expression only depends on \mathcal{S} , and not in the manner of

embedding it into some space-like hypersurface. Using a different set of boundary conditions, another quasi-local quantity was introduced by Kijowski (referred to as a *free energy*). The same quantity was later derived by Liu and Yau, using a different approach [69]. We will refer to the resulting quasi-local energy as the Kijowski-Liu-Yau energy, having the form

$$E_{\text{KLY}} := \frac{1}{8\pi} \oint_{\mathcal{S}} \left(H^0 - \sqrt{H^2 - L^2} \right). \quad (58)$$

On the other hand, aiming at removing the dependence of Brown–York energy on the space-like hypersurface, Epp [36] proposed the following boost-invariant expression

$$E_{\text{E}} := \frac{1}{8\pi} \oint_{\mathcal{S}} \left(\sqrt{(H^0)^2 - (L^0)^2} - \sqrt{H^2 - L^2} \right). \quad (59)$$

Note that Brown–York energy can be seen as a gravitational field version of the quasi-local matter energy (6), whereas Epp’s expression rather corresponds to the matter mass (7). For further recent work along this approach to quasi-local mass, see [74, 90].

4.2.3 Hawking, Geroch, and Hayward Energies

Hawking Energy

Given a topological sphere \mathcal{S} , its Hawking energy is defined as [49]

$$E_{\text{H}}(\mathcal{S}) = \sqrt{\frac{A(\mathcal{S})}{16\pi}} \left(1 + \frac{1}{8\pi} \oint_{\mathcal{S}} \theta_+ \theta_- \right) \sqrt{q} \, d^2x, \quad (60)$$

where $\theta_+ = q^{\mu\nu} \Theta_{\mu\nu}^{(\ell)}$ and $\theta_- = q^{\mu\nu} \Theta_{\mu\nu}^{(k)}$ are the expansions associated with outgoing and ingoing null normals (cf. notation in Section 1.3). It can be motivated by understanding the mass surrounded by the 2-sphere \mathcal{S} as an estimate of the bending of ingoing at outgoing light rays from \mathcal{S} . An average, boost-independent measure of this convergence-divergence behaviour of light rays is given by $\oint_{\mathcal{S}} \theta_+ \theta_-^2 \epsilon$. Then, from the Ansatz $A + B \oint_{\mathcal{S}} \theta_+ \theta_-^2 \epsilon$, the constants A and B are fixed from round spheres in Minkowski and from the horizon sections in Schwarzschild spacetime.

Hawking energy depends only on the surface \mathcal{S} and not on any particular embedding of it in a space-like hypersurface. In the spherically symmetric case it recovers the standard Misner–Sharp energy (51). For apparent horizons, or more generally for marginally trapped surfaces, it reduces to the *irreducible mass* accounting for the energy that cannot be extracted from a black hole by a Penrose process and that is given entirely in terms of the area. Hawking energy does not satisfy a positivity criterion, since it can be negative even in Minkowski spacetime. However, for large spheres approaching null infinity, $E_{\text{H}}(\mathcal{S})$ recovers Bondi–Sachs energy, whereas for

spheres approaching spatial infinity it tends to the ADM energy. Though it is not monotonic in the generic case, monotonicity can be proved for sequences of spheres obtained from appropriate geometric flows. This has a direct interest for the extension of Huisken & Ilmanen proof [62] of the Riemannian Penrose inequality to the general case.

Geroch Energy

For a surface \mathcal{S} embedded in a space-like hypersurface Σ , Geroch energy [40] is defined as

$$E_G(\mathcal{S}) := \frac{1}{16\pi} \sqrt{\frac{A(\mathcal{S})}{16\pi}} \oint_{\mathcal{S}} (2 {}^2R - H^2) \sqrt{q} d^2x, \quad (61)$$

where H is again the trace of the extrinsic curvature of \mathcal{S} inside Σ . Geroch energy is never larger than Hawking energy, but it can be proved that it also tends to the ADM mass for spheres approaching spatial infinity.

The relevance of Geroch energy lies on its key role in the first proof of the Riemannian Penrose, by Huisken & Ilmanen [62] (see also Section 5.1). In particular, use is made of the monotonicity properties of E_G under an *inverse mean curvature* flow in Σ .

Hayward Energy

Some generalizations of Hawking energy exist. A vanishing expression for flat spacetimes can be obtained by considering the modified expression

$$E'_H(\mathcal{S}) = \sqrt{\frac{A(\mathcal{S})}{16\pi}} \left(1 + \frac{1}{8\pi} \oint_{\mathcal{S}} \theta_+ \theta_- - \frac{1}{2} \sigma_{\mu\nu}^+ \sigma_{\mu\nu}^- \right) \sqrt{q} d^2x, \quad (62)$$

where the shears $\sigma_{\mu\nu}^+$ and $\sigma_{\mu\nu}^-$ are the traceless parts of $\Theta_{\mu\nu}^{(\ell)}$ and $\Theta_{\mu\nu}^{(k)}$, respectively. E'_H still asymptotes to the ADM energy at spatial infinity, but does not recover Bondi–Sachs energy at null infinity (but rather Newman–Unti one; cf. references in [85]). Related to this modified Hawking energy, Hayward has proposed [55] another quasi-local energy expression by taking into account the anholonomicity form Ω_μ , one of the normal fundamental 1-forms introduced in Section 1.3

$$E_{\text{Hay}}(\mathcal{S}) = \sqrt{\frac{A(\mathcal{S})}{16\pi}} \left(1 + \frac{1}{8\pi} \oint_{\mathcal{S}} \theta_+ \theta_- - \frac{1}{2} \sigma_{\mu\nu}^+ \sigma_{\mu\nu}^- - 2\Omega_\mu \Omega^\mu \right) \sqrt{q} d^2x. \quad (63)$$

Though the divergence-free part of Ω_μ can be related to angular momentum (see below), this 1-form is a gauge-dependent object changing by a total differential under a boost transformation. Therefore, some natural gauge for fixing the boost freedom is needed.

4.2.4 Bartnik Mass

Bartnik quasi-local mass is an example of *quasi-localization* of a global quantity, in particular the ADM mass. In very rough terms, the idea in Bartnik’s construction consists in defining the mass of a compact space-like 3-domain D as the ADM mass of that asymptotically Euclidean slice Σ that contains D without any other *source* of energy. The strategy to address this absence of further energy is to consider all plausible extensions of D into Euclidean slices, calculate the ADM mass for all them, and then consider the infimum of this set of ADM masses. In more precise terms, let us consider a compact, connected 3-hypersurface D in spacetime, with boundary S and induced metric γ_{ij} . Bartnik’s construction actually focuses on time-symmetric $K_{ij} = 0$ domains D . Let us also assume that a dominant energy condition (though the original formulation in [13] makes use of a weak-energy-constraint condition) is satisfied. In a time-symmetric context this amounts to the positivity of the Ricci scalar, ${}^3R \geq 0$. One can then define $\mathcal{P}(D)$ as the set of Euclidean time-symmetric initial data sets (Σ, γ_{ij}) satisfying the dominant energy condition, with a single asymptotic end, finite ADM mass $M_{\text{ADM}}(\Sigma)$, not containing horizons (minimal surfaces in this context) and extending D through its boundary S . Then, Bartnik’s mass [13] is defined as

$$M_B(D) := \inf \{M_{\text{ADM}}(\Sigma), \text{ such that } \Sigma \in \mathcal{P}(D)\}. \tag{64}$$

The *no-horizon* condition is needed to avoid extensions (Σ, γ_{ij}) with arbitrarily small ADM mass. There is also a spacetime version of Bartnik’s construction, not relying on an initial data set on D but only on the geometry of 2-surfaces S . Let us define $\mathcal{P}(S)$ as the set of globally hyperbolic spacetimes $(\mathcal{M}, g_{\mu\nu})$ satisfying the dominant energy condition, admitting an asymptotically Euclidean Cauchy hypersurface Σ with finite ADM mass, not presenting an event horizon and such that S is embedded (i.e., both its intrinsic and extrinsic geometry) in $(\mathcal{M}, g_{\mu\nu})$. Then, one defines

$$M_B(S) := \inf \{M_{\text{ADM}}(\mathcal{M}), \text{ such that } \mathcal{M} \in \mathcal{P}(S)\}. \tag{65}$$

The comparison between $M_B(D)$ and $M_B(\partial D)$ is not straightforward, due to issues regarding the horizon characterization. From the positivity of the ADM mass it follows the nonnegativity of the Bartnik mass $M_B(D)$. In fact, $M_B(D) = 0$ characterizes D as locally flat. From the definition (64) it also follows the monotonicity of $M_B(D)$, i.e. if $D_1 \subset D_2$ then $M_B(D_1) \leq M_B(D_2)$. Bartnik mass tends to the ADM mass, as domains D tend to Euclidean slices (the proof makes use of the Hawking energy introduced above). Another interesting feature, consequence of the proof of the Riemannian Penrose conjecture [62], is that Bartnik mass reduces to the *standard form* $E(S)$ in Eq. 51 for round spheres. However, the explicit calculation of the Bartnik mass is problematic. An approach to its practical computability is provided by Bartnik’s conjecture stating that the infimum in Eq. 64 is actually a minimum realized by an element in $\mathcal{P}(D)$ characterized by its stationarity outside D . Further developments of these ideas have been proposed by Bray (cf. [22]).

Before concluding this subsection, we mention the explicit construction in [93] of quantum analogs for some of the previous quasi-local gravitational energies (specifically for Brown–York, Liu–Yau, Hawking, and Geroch energies), as operators acting on the appropriate representation Hilbert space in a particular approach to quantum gravity (namely, loop quantum gravity).

4.3 Some Remarks on Quasi-Local Angular Momentum

Spinorial techniques provide a natural setting for the discussion of angular momentum. This does not only apply to angular momentum, since spinorial and also twistor techniques define a framework where further quasi-local mass notions can be introduced (e.g., Penrose mass), and known results can be reformulated in particularly powerful formulations (e.g., the discussion of positive mass theorems using the Nester–Witten form). However, in this article we will not discuss these approaches and we refer the reader to the relevant sections in Ref. [85]. We will focus on certain aspects of quasi-local expressions for angular momentum of Komar-like type. As it was shown in Section 2.2, choosing a two-sphere \mathcal{S} in a 3-slice adapted to the axial symmetry ϕ^μ , a 1-form L_μ can be found such that the Komar angular momentum is expressed as

$$J(\phi^\mu) = \frac{1}{8\pi} \oint_{\mathcal{S}} L_\nu \phi^\nu \sqrt{q} \, d^2x. \quad (66)$$

In particular, in the Komar expression (18) we have $L_\mu = q_\mu{}^\nu K_{\nu\rho} s^\rho$, whereas in the spatial infinity expression (45) this is modified by a term proportional to the trace K of the extrinsic curvature. The same applies for an angular momentum defined from the Brown–York charge (53) when plugging the expression for j_μ in Eq. 54 into Eq. 55, where ϕ^μ does not need to be a symmetry. The normal fundamental 1-forms Ω_μ on \mathcal{S} (cf. Section 1.3) provide another avenue to L_μ . In this section we assume the form in Eq. 66 for the angular momentum and comment on some approaches to the determination of the (*quasi-symmetry*) axial vector ϕ^μ .

Divergence-Free and Quasi-Killing Axial Vectors

No ambiguity for ϕ^μ is present when an axial symmetry exists on \mathcal{S} : ϕ^μ is taken as the corresponding Killing vector. In the absence of such a symmetry, we must address two issues. First, expression (66) depends on the space-like 3-slice in which \mathcal{S} is embedded. This follows from the modification of the 1-form L_μ by a total differential under a boost transformation: $L_\mu \rightarrow L_\mu + {}^2D_\mu f$ (cf. boost/normalization transformation of $\Omega_\mu^{(\ell)}$ in Section 1.3). Angular momentum can be associated with \mathcal{S} , independently of any hypersurface Σ , by demanding the axial vector to be divergence-free: ${}^2D_\mu \phi^\mu = 0$. Then, the boost-induced modification vanishes under integration. Second, the physical meaning of $J(\phi^\mu)$ is unclear if references

to a symmetry notion are completely dropped. In this sense, different approaches exist aiming at defining appropriate *quasi-Killing* notions. We simply mention here some recent works along these lines. In the context of isolated horizons (IH) (see next subsection) a prescription for the determination of a quasi-Killing axial vector on black hole horizons has been proposed in [35], though the divergence-free character is not guaranteed. Ref. [29] presents an approach for finding an approximate Killing vector by means of a minimization variational prescription that respects the divergence-free character of ϕ^μ . In the context of dynamical or trapping horizons [9, 10, 54], a unique divergence-free vector ϕ^μ can be chosen such that it is preserved by the unique slicing of the (space-like) horizon worldtube by marginally outer trapped surfaces [57]. Also in the context of dynamical horizons, a proposal for ϕ^μ has been made in [68] relying on a conformal decomposition of the metric $q_{\mu\nu}$ on \mathcal{S} . See Ref. [86] for a discussion of the divergence-free character of vector fields associated with quasi-local observables on \mathcal{S} .

Equation 66 only provides the expression for the component of the angular momentum vector that is associated with the vector ϕ^μ . If we are interested in determining the total angular momentum vector, a sensible prescription for the other two components is needed. This is an important practical issue in numerical simulations (see e.g., [24]).

4.4 A Study Case: *Quasi-Local Mass of Black Hole IHs*

The need of introducing some additional structure has been discussed above in different settings (e.g., symmetries for Komar quantities and asymptotic flatness for ADM and Bondi masses). We illustrate now this issue in a quasi-local context related to equilibrium black hole horizons.

4.4.1 A Brief Review of IHs

The IH framework introduced by Ashtekar and collaborators [10] provides a quasi-local setting for characterizing black hole horizons in quasi-equilibrium inside an otherwise dynamical spacetime. It presents a hierarchical structure with different quasi-equilibrium levels. The minimal notion of quasi-equilibrium is provided by the so-called non-expanding horizons (NEH). Given a Lorentzian manifold, a NEH is a hypersurface \mathcal{H} such that:

- (i) \mathcal{H} is a null hypersurface of topology $S^2 \times \mathbb{R}$ that is sliced by marginally (outer) trapped surfaces, that is, the expansion of the null congruence associated with the null generator ℓ^μ vanishes on \mathcal{H} : $\theta^{(\ell)} = q^{\mu\nu} \Theta_{\mu\nu}^{(\ell)} = 0$.
- (ii) Einstein equation is satisfied on \mathcal{H} .
- (iii) The vector $-T^\mu{}_\nu \ell^\nu$ is future directed.

The geometry of a NEH is characterized by the pair $(q_{\mu\nu}, \hat{\nabla}_\mu)$, where $q_{\mu\nu}$ is the induced null metric on \mathcal{H} and $\hat{\nabla}_\mu$ is the unique connection (not a Levi–Civita one) induced from the ambient spacetime connection. $\hat{\nabla}_\mu$ characterizes the *extrinsic geometry* of the NEH. A certain combination of components in $\hat{\nabla}_\mu$ can be put together to define an intrinsic object on \mathcal{H} , namely the 1-form ω_μ characterized by

$$\hat{\nabla}_\mu \ell^\nu = \omega_\mu \ell^\nu. \quad (67)$$

Defining a *surface gravity* as $\kappa_{(\ell)} := \ell^\mu \omega_\mu$, the acceleration expression for ℓ^μ is given by: $\hat{\nabla}_\ell \ell^\mu = \kappa_{(\ell)} \ell^\mu$. On the other hand, the projection of ω_μ on \mathcal{S} recovers the fundamental normal 1-form: $\Omega_\mu^{(\ell)} = q_\mu^\rho \omega_\rho$. The quasi-equilibrium hierarchy is introduced by demanding the invariance of the null hypersurface geometry under the ℓ^μ (evolution) flow in a progressive manner:

1. A NEH is characterized by the *time-invariance* of the intrinsic geometry $q_{\mu\nu}$: $\mathcal{L}_\ell^{\mathcal{H}} q_{\mu\nu} = 0$.
2. A weakly isolated horizon (WIH) is a NEH, together with an equivalence class of null normals $[\ell^\mu]$, for which the 1-form ω_μ is time-invariant: $\mathcal{L}_\ell^{\mathcal{H}} \omega_\mu = 0$. This is equivalent to the (time and angular) constancy of the surface gravity: $\hat{\nabla}_\mu \kappa_{(\ell)} = 0$.
3. An IH is a WIH on which the whole extrinsic geometry is time-invariant: $[\mathcal{L}_\ell^{\mathcal{H}}, \hat{\nabla}_\mu] = 0$.

The NEH and IH quasi-equilibrium levels represent genuine restrictions on the geometry of \mathcal{H} as a hypersurface in the ambient spacetime. On the contrary, a WIH structure can always be implemented on a NEH by an appropriate choice of the null normal ℓ^μ normalization. In this sense, a WIH does not represent a higher level of quasi-equilibrium than a NEH. However, from the point of view of the Hamiltonian analysis of spacetimes with a black hole in quasi-equilibrium as an inner boundary, the WIH notion proves to be crucial for the correct definition of the phase space symplectic structure and, more concretely, for the sound formulation of the quasi-local mass and angular momentum of the horizon.

4.4.2 An Overview of the Hamiltonian Analysis of IHs

Conserved Quantities Under Horizon Symmetries

As in the presentation of ADM quantities in Section 3.3, mass and angular momentum of IHs are introduced as conserved quantities under appropriate symmetries (see [6, 7] and the outline in Appendix C of [44] for further details on the following discussion). One starts from a symmetry of the horizon structure in the Lorentzian spacetime manifold and then constructs an associated canonical transformation in the phase or solution space of the system. The conserved quantity under this canonical transformation provides the relevant physical quantity. In view of the

variational problem (see below), a WIH is the relevant horizon structure to be considered in this context. A vector field W^μ preserves the WIH structure (W^μ is a WIH-symmetry) if

$$\mathcal{L}_W^\mathcal{H} \ell^\mu = \text{const} \cdot \ell^\mu, \quad \mathcal{L}_W^\mathcal{H} q_{\mu\nu} = 0, \quad \mathcal{L}_W^\mathcal{H} \omega_\mu = 0. \quad (68)$$

WIH-symmetries are of the form $W^\mu = c_W \ell^\mu + b_W S^\mu$, where c_W and b_W are constants on \mathcal{H} and S^μ is a Killing vector of any spatial section \mathcal{S} of \mathcal{H} .

Variational Problem for Spacetimes Containing WIHs

In order to set up the Hamiltonian treatment, we need first to define a well-posed variational problem. Here we are interested in the variational problem for asymptotically flat spacetimes containing a WIH. We will furthermore demand this WIH to contain an axial Killing vector ϕ^μ . The variational problem is then set in the region contained between two asymptotically Euclidean slices Σ_- and Σ_+ , spatial infinity i^0 , and the part of \mathcal{H} between an initial horizon slice $S_- = \mathcal{H} \cap \Sigma_-$ and a final one $S_+ = \mathcal{H} \cap \Sigma_+$. The action, as in Eq. 21, can be written [6, 7] as the sum of a bulk and a boundary term at spatial infinity, and the variation of the dynamical fields is set to vanish on the slices Σ_- and Σ_+ . No boundary term associated with the inner boundary \mathcal{H} is introduced. The variational problem is well-posed, in particular the Einstein equation is recovered, as long as the condition

$$\int_{\mathcal{H}} \delta\omega \wedge {}^2\epsilon = 0, \quad (69)$$

holds, where ω_μ has been introduced in Eq. 67 and ${}^2\epsilon_{\mu\nu}$ is the volume 2-form on sections \mathcal{S} of \mathcal{H} . The crucial ingredient in the well-posedness of the problem is precisely the WIH structure. This is the additional structure needed in order to guarantee the vanishing of Eq. 69, so that the variational problem is correctly posed and quasi-local quantities can be defined.

Phase Space, Canonical Transformations, and Physical Quantities

The phase space is defined by the couple (Γ, \mathbf{J}) where Γ is an infinite-dimensional manifold where each point represents a solution to the Einstein equation containing a WIH, and \mathbf{J} is a symplectic form (a closed 2-form) on Γ in terms of which the Poisson bracket is defined. In particular, a vector field X on Γ generates a canonical transformation if it leaves the symplectic form invariant: $\mathcal{L}_X^\Gamma \mathbf{J} = 0$. Using the closedness of \mathbf{J} this is locally equivalent to the exactness of the 1-form $i_X \mathbf{J}$, that is, to the (local) existence of a function H_X such that $i_X \mathbf{J} = \delta H_X$ (where δ denotes the differential in Γ). In particular, the quantity H_X defined on the phase space is preserved along the flow of X . In this context, first, the symplectic form can be obtained

from the action by using the *conserved symplectic current* method [30] and, second, a vector field X_W on Γ can be constructed from a WIH-symmetry W^μ on \mathcal{H} (cf. [6, 7] for details). For the correct definition of a physical parameter associated with a given WIH-symmetry W^μ , we must assess if the corresponding X_W preserves the canonical form, that is, if $i_{X_W}\mathbf{J}$ is locally exact. If this is the case, the conserved quantity is simply read from the associated explicit expression of the Hamiltonian H_{X_W} . When this scheme is applied to the axial symmetry ϕ^μ on \mathcal{H} , the corresponding X_ϕ turns out to be automatically an infinitesimal canonical transformation and the conserved quantity has the form

$$J_{\mathcal{H}} := X_\phi = \frac{1}{8\pi} \oint_{S_t} \omega_\mu \phi^\mu \sqrt{q} d^2x = \frac{1}{8\pi} \oint_{S_t} \Omega_\mu^{(\ell)} \phi^\mu \sqrt{q} d^2x, \quad (70)$$

where S_t is any spatial section of \mathcal{H} . This prescription for $J_{\mathcal{H}}$ exactly coincides with the Komar expression (18). The mass discussion is more subtle. In this case the WIH-symmetry t^μ associated with *time evolution* is chosen as an appropriate linear combination of the null normal ℓ^μ and the axial vector ϕ^μ . It is then found

$$i_{X_t}\mathbf{J} = \delta E_{\text{ADM}} - \left(\frac{\kappa_{(t)}}{8\pi} \delta A_{\mathcal{H}} + \Omega_{(t)} \delta J_{\mathcal{H}} \right), \quad (71)$$

where $\kappa_{(t)}$ and $\Omega_{(t)}$ are functions on Γ and $A_{\mathcal{H}}$ and $J_{\mathcal{H}}$ correspond, respectively, to the area of any section of \mathcal{H} and to the horizon angular momentum in Eq. 70. The right-hand-side expression is (locally) exact if functions $\kappa_{(t)}$ and $\Omega_{(t)}$ depend only on $A_{\mathcal{H}}$ and $J_{\mathcal{H}}$, and satisfy: $\frac{\partial \kappa_{(t)}}{\partial J_{\mathcal{H}}} = 8\pi \frac{\partial \Omega_{(t)}}{\partial A_{\mathcal{H}}}$. A function $E_{\mathcal{H}}^t$ only depending on $A_{\mathcal{H}}$ and $J_{\mathcal{H}}$ then exists, such that we can write

$$\delta E_{\mathcal{H}}^t = \frac{\kappa_{(t)}(A_{\mathcal{H}}, J_{\mathcal{H}})}{8\pi} \delta A_{\mathcal{H}} + \Omega_{(t)}(A_{\mathcal{H}}, J_{\mathcal{H}}) \delta J_{\mathcal{H}}. \quad (72)$$

To finally determine the quasi-local mass $M_{\mathcal{H}}$, the functional form of $E_{\mathcal{H}}^t(A_{\mathcal{H}}, J_{\mathcal{H}})$ is normalized to the one in the stationary Kerr family in Γ . Note that this is only justified once $E_{\mathcal{H}}^t$ has been shown to depend *only* on $A_{\mathcal{H}}$ and $J_{\mathcal{H}}$, a nontrivial result. In sum, for IHs $M_{\mathcal{H}}(A_{\mathcal{H}}, J_{\mathcal{H}}) := M_{\text{Kerr}}(A_{\mathcal{H}}, J_{\mathcal{H}})$, given by the Christodoulou mass expression [26].

Remark 8 (Quasi-local first law of black hole dynamics). Expression (72) extends the first law of black hole dynamics (see Section 5.2) from the stationary setting to dynamical spacetimes where only the black hole horizon is in equilibrium.

5 Global and Quasi-Local Quantities in Black Hole Physics

As an application, we briefly comment on some relevant issues concerning mass and angular momentum in the particular case of black hole spacetimes.

5.1 Penrose Inequality: a Claim for an Improved Mass Positivity Result for Black Holes

In the context of the established gravitational collapse picture, Penrose [78] proposed an inequality providing an upper bound for the area of the spatial sections of black hole event horizons in terms of the square of the ADM mass. This conjecture followed from a heuristic chain of arguments including rigorous results (singularity and black hole uniqueness theorems), together with conjectures such as weak cosmic censorship and the stationarity of the final state of the evolution of a black hole spacetime. A local-in-time version of the Penrose inequality can be formulated in terms of data on a Euclidean slice. In this version Penrose conjecture states that given an asymptotically Euclidean slice Σ containing a black hole under the dominant energy condition, the following inequality should be satisfied

$$A_{\min} \leq 16\pi M_{\text{ADM}}^2, \quad (73)$$

where A_{\min} is the minimal area enclosing the apparent horizon. In addition, equality is only attained by a slice of Schwarzschild spacetime. Though this was originally proposed in an attempt to construct counter-examples to the weak cosmic censorship conjecture, growing evidence has accumulated supporting its generic validity. Beyond spherical symmetry [70], a formal proof only exists in the Riemannian case, $K_{\mu\nu} = 0$, where the original derivation [62] (see also [21]) makes use of some of the quasi-local expressions presented in Section 4 (cf. discussion about Geroch and Hawking energies that coincide in this time-symmetric case $K_{\mu\nu} = 0$). The intrinsic geometric relevance of the Penrose inequality is reflected in its alternative name as the *isoperimetric inequality for black holes* [42].

Penrose inequality can also be seen as strengthening the positive ADM mass theorem in Section 3.3, for the case of black hole spacetimes: the ADM mass is not only positive but must be larger than a certain positive-definite quantity. Though it is tempting to identify this positive quantity with some quasi-local mass associated with the black hole, for example, with its irreducible mass $A =: 16\pi M_{\text{irr}}^2$ related to the Hawking mass (60), a caveat follows from the fact that the relevant minimal surface of area A_{\min} does not necessarily coincide with the apparent horizon, as examples in [17] show. In any case, this geometric inequality represents a bridge between global and quasi-local properties in black hole spacetimes and has become one of the current geometric and physical/conceptual main challenges in General Relativity.

5.2 Black Hole (Thermo-)dynamics

A set of four laws was established in [12] for stationary black holes. These black hole laws are analogous in form to the standard thermodynamical laws. Though this analogy is compelling, the fundamental nature of such relation was only

acknowledged under the light of Hawking's discovery [52] of the (semiclassical) thermal emission of particles from the event horizon (Hawking radiation). Given a stationary black hole spacetime with stationary Killing vector t^μ , black hole rigidity theorems [53] imply the existence of a second Killing vector k^μ that coincides with the null generators ℓ^μ on the horizon. We can write $k^\mu = t^\mu + \Omega_H \phi^\mu$, where ϕ^μ is an axial Killing vector and Ω_H is a constant referred to as the angular velocity of the horizon (see also [87]). We can write $k^\nu \nabla_\nu k^\mu = \kappa k^\mu$ on the horizon, which defines the surface gravity function κ . The zeroth law of black hole mechanics then states the constancy of the surface gravity on the event horizon. The second law, namely, Hawking's area theorem [50, 51], guarantees that the area of the event horizon never decreases, whereas the third law states that the surface gravity κ cannot be reduced to zero in a finite (advanced) time (see [63] for a precise statement). In the present context, we are particularly interested in the first law, since it relates the variations of some of the quasi-local and global quantities we have introduced in the text, in the particular black hole context. First law provides an expression for the change of the total mass M of the black hole (a well-defined notion since we deal with asymptotically flat spacetimes) under a small stationary and axisymmetric change in the solution space

$$\delta M = \frac{1}{8\pi} \kappa \delta A + \Omega_H J_H, \quad (74)$$

where A is the area of a spatial section of the horizon, and J_H is the Komar angular momentum associated with the axial Killing ϕ^μ . Equation 74 relates the variation of a global quantity $M = M_{\text{ADM}}$ at spatial infinity on the left-hand-side, to the variation of quantities locally defined at the horizon, on the right-hand-side. In particular, we could express the variation of the horizon area in terms of the variation of the irreducible local mass M_{irr} , as $\delta A = 32\pi M_{\text{irr}} \delta M_{\text{irr}}$. Such a formulation actually plays a role in the criterion for constructing sequences of binary black hole initial data corresponding to quasi-circular adiabatic inspirals (cf. [46] and the first law of binary black holes in [38]). Derivation of Eq. 74 involves the notions of ADM mass, as well as the generalization to stationarity of the Smarr formula for Kerr mass [stating $M = 2\Omega_H J_H + \kappa A/(4\pi)$] by using the Komar mass expression. Result (72) in Section 4.4 provides an *extension* of this law to black hole spacetimes non-necessarily stationary, but containing an IH for which an unambiguous notion of black hole mass can be introduced. Quasi-local attempts to extend the first law to the fully dynamical case have been explored in the dynamical and trapping horizon framework [9, 10, 19, 54]. However, the lack of a general unambiguous notion of quasi-local mass prevents the derivation of a *result* analogous to Eq. 74 or 72, that is, the equality between the variation of two *independent* well-defined quantities. In the quasi-local dynamical context, an unambiguous law for the area evolution can be determined (see, e.g., [20, 45] and references therein). The latter can then be used to *define* a flux of energy through the horizon by comparison with Eq. 74.

5.3 *Black Hole Extremality: a Mass–Angular Momentum Inequality*

Subextremal Kerr black holes are characterized by presenting angular momenta bounded by their total masses. Keeping axisymmetry, it has been recently shown [27, 28, 31] (see also [32]) that the inequality

$$|J_K| \leq M_{\text{ADM}}^2, \quad (75)$$

holds also for vacuum, maximal ($K = 0$), axisymmetric Euclidean data. Moreover, equality only holds for slices of extremal Kerr. Inequality (75) provides a nontrivial relation for black hole spacetimes between precisely the two physical quantities we are focusing on in this review. It is natural to explore if some analogous inequality holds when moving away from axisymmetry and when considering only the local region around the black hole. Attempts have been done in this sense, but they all must face the ambiguities resulting from the absence of canonical expressions for quasi-local masses and angular momenta. In order to illustrate the caveats to keep in mind when undertaking this kind of discussion, one can consider the case in which Komar quantities are used for constructing a truly quasi-local analogue of expression (75) for axisymmetric stationary data: initial data have been constructed [2] where the quotient $|J_K|/M_K^2$ on the black hole horizon can become arbitrarily large. Interestingly, these studies have led to the formulation [3] in axisymmetry of the related conjecture $8\pi|J_K| \leq A$, only involving intrinsic quantities on the horizon. This inequality has been proved to hold in the stationary axisymmetric case [59], as well as in a generalization including the electromagnetic field and the associated electric charge on the horizon [60].

6 Conclusions

The problem of characterizing the energy–momentum and angular momentum associated with the gravitational field in General Relativity has been present since the birth of the theory and controversies have plagued its already long-standing history. Understanding that no local density of energy–momentum can be identified for the gravitational field has challenged the validity of the mass and angular momentum *cherished* notions from nongravitational physics, when trying to perform a straightforward extension of these concepts to the gravitational field in a general relativistic setting.

The study of specific problems suggests concrete and/or partial solutions. In this spirit, at low velocities and weak self-gravities post-Newtonian approaches handle consistent notions of mass and angular momentum and the same holds in perturbative approaches around exact solutions, for which physical quantities can be identified unambiguously. In the same line, a (quasi-local) notion of the energy carried by a gravitational wave can be introduced as an *average* along the wavelength,

proving to provide a useful notion in practical applications. A particular setting of singular conceptual importance is that of isolated systems in General Relativity, specifically through their characterization as asymptotically flat spacetimes. The notions of *total* ADM and Bondi–Sachs energy–momentum provide well-defined quantities that, on the one hand, have clarified important conceptual issues such as the capability of gravitational waves to actually carry energy away from a system and, on the other hand, they also represent inestimable tools in practical applications due to their intrinsic/geometric character. Positivity theorems for the total mass represent without any doubt some of the most important and profound results in General Relativity. The combination of the success in isolated systems, together with the absence of a gravitational local energy–momentum density, has led to the consideration that the whole effort for the search of a *local* expression for the gravitational energy represents an ill-posed or pseudo-problem (see, e.g., [73]). But at the same time, and motivated by practical needs and/or fundamental physical reasons (cf. in this sense [48] for a related discussion on *quasi-local* issues regarding observables in Quantum Field Theory), important efforts have been devoted to the introduction of quasi-local notions of gravitational energy–momentum associated with extended but finite regions of the spacetime. In this respect, significant insights into the structure of the gravitational field have been achieved, with applications in diverse conceptual and practical contexts. But it must be acknowledged, as it is referred in [85], that the status of the quasi-local mass studies is in a kind of *post-modern* situation in which the devoted intensive efforts have resulted in a plethora of proposals with no obvious definitive and entirely satisfying candidate.

A moderate (intermediate) position that avoids radical skepticism against the quasi-local approach would consist in assuming that mass and (in a more restricted sense) angular momentum can be unambiguously defined only as global quantities for isolated systems. But accepting, at the same time, that quasi-local expressions provide meaningful and insightful quantities that are inextricably subject to the need of making systematically explicit the specific setting in which they are defined (one can make the analogy with the notion of *effective mass* in solid state physics, where different masses can be *simultaneously* employed for the same particle as long as their specific purposes are clearly stated¹). The moral of the whole discussion in this article is that the formulation of meaningful global or quasi-local mass and angular momentum notions in General Relativity *always* needs the introduction of some additional structure in the form of symmetries, quasi-symmetries, or some other background structure. This point must be explicitly kept in mind whenever employing the so-defined physical quantities, specially when extrapolating or performing compared analysis.

Acknowledgements The authors wish to thank the organizers of the Orléans School on Mass for their kind invitation and encouragement.

¹ We thank B. Carter for his many comments and insights in this discussion, and in particular for bringing us to the *solid state analogy* for gravitational masses.

References

1. L.F. Abbott, S. Deser, Nucl. Phys. B **195**, 76 (1982)
2. M. Ansorg, D. Petroff, Class. Q. Grav. **23**, L81 (2006)
3. M. Ansorg, H. Pfister, Class. Q. Grav. **25**, 035009 (2008)
4. R. Arnowitt, S. Deser, C.W. Misner, in *Gravitation: an Introduction to Current Research*, ed. by L. Witten (Wiley, New York, 1962), pp. 227–265
5. A. Ashtekar, in *General Relativity and Gravitation, One Hundred Years After the Birth of Albert Einstein*, ed. by A. Held, vol. 2 (Plenum, New York, 1980), p. 37
6. A. Ashtekar, C. Beetle, J. Lewandowski, Phys. Rev. D **64**, 044016 (2001)
7. A. Ashtekar, S. Fairhurst, B. Krishnan, Phys. Rev. D **62**, 104025 (2000)
8. A. Ashtekar, R.O. Hansen, J. Math. Phys. **19**, 1542 (1978)
9. A. Ashtekar, B. Krishnan, Phys. Rev. Lett. **89**, 261101 (2002)
10. A. Ashtekar, B. Krishnan, Living Rev. Rel. **7**, URL (cited on 8 May 2009): <http://www.livingreviews.org/lrr-2004-10>
11. A. Ashtekar, A. Magnon-Ashtekar, J. Math. Phys. **20**, 793 (1979)
12. J.M. Bardeen, B. Carter, S.W. Hawking, Commun. Math. Phys. **31**, 161 (1973)
13. R. Bartnik, Phys. Rev. Lett. **62**, 2346 (1989)
14. R. Beig, Phys. Lett. A **69**, 153 (1978)
15. J. Belinfante, Physica **6**, 887 (1939)
16. J. Belinfante, Physica **7**, 449 (1940)
17. I. Ben-Dov, Phys. Rev. D **70**, 124031 (2004)
18. H. Bondi, M.G.J. van der Burg, A.W.K. Metzner, Proc. R. Soc. Lond. A **269**, 21 (1962)
19. I. Booth, S. Fairhurst, Phys. Rev. Lett. **92**, 011102 (2004)
20. I. Booth, S. Fairhurst, Phys. Rev. D **75**, 084019 (2007)
21. H. Bray, J. Diff. Geom. **59**, 177 (2001)
22. H.L. Bray, P.T. Chruściel, in *The Einstein Equations and the Large Scale Behavior of Gravitational Fields*, ed. by P.T. Chruściel, H. Friedrich (Birkhäuser, Basel, 2004), p. 39
23. J.D. Brown, J.W. York, Phys. Rev. D **47**, 1407 (1993)
24. M. Campanelli, C.O. Lousto, Y. Zlochower, B. Krishnan, D. Merritt, Phys. Rev. D **75**, 064030 (2007)
25. C.C. Chang, J.M. Nester, C.M. Chen, Phys. Rev. Lett. **83**, 1897 (1999)
26. D. Christodoulou, Phys. Rev. Lett. **25**, 1596 (1970)
27. P.T. Chruściel, Ann. Phys. (N.Y.) **323**, 2566 (2008)
28. P.T. Chruściel, Y. Li, G. Weinstein, Ann. Phys. (N.Y.) **323**, 2591 (2008)
29. G.B. Cook, B.F. Whiting, Phys. Rev. D **76**, 041501 (2007)
30. C. Crnković, E. Witten, in *Three Hundred Years of Gravitation*, ed. by S. Hawking, W. Israel (Cambridge University Press, Cambridge, 1987), p. 676
31. S. Dain, Phys. Rev. Lett. **96**, 101101 (2006)
32. S. Dain, O. Ortiz, Phys. Rev. D **80**, 024045 (2009)
33. S. Deser, C. Teitelboim, Phys. Rev. Lett. **39**, 249 (1977)
34. S. Deser, B. Tekin, Phys. Rev. D **67**, 084009 (2003)
35. O. Dreyer, B. Krishnan, D. Shoemaker, E. Schnetter, Phys. Rev. D **67**, 024018 (2003)
36. R.J. Epp, Phys. Rev. D **62**, 124018 (2000)
37. J. Frauendiener, Living Rev. Rel. **7**, URL (cited on 8 May 2009): <http://www.livingreviews.org/lrr-2004-1>
38. J.L. Friedman, K. Uryu, M. Shibata, Phys. Rev. D **65**, 064035 (2002)
39. R.P. Geroch, J. Math. Phys. **13**, 956 (1972)
40. R. Geroch, Ann. N.Y. Acad. Sci. **224**, 108 (1973)
41. R. Geroch, J. Winicour, J. Math. Phys. **32**, 803 (1981)
42. G. Gibbons, in *Global Riemannian Geometry*, ed. by T. Willmore, N. Hitchin (Halsted Press, Chichester, 1984) pp. 194–202
43. E.ourgoulhon, arXiv: gr-qc/0703035v1 (2007)
44. E.ourgoulhon, J.L. Jaramillo, Phys. Rep. **423**, 159 (2006)

45. E.ourgoulhon, J.L. Jaramillo, Phys. Rev. D **74**, 087502 (2006)
46. P. Grandclement, E.ourgoulhon, S. Bonazzola, Phys. Rev. D **65**, 044021 (2002)
47. W. Greiner, J. Reinhardt, *Field Quantization* (Springer, Berlin, 1996)
48. R. Haag, *Local Quantum Physics, Fields, Particles, Algebras*, Text and Monographs in Physics (Springer-Verlag, Berlin, 1992)
49. S. Hawking, J. Math. Phys. **9**, 598 (1968)
50. S.W. Hawking, Phys. Rev. Lett. **26**, 1344 (1971)
51. S.W. Hawking, Commun. Math. Phys. **25**, 152 (1972)
52. S.W. Hawking, Commun. Math. Phys. **43**, 199 (1975)
53. S.W. Hawking, G.F.R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge University Press, London, 1973)
54. S. Hayward, Phys. Rev. D **49**, 6467 (1994)
55. S.A. Hayward, Phys. Rev. D **49**, 831 (1994)
56. S.A. Hayward, Phys. Rev. D **68**, 104015 (2003)
57. S.A. Hayward, Phys. Rev. D **74**, 104013 (2006)
58. M. Henneaux, C. Teitelboim, in *Quantization of Gauge Systems* (Princeton University, Princeton, 1992), p. 520
59. J. Hennig, M. Ansorg, C. Cederbaum, Class. Q. Grav. **25**, 162002 (2008)
60. J. Hennig, C. Cederbaum, M. Ansorg, Commun. Math. Phys. **293**, 449 (2010)
61. G.T. Horowitz, M.J. Perry, Phys. Rev. Lett. **48**, 371 (1982)
62. G. Huisken, T. Ilmanen, J. Diff. Geom. **59**, 353 (2001)
63. W. Israel, Phys. Rev. Lett. **57**, 397 (1986)
64. J.L. Jaramillo, J.A. Valiente Kroon, E.ourgoulhon, Class. Q. Grav. **25**, 093001 (2008)
65. J. Katz, Class. Q. Grav. **2**, 423 (1985)
66. J. Kijowski, Gen. Rel. Grav. **29**, 307 (1997)
67. A. Komar, Phys. Rev. **113**, 934 (1959)
68. M. Korzynski, Class. Q. Grav. **24**, 5935 (2007)
69. C.C.M. Liu, S.T. Yau, Phys. Rev. Lett. **90**, 231002 (2003)
70. E. Malec, N. O'Murchadha, Phys. Rev. D **49**, 6931 (1994)
71. C.W. Misner, Phys. Rev. **130**, 1590 (1963)
72. C.W. Misner, D.H. Sharp, Phys. Rev. **136**, B571 (1964)
73. C.W. Misner, K.S. Thorne, J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973)
74. N.Ó. Murchadha, R.-S. Tung, N. Xie, arXiv:0905.0647v2 [gr-qc] (2010)
75. E. Noether, Nachr. Ges. Wiss. Göttingen **2**, 235 (1918)
76. R. Penrose, Phys. Rev. Lett. **10**, 66 (1963)
77. R. Penrose, Proc. R. Soc. Lond. A **284**, 159 (1965)
78. R. Penrose, Annals N. Y. Acad. Sci. **224**, 125 (1973)
79. E. Poisson, *A Relativist's Toolkit, The Mathematics of Black-Hole Mechanics* (Cambridge University Press, Cambridge, 2004)
80. R.K. Sachs, Proc. R. Soc. Lond. A **270**, 103 (1962)
81. R. Schoen, S.T. Yau, Comm. Math. Phys. **65**, 45 (1979)
82. R. Schoen, S.T. Yau, Comm. Math. Phys. **79**, 231 (1981)
83. R. Schoen, S.T. Yau, Phys. Rev. Lett. **48**, 369 (1982)
84. J.M.M. Senovilla, Class. Q. Grav. **17**, 2799 (2000)
85. L.B. Szabados, Living Rev. Rel. **7**, URL (cited on 8 May 2009): <http://www.livingreviews.org/lrr-2004-4>
86. L.B. Szabados, Class. Q. Grav. **23**, 2291 (2006)
87. R.M. Wald, *General Relativity* (Chicago University Press, Chicago, 1984)
88. R.M. Wald, Living Rev. Rel. **6**, URL (cited on 8 May 2009): <http://www.livingreviews.org/lrr-2001-6>
89. R.M. Wald, A. Zoupas, Phys. Rev. D **61**, 084027 (2000)
90. M.T. Wang, S.T. Yau, Phys. Rev. Lett. **102**, 021101 (2009)
91. J. Winicour, L. Tamburino, Phys. Rev. Lett. **15**, 601 (1965)
92. E. Witten, Comm. Math. Phys. **80**, 381 (1981)
93. J. Yang, Y. Ma, Phys. Rev. D **80**, 084027 (2009)
94. J.W. York Jr, in *Sources of Gravitational Radiation*, ed. by L.L. Smarr (Cambridge University Press, Cambridge, 1979), pp. 83–126

Post-Newtonian Theory and the Two-Body Problem

Luc Blanchet

Abstract Reliable predictions of general relativity theory are extracted using approximation methods. Among these, the powerful post-Newtonian approximation provides us with our best insights into the problems of motion and gravitational radiation of systems of compact objects. This approximation has reached an impressive mature status, because of important progress regarding its theoretical foundations, and the successful construction of templates of gravitational waves emitted by inspiralling compact binaries. The post-Newtonian predictions are routinely used for searching and analyzing the very weak signals of gravitational waves in current generations of detectors. High-accuracy comparisons with the results of numerical simulations for the merger and ring-down of binary black holes are going on. In this article we give an overview on the general formulation of the post-Newtonian approximation and present up-to-date results for the templates of compact binary inspiral.

1 Introduction

Although relativists admire the mathematical coherence – and therefore beauty – of Einstein’s general relativity, this theory is not easy to manage when drawing firm predictions for the outcome of laboratory experiments and astronomical observations. Indeed only few exact solutions of the Einstein field equations are known, and one is obliged in most cases to resort to approximation methods. It is beyond question that approximation methods in general relativity do work, and in some cases with some incredible precision. Many of the great successes of general relativity were in fact obtained using approximation methods. However, because of the complexity of the field equations, such methods become awfully intricate at high approximation orders. On the other hand it is difficult to set up a formalism in which the approximation method would be perfectly well defined and based on

L. Blanchet (✉)

IRéCé, Institut d’Astrophysique de Paris, C.N.R.S. & Université Pierre et Marie Curie,
98^{bis} boulevard Arago, 75014 Paris, France

e-mail: blanchet@iap.fr

clear premises. Sometimes it is impossible to relate the approximation method to the exact framework of the theory. In this case the only thing one can do is to rely on the approximation method as the only representation of real phenomena, and to discover “empirically” that the approximation works well.

The most important approximation scheme in general relativity is the post-Newtonian expansion, which can be viewed as an expansion when the speed of light c tends to infinity, and is physically valid under the assumptions of weak gravitational field inside the source and of slow internal motion. The post-Newtonian approximation makes sense only in the *near zone* of an isolated matter source, defined as $r \ll \lambda$, where $\lambda \equiv c T$ is the wavelength of the emitted gravitational radiation, with T being a characteristic time scale of variation of the source. The approximation has been formalized in the early days of general relativity by Einstein [51], de Sitter [47, 48], and Lorentz and Droste [71]. It was subsequently developed notably by Einstein, Infeld and Hoffmann [52], Fock [55, 56], Plebanski and Bazanski [80], Chandrasekhar and collaborators [32–34], Ehlers and his school [50, 61, 62], and Papapetrou and coworkers [74, 76].

Several long-standing problems with the post-Newtonian approximation for general isolated slowly-moving systems have hindered progress until recently. At high post-Newtonian orders some *divergent* Poisson-type integrals appear, casting a doubt on the physical soundness of the approximation. Linked to that, the domain of validity of the post-Newtonian approximation is limited to the near zone of the source, making it a priori difficult to incorporate into the scheme the condition of no-incoming radiation, to be imposed at past null infinity from an isolated source. In addition, from a mathematical point of view, we do not know what the “reliability” of the post-Newtonian series is, that is, if it comes from the Taylor expansion of a family of exact solutions.

The post-Newtonian approximation gives wonderful answers to the problems of motion and gravitational radiation, two of general relativity’s corner stones. Three crucial applications are:

1. The motion of N point-like objects at the first post-Newtonian level [52], that is, 1PN,¹ is taken into account to describe the Solar System dynamics (motion of the centers of mass of planets).
2. The gravitational radiation-reaction force, which appears in the equations of motion at the 2.5PN order [39–42], has been experimentally verified by the observation of the secular acceleration of the orbital motion of the Hulse–Taylor binary pulsar PSR 1913+16 [90–92].
3. The analysis of gravitational waves emitted by inspiralling compact binaries – two neutron stars or black holes driven into coalescence by emission of gravitational radiation – necessitate the prior knowledge of the equations of motion and radiation field up to high post-Newtonian order.

¹ As usual, we refer to n PN as the order equivalent to terms $\sim (v/c)^{2n}$ in the equations of motion beyond the Newtonian acceleration, and in the asymptotic waveform beyond the Einstein quadrupole formula, where v denotes the binary’s orbital velocity and c is the speed of light.

Strategies to detect and analyze the very weak signals from inspiralling compact binaries involve matched filtering of a set of accurate theoretical template waveforms against the output of the detectors. Measurement-accuracy analyses have shown that in order to get sufficiently accurate theoretical templates, one must at least include conservative post-Newtonian effects up to the 3PN level, and radiation-reaction effects up to 5.5PN order, that is 3PN beyond the leading order of radiation reaction, which is 2.5PN [36–38, 54, 81, 86].

The appropriate description of inspiralling compact binaries is by two structureless *point-particles*, characterized solely by their masses m_1 or m_2 (and possibly their spins). Indeed, most of the nongravitational effects usually plaguing the dynamics of binary star systems, such as the effects of a magnetic field, of an interstellar medium, the influence of the internal structure of the compact bodies, are dominated by purely gravitational effects. Inspiralling compact binaries are very clean systems which can essentially be described in pure general relativity.

Although point-particles are ill defined in the exact theory, they are admissible in post-Newtonian approximations. Furthermore the model of point particles can be pushed to high post-Newtonian order, where an a priori more realistic model involving the internal structure of compact bodies would fail through becoming intractable. However there is an important worry: a process of regularization to deal with the infinite self-field of point particles is crucially needed. The regularization should be carefully defined to be implemented at high orders. It should hopefully be followed by a renormalization.

The orbit of inspiralling compact binaries can be considered to be circular, apart from the gradual inspiral, with an excellent approximation. Indeed, gravitational radiation-reaction forces tend to circularize rapidly the orbital motion. At each instant during the gradual inspiral, the eccentricity e of the orbit is related to the instantaneous frequency $\omega \equiv 2\pi/P$ by [79]

$$e^2 \simeq \text{const } \omega^{-19/9} \quad (\text{for } e \ll 1). \quad (1)$$

For instance one can check that the eccentricity of a system like the binary pulsar PSR 1913+16 will be $e \simeq 5 \cdot 10^{-6}$ when the gravitational waves become visible by the detectors, that is, when the signal frequency after its long chirp reaches $f \equiv \omega/\pi \simeq 10$ Hz. Only systems formed very late, near their final coalescence, could have a non-negligible residual eccentricity.

The intrinsic rotations or spins of the compact bodies could play an important role, yielding some relativistic spin-orbit and spin-spin couplings, both in the binary's equations of motion and radiation field. The spin of a rotating body is of the order of $S \sim m a v_{\text{spin}}$, where m and a denote the mass and typical size of the body, and where v_{spin} represents the velocity of the body's surface. In the case of compact bodies we have $a \sim G m/c^2$, and for maximal rotation $v_{\text{spin}} \sim c$; for such objects the magnitude of the spin is roughly $S \sim G m^2/c$. It is thus customary to introduce a dimensionless spin parameter, generally denoted by χ , defined by

$$S = \frac{G m^2}{c} \chi. \quad (2)$$

We have $\chi \leq 1$ for black holes, and $\chi \lesssim 0.63 - 0.74$ for neutron stars (depending on the equation of state of nuclear matter inside the neutron star). For binary pulsars such as PSR 1913 + 16, we have $\chi \lesssim 5 \cdot 10^{-3}$. Considering models of evolution of observed binary pulsar systems when they become close to the coalescence we expect that the spins will make a negligible contribution to the accumulated phase in this case. However, astrophysical observations suggest that black holes can have non-negligible spins, due to spin up driven by accretion from a companion during some earlier phase of the binary evolution. For a few black holes surrounded by matter, observations indicate a significant intrinsic angular momentum and the spin may even be close to its maximal value. However, very little is known about the black-hole spin magnitudes in binary systems. For black holes rotating near the maximal value the templates of gravitational waves need to take into account the effects of spins, both for a successful detection and an accurate parameter estimation.

We devote Section 2 of this article, including Sections 2.1–2.6, to a general overview of the formulation of the post-Newtonian approximation for isolated sources. We emphasize that the approximation can be carried out up to any post-Newtonian order, without the aforesaid problem of divergences. The main technique used is the matching of asymptotic expansions, which permits to obtain a complete post-Newtonian solution incorporating the correct boundary conditions at infinity. Then Section 3, that is, Sections 3.1–3.6, will deal with the application to systems composed of compact objects. The subtle issues linked with the self-field regularization of point-particles are discussed in some details. The results for the two-body equations of motion and radiation field at the state-of-the-art 3PN level are presented in the case of circular orbits appropriate to inspiralling compact binaries. We put the accent on the description of spinning particles and particularly on the spin-orbit coupling effect on the binary's internal energy and gravitational-wave flux.

2 Post-Newtonian Formalism

2.1 Einstein Field Equations

General relativity is based on two independent tenets, the first one concerned with the dynamics of the gravitational field, the second one dealing with the coupling of all the matter fields with the gravitational field. Accordingly the action of general relativity is made of two terms,

$$S = \frac{c^3}{16\pi G} \int d^4x \sqrt{-g} R + S_{\text{matter}}[\Psi, g_{\alpha\beta}]. \quad (3)$$

The first term represents the kinetic Einstein–Hilbert action for gravity and tells that the gravitational field $g_{\alpha\beta}$ propagates like a pure spin-2 field. Here R is the Ricci scalar and $g = \det(g_{\alpha\beta})$ is the determinant of the metric. The second term expresses the fact that all matter fields (collectively denoted by Ψ) are minimally coupled to

the metric $g_{\alpha\beta}$, which defines the physical lengths and times as measured in local laboratory experiments. The field equations are obtained by varying the action with respect to the metric (such that $\delta g_{\alpha\beta} = 0$ when $|x^\mu| \rightarrow \infty$) and form a system of ten differential equations of second order,

$$G^{\alpha\beta}[g, \partial g, \partial^2 g] = \frac{8\pi G}{c^4} T^{\alpha\beta}[g]. \quad (4)$$

The Einstein tensor $G^{\alpha\beta} \equiv R^{\alpha\beta} - \frac{1}{2}R g^{\alpha\beta}$ is generated by the matter stress–energy tensor $T^{\alpha\beta} \equiv (2/\sqrt{-g})(\delta S_{\text{matter}}/\delta g_{\alpha\beta})$. Four equations give, via the contracted Bianchi identity, the conservation equation for the matter system as

$$\nabla_\mu G^{\alpha\mu} \equiv 0 \quad \implies \quad \nabla_\mu T^{\alpha\mu} = 0, \quad (5)$$

which must be solved conjointly with the Einstein field equations for the gravitational field. The matter Eq. 5 reads also

$$\partial_\mu (\sqrt{-g} T^\mu_\alpha) = \frac{1}{2} \sqrt{-g} T^{\mu\nu} \partial_\alpha g_{\mu\nu}. \quad (6)$$

Let us introduce an asymptotically Minkowskian coordinate system such that the gravitational-wave amplitude, defined by $h^{\alpha\beta} \equiv \sqrt{-g} g^{\alpha\beta} - \eta^{\alpha\beta}$,² is divergenceless, that is, satisfies the de Donder or harmonic gauge condition

$$\partial_\mu h^{\alpha\mu} = 0. \quad (7)$$

With this coordinate choice the Einstein field equations can be recast into the d’Alembertian equation

$$\square h^{\alpha\beta} = \frac{16\pi G}{c^4} \tau^{\alpha\beta}. \quad (8)$$

Here $\square = \eta^{\mu\nu} \partial_\mu \partial_\nu$ is the usual (*flat*-spacetime) d’Alembertian operator. The source term $\tau^{\alpha\beta}$ can rightly be interpreted as the “effective” stress–energy distribution of the matter and gravitational fields in harmonic coordinates; note that $\tau^{\alpha\beta}$ is not a tensor and we shall often call it a pseudo-tensor. It is conserved in the sense that

$$\partial_\mu \tau^{\alpha\mu} = 0. \quad (9)$$

This is equivalent to the condition of harmonic coordinates (7) and to the covariant conservation (5) of the matter tensor. The pseudo-tensor is made of the contribution of matter fields described by $T^{\alpha\beta}$, and of the gravitational contribution $\Lambda^{\alpha\beta}$ that is a

² Here, $g^{\alpha\beta}$ denotes the contravariant metric, inverse of the covariant metric $g_{\alpha\beta}$, and $\eta^{\alpha\beta}$ represents an auxiliary Minkowski metric. We assume that our spatial coordinates are Cartesian so that $(\eta^{\alpha\beta}) = (\eta_{\alpha\beta}) = \text{diag}(-1, 1, 1, 1)$.

complicated function of the gravitational field variable $h^{\mu\nu}$ and its first and second derivatives. Thus,

$$\tau^{\alpha\beta} = |g|T^{\alpha\beta} + \frac{c^4}{16\pi G} \Lambda^{\alpha\beta}[h, \partial h, \partial^2 h]. \quad (10)$$

The point is that $\Lambda^{\alpha\beta}$ is at least quadratic in the field strength $h^{\mu\nu}$, so the field equations (8) are naturally amenable to a perturbative nonlinear treatment. The general expression is

$$\begin{aligned} \Lambda^{\alpha\beta} = & -h^{\mu\nu} \partial_\mu \partial_\nu h^{\alpha\beta} + \partial_\mu h^{\alpha\nu} \partial_\nu h^{\beta\mu} + \frac{1}{2} g^{\alpha\beta} g_{\mu\nu} \partial_\rho h^{\mu\sigma} \partial_\sigma h^{\nu\rho} \\ & - g^{\alpha\mu} g_{\nu\sigma} \partial_\rho h^{\beta\sigma} \partial_\mu h^{\nu\rho} - g^{\beta\mu} g_{\nu\sigma} \partial_\rho h^{\alpha\sigma} \partial_\mu h^{\nu\rho} + g_{\mu\nu} g^{\rho\sigma} \partial_\rho h^{\alpha\mu} \partial_\sigma h^{\beta\nu} \\ & + \frac{1}{4} (2g^{\alpha\mu} g^{\beta\nu} - g^{\alpha\beta} g^{\mu\nu}) \left(g_{\rho\sigma} g_{\epsilon\pi} - \frac{1}{2} g_{\sigma\epsilon} g_{\rho\pi} \right) \partial_\mu h^{\rho\pi} \partial_\nu h^{\sigma\epsilon}. \end{aligned} \quad (11)$$

To select a physically sensible solution of the field equations in the case of a bounded matter system, we must impose a boundary condition at infinity, namely, the famous no-incoming radiation condition, which ensures that the system is truly isolated from other bodies in the Universe. In principle the no-incoming radiation condition is to be formulated at past null infinity \mathcal{S}^- . Here, we can simplify the formulation by taking advantage of the presence of the Minkowski background $\eta_{\alpha\beta}$ to define the no-incoming radiation condition with respect to Minkowskian past null infinity say \mathcal{S}_η^- . Within approximate methods this is legitimate as we can view the gravitational field as propagating on the flat background $\eta_{\alpha\beta}$; indeed $\eta_{\alpha\beta}$ does exist at any finite order of approximation.

The no-incoming radiation condition should be such that it suppresses any homogeneous (regular in \mathbb{R}^4) solution of the d'Alembertian equation $\square h_{\text{hom}} = 0$ in a neighborhood of \mathcal{S}_η^- . We have at our disposal the Kirchhoff formula that expresses h_{hom} at some field point (\mathbf{x}', t') in terms of its values and its derivatives on a sphere centered on \mathbf{x}' with radius $\rho \equiv |\mathbf{x}' - \mathbf{x}|$ and at retarded time $t \equiv t' - \rho/c$,

$$h_{\text{hom}}(\mathbf{x}', t') = \int \frac{d\Omega}{4\pi} \left[\frac{\partial}{\partial \rho} (\rho h_{\text{hom}}) + \frac{1}{c} \frac{\partial}{\partial t} (\rho h_{\text{hom}}) \right] (\mathbf{x}, t), \quad (12)$$

where $d\Omega$ is the solid angle spanned by the unit direction $\mathbf{N} \equiv (\mathbf{x} - \mathbf{x}')/\rho$. From this formula we deduce the no-incoming radiation condition as the following limit at \mathcal{S}_η^- , that is, when $r \equiv |\mathbf{x}| \rightarrow +\infty$ with $t + r/c = \text{const.}$ ³

$$\lim_{\mathcal{S}_\eta^-} \left[\frac{\partial}{\partial r} (r h^{\alpha\beta}) + \frac{1}{c} \frac{\partial}{\partial t} (r h^{\alpha\beta}) \right] = 0. \quad (13)$$

³ In fact we obtain also the auxiliary condition that $r \partial_\mu h^{\alpha\beta}$ should be bounded near \mathcal{S}_η^- . This comes from the fact that ρ differs from r and we have $\rho = r - \mathbf{x}' \cdot \mathbf{n} + \mathcal{O}(1/r)$ with $\mathbf{n} = \mathbf{x}/r$.

Now if $h^{\alpha\beta}$ satisfies (13), so does the pseudo-tensor $\tau^{\alpha\beta}$ built on it, and then it is clear that the retarded integral of $\tau^{\alpha\beta}$ does satisfy the same condition. Thus we infer that the unique solution of the Einstein equation (8) reads

$$h^{\alpha\beta} = \frac{16\pi G}{c^4} \square_{\mathbb{R}}^{-1} \tau^{\alpha\beta}, \quad (14)$$

where the retarded integral takes the standard form

$$(\square_{\mathbb{R}}^{-1} \tau^{\alpha\beta})(\mathbf{x}, t) \equiv -\frac{1}{4\pi} \int \frac{d^3 \mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|} \tau^{\alpha\beta}(\mathbf{x}', t - |\mathbf{x} - \mathbf{x}'|/c). \quad (15)$$

Notice that since $\tau^{\alpha\beta}$ depends on $h^{\mu\nu}$ and its derivatives, Eq. 14 is to be viewed as an integro-differential equation equivalent to the Einstein equation (8) with no-incoming radiation condition.

2.2 Post-Newtonian Iteration in the Near Zone

In this section we proceed with the post-Newtonian iteration of the field equations in harmonic coordinates in the near zone of an isolated matter distribution. We have in mind a general hydrodynamical fluid, whose stress–energy tensor is smooth, that is, $T^{\alpha\beta} \in C^\infty(\mathbb{R}^4)$. Thus the scheme a priori excludes the presence of singularities; these will be dealt with in later sections.

Let us remind that the post-Newtonian approximation in “standard” form (e.g., [1, 61, 62]) is plagued with some apparently inherent difficulties, which crop up at some high post-Newtonian order like 3PN. Up to the 2.5PN order the approximation can be worked out without problems, and at the 3PN order the problems can generally be solved for each case at hand; but the problems worsen at higher orders. Historically these difficulties, even appearing at higher approximations, have cast a doubt on the actual soundness, from a theoretical point of view, of the post-Newtonian expansion. Practically speaking, they posed the question of the reliability of the approximation, when comparing the theory’s predictions with very precise experimental results. In this section and the next one we assess the nature of these difficulties – are they purely technical or linked with some fundamental drawback of the approximation scheme? – and eventually resolve them.

We first distinguish the problem of *divergences* in the post-Newtonian expansion: in higher approximations some divergent Poisson-type integrals appear. Recall that the post-Newtonian expansion replaces the resolution of an hyperbolic-like d’Alembertian equation by a perturbatively equivalent hierarchy of elliptic-like Poisson equations. Rapidly it is found during the post-Newtonian iteration that the right-hand-sides of the Poisson equations acquire a non-compact support (it is distributed all over space), and that the standard Poisson integral diverges because of the bound of the integral at spatial infinity, that is $r \equiv |\mathbf{x}| \rightarrow +\infty$, with $t = \text{const}$.

The divergencies are linked to the fact that the post-Newtonian expansion is actually a singular perturbation, in the sense that the coefficients of the successive powers of $1/c$ are not uniformly valid in space, typically blowing up at spatial infinity like some positive powers of r . We know for instance that the post-Newtonian expansion cannot be “asymptotically flat” starting at the 2PN or 3PN level, depending on the adopted coordinate system [83]. The result is that the standard Poisson integrals are in general badly behaving at infinity. Trying to solve the post-Newtonian equations by means of the Poisson integral does not a priori make sense. This does not mean that there are no solutions to the problem, but simply that the Poisson integral does not constitute the good solution of the Poisson equation in the context of post-Newtonian expansions. So the difficulty is purely of a technical nature, and will be solved once we succeed in finding the appropriate solution to the Poisson equation.

We shall now prove (following [82]) that the post-Newtonian expansion can be *indefinitely* iterated without divergencies.⁴ Let us denote by means of an overline the formal (infinite) post-Newtonian expansion of the field inside the source’s near zone, which is of the form

$$\bar{h}^{\alpha\beta}(\mathbf{x}, t, c) = \sum_{n=2}^{+\infty} \frac{1}{c^n} \bar{h}_n^{\alpha\beta}(\mathbf{x}, t, \ln c). \quad (16)$$

The n -th post-Newtonian coefficient is naturally the factor of the n -th power of $1/c$; however, we know [14] that the post-Newtonian expansion also involves some logarithms of c , included for convenience here into the definition of the coefficients \bar{h}_n . For the stress–energy pseudo-tensor (10) we have the same type of expansion,

$$\bar{\tau}^{\alpha\beta}(\mathbf{x}, t, c) = \sum_{n=-2}^{+\infty} \frac{1}{c^n} \bar{\tau}_n^{\alpha\beta}(\mathbf{x}, t, \ln c). \quad (17)$$

The expansion starts with a term of order c^2 corresponding to the rest mass–energy ($\tau^{\alpha\beta}$ has the dimension of an energy density). Here we shall always understand the infinite sums such as (16)–(17) in the sense of *formal series*, that is, merely as an ordered collection of coefficients. Because of our consideration of regular extended matter distributions the post-Newtonian coefficients are smooth functions of space-time.

Inserting the post-Newtonian ansatz into the Einstein field equation (8) and equating together the powers of $1/c$ results is an infinite set of Poisson-type equations ($\forall n \geq 2$),

$$\Delta \bar{h}_n^{\alpha\beta} = 16\pi G \bar{\tau}_{n-4}^{\alpha\beta} + \partial_{t_{n-2}}^2 \bar{h}_n^{\alpha\beta}. \quad (18)$$

⁴ An alternative solution to the problem of divergencies, proposed in [57, 58], is based on an initial-value formalism, which avoids the appearance of divergencies because of the finiteness of the integration region.

The second term comes from the split of the d'Alembertian into a Laplacian and a second time derivative: $\square = \Delta - \frac{1}{c^2} \partial_t^2$. This term is zero when $n = 2$ and 3. We proceed by induction, that is, fix some post-Newtonian order n , assume that we succeeded in constructing the sequence of previous coefficients ${}_p\bar{h}$ for $p \leq n - 1$, and from this show how to infer the next-order coefficient ${}_n\bar{h}$.

To cure the problem of divergencies we introduce a generalized solution of the Poisson equation with non-compact support source, in the form of an appropriate *finite part* of the usual Poisson integral obtained by regularization of the bound at infinity by means of a specific process of analytic continuation. For any source term like ${}_n\bar{\tau}$, we multiply it by the “regularization” factor

$$|\bar{\mathbf{x}}|^B \equiv \left| \frac{\mathbf{x}}{r_0} \right|^B, \tag{19}$$

where $B \in \mathbb{C}$ is a complex number and r_0 denotes an arbitrary length scale. Only then do we apply the Poisson integral, which therefore defines a certain function of B . The well definedness of that integral heavily relies on the behavior of the integrand at the bound at infinity. There is no problem with the vicinity of the origin inside the source because of the smoothness of the pseudo-tensor. Then one can prove [82] that the latter function of B generates a (unique) analytic continuation down to a neighborhood of the value of interest $B = 0$, except at $B = 0$ itself, at which value it admits a Laurent expansion with multiple poles up to some finite order. Then, we consider the Laurent expansion of that function when $B \rightarrow 0$ and pick up the finite part, or coefficient of the zero-th power of B , of that expansion. This defines our generalized Poisson integral:

$$\Delta^{-1} [{}_n\bar{\tau}^{\alpha\beta}](\mathbf{x}, t) \equiv -\frac{1}{4\pi} \text{FP}_{B=0} \int \frac{d^3\mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|} |\bar{\mathbf{x}}'|^B {}_n\bar{\tau}^{\alpha\beta}(\mathbf{x}', t). \tag{20}$$

The integral extends over all three-dimensional space but with the latter finite-part regularization at infinity denoted $\text{FP}_{B=0}$. The main properties of our generalized Poisson operator is that it does solve the Poisson equation, namely,

$$\Delta \left(\Delta^{-1} [{}_n\bar{\tau}^{\alpha\beta}] \right) = {}_n\bar{\tau}^{\alpha\beta}, \tag{21}$$

and that the so-defined solution $\Delta^{-1} {}_n\bar{\tau}$ owns the same properties as its source ${}_n\bar{\tau}$, that is, the smoothness and the same type of behavior at infinity.

The most general solution of the Poisson equation (18) will be obtained by application of the previous generalized Poisson operator to the right-hand-side of Eq. 18, and augmented by the most general *homogeneous* solution of the Poisson equation. Thus, we can write

$${}_n\bar{h}^{\alpha\beta} = 16\pi G \Delta^{-1} [{}_n\bar{\tau}^{\alpha\beta}] + \partial_t^2 \Delta^{-1} [{}_{n-2}\bar{h}^{\alpha\beta}] + \sum_{\ell=0}^{+\infty} B_n^{\alpha\beta}(\ell) \hat{x}_L. \tag{22}$$

The last term represents the general solution of the Laplace equation, which is regular at the origin $r \equiv |\mathbf{x}| = 0$. It can be written, using the symmetric-trace-free (STF) language, as a multipolar series of terms of the type \hat{x}_L ,⁵ and multiplied by some STF-tensorial functions of time ${}_n B_L(t)$. These functions will be associated with the radiation reaction of the field onto the source; they will depend on which boundary conditions are to be imposed on the gravitational field at infinity from the source.

It is now trivial to iterate the process. We substitute for ${}_{n-2}\bar{h}$ in the right-hand-side of Eq. 22 the same expression but with n replaced by $n-2$, and similarly come down until we stop at either one of the coefficients ${}_0\bar{h} = 0$ or ${}_1\bar{h} = 0$. At this point ${}_n\bar{h}$ is expressed in terms of the “previous” ${}_p\bar{\tau}$ ’s and ${}_p B_L$ ’s with $p \leq n-2$. To finalize the process we introduce what we call the operator of the “instantaneous” potentials \square_I^{-1} . Our notation is chosen to contrast with the standard operators of the retarded and advanced potentials \square_R^{-1} and \square_A^{-1} , see Eq. 15. However, beware of the fact that unlike $\square_{R,A}^{-1}$ the operator \square_I^{-1} will be defined only when acting on a post-Newtonian series such as $\bar{\tau}$. Indeed, we pose

$$\square_I^{-1}[\bar{\tau}^{\alpha\beta}] \equiv \sum_{k=0}^{+\infty} \left(\frac{\partial}{c\partial t} \right)^{2k} \Delta^{-k-1}[\bar{\tau}^{\alpha\beta}], \quad (23)$$

where Δ^{-k-1} is the k -th iteration of the operator (20). It is readily checked that in this way we have a solution of the source-free d’Alembertian equation,

$$\square \left(\square_I^{-1}[\bar{\tau}^{\alpha\beta}] \right) = \bar{\tau}^{\alpha\beta}. \quad (24)$$

On the other hand, the homogeneous solution in Eq. 22 will yield by iteration an homogeneous solution of the d’Alembertian equation, which is necessarily regular at the origin. Hence it should be of the *anti-symmetric* type, that is, be made of the difference between a retarded multipolar wave and the corresponding advanced wave. We shall therefore introduce a new definition for some STF-tensorial functions $A_L(t)$ parametrizing those advanced-minus-retarded free waves. It will not be difficult to relate the post-Newtonian expansion of $A_L(t)$ to the functions ${}_n B_L(t)$, which were introduced in Eq. 22. Finally the most general post-Newtonian solution, iterated ad infinitum and without any divergences, is obtained into the form

$$\bar{h}^{\alpha\beta} = \frac{16\pi G}{c^4} \square_I^{-1}[\bar{\tau}^{\alpha\beta}] - \frac{4G}{c^4} \sum_{\ell=0}^{+\infty} \frac{(-)^\ell}{\ell!} \hat{\partial}_L \left\{ \frac{A_L^{\alpha\beta}(t-r/c) - A_L^{\alpha\beta}(t+r/c)}{2r} \right\}. \quad (25)$$

⁵ Here $L = i_1 \cdots i_\ell$ denotes a multi-index composed of ℓ multipolar spatial indices i_1, \dots, i_ℓ (ranging from 1 to 3); $x_L \equiv x_{i_1} \cdots x_{i_\ell}$ is the product of ℓ spatial vectors x_i ; $\partial_L \equiv \partial_{i_1} \cdots \partial_{i_\ell}$ is the product of ℓ partial derivatives $\partial_i \equiv \partial/\partial x^i$; in the case of summed-up (dummy) multi-indices L , we do not write the ℓ summations from 1 to 3 over their indices; the STF projection is indicated with a hat, that is, $\hat{x}_L \equiv \text{STF}[x_L]$ and similarly $\hat{\partial}_L \equiv \text{STF}[\partial_L]$, or sometimes using brackets surrounding the indices, $x_{<L>} \equiv \hat{x}_L$.

We shall refer to the $A_L(t)$'s as the *radiation-reaction* functions. If we stay at the level of the post-Newtonian iteration, which is confined into the near zone, we cannot do more than Eq. 25; there is no means to compute the radiation-reaction functions $A_L(t)$. We are here touching the second problem faced by the standard post-Newtonian approximation.

2.3 Post-Newtonian Expansion Calculated by Matching

As we now understand this problem is that of the limitation to the near zone. Indeed the post-Newtonian expansion assumes that all retardations r/c are small, so it can be viewed as a formal *near-zone* expansion when $r \rightarrow 0$, valid only in the region surrounding the source that is of small extent with respect to the wavelength of the emitted radiation: $r \ll \lambda$. As we have seen, a consequence is that the post-Newtonian coefficients blow up at infinity, when $r \rightarrow +\infty$. It is thus not possible, a priori, to implement within the post-Newtonian scheme the physical information that the matter system is isolated from the rest of the Universe. The no-incoming radiation condition imposed at past null infinity \mathcal{S}_η^- cannot be taken into account, a priori, within the scheme.

The near-zone limitation can be circumvented to the lowest post-Newtonian orders by considering *retarded* integrals that are formally expanded when $c \rightarrow +\infty$ as series of “instantaneous” Poisson-like integrals [1]. This procedure works well up to the 2.5PN level and has been shown to correctly fix the dominant radiation-reaction term at the 2.5PN order [61, 62]. Unfortunately such a procedure assumes fundamentally that the gravitational field, after expansion of all retardations $r/c \rightarrow 0$, depends on the state of the source at a single time t , in keeping with the instantaneous character of the Newtonian interaction. However, we know that the post-Newtonian field (as well as the source's dynamics) will cease at some stage to be given by a functional of the source-parameters at a single time, because of the imprint of gravitational-wave tails in the near-zone field, in the form of some modification of the radiation reaction-force at the 1.5PN relative order [10, 15]. Since the reaction force is itself of order 2.5PN this means that the formal post-Newtonian expansion of retarded Green functions is no longer valid starting at the 4PN order.

The solution of the problem resides in the matching of the near-zone field to the exterior field, a solution of the vacuum equations outside the source, which has been developed in previous works using some post-*Minkowskian* and multipolar expansions [14, 17]. In the case of post-Newtonian sources, the near zone, that is $r \ll \lambda$, covers entirely the source, because the source's radius itself is such that $a \ll \lambda$. Thus the near-zone overlaps with the exterior zone where the multipole expansion is valid. Matching together the post-Newtonian and multipolar-post-Minkowskian solutions in this overlapping region is an application of the method of matched asymptotic expansions, which has frequently been applied in the present context, both for radiation-reaction [10, 15, 30, 31] and wave-generation [11, 16, 43] problems.

In the previous section we obtained the most general solution (25) for the post-Newtonian expansion, as parametrized by the set of unknown radiation-reaction functions $A_L(t)$. We shall now impose the *matching* condition

$$\mathcal{M}(\bar{h}^{\alpha\beta}) \equiv \overline{\mathcal{M}(h^{\alpha\beta})}, \quad (26)$$

telling that the *multipole* decomposition of the post-Newtonian expansion \bar{h} of the inner field, agrees with the *near-zone* expansion of the multipole expansion $\mathcal{M}(h)$ of the external field. Here the calligraphic letter \mathcal{M} stands for the multipole decomposition or far-zone expansion, while the overbar denotes the post-Newtonian or near-zone expansion. The matching equation results from the numerical equality $\bar{h} = \mathcal{M}(h)$, clearly verified in the exterior part of the near-zone, namely, our overlapping region $a < r \ll \lambda$. The left-hand-side is expanded when $r \rightarrow +\infty$ yielding $\mathcal{M}(\bar{h})$ while the right-hand-side is expanded when $r \rightarrow 0$ leading to $\overline{\mathcal{M}(h)}$. The matching equation is thus physically justified only for post-Newtonian sources, for which the exterior near zone exists. It is actually a *functional* identity; it identifies, *term-by-term*, two asymptotic expansions, each of them being formally taken outside its own domain of validity. In the present context, the matching equation insists that the infinite *far-zone* expansion ($r \rightarrow \infty$) of the inner post-Newtonian field is identical to the infinite *near-zone* expansion ($r \rightarrow 0$) of the exterior multipolar field. Let us now state that Eq. 26, plus the condition of no-incoming radiation, permits determining all the unknowns of the problem: that is, at once, the external multipolar decomposition $\mathcal{M}(h)$ and the radiation-reaction functions A_L and hence the inner post-Newtonian expansion \bar{h} .

When applied to a multipole expansion such as that of the pseudo-tensor, that is, $\mathcal{M}(\tau^{\alpha\beta})$, we have to define a special type of generalized inverse d'Alembertian operator, built on the standard retarded integral (15), viz

$$\square_{\mathbb{R}}^{-1}[\mathcal{M}(\tau^{\alpha\beta})](\mathbf{x}, t) \equiv -\frac{1}{4\pi} \text{FP}_{B=0} \int \frac{d^3\mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|} |\widetilde{\mathbf{x}}|^B \mathcal{M}(\tau^{\alpha\beta})(\mathbf{x}', t - |\mathbf{x} - \mathbf{x}'|/c). \quad (27)$$

Like Eq. 15 this integral extends over the whole three-dimensional space, but a regularization factor $|\widetilde{\mathbf{x}}|^B$ given by Eq. 19 has been “artificially” introduced for application of the finite part operation $\text{FP}_{B=0}$. The reason for introducing such regularization is to cure the divergencies of the integral when $|\mathbf{x}'| \rightarrow 0$; these are coming from the fact that the multipolar expansion is singular at the origin. We notice that this regularization factor is the same as the one entering the generalized Poisson integral (20), however, its role is different, as it takes care of the bound at $|\mathbf{x}'| = 0$ rather than at infinity. We easily find that this new object is a particular retarded solution of the wave equation

$$\square \left(\square_{\mathbb{R}}^{-1}[\mathcal{M}(\tau^{\alpha\beta})] \right) = \mathcal{M}(\tau^{\alpha\beta}). \quad (28)$$

Therefore $\mathcal{M}(h^{\alpha\beta})$ should be given by that solution plus a *retarded* homogeneous solution of the d'Alembertian equation (imposing the no-incoming radiation condition), that is, be of the type

$$\mathcal{M}(h^{\alpha\beta}) = \frac{16\pi G}{c^4} \square_{\mathbb{R}}^{-1}[\mathcal{M}(\tau^{\alpha\beta})] - \frac{4G}{c^4} \sum_{\ell=0}^{+\infty} \frac{(-)^\ell}{\ell!} \hat{\partial}_L \left\{ \frac{F_L^{\alpha\beta}(t - r/c)}{r} \right\}. \quad (29)$$

Now the matching equation (26) will determine both the A_L 's in Eq. 25 and the F_L 's in Eq. 29. We summarize the results that have been obtained in [82].

The functions $F_L(t)$ will play an important role in the following, because they appear as the multipole moments of a general post-Newtonian source as seen from its exterior near zone. Their closed-form expression obtained by matching reads

$$F_L^{\alpha\beta}(t) = \text{FP}_{B=0} \int d^3\mathbf{x} \tilde{|\mathbf{x}|}^B \hat{x}_L \int_{-1}^{+1} dz \delta_\ell(z) \bar{\tau}^{\alpha\beta}(\mathbf{x}, t - z|\mathbf{x}|/c). \quad (30)$$

Again the integral extends over all space but the bound at infinity (where the post-Newtonian expansion becomes singular) is regularized by means of the same finite part. The z -integration involves a weighting function $\delta_\ell(z)$ defined by

$$\delta_\ell(z) = \frac{(2\ell + 1)!!}{2^{\ell+1}\ell!} (1 - z^2)^\ell. \quad (31)$$

The integral of that function is normalized to one: $\int_{-1}^{+1} dz \delta_\ell(z) = 1$. Furthermore it approaches the Dirac function in the limit of large multipoles: $\lim_{\ell \rightarrow \infty} \delta_\ell(z) = \delta(z)$. The multipole moments (30) are physically valid only for post-Newtonian sources. As such, they must be considered only in a perturbative post-Newtonian sense. With the result (30) the multipole expansion (29) is fully determined and will be exploited in the next section.

Concerning the near-zone field (25) we find that the radiation-reaction functions A_L are composed of the multipole moments F_L that will also characterize "linear-order" radiation reaction effects starting at 2.5PN order, and of an extra contribution R_L , which will be due to nonlinear effects in the radiation reaction and turn out to arise at 4PN order. Thus,

$$A_L^{\alpha\beta} = F_L^{\alpha\beta} + R_L^{\alpha\beta}, \quad (32)$$

where F_L is given by Eq. 30 and where R_L is defined from the multipole expansion of the pseudo-tensor as

$$R_L^{\alpha\beta}(t) = \text{FP}_{B=0} \int d^3\mathbf{x} \tilde{|\mathbf{x}|}^B \hat{x}_L \int_1^{+\infty} dz \gamma_\ell(z) \mathcal{M}(\tau^{\alpha\beta})(\mathbf{x}, t - z|\mathbf{x}|/c). \quad (33)$$

Here the regularization deals with the bound of the integral at $|\mathbf{x}| = 0$. Since the variable z extends up to infinity these functions truly depend on the whole past history of the source. The weighting function therein is simply given by $\gamma_\ell(z) \equiv -2\delta_\ell(z)$, this definition being motivated by the fact that the integral of that

function is normalized to one: $\int_1^{+\infty} dz \gamma_\ell(z) = 1$.⁶ The specific contributions due to R_L in the post-Newtonian metric (25) are associated with tails of waves [10, 15]. The fact that the external multipolar expansion $\mathcal{M}(\tau)$ is the source term for the function R_L , and therefore will enter the expression of the near-zone metric (25), is a result of the matching condition (26) and reflects of course the no-incoming radiation condition imposed at \mathcal{S}_η^- .

The post-Newtonian metric (25) is now fully determined. However, let us now derive an interesting alternative formulation of it [24]. To this end we introduce still another object that will be made of the expansion of the standard retarded integral (15) when $c \rightarrow \infty$, but acting on a post-Newtonian source term $\bar{\tau}$,

$$\square_{\mathbb{R}}^{-1}[\bar{\tau}^{\alpha\beta}](\mathbf{x}, t) \equiv -\frac{1}{4\pi} \sum_{n=0}^{+\infty} \frac{(-)^n}{n!} \left(\frac{\partial}{c \partial t} \right)^n \text{FP}_{B=0} \int d^3\mathbf{x}' |\bar{\mathbf{x}}'|^B |\mathbf{x}-\mathbf{x}'|^{n-1} \bar{\tau}^{\alpha\beta}(\mathbf{x}', t). \quad (34)$$

Each of the terms is regularized by means of the finite part to deal with the bound at infinity where the post-Newtonian expansion is singular. This regularization is crucial and the object should carefully be distinguished from the “global” solution $\square_{\mathbb{R}}^{-1}[\tau]$ defined by Eq. 14 and in which the pseudo-tensor is *not* expanded in post-Newtonian fashion. We emphasize that Eq. 34 constitutes merely the *definition* of a (formal) post-Newtonian expansion, each term of which being built from the post-Newtonian expansion of the pseudo-tensor. Such a definition is of interest because it corresponds to what one would intuitively think as the “natural” way of performing the post-Newtonian iteration, that is, by Taylor expanding the retardations as in [1]. Moreover, each of the terms of the series (34) is mathematically well defined thanks to the finite part, and can therefore be implemented in practical computations. The point is that Eq. 34 solves, in a post-Newtonian sense, the wave equation,

$$\square \left(\square_{\mathbb{R}}^{-1}[\bar{\tau}^{\alpha\beta}] \right) = \bar{\tau}^{\alpha\beta}, \quad (35)$$

so constitutes a good prescription for a particular solution of the wave equation – as legitimate a prescription as Eq. 23. Therefore Eqs. 23 and 34 should differ by an homogeneous solution of the wave equation, which is necessarily of the anti-symmetric type. Detailed investigations yield

$$\square_{\mathbb{R}}^{-1}[\bar{\tau}^{\alpha\beta}] = \square_{\mathbb{I}}^{-1}[\bar{\tau}^{\alpha\beta}] - \frac{1}{4\pi} \sum_{\ell=0}^{+\infty} \frac{(-)^{\ell}}{\ell!} \hat{\partial}_L \left\{ \frac{F_L^{\alpha\beta}(t-r/c) - F_L^{\alpha\beta}(t+r/c)}{2r} \right\}, \quad (36)$$

in which the homogeneous solution is parametrized precisely by the multipole-moment functions $F_L(t)$. This formula is the basis of our writing of the new

⁶ This integral is a priori divergent, however, its value can be obtained by invoking complex analytic continuation in $\ell \in \mathbb{C}$.

form of the post-Newtonian expansion. Indeed, by combining Eqs. 25 and 36, we nicely get

$$\bar{h}^{\alpha\beta} = \frac{16\pi G}{c^4} \square_{\mathbb{R}}^{-1}[\bar{\tau}^{\alpha\beta}] - \frac{4G}{c^4} \sum_{\ell=0}^{+\infty} \frac{(-)^\ell}{\ell!} \hat{\partial}_L \left\{ \frac{R_L^{\alpha\beta}(t-r/c) - R_L^{\alpha\beta}(t+r/c)}{2r} \right\}, \quad (37)$$

which is our final expression for the general solution of the post-Newtonian field in the near-zone of any isolated matter distribution. This expression is probably most convenient and fruitful when doing practical applications.

We recognize in the first term of Eq. 37 (notwithstanding the finite part therein) the old way of performing the post-Newtonian expansion as it was advocated by Anderson and DeCanio [1]. For computations limited to the 3.5PN order, that is, up to the level of the 1PN correction to the radiation-reaction force, such a first term is sufficient. However, at the 4PN order there is a fundamental breakdown of this scheme and it becomes necessary to take into account the second term in Eq. 37, which corresponds to nonlinear radiation-reaction effects associated with tails.

Note that the post-Newtonian solution, in either form Eq. 25 or Eq. 37, has been obtained without imposing the condition of harmonic coordinates in an explicit way, see Eq. 7. We have simply matched together the post-Newtonian and multipolar expansions, satisfying the “relaxed” Einstein field equations (8) in their respective domains, and found that the matching determines uniquely the solution. An important check (carried out in [24, 82]) is therefore to verify that the harmonic coordinate condition (7) is indeed satisfied as a consequence of the conservation of the pseudo-tensor (9), so that we really grasp a solution of the full Einstein field equations.

2.4 Multipole Moments of a Post-Newtonian Source

The multipole expansion of the field outside a general post-Newtonian source has been obtained in the previous section as⁷

$$\mathcal{M}(h^{\alpha\beta}) = -\frac{4G}{c^4} \sum_{\ell=0}^{+\infty} \frac{(-)^\ell}{\ell!} \hat{\partial}_L \left\{ \frac{F_L^{\alpha\beta}(t-r/c)}{r} \right\} + u^{\alpha\beta}, \quad (38)$$

where the multipole moments are explicitly given by Eq. 30, and the second piece reflects the nonlinearities of the Einstein field equations and reads

$$u^{\alpha\beta} = \square_{\mathbb{R}}^{-1}[\mathcal{M}(\Lambda^{\alpha\beta})]. \quad (39)$$

⁷ An alternative formulation of the multipole expansion for a post-Newtonian source, with non-STF multipole moments, has been developed by Will and collaborators [77, 78, 99].

To write the latter expression we have used the fact that since the matter tensor $T^{\alpha\beta}$ has a spatially compact support we have $\mathcal{M}(T^{\alpha\beta}) = 0$. Thus $u^{\alpha\beta}$ is indeed generated by the nonlinear gravitational source term (11). We notice that the divergence of this piece, say $w^\alpha \equiv \partial_\mu u^{\alpha\mu}$, is a retarded homogeneous solution of the wave equation, that is, of the same type as the first term in Eq. 38. Now from w^α we can construct a secondary object $v^{\alpha\beta}$, which is also a retarded homogeneous solution of the wave equation, and furthermore whose divergence satisfies $\partial_\mu v^{\alpha\mu} = -w^\alpha$, so that it will cancel the divergence of $u^{\alpha\beta}$ (see [11] for details). With the above construction of $v^{\alpha\beta}$ we are able to define the following combination,

$$G h_{(1)}^{\alpha\beta} \equiv -\frac{4G}{c^4} \sum_{\ell=0}^{+\infty} \frac{(-)^\ell}{\ell!} \hat{\partial}_L \left\{ \frac{F_L^{\alpha\beta}(t-r/c)}{r} \right\} - v^{\alpha\beta}, \quad (40)$$

which will constitute the *linearized* approximation to the multipolar expansion $\mathcal{M}(h^{\alpha\beta})$ outside the source. Then we have

$$\mathcal{M}(h^{\alpha\beta}) = G h_{(1)}^{\alpha\beta} + u^{\alpha\beta} + v^{\alpha\beta}. \quad (41)$$

Having singled out such linearized part, it is clear that the sum of the second and third terms should represent the nonlinearities in the external field. If we index those nonlinearities by Newton's constant G , then we can prove indeed that $u^{\alpha\beta} + v^{\alpha\beta} = \mathcal{O}(G^2)$. More precisely we can decompose $u^{\alpha\beta} + v^{\alpha\beta}$ as a complete nonlinearity or “post-Minkowskian” expansion of the type

$$u^{\alpha\beta} + v^{\alpha\beta} = \sum_{m=2}^{+\infty} G^m h_{(m)}^{\alpha\beta}. \quad (42)$$

One can effectively define a post-Minkowskian “algorithm” [11, 14] able to construct the nonlinear series up to any postMinkowskian order m . The post-Minkowskian expansion represents the most general solution of the Einstein field equations in harmonic coordinates valid in the vacuum region outside an isolated source.

The above linearized approximation $h_{(1)}$ solves the linearized vacuum Einstein field equations in harmonic coordinates and it can be decomposed into multipole moments in a standard way [94]. Modulo an infinitesimal gauge transformation preserving the harmonic gauge, namely,

$$h_{(1)}^{\alpha\beta} = k_{(1)}^{\alpha\beta} + \partial^\alpha \varphi_{(1)}^\beta + \partial^\beta \varphi_{(1)}^\alpha - \eta^{\alpha\beta} \partial_\mu \varphi_{(1)}^\mu, \quad (43)$$

where the infinitesimal gauge vector $\varphi_{(1)}^\alpha$ satisfies $\square \varphi_{(1)}^\alpha = 0$, we can decompose⁸

⁸ The superscript (k) refers to k time derivatives of the moments; ε_{abc} is the Levi-Civita antisymmetric symbol such that $\varepsilon_{123} = 1$. From here on the spatial indices such as i, j, \dots will be raised

$$k_{(1)}^{00} = -\frac{4}{c^2} \sum_{\ell \geq 0} \frac{(-)^\ell}{\ell!} \partial_L \left(\frac{1}{r} \mathbf{I}_L \right), \quad (44a)$$

$$k_{(1)}^{0i} = \frac{4}{c^3} \sum_{\ell \geq 1} \frac{(-)^\ell}{\ell!} \left\{ \partial_{L-1} \left(\frac{1}{r} \mathbf{I}_{iL-1}^{(1)} \right) + \frac{\ell}{\ell+1} \varepsilon_{iab} \partial_{aL-1} \left(\frac{1}{r} \mathbf{J}_{bL-1} \right) \right\}, \quad (44b)$$

$$k_{(1)}^{ij} = -\frac{4}{c^4} \sum_{\ell \geq 2} \frac{(-)^\ell}{\ell!} \left\{ \partial_{L-2} \left(\frac{1}{r} \mathbf{I}_{ijL-2}^{(2)} \right) + \frac{2\ell}{\ell+1} \partial_{aL-2} \left(\frac{1}{r} \varepsilon_{ab(i} \mathbf{J}_{j)L-2}^{(1)} \right) \right\}. \quad (44c)$$

This decomposition defines two types of multipole moments, both assumed to be STF: the mass-type $\mathbf{I}_L(u)$ and the current-type $\mathbf{J}_L(u)$. These moments can be arbitrary functions of the retarded time $u \equiv t - r/c$, except that the monopole and dipole moments (having $\ell \leq 1$) satisfy standard conservation laws, namely,

$$\mathbf{I}^{(1)} = \mathbf{I}_i^{(2)} = \mathbf{J}_i^{(1)} = 0. \quad (45)$$

The gauge transformation vector admits a decomposition in similar fashion,

$$\varphi_{(1)}^0 = \frac{4}{c^3} \sum_{\ell \geq 0} \frac{(-)^\ell}{\ell!} \partial_L \left(\frac{1}{r} \mathbf{W}_L \right), \quad (46a)$$

$$\varphi_{(1)}^i = -\frac{4}{c^4} \sum_{\ell \geq 0} \frac{(-)^\ell}{\ell!} \partial_{iL} \left(\frac{1}{r} \mathbf{X}_L \right) \quad (46b)$$

$$-\frac{4}{c^4} \sum_{\ell \geq 1} \frac{(-)^\ell}{\ell!} \left\{ \partial_{L-1} \left(\frac{1}{r} \mathbf{Y}_{iL-1} \right) + \frac{\ell}{\ell+1} \varepsilon_{iab} \partial_{aL-1} \left(\frac{1}{r} \mathbf{Z}_{bL-1} \right) \right\}. \quad (46c)$$

The six sets of STF multipole moments \mathbf{I}_L , \mathbf{J}_L , \mathbf{W}_L , \mathbf{X}_L , \mathbf{Y}_L , and \mathbf{Z}_L will collectively be called the multipole moments of the source. They contain the full physical information about any isolated source as seen from its exterior near zone. Actually it should be clear that the main moments are \mathbf{I}_L and \mathbf{J}_L because the other moments $\mathbf{W}_L, \dots, \mathbf{Z}_L$ parametrize a linear gauge transformation and thus have no physical implications at the linearized order. However, because the theory is covariant with respect to nonlinear diffeomorphisms and not merely with respect to linear gauge transformations, the moments $\mathbf{W}_L, \dots, \mathbf{Z}_L$ do play a physical role starting at the nonlinear level. We shall occasionally refer to the moments \mathbf{W}_L , \mathbf{X}_L , \mathbf{Y}_L , and \mathbf{Z}_L as the gauge moments.

and lowered with the Kronecker metric δ_{ij} . They will be located lower or upper depending on context.

To express in the best way the source multipole moments, we introduce the following notation for combinations of components of the pseudo-tensor $\bar{\tau}^{\alpha\beta}$,

$$\bar{\Sigma} \equiv \frac{\bar{\tau}^{00} + \bar{\tau}^{ii}}{c^2}, \quad (47a)$$

$$\bar{\Sigma}_i \equiv \frac{\bar{\tau}^{0i}}{c}, \quad (47b)$$

$$\bar{\Sigma}_{ij} \equiv \bar{\tau}^{ij}, \quad (47c)$$

where $\bar{\tau}^{ii} \equiv \delta_{ij} \bar{\tau}^{ij}$. Here the overbar reminds us that we are exclusively dealing with post-Newtonian-expanded expressions, that is, formal series of the type (17). Then the general expressions of the “main” source multipole moments I_L and J_L in the case of the time-varying moments for which $\ell \geq 2$, are

$$\begin{aligned} I_L(u) = \text{FP}_{B=0} \int d^3\mathbf{x} |\tilde{\mathbf{x}}|^B \int_{-1}^1 dz \left\{ \delta_\ell \hat{x}_L \bar{\Sigma} - \frac{4(2\ell+1)}{c^2(\ell+1)(2\ell+3)} \delta_{\ell+1} \hat{x}_{iL} \bar{\Sigma}_i^{(1)} \right. \\ \left. + \frac{2(2\ell+1)}{c^4(\ell+1)(\ell+2)(2\ell+5)} \delta_{\ell+2} \hat{x}_{ijL} \bar{\Sigma}_{ij}^{(2)} \right\}, \end{aligned} \quad (48a)$$

$$\begin{aligned} J_L(u) = \text{FP}_{B=0} \int d^3\mathbf{x} |\tilde{\mathbf{x}}|^B \int_{-1}^1 dz \varepsilon_{ab(i\ell} \left\{ \delta_\ell \hat{x}_{L-1)a} \bar{\Sigma}_b \right. \\ \left. - \frac{2\ell+1}{c^2(\ell+2)(2\ell+3)} \delta_{\ell+1} \hat{x}_{L-1)ac} \bar{\Sigma}_{bc}^{(1)} \right\}. \end{aligned} \quad (48b)$$

The integrands are computed at the spatial point \mathbf{x} and at time $u + z|\mathbf{x}|/c$, where $u = t - r/c$ is the retarded time at which are evaluated the moments. We recall that z is the argument of the function δ_ℓ defined in Eq. 31. Similarly we can write the expressions of the gauge-type moments W_L, \dots, Z_L . Notice that the source multipole moments (48) have no invariant meaning; they are defined for the harmonic coordinate system we have chosen.

Of what use are these results for the multipole moments I_L and J_L ? From Eq. 44 these moments parametrize the linearized metric $h_{(1)}$, which is the “seed” of an infinite post-Minkowskian algorithm symbolized by Eq. 42. For a specific application, that is, for a specific choice of matter tensor like the one we shall describe in Section 3.1, the expressions (48) have to be worked out up to a given post-Newtonian order. The moments should then be inserted into the post-Minkowskian series (42) for the computation of the nonlinearities. The result will be in the form of a nonlinear multipole decomposition depending on the source moments I_L, J_L, \dots, Z_L , say

$$\mathcal{M}(h^{\alpha\beta}) = \sum_{m=1}^{+\infty} G^m h_{(m)}^{\alpha\beta}[I_L, J_L, \dots]. \quad (49)$$

In the next section we shall expand this metric at (retarded) infinity from the source in order to obtain the observables of the gravitational radiation field.

2.5 Radiation Field and Polarization Waveforms

The asymptotic waveform at future null infinity from an isolated source is the transverse-traceless (TT) projection of the metric deviation at the leading order $1/R$ in the distance $R = |\mathbf{X}|$ to the source, in a radiative coordinate system $X^\mu = (cT, \mathbf{X})$.⁹ The waveform can be uniquely decomposed [94] into radiative multipole components parametrized by mass-type moments U_L and current-type ones V_L . We shall define the radiative moments in such a way that they agree with the ℓ -th time derivatives of the source moments I_L and J_L at the linear level, that is,

$$U_L = I_L^{(\ell)} + \mathcal{O}(G), \quad (50a)$$

$$V_L = J_L^{(\ell)} + \mathcal{O}(G). \quad (50b)$$

At the nonlinear level the radiative moments will crucially differ from the source moments; the relations between these two types of moments will be discussed in the next section. The radiative moments $U_L(U)$ and $V_L(U)$ are functions of the retarded time $U \equiv T - R/c$ in radiative coordinates.

The asymptotic waveform at distance R and retarded time U is then given by

$$h_{ij}^{\text{TT}} = \frac{4G}{c^2 R} \mathcal{P}_{ijkl} \sum_{\ell=2}^{+\infty} \frac{1}{c^\ell \ell!} \left\{ N_{L-2} U_{klL-2} - \frac{2\ell}{c(\ell+1)} N_{aL-2} \varepsilon_{ab(k} V_{l)bL-2} \right\}. \quad (51)$$

We denote by $\mathbf{N} = \mathbf{X}/R = (N_i)$ the unit vector pointing from the source to the far-away detector. The TT projection operator reads $\mathcal{P}_{ijkl} = \mathcal{P}_{ik}\mathcal{P}_{jl} - \frac{1}{2}\mathcal{P}_{ij}\mathcal{P}_{kl}$ where $\mathcal{P}_{ij} = \delta_{ij} - N_i N_j$ is the projector orthogonal to the unit direction \mathbf{N} . We project out the asymptotic waveform (51) on polarization directions in a standard way. We denote the two unit polarization vectors by \mathbf{P} and \mathbf{Q} , which are orthogonal and transverse to the direction of propagation \mathbf{N} (hence $\mathcal{P}_{ij} = P_i P_j + Q_i Q_j$). Our conventions and choice for \mathbf{P} and \mathbf{Q} will be specified in Section 3.5. Then the two “plus” and “cross” polarization states of the waveform are

$$h_+ \equiv \frac{P_i P_j - Q_i Q_j}{2} h_{ij}^{\text{TT}}, \quad (52a)$$

$$h_\times \equiv \frac{P_i Q_j + P_j Q_i}{2} h_{ij}^{\text{TT}}. \quad (52b)$$

⁹ Radiative coordinates T and \mathbf{X} , also called Bondi-type coordinates [28], are such that the metric coefficients admit an expansion when $R \rightarrow +\infty$ with $U \equiv T - R/c$ being constant, in simple powers of $1/R$, without the logarithms of R plaguing the harmonic coordinate system. Here U is a null or asymptotically null characteristic. It is known that the “far-zone” logarithms in harmonic coordinates can be removed order-by-order by going to radiative coordinates [9].

Although the multipole decomposition (51) entirely describes the waveform, it is also important, especially having in mind the comparison between the post-Newtonian results and numerical relativity [29], to consider separately the modes (ℓ, m) of the waveform as defined with respect to a basis of spin-weighted spherical harmonics. To this end we decompose h_+ and h_\times as (see, e.g., [29, 64])

$$h_+ - ih_\times = \sum_{\ell=2}^{+\infty} \sum_{m=-\ell}^{\ell} h^{\ell m} Y_{-2}^{\ell m}(\Theta, \Phi), \quad (53)$$

where the spin-weighted spherical harmonics of weight -2 is a function of the spherical angles (Θ, Φ) defining the direction of propagation \mathbf{N} and reads

$$Y_{-2}^{\ell m} = \sqrt{\frac{2\ell+1}{4\pi}} d^{\ell m}(\Theta) e^{im\Phi}, \quad (54a)$$

$$d^{\ell m} \equiv \sum_{k=k_1}^{k_2} D_k^{\ell m} \left(\cos \frac{\Theta}{2} \right)^{2\ell+m-2k-2} \left(\sin \frac{\Theta}{2} \right)^{2k-m+2}, \quad (54b)$$

$$D_k^{\ell m} \equiv \frac{(-)^k}{k!} \frac{\sqrt{(\ell+m)!(\ell-m)!(\ell+2)!(\ell-2)!}}{(k-m+2)!(\ell+m-k)!(\ell-k-2)!}. \quad (54c)$$

Here $k_1 = \max(0, m-2)$ and $k_2 = \min(\ell+m, \ell-2)$. Using the orthonormality properties of these harmonics we obtain the separate modes $h^{\ell m}$ from the surface integral (with the overline denoting the complex conjugate)

$$h^{\ell m} = \int d\Omega \left[h_+ - ih_\times \right] \overline{Y}_{-2}^{\ell m}(\Theta, \Phi). \quad (55)$$

On the other hand, we can also write $h^{\ell m}$ directly in terms of the radiative multipole moments U_L and V_L , with result

$$h^{\ell m} = -\frac{G}{\sqrt{2} R c^{\ell+2}} \left[U^{\ell m} - \frac{i}{c} V^{\ell m} \right], \quad (56)$$

where $U^{\ell m}$ and $V^{\ell m}$ are the radiative mass and current moments in non-STF guise. These are given in terms of the STF moments by

$$U^{\ell m} = \frac{4}{\ell!} \sqrt{\frac{(\ell+1)(\ell+2)}{2\ell(\ell-1)}} \alpha_L^{\ell m} U_L, \quad (57a)$$

$$V^{\ell m} = -\frac{8}{\ell!} \sqrt{\frac{\ell(\ell+2)}{2(\ell+1)(\ell-1)}} \alpha_L^{\ell m} V_L. \quad (57b)$$

Here $\alpha_L^{\ell m}$ denotes the STF tensor connecting together the usual basis of spherical harmonics $Y^{\ell m}$ to the set of STF tensors $\hat{N}_L \equiv \text{STF}(N_L)$, recalling that $Y^{\ell m}$ and \hat{N}_L represent two basis of an irreducible representation of weight ℓ of the rotation group. They are related by

$$\hat{N}_L(\Theta, \Phi) = \sum_{m=-\ell}^{\ell} \alpha_L^{\ell m} Y^{\ell m}(\Theta, \Phi), \quad (58a)$$

$$Y^{\ell m}(\Theta, \Phi) = \frac{(2\ell + 1)!!}{4\pi\ell!} \bar{\alpha}_L^{\ell m} \hat{N}_L(\Theta, \Phi), \quad (58b)$$

with the STF tensorial coefficient being

$$\alpha_L^{\ell m} = \int d\Omega \hat{N}_L \bar{Y}^{\ell m}. \quad (59)$$

The decomposition in spherical harmonic modes is especially useful if some of the radiative moments are known to higher post-Newtonian order than others. In this case the comparison with the numerical calculation [29, 64] can be made for these individual modes with higher post-Newtonian accuracy.

2.6 Radiative Moments Versus Source Moments

The basis of our computation is the general solution of the Einstein field equations outside an isolated matter system computed iteratively in the form of a post-Minkowskian or nonlinearity expansion (49) (see details in [14, 17]). Here we give some results concerning the relation between the set of radiative moments $\{\mathbf{U}_L, \mathbf{V}_L\}$ and the sets of source moments $\{\mathbf{I}_L, \mathbf{J}_L\}$ and gauge moments $\{\mathbf{W}_L, \dots, \mathbf{Z}_L\}$. Complete results up to 3PN order are available and have recently been used to control the 3PN waveform of compact binaries [23].

Armed with definitions for all those moments, we proceed in a modular way. We express the radiative moments $\{\mathbf{U}_L, \mathbf{V}_L\}$ in terms of some convenient intermediate constructs $\{\mathbf{M}_L, \mathbf{S}_L\}$ called the canonical moments. Essentially these canonical moments take into account the effect of the gauge transformation present in Eq. 43. Therefore they differ from the source moments $\{\mathbf{I}_L, \mathbf{J}_L\}$ only at nonlinear order. We shall see that in terms of a post-Newtonian expansion the canonical and source moments agree with each other up to 2PN order. The canonical moments are then connected to the actual source multipole moments $\{\mathbf{I}_L, \mathbf{J}_L\}$ and $\{\mathbf{W}_L, \dots, \mathbf{Z}_L\}$. The point of the above strategy is that the source moments (including gauge moments) admit closed-form expressions as integrals over the stress-energy distribution of matter and gravitational fields in the source, as shown in Eq. 48.

The mass quadrupole moment U_{ij} (having $\ell = 2$) is known up to the 3PN order [12]. At that order it is made of quadratic and cubic nonlinearities, and we have

$$\begin{aligned}
U_{ij}(U) = & M_{ij}^{(2)}(U) + \frac{2GM}{c^3} \int_{-\infty}^U du \left[\ln \left(\frac{U-u}{2u_0} \right) + \frac{11}{12} \right] M_{ij}^{(4)}(u) \\
& + \frac{G}{c^5} \left\{ -\frac{2}{7} \int_{-\infty}^U du M_{a(i}^{(3)}(u) M_{j)a}^{(3)}(u) + \frac{1}{7} M_{a(i}^{(5)} M_{j)a} - \frac{5}{7} M_{a(i}^{(4)} M_{j)a}^{(1)} \right. \\
& \quad \left. - \frac{2}{7} M_{a(i}^{(3)} M_{j)a}^{(2)} + \frac{1}{3} \varepsilon_{ab(i} M_{j)a}^{(4)} S_b \right\} + 2 \left(\frac{GM}{c^3} \right)^2 \\
& \times \int_{-\infty}^U du \left[\ln^2 \left(\frac{U-u}{2u_0} \right) + \frac{57}{70} \ln \left(\frac{U-u}{2u_0} \right) + \frac{124627}{44100} \right] M_{ij}^{(5)}(u) \\
& + \mathcal{O} \left(\frac{1}{c^7} \right). \tag{60}
\end{aligned}$$

Notice the quadratic tail integral at 1.5PN order, the cubic tail-of-tail integral at 3PN order, and the nonlinear memory integral at 2.5PN order [17, 35, 95, 100]. The tail is composed of the coupling between the mass quadrupole moment M_{ij} and the mass monopole moment or total mass M ; the tail-of-tail is a coupling between M_{ij} and two monopoles $M \times M$; the nonlinear memory is a coupling $M_{ij} \times M_{kl}$. All these “hereditary” integrals imply a dependence of the waveform on the complete history of the source, from infinite past up to the current retarded time $U \equiv T - R/c$. The constant u_0 in the tail integrals is defined by $u_0 \equiv r_0/c$, where r_0 is the arbitrary length scale introduced in Eq. 19.

Note that the dominant hereditary integral is the tail arising at 1.5PN order in all radiative moments. For general ℓ we have at that order

$$U_L = M_L^{(\ell)} + \frac{2GM}{c^3} \int_{-\infty}^U du \left[\ln \left(\frac{U-u}{2u_0} \right) + \kappa_\ell \right] M_L^{(\ell+2)}(u) + \mathcal{O} \left(\frac{1}{c^5} \right), \tag{61a}$$

$$V_L = S_L^{(\ell)} + \frac{2GM}{c^3} \int_{-\infty}^U du \left[\ln \left(\frac{U-u}{2u_0} \right) + \pi_\ell \right] S_L^{(\ell+2)}(u) + \mathcal{O} \left(\frac{1}{c^5} \right), \tag{61b}$$

where the constants κ_ℓ and π_ℓ are given by

$$\kappa_\ell = \frac{2\ell^2 + 5\ell + 4}{\ell(\ell+1)(\ell+2)} + \sum_{k=1}^{\ell-2} \frac{1}{k}, \tag{62a}$$

$$\pi_\ell = \frac{\ell-1}{\ell(\ell+1)} + \sum_{k=1}^{\ell-1} \frac{1}{k}. \tag{62b}$$

Now it can be proved that the retarded time U in radiative coordinates reads

$$U = t - \frac{r}{c} - \frac{2GM}{c^3} \ln \left(\frac{r}{r_0} \right) + \mathcal{O} \left(\frac{1}{c^5} \right), \tag{63}$$

where (t, r) are the harmonic coordinates. Inserting U into Eq. 61 we obtain the radiative moments expressed in terms of local source-rooted coordinates (t, r) , for example,

$$U_L = M_L^{(\ell)}(t - r/c) + \frac{2GM}{c^3} \int_{-\infty}^{t-r/c} du \left[\ln \left(\frac{t - u - r/c}{2r/c} \right) + \kappa_\ell \right] M_L^{(\ell+2)}(u) + \mathcal{O} \left(\frac{1}{c^5} \right). \quad (64)$$

This no longer depends on the constant u_0 —that is, the u_0 gets replaced by the retardation time r/c . More generally it can be checked that u_0 always disappears from physical results at the end. On the other hand we can be convinced that the constant κ_ℓ (and π_ℓ as well) depends on the choice of source-rooted coordinates (t, r) . For instance if we change the harmonic coordinate system (t, r) to some “Schwarzschild-like” coordinates (t', r') such that $t' = t$ and $r' = r + GM/c^2$, we get a new constant $\kappa'_\ell = \kappa_\ell + 1/2$. Thus we have $\kappa_2 = 11/12$ in harmonic coordinates [as shown in Eq. 60], but $\kappa'_2 = 17/12$ in Schwarzschild coordinates.

We still have to relate the canonical moments $\{M_L, S_L\}$ to the source multipole moments. As we said the difference between these two types of moments comes from the gauge transformation (46) and arises only at the small 2.5PN order. The consequence is that we have to worry about this difference only for high post-Newtonian waveforms. For the mass quadrupole moment M_{ij} , the requisite correction is given by

$$M_{ij} = I_{ij} + \frac{4G}{c^5} \left[W^{(2)} I_{ij} - W^{(1)} I_{ij}^{(1)} \right] + \mathcal{O} \left(\frac{1}{c^7} \right), \quad (65)$$

where I_{ij} denotes the source mass quadrupole, and where W is the monopole corresponding to the gauge moments W_L (i.e., the moment having $\ell = 0$). Up to 3PN order, W is only needed at Newtonian order. The expression (65) is valid in a mass-centered frame defined by the vanishing of the conserved mass dipole moment: $I_i = 0$.

Note that closed-form formulas generalizing (65) and similar expressions to all post-Newtonian orders (and all multipole interactions), if they exist are not known; they need to be investigated anew for specific cases. Thus it is convenient in the present approach to systematically keep the source moments $\{I_L, J_L, W_L, X_L, Y_L, Z_L\}$ as the fundamental variables describing the source.

3 Inspiralling Compact Binaries

3.1 Stress–Energy Tensor of Spinning Particles

So far the post-Newtonian formalism has been developed for arbitrary matter distributions. We want now to apply it to material systems made of compact objects (neutron stars or black holes), which can be described with great precision by point

masses. We thus discuss the modeling of point-particles possibly carrying some intrinsic rotation or spin. This means finding the appropriate stress–energy tensor that will have to be inserted into the general post-Newtonian formulas such as the expressions of the source moments (48).

In the general case the stress–energy tensor will be the sum of a “monopolar” piece, which is a linear combination of monopole sources, that is, made of Dirac delta-functions, plus the “dipolar” or spin piece, made of *gradients* of Dirac delta-functions. Hence we write

$$T^{\alpha\beta} = T_{\text{mono}}^{\alpha\beta} + T_{\text{spin}}^{\alpha\beta}. \quad (66)$$

The monopole part takes the form of the stress–energy tensor for N particles (labeled by $A = 1, \dots, N$) without spin, reading in a four-dimensional picture

$$T_{\text{mono}}^{\alpha\beta} = c \sum_{A=1}^N \int_{-\infty}^{+\infty} d\tau_A p_A^{(\alpha} u_A^{\beta)} \frac{\delta^{(4)}(x - y_A)}{\sqrt{-(g)_A}}. \quad (67)$$

Here $\delta^{(4)}$ is the four-dimensional Dirac function. The world-line of particle A , denoted y_A^α , is parametrized by the particle’s proper time τ_A . The four-velocity is given by $u_A^\alpha = dy_A^\alpha/d\tau_A$ and is normalized to $(g_{\mu\nu})_A u_A^\mu u_A^\nu = -c^2$, where $(g_{\mu\nu})_A$ denotes the metric at the particle’s location. The four-vector p_A^α is the particle’s linear momentum. For particles without spin we shall simply have $p_A^\alpha = m_A u_A^\alpha$. However, for spinning particles p_A^α will differ from that and include some contributions from the spins as given by Eq. 75 below.

The dipolar part of the stress–energy tensor depends specifically on the spins, and reads in the classic formalism of spinning particles (due to Tulczyjew [97, 98], Trautman [96], Dixon [49], Bailey and Israel [6]),

$$T_{\text{spin}}^{\alpha\beta} = - \sum_{A=1}^N \nabla_\mu \left[\int_{-\infty}^{+\infty} d\tau_A S_A^{\mu(\alpha} u_A^{\beta)} \frac{\delta^{(4)}(x - y_A)}{\sqrt{-(g)_A}} \right], \quad (68)$$

where ∇_μ is the covariant derivative, and the anti-symmetric tensor $S_A^{\alpha\beta}$ represents the spin angular momentum of particle A . In this formalism the momentum-like quantity p_A^α [entering Eq. 67] is a time-like solution of the equation

$$\frac{D S_A^{\alpha\beta}}{d\tau_A} = \left(p_A^\alpha u_A^\beta - p_A^\beta u_A^\alpha \right), \quad (69)$$

where $D/d\tau_A$ denotes the covariant proper time derivative. The equation of translational motion of the spinning particle, equivalent to the covariant conservation $\nabla_\mu T^{\alpha\mu} = 0$ of the total stress–energy tensor (66), is the Papapetrou equation [75] involving a coupling to curvature,

$$\frac{D p_A^\alpha}{d\tau_A} = -\frac{1}{2} S_A^{\mu\nu} u_A^\rho (R^\alpha{}_{\rho\mu\nu})_A. \quad (70)$$

The Riemann tensor is evaluated at the particle's position A . This equation can also be derived directly from an action principle [6].

It is well known that a choice must be made for a supplementary spin condition in order to fix unphysical degrees of freedom associated with an arbitrariness in the definition of the spin tensor $S^{\alpha\beta}$. This arbitrariness can be interpreted, in the case of extended bodies, as a freedom in the choice for the location of the center-of-mass world-line of the body, with respect to which the angular momentum is defined (see, e.g., [63]). An elegant spin condition is the covariant one,

$$S_A^{\alpha\mu} p_\mu^A = 0, \quad (71)$$

which allows a natural definition of a spin four-(co)vector S_α^A such that

$$S_A^{\alpha\beta} = -\frac{1}{\sqrt{-(g)_A}} \varepsilon^{\alpha\beta\mu\nu} \frac{p_\mu^A}{m_{AC}} S_\nu^A. \quad (72)$$

For the spin vector S_α^A itself, we can choose a four-vector that is purely spatial in the particle's instantaneous rest frame, where $u_A^\alpha = (1, \mathbf{0})$. Therefore, we deduce that in any frame

$$S_\mu^A u_A^\mu = 0. \quad (73)$$

As a consequence of the covariant spin condition (71), we easily verify that the spin scalar is conserved along the trajectories, that is,

$$S_A^{\mu\nu} S_{\mu\nu}^A = \text{const.} \quad (74)$$

Furthermore, we can check, using Eq. 71 and also the law of motion (70), that the mass defined by $m_A^2 c^2 = -p_A^\mu p_\mu^A$ is also constant along the trajectories: $m_A = \text{const.}$ Finally, the relation linking the four-momentum p_A^α and the four-velocity u_A^α is readily deduced from the contraction of Eq. 69 with the four-momentum, which results in

$$p_A^\alpha (pu)_A + m_A^2 c^2 u_A^\alpha = \frac{1}{2} S_A^{\alpha\mu} S_A^{\nu\rho} u_A^\sigma (R_{\mu\sigma\nu\rho})_A, \quad (75)$$

where $(pu)_A \equiv p_\mu^A u_A^\mu$. Contracting further this relation with the four-velocity one deduces the expression of $(pu)_A$ and inserting this back into Eq. 75 yields the desired relation between p_A^α and u_A^α .

Focusing our attention on spin-orbit interactions, which are *linear* in the spins, we can neglect quadratic and higher spin corrections denoted $\mathcal{O}(S^2)$; drastic simplifications of the formalism occur in this case. Since the right-hand-side of Eq. 75 is quadratic in the spins, we find that the four-momentum is linked to the four-velocity by the simple proportionality relation

$$p_A^\alpha = m_A u_A^\alpha + \mathcal{O}(S^2). \quad (76)$$

Hence, the spin condition (71) becomes

$$S_A^{\alpha\mu} u_\mu^A = \mathcal{O}(S^3). \quad (77)$$

Also, the equation of evolution for the spin, sometimes called the *precessional* equation, follows immediately from the relationship (69) together with the law (76) as

$$\frac{DS_A^{\alpha\beta}}{d\tau_A} = \mathcal{O}(S^2) \iff \frac{DS_\alpha^A}{d\tau_A} = \mathcal{O}(S^2). \quad (78)$$

Hence the spin vector S_α^A satisfies the equation of parallel transport, which means that it remains constant in a freely falling frame, as could have been expected beforehand. Of course the norm of the spin vector is preserved along trajectories,

$$S_\mu^A S_A^\mu = \text{const.} \quad (79)$$

In Section 3.6 we shall apply this formalism to the study of spin–orbit effects in the equations of motion and energy flux of compact binaries.

3.2 Hadamard Regularization

The stress–energy tensor of point masses has been defined in the previous section by means of Dirac functions, and involves metric coefficients evaluated at the locations of the point particles, namely, $(g_{\alpha\beta})_A$. However, it is clear that the metric $g_{\alpha\beta}$ becomes *singular* at the particles. Indeed this is already true at Newtonian order where the potential generated by N particles reads

$$U = \sum_{B=1}^N \frac{Gm_B}{r_B}, \quad (80)$$

where $r_B \equiv |\mathbf{x} - \mathbf{y}_B|$ is the distance between the field point and the particle B . Thus the values of U and hence of the metric coefficients [recall that $g_{00} = -1 + 2U/c^2 + \mathcal{O}(c^{-4})$], are ill-defined at the locations of the particles. What we need is a *self-field regularization*, that is, a prescription for removing the infinite self-field of the point masses. Arguably the choice of a particular regularization constitutes a fully qualified element of our physical modeling of compact objects. At Newtonian order the regularization of the potential (80) should give the well-known result

$$(U)_A = \sum_{B \neq A} \frac{Gm_B}{r_{AB}}, \quad (81)$$

where $r_{AB} \equiv |\mathbf{y}_A - \mathbf{y}_B|$ and the infinite self-interaction term has simply been discarded from the summation. At high post-Newtonian orders the problem is not trivial and the self-field regularization must be properly defined.

The post-Newtonian formalism reviewed in Sections 2.1–2.6 assumed from the start a continuous (smooth) matter distribution. Actually this formalism will be applicable to singular point-mass sources, described by the stress–energy tensor of Section 3.1, provided that we supplement the scheme by the self-field regularization. Note that this regularization has nothing to do with the finite-part process $\text{FP}_{B=0}$ extensively used in the case of extended matter sources. The latter finite part was an ingredient of the rigorous derivation of the general post-Newtonian solution [see Eq. 20], while the self-field regularization is an assumption regarding a particular type of singular source.

Our aim is to compute up to 3PN order the metric coefficients at the location of one of the particles: $(g_{\alpha\beta})_A$. At this stage different self-field regularizations are possible. We first review Hadamard’s regularization [60, 85], that has proved to be very efficient for doing practical computations, but suffers from the important drawback of yielding some “ambiguity parameters,” which cannot be determined within the regularization, starting at the 3PN order.

Iterating the Einstein field equations with point-like matter sources (delta-functions with spatial supports localized on \mathbf{y}_A) yields a generic form of functions representing the metric coefficients in successive post-Newtonian approximations. The generic functions, say $F(\mathbf{x})$, are smooth except at the points \mathbf{y}_A , around which they admit singular Laurent expansions in powers and inverse powers of $r_A \equiv |\mathbf{x} - \mathbf{y}_A|$. When $r_A \rightarrow 0$ we have (say, for any $P \in \mathbb{N}$)

$$F(\mathbf{x}) = \sum_{p=p_0}^P r_A^p {}_A f_p(\mathbf{n}_A) + o(r_A^P). \quad (82)$$

The coefficients ${}_A f_p$ of the various powers of r_A depend on the unit direction $\mathbf{n}_A \equiv (\mathbf{x} - \mathbf{y}_A)/r_A$ of approach to the singular point A . The powers p are relative integers, and are bounded from below by $p_0 \in \mathbb{Z}$. The Landau o -symbol for remainders takes its standard meaning. The ${}_A f_p$ ’s depend also on the (coordinate) time t , through their dependence on velocities $\mathbf{v}_B(t)$ and relative positions $\mathbf{y}_{BC}(t) \equiv \mathbf{y}_B(t) - \mathbf{y}_C(t)$; however the time t is purely “spectator” in the regularization process, and thus will not be indicated. The coefficients ${}_A f_p$ for which $p < 0$ are referred to as the *singular* coefficients of F around A .

The function F being given that way, we define the *Hadamard partie finie* as the following value of F at the location of the particle A ,¹⁰

$$(F)_A = \langle f_0 \rangle_A \equiv \int \frac{d\Omega_A}{4\pi} f_0(\mathbf{n}_A), \quad (83)$$

¹⁰ With this definition it is immediate to check that the previous Newtonian result (81) will hold.

where $d\Omega_A$ denotes the solid angle element centered on \mathbf{y}_A and sustained by \mathbf{n}_A . The brackets $\langle \rangle$ mean the angular average. The second notion of Hadamard partie finie concerns the integral $\int d^3\mathbf{x} F$, which is generically divergent at the points \mathbf{y}_A . Its partie finie (in short Pf) is defined by

$$\text{Pf}_{s_1 \dots s_N} \int d^3\mathbf{x} F = \lim_{s \rightarrow 0} \left\{ \int_{\mathbb{R}^3 \setminus \bigcup_{A=1}^N B_A(s)} d^3\mathbf{x} F + 4\pi \sum_{A=1}^N D_A(s) \right\}. \quad (84)$$

The first term integrates over the domain $\mathbb{R}^3 \setminus \bigcup_{A=1}^N B_A(s)$ defined as \mathbb{R}^3 deprived from the N spherical balls $B_A(s) \equiv \{\mathbf{x}; r_A \leq s\}$ of radius s and centered on the points \mathbf{y}_A . The second term is the opposite of the sum of divergent parts associated with the first term around each of the particles in the limit where $s \rightarrow 0$. We have

$$D_A(s) = \sum_{p=p_0}^{-4} \frac{s^{p+3}}{p+3} \langle f_p \rangle_A + \ln \left(\frac{s}{r_A} \right) \langle f_{-3} \rangle_A. \quad (85)$$

Since the divergent parts are canceled (by definition) the Hadamard partie finie is obtained in the limit $s \rightarrow 0$. Notice that as indicated in Eq. 84 the Hadamard partie finie integral is not fully specified: it depends on N strictly positive and a priori arbitrary constants s_1, \dots, s_N parametrizing the logarithms in Eq. 85.

We have seen that the post-Newtonian scheme consists of breaking the hyperbolic d'Alembertian operator \square into the elliptic Laplacian Δ and the retardation term $c^{-2}\partial_t^2$ considered to be small, and put in the right-hand-side of the equation where it can be iterated; see Eq. 18. We thus have to deal with the regularization of Poisson integrals, or iterated Poisson integrals, of some generic function F . The Poisson integral will be divergent¹¹ and we apply the prescription (84). Thus,

$$P(\mathbf{x}') = -\frac{1}{4\pi} \text{Pf}_{s_1 \dots s_N} \int \frac{d^3\mathbf{x}}{|\mathbf{x} - \mathbf{x}'|} F(\mathbf{x}). \quad (86)$$

This definition is valid for each field point \mathbf{x}' *different* from the \mathbf{y}_A 's, and we want to investigate the singular limit when \mathbf{x}' tends to one of the source points \mathbf{y}_A , so as to define the object $(P)_A$. The definition (83) is not directly applicable because the expansion of the Poisson integral $P(\mathbf{x}')$ when $\mathbf{x}' \rightarrow \mathbf{y}_A$ will involve besides the normal powers of $r'_A \equiv |\mathbf{x}' - \mathbf{y}_A|$ some *logarithms* of r'_A . The proper way to define the Hadamard partie finie in this case is to include the $\ln r'_A$ into the definition (83) as if it were a mere constant parameter. With this definition we arrive at [20]

$$(P)_A = -\frac{1}{4\pi} \text{Pf}_{s_1 \dots s_N} \int \frac{d^3\mathbf{x}}{r_A} F(\mathbf{x}) + \left[\ln \left(\frac{r'_A}{s_A} \right) - 1 \right] \langle f_{-2} \rangle_A. \quad (87)$$

¹¹ We consider only the local divergencies due to the singular points \mathbf{y}_A . The problem of divergencies of Poisson integrals at infinity is part of the general post-Newtonian formalism and has been treated in Section 2.2.

The first term is given by a partie-finie integral following the definition (84); the second involves the logarithm of r'_A . The constants s_1, \dots, s_N come from Eq. 86. Since r'_A is actually tending to zero, $\ln r'_A$ represents a formally infinite “constant,” which will ultimately parametrize the final Hadamard regularized 3PN equations of motion. In the two-body case we shall find that the constants r'_A are unphysical in the sense that they can be removed by a coordinate transformation [21]. Note that the apparent dependence of Eq. 87 on the constant s_A is illusory. Indeed the dependence on s_A cancels out between the first and the second terms in the right-hand-side of Eq. 87, so the result depends only on r'_A and the s_B 's for $B \neq A$. We thus have a simpler rewriting of Eq. 87 as

$$(P)_A = -\frac{1}{4\pi} \text{Pf}_{s_1, \dots, r'_A, \dots, s_N} \int \frac{d^3 \mathbf{x}}{r_A} F(\mathbf{x}) - \langle f_{-2} \rangle_A. \quad (88)$$

Unfortunately, the constants s_B for $B \neq A$ remaining in the result (88) will be the source of a genuine ambiguity. This ambiguity can in fact be traced back to the so-called non-distributivity of the Hadamard partie finie, a consequence of the presence of the angular integration in Eq. 83, and implying that $(FG)_A \neq (F)_A(G)_A$ in general. The non-distributivity arises precisely at the 3PN order both in the equations of motion and radiation field of point-mass binaries. At that order we are loosing with Hadamard's regularization an elementary rule of ordinary calculus. Consequently we expect that some basic symmetries of general relativity such as diffeomorphism invariance will be lost. However Hadamard's regularization can still be efficiently used to compute most of the terms in the equations of motion and radiation field at 3PN order; only a few ambiguous terms will show up that have then to be determined by another method.

3.3 Dimensional Regularization

Dimensional regularization is an extremely powerful regularization, which is free of ambiguities (at least up to the 3PN order). The main reason is that it is able to preserve the symmetries of classical general relativity; in fact dimensional regularization was invented [27, 93] as a means to preserve the gauge symmetry of perturbative quantum field theories. In the present context we shall show that dimensional regularization permits to resolve the problem of ambiguities arising at the 3PN order in Hadamard's regularization. We shall employ dimensional regularization not merely as a trick to compute some particular integrals that would otherwise be divergent or ambiguous, but as a fundamental tool for solving in a consistent way the Einstein field equations with singular sources. We therefore assume that the correct theory is general relativity in $D = d + 1$ space-time dimensions. As usual, any intermediate formulas will be interpreted by analytic continuation for a general complex spatial dimension $d \in \mathbb{C}$. In particular we shall analytically continue d down to the value of interest 3 and pose

$$d = 3 + \varepsilon. \quad (89)$$

The Einstein field equations in d spatial dimensions take the same form as presented in Section 2.1, with the exception that the explicit expression of the gravitational source term $\Lambda^{\alpha\beta}$ now depends on d . We find that only the last term in Eq. 11 acquires a dependence on d ; namely, the factor $\frac{1}{2}$ in $g_{\rho\sigma}g_{\varepsilon\pi} - \frac{1}{2}g_{\sigma\varepsilon}g_{\rho\pi}$ should now read $\frac{1}{d-1}$. In addition the d -dimensional gravitational constant is related to the usual three-dimensional Newton constant G by

$$G^{(d)} = G \ell_0^\varepsilon, \quad (90)$$

where ℓ_0 is a characteristic length associated with dimensional regularization.

In the post-Newtonian iteration performed in d dimensions we shall meet the analogue of the function F , which we denote by $F^{(d)}(\mathbf{x})$ where $\mathbf{x} \in \mathbb{R}^d$. It turns out that in the vicinity of the singular points \mathbf{y}_A , the function $F^{(d)}$ admits an expansion richer than in Eq. 82, and of the type

$$F^{(d)}(\mathbf{x}) = \sum_{p=p_0}^P \sum_{q=q_0}^{q_1} r_A^{p+q\varepsilon} f_{p,q}^{(\varepsilon)}(\mathbf{n}_A) + o(r_A^P). \quad (91)$$

The coefficients $f_{p,q}^{(\varepsilon)}(\mathbf{n}_1)$ depend on the dimension through $\varepsilon \equiv d-3$ and also on the scale ℓ_0 . The powers of r_A are now of the type $p+q\varepsilon$ where the two relative integers $p, q \in \mathbb{Z}$ have values limited as indicated. Because $F^{(d)}$ reduces to F when $\varepsilon = 0$ we necessarily have the constraints

$$\sum_{q=q_0}^{q_1} f_{p,q}^{(0)} = f_A^p. \quad (92)$$

To proceed with the iteration we need the Green function of the Laplace operator in d dimensions. Its explicit form is

$$u_A^{(d)} = K r_A^{2-d}. \quad (93)$$

It satisfies $\Delta u_A^{(d)} = -4\pi \delta_A^{(d)}$, where $\delta_A^{(d)} \equiv \delta^{(d)}(\mathbf{x}-\mathbf{y}_A)$ denotes the d -dimensional Dirac function. The constant K is given by

$$K = \frac{\Gamma\left(\frac{d-2}{2}\right)}{\pi^{\frac{d-2}{2}}}, \quad (94)$$

where Γ is the usual Eulerian function. It reduces to 1 when $d \rightarrow 3$. Note that the volume Ω_{d-1} of the sphere with $d-1$ dimensions is related to K by

$$\Omega_{d-1} = \frac{4\pi}{(d-2)K}. \quad (95)$$

With these results the Poisson integral of $F^{(d)}$, constituting the d -dimensional analogue of Eq. 86, reads

$$P^{(d)}(\mathbf{x}') = -\frac{K}{4\pi} \int \frac{d^d \mathbf{x}}{|\mathbf{x} - \mathbf{x}'|^{d-2}} F^{(d)}(\mathbf{x}). \tag{96}$$

In dimensional regularization the singular behavior of this integral is automatically taken care of by analytic continuation in d . Next we evaluate the integral at the singular point $\mathbf{x}' = \mathbf{y}_A$. In contrast with Hadamard's regularization where the result was given by Eq. 87, in dimensional regularization this is quite easy, as we are allowed to simply replace \mathbf{x}' by \mathbf{y}_A into the explicit integral form Eq. 96. So we simply have

$$P^{(d)}(\mathbf{y}_A) = -\frac{K}{4\pi} \int \frac{d^d \mathbf{x}}{r_A^{d-2}} F^{(d)}(\mathbf{x}). \tag{97}$$

The main step of our strategy [18, 44] will now consist of computing the *difference* between the d -dimensional Poisson potential (97) and its three-dimensional counterpart which is defined from Hadamard's regularization as Eq. 87. We shall then add this difference (in the limit $\varepsilon = d - 3 \rightarrow 0$) to the result obtained by Hadamard regularization in order to get the corresponding dimensional regularization result. This strategy is motivated by the fact that as already mentioned most of the terms do not present any problems and have already been correctly computed using Hadamard's regularization. Denoting the difference between the two regularizations by means of the script letter \mathcal{D} , we write

$$\mathcal{D}(P)_A \equiv P^{(d)}(\mathbf{y}_A) - (P)_A. \tag{98}$$

We shall only compute the first two terms of the Laurent expansion of $\mathcal{D}(P)_A$ when $\varepsilon \rightarrow 0$, which will be of the form $\mathcal{D}(P)_A = a_{-1} \varepsilon^{-1} + a_0 + \mathcal{O}(\varepsilon)$. This is the information needed to determine the value of the ambiguity parameters. Notice that the difference $\mathcal{D}(P)_A$ comes exclusively from the contribution of terms developing some *poles* $\propto 1/\varepsilon$ in the d -dimensional calculation. The ambiguity in Hadamard's regularization at 3PN order is reflected by the appearance of poles in d dimensions. The point is that in order to obtain the difference $\mathcal{D}(P)_A$ we do not need the expression of $F^{(d)}$ for an arbitrary source point \mathbf{x} but only in the vicinity of the singular points \mathbf{y}_A . Thus this difference depends only on the singular coefficients of the local expansions of $F^{(d)}$ near the singularities. We find [18]

$$\begin{aligned} \mathcal{D}(P)_A = & -\frac{1}{\varepsilon(1+\varepsilon)} \sum_{q=q_0}^{q_1} \left[\left(\frac{1}{q} + \varepsilon [\ln r'_A - 1] \right) \langle f_A^{(\varepsilon)} \rangle_{-2,q} \right. \\ & \left. + \sum_{B \neq A} \left(\frac{1}{q+1} + \varepsilon \ln s_B \right) \sum_{\ell=0}^{+\infty} \frac{(-)^\ell}{\ell!} \partial_L \left(\frac{1}{r_{AB}^{1+\varepsilon}} \right) \langle n_B^L f_B^{(\varepsilon)} \rangle_{-\ell-3,q} \right] + \mathcal{O}(\varepsilon). \end{aligned} \tag{99}$$

We still use the bracket notation to denote the angular average but this time performed in d dimensions, that is,

$$\langle f_{p,q}^{(\varepsilon)} \rangle_A \equiv \int \frac{d\Omega_{d-1}(\mathbf{n}_A)}{\Omega_{d-1}} f_{p,q}^{(\varepsilon)}(\mathbf{n}_A). \quad (100)$$

The above differences for all the Poisson and iterated Poisson integrals composing the equations of motion (i.e., the accelerations of the point masses) are added to the corresponding results of the Hadamard regularization in the variant of it called the “pure Hadamard-Schwartz” regularization (see [18] for more details). In this way we find that the equations of motion in dimensional regularization are composed of a pole part $\propto 1/\varepsilon$ that is purely 3PN, followed by a finite part when $\varepsilon \rightarrow 0$, plus the neglected terms $\mathcal{O}(\varepsilon)$. It has been shown (in the two-body case $N = 2$) that:

1. The pole part $\propto 1/\varepsilon$ of the accelerations can be renormalized into some *shifts* of the “bare” world-lines by $\mathbf{y}_A \rightarrow \mathbf{y}_A + \boldsymbol{\xi}_A$, with $\boldsymbol{\xi}_A$ containing the poles, so that the result expressed in terms of the “dressed” world-lines is *finite* when $\varepsilon \rightarrow 0$.
2. The renormalized acceleration is physically equivalent to the result of the Hadamard regularization, in the sense that it differs from it only by the effects of shifts $\boldsymbol{\xi}_A$, if and only if the ambiguities in the Hadamard regularization are fully and uniquely determined (i.e., take specific values).

These results [18] provide an unambiguous determination of the equations of motion of compact binaries up to the 3PN order. A related strategy with similar complete results has been applied to the problem of multipole moments and radiation field of point-mass binaries [19]. This finally completed the derivation of the general relativistic prediction for compact binary inspiral up to 3PN order (and even to 3.5PN order). In later sections we shall review some features of the 3.5PN gravitational-wave templates of inspiralling compact binaries.

Why should the final results of the employed regularization scheme be unique, in agreement with our expectation that the problem is well posed and should possess a unique physical answer? The results can be justified by invoking the “effacing principle” of general relativity [40] – namely, the internal structure of the compact bodies does not influence the equations of motion and emitted radiation until a very high post-Newtonian order. Only the masses m_A of the bodies should drive the motion and radiation, and not for instance their “compactness” $Gm_A/(c^2 a_A)$. A model of point masses should therefore give the correct physical answer, which we expect to be also valid for black holes, provided that the regularization scheme is mathematically consistent.

3.4 Energy and Flux of Compact Binaries

The equations of motion of compact binary sources, up to the highest known post-Newtonian order that is 3.5PN, will serve in the definition of the gravitational-wave templates for two purposes:

1. To compute the center-of-mass energy E appearing in the left-hand-side of the energy balance equation to be used for deducing the orbital phase,

$$\frac{dE}{dt} = -\mathcal{F}. \quad (101)$$

2. To *order-reduce* the accelerations coming from the time derivatives of the source multipole moments required to compute the gravitational-wave energy flux \mathcal{F} in the right-hand-side of the balance equation.

We consider two compact objects moving under purely gravitational mutual interaction. In a first stage we assume that the bodies are non-spinning so the motion takes place in a fixed plane, say the x-y plane. The relative position $\mathbf{x} = \mathbf{y}_1 - \mathbf{y}_2$, velocity $\mathbf{v} = d\mathbf{x}/dt$, and acceleration $\mathbf{a} = d\mathbf{v}/dt$ are given by

$$\mathbf{x} = r \mathbf{n}, \quad (102a)$$

$$\mathbf{v} = \dot{r} \mathbf{n} + r \omega \boldsymbol{\lambda}, \quad (102b)$$

$$\mathbf{a} = (\ddot{r} - r \omega^2) \mathbf{n} + (r \dot{\omega} + 2\dot{r} \omega) \boldsymbol{\lambda}, \quad (102c)$$

The orbital frequency ω is related in the usual way to the orbital phase ϕ by $\omega = \dot{\phi}$ (time derivatives are denoted with a dot). Here the vector $\boldsymbol{\lambda} = \hat{\mathbf{z}} \times \mathbf{n}$ is perpendicular to the unit vector $\hat{\mathbf{z}}$ along the z-direction orthogonal to the orbital plane, and to the binary's separation unit direction $\mathbf{n} \equiv \mathbf{x}/r$.

Through 3PN order, it is possible to model the binary's orbit as a *quasi-circular* orbit decaying by the effect of radiation reaction at the 2.5PN order. The restriction to quasi-circular orbits is both to simplify the presentation,¹² and for physical reasons because the orbit of inspiralling compact binaries detectable by current detectors should be circular (see the discussion in Section 1). The radiation-reaction effect at 2.5PN order yields¹³

$$\dot{r} = -\frac{64}{5} \sqrt{\frac{Gm}{r}} v \gamma^{5/2} + \mathcal{O}\left(\frac{1}{c^7}\right), \quad (103a)$$

$$\dot{\omega} = \frac{96}{5} \frac{Gm}{r^3} v \gamma^{5/2} + \mathcal{O}\left(\frac{1}{c^7}\right), \quad (103b)$$

where γ is defined as the small [i.e., $\gamma = \mathcal{O}(c^{-2})$] post-Newtonian parameter

$$\gamma \equiv \frac{Gm}{rc^2}. \quad (104)$$

¹² However the 3PN equations of motion are known in an arbitrary frame and for general orbits.

¹³ Mass parameters are the total mass $m \equiv m_1 + m_2$, the symmetric mass ratio $\nu \equiv m_1 m_2 / m^2$ satisfying $0 < \nu \leq 1/4$, and for later use the mass difference ratio $\Delta \equiv (m_1 - m_2)/m$.

Substituting these results into Eq. 102, we obtain the expressions for the velocity and acceleration during the inspiral,

$$\mathbf{v} = r \omega \boldsymbol{\lambda} - \frac{64}{5} \sqrt{\frac{Gm}{r}} v \gamma^{5/2} \mathbf{n} + \mathcal{O}\left(\frac{1}{c^7}\right), \quad (105a)$$

$$\mathbf{a} = -\omega^2 \mathbf{x} - \frac{32}{5} \sqrt{\frac{Gm}{r^3}} v \gamma^{5/2} \mathbf{v} + \mathcal{O}\left(\frac{1}{c^7}\right). \quad (105b)$$

Notice that while $\dot{r} = \mathcal{O}(c^{-5})$, we have $\ddot{r} = \mathcal{O}(c^{-10})$, which is of the order of the *square* of radiation-reaction effects and is thus zero with the present approximation.

A central result of post-Newtonian calculations is the expression of the orbital frequency ω in terms of the binary's separation r up to 3PN order. This result has been obtained independently by three groups. Two are working in harmonic coordinates: Blanchet and Faye [18, 21, 46] use a direct post-Newtonian iteration of the equations of motion, while Itoh and Futamase [66–68] apply a variant of the surface-integral approach (à la Einstein-Infeld-Hoffmann [52]) valid for compact bodies without the need of a self-field regularization. The group of Jaranowski and Schäfer [44, 69, 70] employs Arnowitt–Deser–Misner coordinates within the Hamiltonian formalism of general relativity.¹⁴ The 3PN orbital frequency in harmonic coordinates is

$$\begin{aligned} \omega^2 = \frac{Gm}{r^3} & \left\{ 1 + (-3 + \nu)\gamma + \left(6 + \frac{41}{4}\nu + \nu^2\right)\gamma^2 \right. \\ & \left. + \left(-10 + \left[-\frac{75707}{840} + \frac{41}{64}\pi^2 + 22 \ln\left(\frac{r}{r'_0}\right)\right]\nu + \frac{19}{2}\nu^2 + \nu^3\right)\gamma^3 + \mathcal{O}\left(\frac{1}{c^8}\right) \right\}. \end{aligned} \quad (106)$$

Note the logarithm at 3PN order coming from a Hadamard self-field regularization scheme, and depending on a constant r'_0 defined by $m \ln r'_0 \equiv m_1 \ln r'_1 + m_2 \ln r'_2$, where $r'_A \equiv |\mathbf{x}' - \mathbf{y}_A|$ are arbitrary “constants” discussed in Section 3.2. We shall see that r'_0 disappears from final results – it can be qualified as a gauge constant.

To obtain the 3PN energy we need to go back to the equations of motion for general noncircular orbits, and deduce the energy as the integral of the motion associated with a Lagrangian formulation of (the conservative part of) these equations [46]. Once we have the energy for general orbits we can reduce it to quasi-circular orbits. We find

$$\begin{aligned} E = -\frac{Gm^2\nu}{2r} & \left\{ 1 + \left(-\frac{7}{4} + \frac{1}{4}\nu\right)\gamma + \left(-\frac{7}{8} + \frac{49}{8}\nu + \frac{1}{8}\nu^2\right)\gamma^2 \right. \\ & \left. + \left(-\frac{235}{64} + \left[\frac{46031}{2240} - \frac{123}{64}\pi^2 + \frac{22}{3} \ln\left(\frac{r}{r'_0}\right)\right]\nu \right. \right. \\ & \left. \left. + \frac{27}{32}\nu^2 + \frac{5}{64}\nu^3\right)\gamma^3 + \mathcal{O}\left(\frac{1}{c^8}\right) \right\}. \end{aligned} \quad (107)$$

¹⁴ This approach is extensively reviewed in the contribution of Gerhard Schäfer in this volume.

A convenient post-Newtonian parameter $x = \mathcal{O}(c^{-2})$ is now used in place of γ ; it is defined from the orbital frequency as

$$x = \left(\frac{G m \omega}{c^3} \right)^{2/3}. \quad (108)$$

The interest in this parameter stems from its invariant meaning in a large class of coordinate systems including the harmonic and ADM coordinates. By inverting Eq. 106 we find at 3PN order

$$\begin{aligned} \gamma = x \left\{ 1 + \left(1 - \frac{\nu}{3} \right) x + \left(1 - \frac{65}{12} \nu \right) x^2 \right. \\ \left. + \left(1 + \left[-\frac{2203}{2520} - \frac{41}{192} \pi^2 - \frac{22}{3} \ln \left(\frac{r}{r'_0} \right) \right] \nu + \frac{229}{36} \nu^2 + \frac{\nu^3}{81} \right) x^3 + \mathcal{O} \left(\frac{1}{c^8} \right) \right\}. \end{aligned} \quad (109)$$

This is substituted back into Eq. 107 to get the 3PN energy in invariant form. We happily observe that the logarithm and the gauge constant r'_0 cancel out in the process and our final result is

$$\begin{aligned} E = -\frac{m\nu c^2 x}{2} \left\{ 1 + \left(-\frac{3}{4} - \frac{1}{12} \nu \right) x + \left(-\frac{27}{8} + \frac{19}{8} \nu - \frac{1}{24} \nu^2 \right) x^2 \right. \\ \left. + \left(-\frac{675}{64} + \left[\frac{34445}{576} - \frac{205}{96} \pi^2 \right] \nu - \frac{155}{96} \nu^2 - \frac{35}{5184} \nu^3 \right) x^3 + \mathcal{O} \left(\frac{1}{c^8} \right) \right\}. \end{aligned} \quad (110)$$

The conserved energy E corresponds to the Newtonian, 1PN, 2PN, and 3PN conservative orders in the equations of motion; the damping part is associated with radiation reaction and arises at 2.5PN order. The radiation reaction at the dominant 2.5PN level will correspond to the “Newtonian” gravitational-wave flux only. Hence the flying-color 3PN flux \mathcal{F} we are looking for cannot be computed from the 3PN equations of motion alone. Instead we have to apply all the machinery of the post-Newtonian wave generation formalism described in Sections 2.3–2.6. The final result at 3.5PN order is [11, 13, 19]

$$\begin{aligned} \mathcal{F} = \frac{32c^5}{5G} \nu^2 x^5 \left\{ 1 + \left(-\frac{1247}{336} - \frac{35}{12} \nu \right) x + 4\pi x^{3/2} \right. \\ + \left(-\frac{44711}{9072} + \frac{9271}{504} \nu + \frac{65}{18} \nu^2 \right) x^2 + \left(-\frac{8191}{672} - \frac{583}{24} \nu \right) \pi x^{5/2} \\ + \left[\frac{6643739519}{69854400} + \frac{16}{3} \pi^2 - \frac{1712}{105} C - \frac{856}{105} \ln(16x) \right. \\ \left. + \left(-\frac{134543}{7776} + \frac{41}{48} \pi^2 \right) \nu - \frac{94403}{3024} \nu^2 - \frac{775}{324} \nu^3 \right] x^3 \\ \left. + \left(-\frac{16285}{504} + \frac{214745}{1728} \nu + \frac{193385}{3024} \nu^2 \right) \pi x^{7/2} + \mathcal{O} \left(\frac{1}{c^8} \right) \right\}. \end{aligned} \quad (111)$$

Here $C = 0.577\cdots$ is the Euler constant. This result is fully consistent with black-hole perturbation theory: using it Sasaki and Tagoshi [84, 88, 89] obtain Eq. 111 in the small mass-ratio limit $\nu \rightarrow 0$. The generalization of Eq. 111 to arbitrary eccentric (bound) orbits has also been worked out [4, 5].

3.5 *Waveform of Compact Binaries*

We specify our conventions for the orbital phase and polarization vectors defining the polarization waveforms (52) in the case of a non-spinning compact binary moving on a quasi-circular orbit. If the orbital plane is chosen to be the x-y plane as in Section 3.4, with the orbital phase ϕ measuring the direction of the unit separation vector $\mathbf{n} = \mathbf{x}/r$, then

$$\mathbf{n} = \hat{\mathbf{x}} \cos \phi + \hat{\mathbf{y}} \sin \phi, \quad (112)$$

where $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are the unit directions along x and y. Following [3, 14] we choose the polarization vector \mathbf{P} to lie along the x-axis and the observer to be in the y-z plane in the direction

$$\mathbf{N} = \hat{\mathbf{y}} \sin i + \hat{\mathbf{z}} \cos i, \quad (113)$$

where i is the orbit's inclination angle. With this choice \mathbf{P} lies along the intersection of the orbital plane with the plane of the sky in the direction of the *ascending node*, that is, that point at which the bodies cross the plane of the sky moving toward the observer. Hence the orbital phase ϕ is the angle between the ascending node and the direction of body 1. The rotating orthonormal triad $(\mathbf{n}, \boldsymbol{\lambda}, \hat{\mathbf{z}})$ describing the motion of the binary and used in Eq. 102 is related to the fixed polarization triad $(\mathbf{N}, \mathbf{P}, \mathbf{Q})$ by

$$\mathbf{n} = \mathbf{P} \cos \phi + (\mathbf{Q} \cos i + \mathbf{N} \sin i) \sin \phi, \quad (114a)$$

$$\boldsymbol{\lambda} = -\mathbf{P} \sin \phi + (\mathbf{Q} \cos i + \mathbf{N} \sin i) \cos \phi, \quad (114b)$$

$$\hat{\mathbf{z}} = -\mathbf{Q} \sin i + \mathbf{N} \cos i. \quad (114c)$$

The 3.5PN expression of the orbital phase ϕ as function of the orbital frequency or equivalently the x -parameter is obtained from the energy balance equation (101) in which the binary's conservative center-of-mass energy E and total gravitational-wave flux \mathcal{F} have been obtained in Eqs. 110–111. For circular orbits the orbital phase is computed from

$$\phi \equiv \int \omega dt = - \int \frac{\omega}{\mathcal{F}} \frac{dE}{d\omega} d\omega. \quad (115)$$

Various methods (numerical or analytical) are possible for solving Eq. 115 given the expressions (110) and (111). This yields different waveform families all valid at the same 3.5PN order, but which may differ when extrapolated beyond the normal domain of validity of the post-Newtonian expansion, that is, in this case very near the coalescence. Such differences must be taken into account when comparing the post-Newtonian waveforms to numerical results [29].

It is convenient to perform a change of phase, from the actual orbital phase ϕ to the new phase variable

$$\psi = \phi - \frac{2G M \omega}{c^3} \ln \left(\frac{\omega}{\omega_0} \right), \quad (116)$$

where M is the binary's total mass monopole moment¹⁵ and $\omega_0 = \frac{1}{4u_0} \exp[\frac{11}{12} - C]$ is related to the constant $u_0 \equiv r_0/c$ entering the tail integrals in Eq. 60. The logarithmic term in ψ corresponds physically to some spreading of the different frequency components of the wave along the line of sight from the source to the detector, and expresses the tail effect as a small delay in the arrival time of gravitational waves. This effect, although of formal 1.5PN order in Eq. 116, represents in fact a very small modulation of the orbital phase: compared to the dominant phase evolution whose order is that of the inverse of 2.5PN radiation reaction, this modulation is of order 4PN and can thus be neglected with the present accuracy.

The spherical harmonic modes of the polarization waveforms can now be obtained at 3PN order using the angular integration formula (55). We start from the expressions of the wave polarizations h_+ and h_\times as functions of the inclination angle i and of the phase ψ . We use the known dependence of the spherical harmonics (54) on the azimuthal angle. Denoting $h \equiv h_+ - ih_\times = h(i, \psi)$ we find that the latter angular integration becomes

$$h^{\ell m} = (-i)^m e^{-im\psi} \int_0^{2\pi} d\psi' \int_0^\pi di \sin i h(i, \psi') Y_{-2}^{\ell m}(i, \psi'), \quad (117)$$

exhibiting the azimuthal factor $e^{-im\psi}$ appropriate for each mode. Let us introduce a normalized mode coefficient $H^{\ell m}$ starting by definition with one at the Newtonian order for the dominant mode having $(\ell, m) = (2, 2)$. This means posing

$$h^{\ell m} = \frac{2G m v x}{R c^2} \sqrt{\frac{16\pi}{5}} H^{\ell m} e^{-im\psi}. \quad (118)$$

All the modes have been given in [23] up to the 3PN order. The dominant mode $(2, 2)$, which is primarily needed for the comparison between post-Newtonian calculations and numerical simulations, reads at 3PN order

$$\begin{aligned} H^{22} = & 1 + \left(-\frac{107}{42} + \frac{55}{42}v \right) x + 2\pi x^{3/2} \\ & + \left(-\frac{2173}{1512} - \frac{1069}{216}v + \frac{2047}{1512}v^2 \right) x^2 + \left(-\frac{107}{21}\pi + \frac{34}{21}\pi v - 24iv \right) x^{5/2} \end{aligned}$$

¹⁵ The mass monopole M differs from $m = m_1 + m_2$ as it includes the contribution of the gravitational binding energy. At 1PN order it is given for circular orbits by $M = m[1 - \frac{v}{2}\gamma] + \mathcal{O}(c^{-4})$.

$$\begin{aligned}
& + \left(\frac{27027409}{646800} - \frac{856}{105}C + \frac{2}{3}\pi^2 - \frac{428}{105}\ln(16x) + \frac{428}{105}i\pi \right. \\
& \quad \left. + \left[-\frac{278185}{33264} + \frac{41}{96}\pi^2 \right]v - \frac{20261}{2772}v^2 + \frac{114635}{99792}v^3 \right)x^3 + \mathcal{O}\left(\frac{1}{c^7}\right).
\end{aligned} \tag{119}$$

3.6 Spin–Orbit Contributions in the Energy and Flux

To successfully detect the gravitational waves emitted by spinning, precessing binaries, and to estimate the binary parameters, spin effects should be included in the templates. For maximally spinning compact bodies the spin–orbit coupling (linear in the spins) appears dominantly at the 1.5PN order, while the spin–spin one (which is quadratic) appears at 2PN order. The spin effect on the free motion of a test particle was obtained by Papapetrou [75] in the form of a coupling to curvature. Seminal later works by Barker and O’Connell [7, 8] obtained the leading order spin–orbit and spin–spin contributions in the post-Newtonian equations of motion. Based on these works, the spin–orbit and spin–spin terms were obtained in the radiation field [59, 63, 65, 72], enabling the derivation of the orbital phase evolution (the crucial quantity that determines the templates). Finding the 1PN corrections to the leading spin–orbit coupling in both the (translational) equations of motion and radiation field was begun in [73, 87] and completed in [13, 53]. The result [53] for the equations of motion was confirmed by an alternative derivation based on the ADM-Hamiltonian formalism [45].

In Section 3.1 we discussed a covariant formalism for spinning particles [6, 49, 96–98]. We want now to find a convenient three-dimensional variable for the spin. Restricting ourselves to spin–orbit effects, that is, neglecting $\mathcal{O}(S^2)$, we can write the components of the spin tensor $S_A^{\alpha\beta}$ as

$$S_A^{0i} = -\frac{1}{c\sqrt{-(g)_A}} \varepsilon^{ijk} u_j^A S_k^A, \tag{120a}$$

$$S_A^{ij} = -\frac{1}{c\sqrt{-(g)_A}} \varepsilon^{ijk} \left[u_0^A S_k^A + u_k^A \frac{v_A^j}{c} S_l^A \right]. \tag{120b}$$

We have used the momentum–velocity relation (76) and have taken into account the spin condition (73). A first possibility is to adopt as the basic spin variable the *contravariant* components of the spin covector S_i^A in Eq. 120, which are obtained by raising the index by means of the contravariant spatial metric, viz

$$\mathbf{S}_A = (S_A^i) \quad \text{with} \quad S_A^i \equiv (\gamma^{ij})_A S_j^A, \tag{121}$$

where γ^{ij} is the inverse of the covariant spatial metric $\gamma_{ij} \equiv g_{ij}$ that is, satisfies $\gamma^{ik}\gamma_{kj} = \delta_j^i$. The choice of spin variable (121) has been adopted in [73, 87]. However, to express final results (to be used in gravitational-wave templates) it is better

to use a different set of spin variables characterized by having some *conserved* Euclidean lengths. Such spins with constant Euclidean magnitude will be denoted by \mathbf{S}_A^c . They can be computed in a straightforward way at a given post-Newtonian order in terms of the previous variables (121). For instance we find up to 1PN order,

$$\mathbf{S}_A^c = \left[1 + \frac{(U)_A}{c^2} \right] \mathbf{S}_A - \frac{1}{2c^2} (\mathbf{v}_A \cdot \mathbf{S}_A) \mathbf{v}_A + \mathcal{O} \left(\frac{1}{c^4} \right). \quad (122)$$

The (regularized) gravitational potential $(U)_A$ is defined by Eq. 81. The constant-magnitude spin variable \mathbf{S}_A^c obeys a spin precession equation that is necessarily of the form

$$\frac{d\mathbf{S}_A^c}{dt} = \boldsymbol{\Omega}_A \times \mathbf{S}_A^c. \quad (123)$$

Indeed this equation implies that $|\mathbf{S}_A^c| = \text{const}$. The precession angular frequency vector $\boldsymbol{\Omega}_A$ for two-body systems has been computed for the leading spin-orbit and spin-spin contributions [63] and for the 1PN correction to the spin-orbit [13,45,53]. For two bodies we conveniently use the following combinations (introduced in [63]) of the two spins:

$$\mathbf{S}^c \equiv \mathbf{S}_1^c + \mathbf{S}_2^c, \quad (124a)$$

$$\boldsymbol{\Sigma}^c \equiv m \left[\frac{\mathbf{S}_2^c}{m_2} - \frac{\mathbf{S}_1^c}{m_1} \right]. \quad (124b)$$

Furthermore, recalling the orthonormal triad $\{\mathbf{n}, \boldsymbol{\lambda}, \hat{\mathbf{z}}\}$ used in Section 3.4, where $\hat{\mathbf{z}}$ is the unit vector in the direction perpendicular to the orbital plane, we denote by $S_z^c \equiv \mathbf{S}^c \cdot \hat{\mathbf{z}}$ and $\Sigma_z^c \equiv \boldsymbol{\Sigma}^c \cdot \hat{\mathbf{z}}$ the projections along that perpendicular direction.

The spin-orbit terms have been computed at 1PN order both in the equations of motion and in the radiation field. In the equations of motion they correct the orbital frequency and invariant conserved energy with terms at orders 1.5PN and 2.5PN. In the presence of spins the energy gets modified to $E = E_{\text{mono}} + E_{\text{spin}}$ where the monopole part has been obtained in Eq. 110 and where the spin terms read

$$E_{\text{spin}} = -\frac{c^2}{2Gm} \nu x \left\{ x^{3/2} \left[\frac{14}{3} S_z^c + 2\Delta \Sigma_z^c \right] + x^{5/2} \left[\left(11 - \frac{61}{9} \nu \right) S_z^c + \left(3 - \frac{10}{3} \nu \right) \Delta \Sigma_z^c \right] + \mathcal{O} \left(\frac{1}{c^6} \right) \right\}. \quad (125)$$

We recall that $\Delta \equiv (m_1 - m_2)/m$; see footnote 13. This expression is valid for quasi-circular orbits, and we neglect the spin-spin terms. Similarly the gravitational-wave flux will be modified at the same 1.5PN and 2.5PN orders. Posing $\mathcal{F} = \mathcal{F}_{\text{mono}} + \mathcal{F}_{\text{spin}}$ where $\mathcal{F}_{\text{mono}}$ is given by Eq. 111, we find

$$\mathcal{F}_{\text{spin}} = \frac{32c^5}{5G^2 m^2} v^2 x^5 \left\{ x^{3/2} \left[-4S_z^c - \frac{5}{4} \Delta \Sigma_z^c \right] + x^{5/2} \left[\left(-\frac{9}{2} + \frac{272}{9} v \right) S_z^c + \left(-\frac{13}{16} + \frac{43}{4} v \right) \Delta \Sigma_z^c \right] + \mathcal{O} \left(\frac{1}{c^6} \right) \right\}. \quad (126)$$

Having in hand the spin contributions in E and \mathcal{F} , we can deduce the evolution of the orbital phase from the energy balance equation (101). In absence of precession of the orbital plane, for example, for spins aligned or anti-aligned with the orbital angular momentum, the gravitational-wave phase will reduce to the “carrier” phase $\phi_{\text{GW}} \equiv 2\phi$ (keeping only the dominant harmonics), where ϕ is the orbital phase that is obtained by integrating the orbital frequency. However, in the general case of non-aligned spins, we must take into account the effect of precession of the orbital plane induced by spin modulations. Then the gravitational-wave phase is given by $\Phi_{\text{GW}} = \phi_{\text{GW}} + \delta\phi_{\text{GW}}$, where the precessional correction $\delta\phi_{\text{GW}}$ arises from the changing orientation of the orbital plane, and can be computed by standard methods using numerical integration [2]. Thus, the carrier phase ϕ_{GW} constitutes the main theoretical output to be provided for the gravitational-wave templates, and can directly be obtained numerically from using the integration formula (115).

References

1. J.L. Anderson, T.C. DeCanio, *Gen. Rel. Grav.* **6**, 197 (1975)
2. T.A. Apostolatos, C. Cutler, G.J. Sussman, K.S. Thorne, *Phys. Rev. D* **49**, 6274 (1994)
3. K.G. Arun, L. Blanchet, B.R. Iyer, M.S.S. Qusailah, *Class. Q. Grav.* **21**, 3771 (2004); Erratum, *ibidem* **22**, 3115 (2005)
4. K.G. Arun, L. Blanchet, B.R. Iyer, M.S.S. Qusailah, *Phys. Rev. D* **77**, 064034 (2008)
5. K.G. Arun, L. Blanchet, B.R. Iyer, M.S.S. Qusailah, *Phys. Rev. D* **77**, 064035 (2008)
6. I. Bailey, W. Israel, *Ann. Phys. (N.Y.)* **130**, 188 (1980)
7. B.M. Barker, R.F. O’Connell, *Phys. Rev. D* **12**, 329 (1975)
8. B.M. Barker, R.F. O’Connell, *Gen. Rel. Grav.* **11**, 149 (1979)
9. L. Blanchet, *Proc. R. Soc. Lond. A* **409**, 383 (1987)
10. L. Blanchet, *Phys. Rev. D* **47**, 4392 (1993)
11. L. Blanchet, *Class. Q. Grav.* **15**, 1971 (1998)
12. L. Blanchet, *Class. Q. Grav.* **15**, 113 (1998); Erratum, *ibidem* **22**, 3381 (2005)
13. L. Blanchet, A. Buonanno, G. Faye, *Phys. Rev. D* **74**, 104034 (2006); Erratum, *ibidem* **75**, 049903(E) (2007); Erratum, *ibidem* **81**, 089901(E) (2010)
14. L. Blanchet, T. Damour, *Phil. Trans. R. Soc. Lond. A* **320**, 379 (1986)
15. L. Blanchet, T. Damour, *Phys. Rev. D* **37**, 1410 (1988)
16. L. Blanchet, T. Damour, *Ann. Inst. H. Poincaré (A) Phys. Théor.* **50**, 377 (1989)
17. L. Blanchet, T. Damour, *Phys. Rev. D* **46**, 4304 (1992)
18. L. Blanchet, T. Damour, G. Esposito-Farèse, *Phys. Rev. D* **69**, 124007 (2004)
19. L. Blanchet, T. Damour, G. Esposito-Farèse, B.R. Iyer, *Phys. Rev. Lett.* **93**, 091101 (2004)
20. L. Blanchet, G. Faye, *J. Math. Phys.* **41**, 7675 (2000)
21. L. Blanchet, G. Faye, *Phys. Rev. D* **63**, 062005 (2001)
22. L. Blanchet, G. Faye, B.R. Iyer, B. Joguet, *Phys. Rev. D* **65**, 061501(R) (2002); Erratum, *ibidem* **71**, 129902(E) (2005)

23. L. Blanchet, G. Faye, B.R. Iyer, S. Sinha, *Class. Q. Grav.* **25**, 165003 (2008)
24. L. Blanchet, G. Faye, S. Nissanke, *Phys. Rev. D* **72**, 044024 (2005)
25. L. Blanchet, B.R. Iyer, B. Joguet, *Phys. Rev. D* **65**, 064005 (2002); Erratum, *ibidem* **71**, 129903(E) (2005)
26. L. Blanchet, B.R. Iyer, C.M. Will, A.G. Wiseman, *Class. Q. Grav.* **13**, 575 (1996)
27. C.G. Bollini, J.J. Giambiagi, *Phys. Lett. B* **40**, 566 (1972)
28. H. Bondi, M.G.J. van der Burg, A.W.K. Metzner, *Proc. R. Soc. Lond. Ser. A* **269**, 21 (1962)
29. A. Buonanno, G.B. Cook, F. Pretorius, *Phys. Rev. D* **75**, 124018 (2007)
30. W.L. Burke, *J. Math. Phys.* **12**, 401 (1971)
31. W.L. Burke, K.S. Thorne, in *Relativity*, ed. by M. Carmeli, S.I. Fickler, L. Witten 209 (Plenum, New York, 1970)
32. S. Chandrasekhar, *Astrophys. J.* **142**, 1488 (1965)
33. S. Chandrasekhar, F.P. Esposito, *Astrophys. J.* **160**, 153 (1970)
34. S. Chandrasekhar, Y. Nutku, *Astrophys. J.* **158**, 55 (1969)
35. D. Christodoulou, *Phys. Rev. Lett.* **67**, 1486 (1991)
36. C. Cutler, T.A. Apostolatos, L. Bildsten, L.S. Finn, E.E. Flanagan, D. Kennefick, D.M. Markovic, A. Ori, E. Poisson, G.J. Sussman, K.S. Thorne, *Phys. Rev. Lett.* **70**, 2984 (1993)
37. C. Cutler, L.S. Finn, E. Poisson, G.J. Sussman, *Phys. Rev. D* **47**, 1511 (1993)
38. C. Cutler, E.E. Flanagan, *Phys. Rev. D* **49**, 2658 (1994)
39. T. Damour, *Phys. Rev. Lett.* **51**, 1019 (1983)
40. T. Damour, in *Gravitational Radiation*, ed. by N. Deruelle, T. Piran 59 (North-Holland, Amsterdam, 1983)
41. T. Damour, N. Deruelle, *Phys. Lett. A* **87**, 81 (1981)
42. T. Damour, N. Deruelle, *C. R. Acad. Sci. Paris* **293**, 537 (1981)
43. T. Damour, B.R. Iyer, *Ann. Inst. H. Poincaré, (A) Phys. Théor.* **54**, 115 (1991)
44. T. Damour, P. Jaranowski, G. Schäfer, *Phys. Lett. B* **513**, 147 (2001)
45. T. Damour, P. Jaranowski, G. Schäfer, *Phys. Rev. D* **77**, 064032 (2008)
46. V.C. de Andrade, L. Blanchet, G. Faye, *Class. Q. Grav.* **18**, 753 (2001)
47. W. de Sitter, *Mon. Not. R. Astron. Soc.* **76**, 699 (1916)
48. W. de Sitter, *Mon. Not. R. Astron. Soc.* **77**, 155 (1916)
49. W.G. Dixon, in *Isolated Systems in General Relativity*, ed. by J. Ehlers (North Holland, Amsterdam, 1979), p. 156
50. J. Ehlers, *Ann. N.Y. Acad. Sci.* **33**, 279 (1980)
51. A. Einstein, *Sitzungsber. Preuss. Akad. Wiss., Phys. Math. Kl.* 688 (1916); *ibidem* 1111 (1916)
52. A. Einstein, L. Infeld, B. Hoffmann, *Ann. Math.* **39**, 65 (1938)
53. G. Faye, L. Blanchet, A. Buonanno, *Phys. Rev. D* **74**, 104033 (2006)
54. L.S. Finn, D.F. Chernoff, *Phys. Rev. D* **47**, 2198 (1993)
55. V.A. Fock, *Zh. Exp. Teor. Fiz.* **9**, 375 (1939). English translation: *J. Phys. (Moscow)* **1**, 81 (1939)
56. V.A. Fock, *Theory of Space, Time and Gravitation* (Pergamon, London, 1959)
57. T. Futamase, *Phys. Rev. D* **28**, 2373 (1983)
58. T. Futamase, B.F. Schutz, *Phys. Rev. D* **28**, 2363 (1983)
59. L.A. Gergely, *Phys. Rev. D* **61**, 024035 (1999)
60. J. Hadamard, *Le problème de Cauchy et les équations aux dérivées partielles linéaires hyperboliques* (Hermann, Paris, 1932)
61. G.D. Kerlick, *Gen. Rel. Grav.* **12**, 467 (1980)
62. G.D. Kerlick, *Gen. Rel. Grav.* **12**, 521 (1980)
63. L.E. Kidder, *Phys. Rev. D* **52**, 821 (1995)
64. L.E. Kidder, *Phys. Rev. D* **77**, 044016 (2008)
65. L.E. Kidder, C.M. Will, A.G. Wiseman, *Phys. Rev. D* **47**, R4183 (1993)
66. Y. Itoh, *Phys. Rev. D* **69**, 064018 (2004)
67. Y. Itoh, T. Futamase, *Phys. Rev. D* **68**, 121501(R) (2003)
68. Y. Itoh, T. Futamase, H. Asada, *Phys. Rev. D* **63**, 064038 (2001)
69. P. Jaranowski, G. Schäfer, *Phys. Rev. D* **57**, 7274 (1998)

70. P. Jaranowski, G. Schäfer, *Phys. Rev. D* **60**, 124003 (1999)
71. H.A. Lorentz, J. Droste, *Versl. K. Akad. Wet. Amsterdam* **26**, 392 (part I) and 649 (part II) (1917). English translation in H.A. Lorentz, *Collected Papers*, ed. by P. Zeeman, A.D. Fokker, vol. 5 (Nijhoff, The Hague, 1937)
72. B. Mikóczy, M. Vasúth, L.A. Gergely, *Phys. Rev. D* **71**, 124043 (2005)
73. B.J. Owen, H. Tagoshi, A. Ohashi, *Phys. Rev. D* **57**, 6168 (1998)
74. A. Papapetrou, *Proc. Phys. Soc. A* **64**, 57 (1951)
75. A. Papapetrou, *Proc. R. Soc. Lond. A* **209**, 248 (1951)
76. A. Papapetrou, B. Linet, *Gen. Rel. Grav.* **13**, 335 (1981)
77. M.E. Pati, C.M. Will, *Phys. Rev. D* **62**, 124015 (2000)
78. M.E. Pati, C.M. Will, *Phys. Rev. D* **65**, 104008 (2002)
79. P.C. Peters, *Phys. Rev.* **136**, B1224 (1964)
80. J. Plebanski, S.L. Bazanski, *Acta Phys. Polon.* **18**, 307 (1959)
81. E. Poisson, *Phys. Rev. D* **52**, 5719 (1995); Erratum and Addendum, *ibidem* **55**, 7980 (1997)
82. O. Poujade, L. Blanchet, *Phys. Rev. D* **65**, 124020 (2002)
83. A.D. Rendall, *Proc. R. Soc. Lond. A* **438**, 341 (1992)
84. M. Sasaki, *Prog. Theor. Phys.* **92**, 17 (1994)
85. L. Schwartz, *Théorie des distributions* (Hermann, Paris, 1978)
86. H. Tagoshi, T. Nakamura, *Phys. Rev. D* **49**, 4016 (1994)
87. H. Tagoshi, A. Ohashi, B.J. Owen, *Phys. Rev. D* **63**, 044006 (2001)
88. H. Tagoshi, M. Sasaki, *Prog. Theor. Phys.* **92**, 745 (1994)
89. T. Tanaka, H. Tagoshi, M. Sasaki, *Prog. Theor. Phys.* **96**, 1087 (1996)
90. J.H. Taylor, *Class. Q. Grav.* **10**, 167 (1993)
91. J.H. Taylor, L.A. Fowler, P.M. McCulloch, *Nature* **277**, 437 (1979)
92. J.H. Taylor, J.M. Weisberg, *Astrophys. J.* **253**, 908 (1982)
93. G. 't Hooft, M. Veltman, *Nucl. Phys. B* **44**, 139 (1972)
94. K.S. Thorne, *Rev. Mod. Phys.* **52**, 299 (1980)
95. K.S. Thorne, *Phys. Rev. D* **45**, 520 (1992)
96. A. Trautman, *Lectures on general relativity*, issued by King's College of London (1958), published for the Brandeis Univ. Summer Inst. Theoretical Physics of 1964, ed. by A. Trautmann, F.A.E. Pirani, H. Bondi (Prentice-Hall, Englewood Cliffs, 1965), and by *Gen. Rel. Grav.* **34**, 721 (2002)
97. W. Tulczyjew, *Bull. Acad. Polon. Sci. Cl. III* **5**, 279 (1957)
98. W. Tulczyjew, *Acta Phys. Polon.* **18**, 37 (1959)
99. C.M. Will, A.G. Wiseman, *Phys. Rev. D* **54**, 4813 (1996)
100. A.G. Wiseman, C.M. Will, *Phys. Rev. D* **44**, R2945 (1991)

Post-Newtonian Methods: Analytic Results on the Binary Problem

Gerhard Schäfer

Abstract A detailed account is given on approximation schemes to the Einstein theory of general relativity where the iteration starts from the Newton theory of gravity. Two different coordinate conditions are used to represent the Einstein field equations, the generalized isotropic ones of the canonical formalism of Arnowitt, Deser, and Misner and the harmonic ones of the Lorentz-covariant Fock-de Donder approach. Conserved quantities of isolated systems are identified and the Poincaré algebra is introduced. Post-Newtonian (PN) expansions are performed in the near and far (radiation) zones. The natural fitting of multipole expansions to PN schemes is emphasized. The treated matter models are ideal fluids, pure point masses, and point masses with spin and mass-quadrupole moments modeling rotating black holes. Various Hamiltonians of spinning binaries are presented in explicit forms to higher PN orders. The delicate use of black holes in PN expansion calculations and of the Dirac delta function in general relativity find discussions.

1 Introduction

In the weak-field slow-motion limit of any theory of gravity the Newtonian theory of gravity comes into play because it describes the motion and structure of gravitating objects very well in that regime. According to current knowledge from experiments and observations, the most reliable theory of gravity is the Einstein theory of general relativity [72].

Most of the objects in the Universe seem to have velocities v , which are small compared with the speed of light (in the following denoted by c), that is, $v/c \lesssim 1/3$. In those cases it is to be expected that the Newtonian theory is a good starting point for an iteration scheme toward full general relativity thinking in terms of expansion

G. Schäfer (✉)
Friedrich-Schiller-Universität Jena, Theoretisch-Physikalisches Institut,
Max-Wien-Platz 1, D-07743 Jena, Germany
e-mail: G.Schaefer@tpi.uni-jena.de

of general relativity in powers of dimensionless v/c . The conservative nature of most of the phenomena, that is, no measurable gravitational radiation damping, the power series has to be one with ordering parameter $(v/c)^2$ (recall $v^2/c^2 \lesssim 1/10$ in the regime of our interest) because of motion-inversion symmetry in those cases. We shall see later that at the order $(v/c)^5$ the gravitational radiation damping enters for the first time. Calling the order $(v/c)^{2n}$ the n th PN order (n PN), the gravitational dissipation (because of gravitational radiation damping from radiation emission) enters at the 2.5 PN order so the exponent n will also include half-integer numbers. In case of bound systems the virial theorem holds, which tells us that $v^2 \sim GM/r$, where G is the Newtonian gravitational constant, r a typical distance of two bodies or a typical radius of one body, and M the total mass of the system. So an expansion of general relativity in powers of $(v/c)^2$ is at the same time an expansion in powers of GM/c^2r if bound systems are considered. From the dynamics of bound systems the dynamics of low-velocity scattering is straightforwardly obtained.

Depending on the mathematical representation of the Einstein theory (choice of coordinates, choice of variables, etc.) and on the physical aspects under investigation (physics in the near zone, physics in the far zone, etc.), there exist many of those expansions that all are called PN expansions. The crucial ordering parameter in all PN expansions is always $1/c^2$. The Einstein field equations as well as the equations of motion that follow from them by integrability conditions can be formally expanded in powers of $1/c^2$. The solutions of these equations, however, cannot be expanded in this way, in general. Only under conditions where no gravitational radiation is present an all-over-in-space expansion in powers of $1/c^2$ is feasible. This is the case for stationary systems only. Yet for those parts of a radiating system where radiation plays no role, that is, in the conservative parts, an expansion in powers of $1/c^2$ can be carried out too. In general, PN expansions in powers of $1/c$ are feasible in the near zone ($r \ll \lambda$ with λ a typical wavelength of the gravitational radiation and r the radius of a sphere enclosing the matter source) and in the far or radiation zone ($r \gg \lambda$) of a radiating system, but they are not analytic in $1/c$ because c is showing up in log-terms. In the near zone, up to 3.5 PN order, which means order $1/c^7 = (1/c^2)^{3.5}$, a PN expansion in powers of $1/c$ is valid, where the even powers are connected with the conservative dynamics and the odd powers with the dissipative radiation reaction dynamics. From the 4 PN order on (corresponding to 1.5 PN order in the far zone), log- c terms enter via tail effects which emerge from back scattering [4]. Post-Minkowskian (PM) series are global series in space where only the weak-field limit is assumed without restrictions to the velocities of the bodies. Here, the ordering parameter is G or, dimensionless GM/c^2r , which is equivalent to a nonlinearity expansion of the Einstein field equations [4]. If the virial theorem holds, namely, $(GM/c^2r) \sim (v/c)^2$, PM series can be further expanded into PN series. Having much less information on their respective orders, the PN series are much easier to be worked out analytically, compare, for example, the 1 PN point-mass Hamiltonian, see [51], with the 1 PM point-mass Hamiltonian, only recently achieved in [52].

The most extensively performed PN calculations in the literature can be divided into two classes, the class applying harmonic coordinate conditions,

see, for example, [33], and the other class using Arnowitt–Deser–Misner (ADM) generalized isotropic coordinate conditions [1]. Whenever detailed comparisons between the two classes have been made, agreement of the results could be achieved, see, for example, [12, 23]. This is quite an important aspect in view of the use of Dirac delta functions for modeling black holes in explicit calculations. These functions are a-priori problematic in a nonlinear theory like general relativity so sophisticated regularization methods have to be employed. The only regularization method that turned out to be successful to the highest orders of the explicit calculations is the dimensional one [6, 22]. In Section 3.4 a detailed account of black holes together with their representing Dirac delta functions will be given because they blatantly violate the weak-field condition. In this case, only external velocities of the black holes can be small and their relative gravitational interaction weak, and only in this sense black holes and Dirac delta functions can fit into PN schemes.

The binary point-mass dynamics at 3 PN order and the gravitational waves emitted hereof (3 PN wave generation) could have been managed successfully through only dimensional regularization [6, 7, 22]. Extended body calculations have been performed too but they are much more complicated and still not fully under control at 3 PN order [56]. Only a surface-integral-based method of the Einstein–Infeld–Hoffmann type did succeed too [34, 44].

The many-body systems treated most successfully with PN approximation methods are point-mass systems composed from two or more objects and spinning binary systems with even rotationally deformed components; for the most recent results with spin, see [5, 25, 31, 40, 41, 67, 69].

Recently, an effective field theory approach has been advocated for PN calculations [36]. Applications to point-mass systems and to systems with spinning components have already been performed through higher PN orders [35, 57, 58]. An obvious difference in the spin dynamics between [67] and [58] found clarification in [68] showing the correctness of the result in [67].

In this article, Latin indices from the mid-alphabet are mostly running from 1 to 3, $\mathbf{v} = (v^i)$, $\mathbf{v}^2 = v^i v^i$, ∂_t and ∂_i denote the partial derivatives with respect to the time and space coordinates t and x^i , respectively, and the functional derivative (Fréchet derivative) of a functional, say $F[f]$, with respect to a function f takes the form $\delta F[f]/\delta f$. Greek indices are mostly running from 0 to 3, whereby $x^0 = ct$. The signature of the four-dimensional metric $g_{\mu\nu}$ is $+2$. Particles are numbered with indices from the beginning of the Latin alphabet.

2 Systems in Newtonian Gravity in Canonical Form

In the Newtonian theory, the equations governing the motion of gravitating ideal fluids are (i) the equation for the conservation of mass,

$$\partial_t \varrho_* + \operatorname{div}(\varrho_* \mathbf{v}) = 0, \quad (1)$$

where ϱ_* is the mass density and \mathbf{v} the velocity field of the fluid, (ii) the equations of motion,

$$\varrho_* \partial_t \mathbf{v} + \frac{\varrho_*}{2} \text{grad } \mathbf{v}^2 - \varrho_* \mathbf{v} \times \text{curl } \mathbf{v} = -\text{grad } p + \varrho_* \text{grad } U, \quad (2)$$

where p is the pressure in the fluid and U the gravitational potential, and (iii) the equation of state, using internal energy density ϵ and specific enthalpy h ,

$$\epsilon = \epsilon(\varrho_*, s) \quad \text{with} \quad d\epsilon = h d\varrho_* + \varrho_* T ds, \quad \text{or} \quad dp = \varrho_* dh - \varrho_* T ds, \quad (3)$$

with (iv) the conservation law for the specific entropy s along the flow lines,

$$\partial_t s + \mathbf{v} \cdot \text{grad } s = 0. \quad (4)$$

The gravitational potential reads,

$$U(\mathbf{x}, t) = G \int d^3 \mathbf{x}' \frac{\varrho_*(\mathbf{x}', t)}{|\mathbf{x} - \mathbf{x}'|}. \quad (5)$$

It results from (v) the Newtonian field equation,

$$\Delta U = -4\pi G \varrho_*. \quad (6)$$

Δ is the Laplacian and $|\dots|$ means the standard Euclidean distance.

Written in form of Hamilton equations of motion, that is, $\partial_t A(\mathbf{x}, t) = \{A(\mathbf{x}, t), H\}$, the equations above take the forms, (i) the mass conservation equation

$$\frac{\partial \varrho_*}{\partial t} = -\partial_i \left(\frac{\delta H}{\delta \pi_i} \varrho_* \right), \quad (7)$$

notice $v^i = \frac{\delta H}{\delta \pi_i}$, (ii) the equations of motion

$$\frac{\partial \pi_i}{\partial t} = -\partial_j \left(\frac{\delta H}{\delta \pi_j} \pi_i \right) - \partial_i \left(\frac{\delta H}{\delta \pi_j} \right) \pi_j - \partial_i \left(\frac{\delta H}{\delta \varrho_*} \right) \varrho_* + \frac{\delta H}{\delta s} \partial_i s, \quad (8)$$

and (iv) the entropy conservation law

$$\frac{\partial s}{\partial t} = -\frac{\delta H}{\delta \pi_i} \partial_i s, \quad (9)$$

where the Hamiltonian is given by $H = H(\varrho_*, \pi_i, s)$ with π_i the linear momentum density of the fluid, see [42]. Here, use has been made of the kinematical Lie-Poisson bracket relations between the fundamental variables

$$\{\pi_i(\mathbf{x}, t), \varrho_*(\mathbf{x}', t)\} = \frac{\partial}{\partial x'^i} [\varrho_*(\mathbf{x}', t) \delta(\mathbf{x} - \mathbf{x}')], \quad (10)$$

$$\{\pi_i(\mathbf{x}, t), s(\mathbf{x}', t)\} = \frac{\partial s(\mathbf{x}', t)}{\partial x'^i} \delta(\mathbf{x} - \mathbf{x}'), \quad (11)$$

$$\{\pi_i(\mathbf{x}, t), \pi_j(\mathbf{x}', t)\} = \pi_i(\mathbf{x}', t) \frac{\partial}{\partial x'^j} \delta(\mathbf{x} - \mathbf{x}') - \pi_j(\mathbf{x}, t) \frac{\partial}{\partial x^i} \delta(\mathbf{x} - \mathbf{x}'), \quad (12)$$

and zero otherwise, where $\delta(\mathbf{x} - \mathbf{x}')$ denotes the standard Dirac delta function in three-dimensional space. It fulfills $\int d^3x \delta(\mathbf{x}) = 1$.

In the Newtonian theory, the Hamiltonian of the fluid is given by,

$$H = \frac{1}{2} \int d^3\mathbf{x} \frac{\pi_i \pi_i}{\varrho_*} - \frac{G}{2} \int d^3\mathbf{x} d^3\mathbf{x}' \frac{\varrho_*(\mathbf{x}, t) \varrho_*(\mathbf{x}', t)}{|\mathbf{x} - \mathbf{x}'|} + \int d^3\mathbf{x} \epsilon. \quad (13)$$

For point masses, the total momentum and mass densities read, consistent with the Eqs. 10 and 12,

$$\pi_i = \sum_a p_{ai} \delta(\mathbf{x} - \mathbf{x}_a), \quad \varrho_* = \sum_a m_a \delta(\mathbf{x} - \mathbf{x}_a), \quad (14)$$

where the position and momentum variables fulfill the standard Poisson bracket relations,

$$\{x_a^i, p_{aj}\} = \delta_{ij}, \quad \text{zero otherwise}, \quad (15)$$

and the Hamiltonian takes the form,

$$H = \frac{1}{2} \sum_a \frac{p_{ai} p_{ai}}{m_a} - \frac{G}{2} \sum_{a \neq b} \frac{m_a m_b}{|\mathbf{x}_a - \mathbf{x}_b|}, \quad (16)$$

where the self-energy term has been dropped (for regularization techniques, see Section 3.4).

3 Canonical General Relativity and PN Expansions

In curved spacetime the stress–energy tensor of an ideal fluid takes the form

$$T^{\mu\nu} = \varrho(c^2 + h)u^\mu u^\nu + p g^{\mu\nu}, \quad g_{\mu\nu} u^\mu u^\nu = -1, \quad (17)$$

where ϱ denotes the proper rest-mass density, h the specific enthalpy, and u^μ the four-velocity field of the fluid. Using energy density $e = \varrho(c^2 + h) - p$ (also the specific internal energy $\Pi = e/\varrho - c^2$ could be used), the equation of state reads

$$e = e(\varrho, s) \quad \text{with} \quad de = (c^2 + h)d\varrho + \varrho T ds, \quad \text{or} \quad dp = \varrho dh - \varrho T ds. \quad (18)$$

The variables of the canonical formalism are chosen to be

$$\varrho_* = \sqrt{-g} u^0 \varrho, \quad s, \quad \pi_i = \frac{1}{c} \sqrt{-g} T_i^0. \quad (19)$$

They fulfill the same (universal) kinematical Lie–Poisson bracket relations as in the Newtonian theory, see [42], or also [9],

$$\{\pi_i(\mathbf{x}, t), \varrho_*(\mathbf{x}', t)\} = \frac{\partial}{\partial x'^i} [\varrho_*(\mathbf{x}', t) \delta(\mathbf{x} - \mathbf{x}')], \quad (20)$$

$$\{\pi_i(\mathbf{x}, t), s(\mathbf{x}', t)\} = \frac{\partial s(\mathbf{x}', t)}{\partial x'^i} \delta(\mathbf{x} - \mathbf{x}'), \quad (21)$$

$$\{\pi_i(\mathbf{x}, t), \pi_j(\mathbf{x}', t)\} = \pi_i(\mathbf{x}', t) \frac{\partial}{\partial x'^j} \delta(\mathbf{x} - \mathbf{x}') - \pi_j(\mathbf{x}, t) \frac{\partial}{\partial x^i} \delta(\mathbf{x} - \mathbf{x}'). \quad (22)$$

The evolution equations take the form

$$\frac{\partial \varrho_*}{\partial t} = -\partial_i \left(\frac{\delta H}{\delta \pi_i} \varrho_* \right) \iff \partial_\mu (\sqrt{-g} \varrho u^\mu) = 0, \quad (23)$$

$$\frac{\partial s}{\partial t} = -\frac{\delta H}{\delta \pi_i} \partial_i s \iff u^\mu \partial_\mu s = 0, \quad (24)$$

$$\frac{\partial \pi_i}{\partial t} = -\partial_j \left(\frac{\delta H}{\delta \pi_j} \pi_i \right) - \partial_i \left(\frac{\delta H}{\delta \pi_j} \right) \pi_j - \partial_i \left(\frac{\delta H}{\delta \varrho_*} \right) \varrho_* + \frac{\delta H}{\delta s} \partial_i s, \quad (25)$$

corresponding to $\partial_\mu (\sqrt{-g} T_i^\mu) - \frac{1}{2} \sqrt{-g} T^{\mu\nu} \partial_i g_{\mu\nu} = 0$,

$$v^j = \frac{\delta H}{\delta \pi_i}, \quad \text{where} \quad v^j = c \frac{u^j}{u^0}. \quad (26)$$

The linear and angular momenta of the fluid read, respectively,

$$P_i = \int d^3\mathbf{x} \pi_i, \quad J_i = \int d^3\mathbf{x} \epsilon_{ijk} x^j \pi_k. \quad (27)$$

For a system made of point masses simplifications take place,

$$h = p = s = 0, \quad (\text{dusty matter}), \quad (28)$$

and further,

$$\varrho_* = \sum_a m_a \delta(\mathbf{x} - \mathbf{x}_a), \quad \pi_i = \sum_a p_{ai} \delta(\mathbf{x} - \mathbf{x}_a), \quad v_a^i = \frac{dx_a^i}{dt}, \quad (29)$$

where p_{ai} and x_a^i , respectively, are the linear momentum and the position vector of the a th particle. The kinematical Poisson bracket relations are given by

$$\{x_a^i, p_{aj}\} = \delta_{ij}, \quad \text{zero otherwise.} \quad (30)$$

Hereof the standard Hamilton equations result,

$$\frac{dp_{ai}}{dt} = -\frac{\partial H}{\partial x_a^i}, \quad \frac{dx_a^i}{dt} = \frac{\partial H}{\partial p_{ai}}. \quad (31)$$

Remarkably, the difference to the Newtonian theory comes solely from the Hamiltonian, which is thus a dynamical difference and not a kinematical one. This statement refers to the matter only and not to the gravitational field. The latter is quite different in general relativity.

3.1 Canonical Variables of the Gravitational Field

Within the ADM canonical formalism of general relativity, in generalized isotropic coordinates, the independent gravitational field variables h_{ij}^{TT} and π_{TT}^{ij} enter in the form

$$g_{ij} = \left(1 + \frac{1}{8}\phi\right)^4 \delta_{ij} + h_{ij}^{\text{TT}}, \quad (32)$$

$$\pi^{ij} = \tilde{\pi}^{ij} + \pi_{\text{TT}}^{ij}, \quad (33)$$

where $g_{ij} = g_{ji} \equiv \gamma_{ij}$ is the metric of the curved three-dimensional hypersurfaces $t = \text{const}$, $\pi^{ij} c^3 / 16\pi G$ is the canonical conjugate to γ_{ij} , that is, $\pi^{ij} = -\gamma^{1/2} (K^{ij} - \gamma^{ij} K_k^k)$, where $K_{ij} = K_{ji}$ is the extrinsic curvature of the $t = \text{const}$ slices, and $\gamma = \det(\gamma_{ij})$, $\gamma^{il} \gamma_{lj} = \delta_{ij}$, and for $\tilde{\pi}^{ij}$ holds

$$\tilde{\pi}^{ij} = \partial_i \pi^j + \partial_j \pi^i - \frac{2}{3} \delta_{ij} \partial_k \pi^k. \quad (34)$$

Obviously, $\pi^{ii} = 0$, or $\pi_i^i = \pi^{ij} h_{ij}^{\text{TT}}$. The canonical conjugate to h_{ij}^{TT} reads $\pi_{\text{TT}}^{ij} c^3 / 16\pi G$. The index TT means tranverse-traceless, that is, $h_{ii}^{\text{TT}} = \pi_{\text{TT}}^{ii} = 0$, $\partial_j h_{ij}^{\text{TT}} = \partial_j \pi_{\text{TT}}^{ij} = 0$.

Using those variables, the Einstein field equation $\sqrt{-g} G^{00} = \frac{8\pi G}{c^4} \sqrt{-g} T^{00}$ can be put into the form, employing point masses for the source,

$$\gamma^{1/2} \mathbf{R} = \frac{1}{\gamma^{1/2}} \left(\pi_j^i \pi_i^j - \frac{1}{2} \pi_i^i \pi_j^j \right) + \frac{16\pi G}{c^3} \sum_a (m_a^2 c^2 + \gamma^{ij} p_{ai} p_{aj})^{1/2} \delta_a, \quad (35)$$

and the field equations $\sqrt{-g}G_i^0 = \frac{8\pi G}{c^4}\sqrt{-g}T_i^0$ read

$$-2\partial_j\pi_i^j + \pi^{kl}\partial_i\gamma_{kl} = \frac{16\pi G}{c^3}\sum_a p_{ai}\delta_a, \quad (36)$$

where $\delta_a = \delta(\mathbf{x} - \mathbf{x}_a)$. The Eqs. 35 and 36 are the famous four constraint equations of general relativity.

In the gauge Eqs. 32–34 the ADM Hamiltonian can be written, [1],

$$H[x_a^i, p_{ai}, h_{ij}^{\text{TT}}, \pi_{\text{TT}}^{ij}] = -\frac{c^4}{16\pi G}\int d^3\mathbf{x}\Delta\phi[x_a^i, p_{ai}, h_{ij}^{\text{TT}}, \pi_{\text{TT}}^{ij}], \quad (37)$$

resulting from the solution of the four (elliptic-type) constraint equations. The additional Hamilton equations of motion for the gravitational field are given by

$$\frac{\partial\pi_{\text{TT}}^{ij}}{\partial t} = -\frac{16\pi G}{c^3}\frac{\delta H}{\delta h_{ij}^{\text{TT}}}, \quad \frac{\partial h_{ij}^{\text{TT}}}{\partial t} = \frac{16\pi G}{c^3}\frac{\delta H}{\delta\pi_{\text{TT}}^{ij}}. \quad (38)$$

The transition to a Routh functional simplifies a lot the construction of the dynamics of the matter and of the gravitational field. The Routh functional is chosen in the form [46],

$$R[x_a^i, p_{ai}, h_{ij}^{\text{TT}}, \partial_t h_{ij}^{\text{TT}}] = H - \frac{c^3}{16\pi G}\int d^3\mathbf{x}\pi_{\text{TT}}^{ij}\partial_t h_{ij}^{\text{TT}}. \quad (39)$$

The evolution equations for the matter and the gravitational field now read

$$\frac{\delta\int R(t')dt'}{\delta h_{ij}^{\text{TT}}(x^k, t)} = 0, \quad \dot{p}_{ai} = -\frac{\partial R}{\partial x_a^i}, \quad \dot{x}_a^i = \frac{\partial R}{\partial p_{ai}}. \quad (40)$$

The conservative dynamics results from the on-field-shell Routh functional

$$R_{\text{shell}}(t) = R[x_a^i, p_{ai}, h_{ij}^{\text{TT}}[x_a^k, p_{ak}], \partial_t h_{ij}^{\text{TT}}[x_a^k, p_{ak}]], \quad (41)$$

with solved field equations, in the form

$$\dot{p}_{ai}(t) = -\frac{\delta\int R_{\text{shell}}(t')dt'}{\delta x_a^i(t)}, \quad \dot{x}_a^i(t) = \frac{\delta\int R_{\text{shell}}(t')dt'}{\delta p_{ai}(t)}, \quad (42)$$

where

$$\frac{\delta\int R_{\text{shell}}(t')dt'}{\delta z(t)} = \frac{\partial R_{\text{shell}}}{\partial z(t)} - \frac{d}{dt}\frac{\partial R_{\text{shell}}}{\partial\dot{z}(t)} + \dots, \quad z = (x_a^i, p_{ai}). \quad (43)$$

Using the matter equations of motion in the Routhian R_{shell} the Routhian can be brought into the form $R(x_a^i, p_{ai})$. Herein, however, the meaning of the variables x_a^i and p_{ai} has changed, see [19, 27, 64].

3.2 Brill–Lindquist Initial-Value Solution for Binary Black Holes

The Brill–Lindquist solution for multiple black holes is a pure vacuum solution of the constraint equations at initial time t under the conditions of time symmetry, that is, $p_{ai} = 0 = \pi^{ij}$, and of conformal flatness, that is, $h_{ij}^{\text{TT}} = 0$ [16]. A related vacuum solution is the Misner–Lindquist solution where an additional isometry condition is imposed [53, 55]. Under those conditions, the only remaining constraint equation reads, not using vacuum but (point-mass) sources,

$$-\left(1 + \frac{1}{8}\phi\right)\Delta\phi = \frac{16\pi G}{c^2}\sum_a m_a\delta_a, \quad (h_{ij}^{\text{TT}} = 0 = p_{ai} = \pi^{ij}). \quad (44)$$

In the case of two black holes, its solution is given by, see [47],

$$\phi = \frac{4G}{c^2}\left(\frac{\alpha_1}{r_1} + \frac{\alpha_2}{r_2}\right) \quad (45)$$

with $(a, b = 1, 2$ and $b \neq a)$

$$\alpha_a = \frac{m_a - m_b}{2} + \frac{c^2 r_{ab}}{G} \left(\sqrt{1 + \frac{m_a + m_b}{c^2 r_{ab}/G} + \left(\frac{m_a - m_b}{2c^2 r_{ab}/G}\right)^2} - 1 \right), \quad (46)$$

resulting into the Brill–Lindquist solution for binary black holes. Obviously, each Brill–Lindquist black hole is represented by a Dirac delta function (fictitious image mass-point; see Section 3.4). In the Misner–Lindquist case, infinite many fictitious image mass-points are needed for each black hole [47, 53, 55].

The energy of the Brill–Lindquist solution simply reads

$$H_{\text{BL}} = (\alpha_1 + \alpha_2)c^2 = (m_1 + m_2)c^2 - G \frac{\alpha_1\alpha_2}{r_{12}}. \quad (47)$$

The methods that have been used for the obtention of the Brill–Lindquist solution from sources (notice, in the original work of Brill and Lindquist this solution has been obtained without any regularization as a purely vacuum solution) are analytical Hadamard regularization and mass renormalization [47], as well as dimensional regularization based on the d -dimensional metric

$$\gamma_{ij} = \left(1 + \frac{1}{4}\frac{d-2}{d-1}\phi\right)^{\frac{4}{d-2}}\delta_{ij} \quad (48)$$

with solution (Γ denotes the Euler gamma function)

$$\phi = \frac{4G}{c^2} \frac{\Gamma(\frac{d-2}{2})}{\pi^{\frac{d-2}{2}}} \left(\frac{\alpha_1}{r_1^{d-2}} + \frac{\alpha_2}{r_2^{d-2}} \right) \quad (49)$$

(for more details see Section 3.4). The PN expansion of the Brill–Lindquist initial energy expression is straightforward to all orders of $1/c^2$. Once it had fixed the static ambiguity parameter ω_{static} (see [19]) in nondimensional-regularization calculations to the correct value of zero [47]. At that time, however, it was not quite clear that the Brill–Lindquist solution delivers the correct boundary conditions for the point-mass model.

The truncation of the constraint equations in the form $h_{ij}^{\text{TT}} \equiv 0$ as well as dropping an additional term in the Hamiltonian constraint connected with the energy density of the field momentum results in a remarkable, fully explicitly solvable conservative so-called skeleton dynamics that allows a PN expansion of the Hamiltonian, and of all the metric coefficients too, to all orders, see next section.

3.3 Skeleton Hamiltonian

In Ref. [32] the skeleton dynamics has been developed. The skeleton approach to general relativity requires the conformal flat condition for the spatial three-metric for all times (not only initially as for the Brill–Lindquist solution)

$$\gamma_{ij} = \left(1 + \frac{1}{8}\phi \right)^4 \delta_{ij}. \quad (50)$$

Hereof, in our coordinate system, maximal slicing follows,

$$\pi^{ij} \gamma_{ij} = 2\sqrt{\gamma} \gamma^{ij} K_{ij} = 0. \quad (51)$$

Under the conformal flat condition for the spatial three-metric, the momentum constraint equations become

$$\pi_{i,j}^j = -\frac{8\pi G}{c^3} \sum_a p_{ai} \delta_a. \quad (52)$$

The solution of these equations is constructed under the condition that π_i^j (and not π^{ij} , see Eqs. 33 and 34) is purely longitudinal, that is,

$$\pi_i^j = \partial_i V_j + \partial_j V_i - \frac{2}{3} \delta_{ij} \partial_l V_l. \quad (53)$$

This condition is part of the definition of the skeleton model. At spacelike infinity, the surface-area integrals of π_i^j or π^{ij} are proportional to the total linear momentum of the binary system.

Furthermore, in the Hamilton constraint equation, which in our case reads

$$\Delta\phi = -\frac{\pi_i^j \pi_j^i}{(1 + \frac{1}{8}\phi)^7} - \frac{16\pi G}{c^2} \sum_a \frac{m_a \delta_a}{(1 + \frac{1}{8}\phi)} \left(1 + \frac{p_a^2}{(1 + \frac{1}{8}\phi)^4 m_a^2 c^2}\right)^{1/2}, \quad (54)$$

we perform a truncation of the numerator of the first term in the following way

$$\pi_i^j \pi_j^i \equiv -2V_j \partial_i \pi_j^i + \partial_i (2V_j \pi_j^i) \rightarrow -2V_j \partial_i \pi_j^i = \frac{16\pi G}{c^3} \sum_a p_{aj} V_j \delta_a, \quad (55)$$

that is, we drop from $\pi_i^j \pi_j^i$ the term $\partial_i (2V_j \pi_j^i)$. This is the second crucial truncation condition additional to the conformal flat one. Without this truncation neither an explicit solution can be achieved nor a PN expansion is feasible. From [46] we know that at the 3 PN level the h_{ij}^{TT} -field is needed to make the sum of the corresponding terms from $\pi_i^j \pi_j^i$ analytic in $1/c$.

With the aid of the ansatz

$$\phi = \frac{4G}{c^2} \sum_a \frac{\alpha_a}{r_a} \quad (56)$$

and by making use of dimensional regularization, the energy and momentum constraint equations result in an algebraic equation of the form, [32],

$$\alpha_a = \frac{m_a}{1 + A \frac{\alpha_b}{r_{ab}}} \left[1 + \frac{p_a^2 / (m_a^2 c^2)}{(1 + A \alpha_b / r_{ab})^4}\right]^{\frac{1}{2}} + \frac{p_{ai} V_{ai} / c}{(1 + A \alpha_b / r_{ab})^7}, \quad (57)$$

where $A \equiv G/(2c^2)$ and $b \neq a$.

With these inputs the skeleton Hamiltonian for binary black holes becomes (at least initially, for $p_a = 0$, the solution is consistent with general relativity)

$$H_{\text{sk}} \equiv -\frac{c^4}{16\pi G} \int d^3\mathbf{x} \Delta\phi = c^2 \sum_a \alpha_a. \quad (58)$$

The Hamilton equations of motion read

$$\dot{\mathbf{x}}_a = \frac{\partial H}{\partial \mathbf{p}_a}, \quad \dot{\mathbf{p}}_a = -\frac{\partial H}{\partial \mathbf{x}_a}. \quad (59)$$

In the center-of-mass frame of the binary system, we define

$$\mathbf{p} \equiv \mathbf{p}_1 = -\mathbf{p}_2, \quad \mathbf{r} \equiv \mathbf{x}_1 - \mathbf{x}_2, \quad r^2 = (\mathbf{x}_1 - \mathbf{x}_2) \cdot (\mathbf{x}_1 - \mathbf{x}_2). \quad (60)$$

Further, we will employ the following convenient dimensionless quantities

$$\hat{t} = \frac{tc^3}{Gm}, \quad \hat{r} = \frac{rc^2}{Gm}, \quad \hat{\mathbf{p}} = \frac{\mathbf{p}}{\mu c}, \quad \hat{H}_{\text{sk}} = \frac{H_{\text{sk}}}{\mu c^2}, \quad (61)$$

$$\hat{j} = \frac{Jc}{Gm\mu}, \quad \hat{p}_r = \frac{p_r}{\mu c}, \quad \hat{\mathbf{p}}^2 = \hat{p}_r^2 + \hat{j}^2/\hat{r}^2, \quad (62)$$

where $\mathbf{J} = \mathbf{r} \times \mathbf{p}$ is the orbital angular momentum in the center-of-mass frame and $p_r = \mathbf{p} \cdot \mathbf{r}/r$ the radial momentum. The total rest-mass is denoted by $m = m_1 + m_2$ and the reduced mass by $\mu = m_1 m_2 / m$. The binary skeleton Hamiltonian \hat{H}_{sk} can be put into the following form [37],

$$\hat{H}_{\text{sk}} = 2\hat{r} \left(\psi_1 + \psi_2 - 2 \right) \quad \text{with} \quad (63)$$

$$\psi_1 = 1 + \frac{\chi_-}{4\hat{r}\psi_2} \left(1 + \frac{4v^2(\hat{p}_r^2 + \hat{j}^2/\hat{r}^2)}{\chi_-^2\psi_2^4} \right)^{1/2} - \frac{(8\hat{p}_r^2 + 7\hat{j}^2/\hat{r}^2)v^2}{8\hat{r}^2\psi_2^7}, \quad (64)$$

$$\psi_2 = 1 + \frac{\chi_+}{4\hat{r}\psi_1} \left(1 + \frac{4v^2(\hat{p}_r^2 + \hat{j}^2/\hat{r}^2)}{\chi_+^2\psi_1^4} \right)^{1/2} - \frac{(8\hat{p}_r^2 + 7\hat{j}^2/\hat{r}^2)v^2}{8\hat{r}^2\psi_1^7}, \quad (65)$$

where $\chi_- = (1 - \sqrt{1 - 4v})$ and $\chi_+ = (1 + \sqrt{1 - 4v})$ with $v = \mu/m$.

The conservative skeleton Hamiltonian has the following nice properties. It is exact in the test-body limit where it describes the motion of a test particle in the Schwarzschild spacetime. It is identical to the 1 PN accurate Hamiltonian for the binary dynamics in general relativity. Further, as explained earlier, when point particles are at rest, the Brill–Lindquist initial value solution is reproduced. It is remarkable that the skeleton Hamiltonian allows a PN expansion in powers of $1/c^2$ to arbitrary orders. The skeleton Hamiltonian thus describes the evolution of a kind of black hole under both conformal flat conditions for the three-metric and analyticity conditions in $1/c^2$ for the Hamiltonian. Of course, gravitational radiation emission is not included. It can, however, be added to some reasonable extent, see [37].

Restricting to circular orbits and defining $x = (Gm\omega/c^3)^{2/3}$, where ω is the orbital angular frequency, the skeleton Hamiltonian reads explicitly to 3 PN order,

$$\begin{aligned} \hat{H}_{\text{sk}} = & -\frac{x}{2} + \left(\frac{3}{8} + \frac{v}{24} \right) x^2 + \left(\frac{27}{16} + \frac{29}{16}v - \frac{17}{48}v^2 \right) x^3 \\ & + \left(\frac{675}{128} + \frac{8585}{384}v - \frac{7985}{192}v^2 + \frac{1115}{10368}v^3 \right) x^4 + \mathcal{O}(x^5). \end{aligned} \quad (66)$$

In Ref. [32] the coefficients are given to the order x^{11} inclusively.

In the Isenberg–Wilson–Mathews approach to general relativity only the conformal flat condition is employed. Thus the energy stops being analytic in $1/c$ at 3 PN. Through 2 PN order, the Isenberg–Wilson–Mathews energy of a binary is given by

$$\hat{H}_{\text{IWM}} = -\frac{x}{2} + \left(\frac{3}{8} + \frac{\nu}{24}\right)x^2 + \left(\frac{27}{16} - \frac{39}{16}\nu - \frac{17}{48}\nu^2\right)x^3. \quad (67)$$

We already quote here the 3 PN result of general relativity. It reads, see Eq. 129,

$$\begin{aligned} \hat{H}_{3\text{PN}} = & -\frac{x}{2} + \left(\frac{3}{8} + \frac{\nu}{24}\right)x^2 + \left(\frac{27}{16} - \frac{19}{16}\nu + \frac{1}{48}\nu^2\right)x^3 \\ & + \left(\frac{675}{128} + \left(\frac{205}{192}\pi^2 - \frac{34445}{1152}\right)\nu + \frac{155}{192}\nu^2 + \frac{35}{10368}\nu^3\right)x^4. \end{aligned} \quad (68)$$

The difference between \hat{H}_{IWM} and \hat{H}_{sk} through 2 PN order shows the effect of truncation in the field-momentum part of \hat{H}_{sk} and the difference between \hat{H}_{IWM} and $\hat{H}_{3\text{PN}}$ reveals the effect of conformal flat truncation. In the test-body limit, $\nu = 0$, all the Hamiltonians coincide and for the equal-mass case, $\nu = 1/4$, their differences are largest.

3.4 Functional Representation of Compact Objects

Before going on with the presentation of more dynamical expressions we will discuss in more detail the δ -function-source model we are employing. Although we are interested in both neutron stars and black holes, our matter model will be based on black holes because these are the simplest objects in general relativity and neutron stars resemble them very much as seen from outside. The simplest black holes are the isolated nonrotating ones. Their solution is the Schwarzschild metric which solves the Einstein field equations for all time. In isotropic coordinates, the Schwarzschild metric reads

$$ds^2 = -\left(\frac{1 - \frac{MG}{2rc^2}}{1 + \frac{MG}{2rc^2}}\right)^2 c^2 dt^2 + \left(1 + \frac{MG}{2rc^2}\right)^4 d\mathbf{x}^2, \quad (69)$$

where M is the gravitating mass of the black hole and $r^2 = (x^1)^2 + (x^2)^2 + (x^3)^2$, $d\mathbf{x}^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2$. It should be pointed out that the origin of the coordinate system $r = 0$ is not located where the Schwarzschild singularity $R = 0$ (radial Schwarzschild coordinate R) in Schwarzschild coordinates is located, rather it is located on the other side of the Einstein–Rosen bridge, at infinity. The relation between isotropic coordinates and Schwarzschild coordinates reads $R = r(1 + \frac{MG}{2rc^2})^2$ if $MG/2c^2 \leq r$ and $R' = r(1 + \frac{MG}{2rc^2})^2$ if $0 \leq r \leq MG/2c^2$,

where R' is another Schwarzschild radial coordinate appropriate for the geometry of the other side of the Einstein–Rosen bridge. The regimes $0 \leq R < 2GM/c^2$ and $0 \leq R' < 2GM/c^2$ are not accessible to isotropic coordinates. The harmonic radial coordinate, say here, ϱ , relates to the Schwarzschild radial coordinate through $R = \varrho + MG/c^2$. Evidently, the origin of the harmonic coordinates is located at $R = MG/c^2$ which is a spacelike curve in the region between event horizon and Schwarzschild singularity.

For two black holes, the metric for maximally sliced Brill–Lindquist initial time-symmetric data reads

$$ds^2 = - \left(\frac{1 - \frac{\beta_1 G}{2r_1 c^2} - \frac{\beta_2 G}{2r_2 c^2}}{1 + \frac{\alpha_1 G}{2r_1 c^2} + \frac{\alpha_2 G}{2r_2 c^2}} \right)^2 c^2 dt^2 + \left(1 + \frac{\alpha_1 G}{2r_1 c^2} + \frac{\alpha_2 G}{2r_2 c^2} \right)^4 d\mathbf{x}^2, \quad (70)$$

where the α_a coefficients are given in Eq. 46 and where the β_a coefficients can be found in [48] (notice $\partial_t r_a = 0$, initially).

The total energy results from the ADM mass-energy expression

$$E_{ADM} = - \frac{c^4}{2\pi G} \oint_{i_0} ds_i \partial_i \Psi = - \frac{c^4}{2\pi G} \int d^3\mathbf{x} \Delta \Psi = (\alpha_1 + \alpha_2) c^2, \quad (71)$$

where $\Psi = 1 + \frac{\alpha_1 G}{2r_1 c^2} + \frac{\alpha_2 G}{2r_2 c^2}$ and $ds_i = n^i r^2 d\Omega$ is a two-dimensional surface-area element with unit radial vector $n^i = x^i/r$ and solid angle element $d\Omega$.

Introducing the inversion map $r'_1 = \alpha_1^2 G^2 / 4c^4 r_1$ or, $\mathbf{r}'_1 = \mathbf{r}_1 \alpha_1^2 G^2 / 4c^4 r_1^2$ and $\mathbf{r}_1 = \mathbf{r}'_1 \alpha_1^2 G^2 / 4c^4 r_1'^2$, where $\mathbf{r}_1 = \mathbf{x} - \mathbf{x}_1$, $r_1 = |\mathbf{x} - \mathbf{x}_1|$, $\mathbf{r}'_1 = \mathbf{x}' - \mathbf{x}_1$, $r'_1 = |\mathbf{x}' - \mathbf{x}_1|$, the three-metric at the throat of black hole 1, $dl^2 = \Psi^4 d\mathbf{x}^2$, transforms into

$$dl^2 = \Psi'^4 d\mathbf{x}'^2 = \left(1 + \frac{\alpha_1 G}{2r'_1 c^2} + \frac{\alpha_1 \alpha_2 G^2}{4r_2 r'_1 c^4} \right)^4 d\mathbf{x}'^2, \quad (72)$$

where $\mathbf{r}_2 = \frac{\alpha_1^2 G^2}{4c^4} \frac{r'_1}{r_1^2} + \mathbf{r}_{12}$ with $\mathbf{r}_{12} = \mathbf{r}_1 - \mathbf{r}_2$. From the new metric function $\Psi' = 1 + \frac{\alpha_1 G}{2r'_1 c^2} + \frac{\alpha_1 \alpha_2 G^2}{4r_2 r'_1 c^4}$ the proper mass of the throat 1 results in, taking into account $\mathbf{r}_2 = \mathbf{r}'_1 \alpha_1^2 G^2 / 4c^4 r_1'^2 + \mathbf{r}_{12}$,

$$\begin{aligned} m_1 &= - \frac{c^2}{2\pi G} \oint_{i_0} ds'_i \partial'_i \Psi' = - \frac{c^2}{2\pi G} \int d^3\mathbf{x}' \Delta' \Psi' \\ &= \alpha_1 + \frac{\alpha_1 \alpha_2 G}{2r_{12} c^2}. \end{aligned} \quad (73)$$

This construction as performed in Ref. [16] is a purely geometrical or vacuum one without touching singularities. Thus having the two individual masses m_1 and m_2 the gravitational interaction energy is obtained as $E = E_{ADM} - (m_1 + m_2)c^2$.

Recall that this energy belongs to an initial value solution of the Einstein constraint equations with vanishing of both h_{ij}^{TT} and particle and field momenta. In this initial condition spurious gravitational waves are included.

Let us introduce now point masses as sources for the Schwarzschild black hole. The stress–energy tensor density of test-mass-point particles in a $(d+1)$ - dimensional curved spacetime reads

$$\mathcal{T}_\mu^\nu(x^\sigma) = c^2 \sum_a m_a u_{a\mu}(t) v_a^\nu(t) \delta_a, \quad (74)$$

where $g_a^{\mu\nu} u_{a\mu} u_{a\nu} = g_{a\mu\nu} u_a^\mu u_a^\nu = -1$, $v_a^\mu = u_a^\mu / u_a^0$, and $\delta_a = \delta(\mathbf{x} - \mathbf{x}_a(t))$ is the usual Dirac delta functions in d -dimensional flat space. In canonical framework, the momentum density π_i of the matter is given by

$$\pi_i = \frac{1}{c} \mathcal{T}_i^0 = c \sum_a m_a u_{ai} \delta_a. \quad (75)$$

It is the source term in the momentum constraint. The important energy density in the Hamilton constraint reads

$$-\mathcal{T}_\mu^0 n^\mu = -c^2 \sum_a m_a u_{a\mu} n^\mu \delta_a = c^2 \sum_a m_a u_a^0 N \delta_a, \quad (76)$$

where n^ν is the timelike unit vector, $n^\mu n_\mu = -1$ orthogonal to time slices $t = \text{const}$, $n_\mu = (-N, 0, \dots)$; N is the lapse function, see [1]. Independently from the form of the metric, the equations of motion are fulfilled and can be put into the form

$$v_a^\nu \nabla_\nu u_{a\mu} = 0, \quad (77)$$

which means geodesic motion for each particle.

The formal insertion of the stress–energy density into the Einstein field equations yields the following equations for the metric functions, for $u_{ai} = 0$,

$$-\Psi \Delta \phi = \frac{16\pi G}{c^2} \sum_a m_a \delta_a, \quad (78)$$

where

$$\gamma_{ij} = \Psi^{\frac{4}{d-2}} \delta_{ij}, \quad \Psi = 1 + \frac{d-2}{4(d-1)} \phi. \quad (79)$$

If the lapse function N is represented by

$$N = \frac{\chi}{\Psi}, \quad (80)$$

an equation for χ (be aware of the difference with χ_{\pm} in the Eqs. 64 and 65) results of the form,

$$\Psi^2 \Delta \chi = \frac{4\pi G}{c^2} \frac{d-2}{d-1} \sum_a m_a \chi \delta_a. \quad (81)$$

With the aid of the relation,

$$-\Delta^{-1} \delta = \frac{\Gamma((d-2)/2)}{4\pi^{d/2}} r^{2-d}, \quad (82)$$

it is easy to show that for $1 < d < 2$ the equations for Ψ and χ do have well-defined solutions. Plugging in the ansatz, for d dimensions,

$$\Psi = 1 + \frac{G(d-2)\Gamma((d-2)/2)}{c^2(d-1)\pi^{(d-2)/2}} \left(\frac{\alpha_1}{r_1^{d-2}} + \frac{\alpha_2}{r_2^{d-2}} \right), \quad (83)$$

gives, for “mass-point 1” (mass-point seems the better notion compared with point particle or point mass because it is a fictitious particle only),

$$\left(1 + \frac{G(d-2)\Gamma((d-2)/2)}{c^2(d-1)\pi^{(d-2)/2}} \left(\frac{\alpha_1}{r_1^{d-2}} + \frac{\alpha_2}{r_2^{d-2}} \right) \right) \alpha_1 \delta_1 = m_1 \delta_1 \quad (84)$$

or, taking $1 < d < 2$, and then taking the limit $r_1 \rightarrow 0$,

$$\left(1 + \frac{G(d-2)\Gamma((d-2)/2)}{c^2(d-1)\pi^{(d-2)/2}} \frac{\alpha_2}{r_{12}^{d-2}} \right) \alpha_1 \delta_1 = m_1 \delta_1. \quad (85)$$

Going over to $d = 3$ by arguing that the solutions are analytic in d just results in the equation, cf. Eq. 57,

$$\alpha_a = \frac{m_a}{1 + A \frac{\alpha_b}{r_{ab}}}, \quad (86)$$

where $b \neq a$ and $a, b = 1, 2$, with the solution shown in Eq. 46. The ADM energy is again given by, in the limit $d = 3$, see Eq. 71,

$$E_{ADM} = (\alpha_1 + \alpha_2)c^2. \quad (87)$$

Here we recognize the important property that although the Eqs. 46 and 47 may describe close binary black holes with strongly deformed apparent horizons, both black holes can still be generated by mass-points in conformally related flat space. This is the justification for our particle model to be taken as model for orbiting black holes. We also will argue that binary black holes generated by mass-points are

orbiting black holes without spin, that is, binary Schwarzschild-type black holes. We wish to point out that at the support of our δ -functions the physical spacetime is completely flat so they cannot be interpreted as local sources of gravity. They rather represent wormhole geometries. The geometrical vacuum calculations in [16] are completely finite, no infinite energies enter. The same holds with dimensional regularization where the formally infinite self-energies turn out to be zero. This is nicely consistent with the fact that the Dirac delta functions are living in flat space.

Working in the sense of distributions, the dimensional regularization procedure preserves the important law of “tweedling of products,” [43], $F_{reg}G_{reg} = (FG)_{reg}$, and gives all integrals, particularly the inverse Laplacian, a unique definition. In the sense of analytic functions, all integrals are well defined. A famous formula derived in [62] plays an all-over important role in PN calculations,

$$\int d^d \mathbf{x} r_1^\alpha r_2^\beta = \pi^{d/2} \frac{\Gamma(\frac{\alpha+d}{2})\Gamma(\frac{\beta+d}{2})\Gamma(-\frac{\alpha+\beta+d}{2})}{\Gamma(-\frac{\alpha}{2})\Gamma(-\frac{\beta}{2})\Gamma(\frac{\alpha+\beta+2d}{2})} r_{12}^{\alpha+\beta+d}. \quad (88)$$

It is well known that distributions or generalized functions like the Dirac delta function are boundary-value functions. To overcome distributional derivations like in

$$\partial_i \partial_j r^{2-d} = \text{Pf} \left((d-2) \frac{dn^i n^j - \delta_{ij}}{r^d} \right) - \frac{4\pi^{d/2}}{d\Gamma(d/2-1)} \delta_{ij} \delta, \quad (89)$$

where Pf denotes the Hadamard partie finie, it is very convenient to resort on the class of analytic functions introduced in [62],

$$\delta_\epsilon = \frac{\Gamma((d-\epsilon)/2)}{\pi^{d/2} 2^\epsilon \Gamma(\epsilon/2)} r^{\epsilon-d}, \quad (90)$$

resulting in the Dirac delta function in the limit

$$\delta = \lim_{\epsilon \rightarrow 0} \delta_\epsilon. \quad (91)$$

On this class of functions, the inverse Laplacian operates as

$$-\Delta^{-1} \delta_\epsilon = \frac{\Gamma((d-2-\epsilon)/2)}{4\pi^{d/2} 2^\epsilon \Gamma(\epsilon/2+1)} r^{\epsilon+2-d} = \delta_{\epsilon+2}, \quad (92)$$

which is a special case of the convolution property $\delta_\epsilon * \delta_{\epsilon'} = \delta_{\epsilon+\epsilon'}$ that also results in the formula of Eq. 88, and the second partial derivatives read

$$\partial_i \partial_j r^{\epsilon+2-d} = \text{Pf} \left((d-2-\epsilon) \frac{(d-\epsilon)n^i n^j - \delta_{ij}}{r^{d-\epsilon}} \right). \quad (93)$$

No delta-function distributions are involved. Though the replacement in the stress-energy tensor density of δ through δ_ϵ does destroy the divergence freeness of the

stress–energy tensor and thus the integrability conditions of the Einstein theory, the relaxed Einstein field equations (the ones that result after imposing coordinate conditions) do not force the stress–energy tensor to be divergence free and can thus be solved without problems. The solutions one gets do not fulfill the Einstein field equations but in the final limits of the ϵ_a going to zero the general coordinate covariance of the theory is recovered. This property, however, only holds if these limits were taken *before* the limit $d = 3$ is performed [24]. For completeness we give here the terms that violate the contracted Bianchi identities,

$$\nabla_\nu T_\mu^\nu = \frac{c^2}{2} \sum_a m_a (g_{\rho\sigma,\mu} - (g_{\rho\sigma,\mu})_a) v_a^\rho u_a^\sigma \delta_{\epsilon_a}. \quad (94)$$

We wish to point out here a difference between ADM formalism and the harmonic coordinates approach. If in the harmonic coordinates approach the stress–energy tensor is not divergence free the relaxed field equations can be solved, but the harmonic coordinate conditions will not be satisfied any further. This is different with the form we use the ADM formalism where the coordinate conditions are kept valid when solving the relaxed field equations. The relaxed field equations in the harmonic case include ten functions, which are just the metric coefficients, and the divergence freeness of the stress–energy tensor is achieved if on the solution space the harmonic coordinate conditions are imposed. In the ADM formalism in Routhian form the ten metric functions do fulfill the ADM coordinate conditions and equations of motion do follow from the Routhian. They, however, will not be the ones resulting from the Einstein field equations. Those will be obtained in the limits of $\epsilon_a \rightarrow 0$ only.

The method of dimensional regularization has proven fully successful in both approaches, the Hamiltonian one and the one using the Einstein field equations in harmonic coordinates. However, another important difference between both approaches should be mentioned. Whereas in the ADM approach all poles of the type $1/(d-3)$ cancel each other and no regularization constants are left, in the harmonic gauge approach poles survive with uncancelled constants [6, 22]. As found out, the difference is of gauge type only and can thus be eliminated by redefinition of the particle positions. On the other side, it shows that the positions of the mass-points in the Hamiltonian formalism are excellently chosen. Resorting to the maximally extended Schwarzschild metric, the spatial origin of the harmonic coordinates has Schwarzschild coordinate $R = MG/c^2$ inside horizon, which can be reached by observers whereas the spatial origin of the ADM coordinates is located on a spacelike hypersurface at $R' = \infty$ beyond horizon. The location of the origin of the ADM coordinates allows quite a nice control of the motion of the objects.

The ADM coordinate system we are using in our article is called asymptotically maximal slicing because the trace of the extrinsic curvature of the $t = \text{const}$ spacelike slices is not zero but decays as $1/r^3$ (in four spacetime dimensions) at spacelike infinity. It is closely related with the Dirac coordinate conditions, $(\gamma^{1/3} \gamma^{ij})_{,j} = 0$, $K_i^i = 0$, which introduce maximal slicing. Recently, maximal slicing coordinates of the type introduced in Ref. [30] have proved useful in numerical

relativity using moving punctures [38]. These coordinates are completely different from both the Dirac and ADM coordinates because those slices, for example, for Schwarzschild black holes, show crossings of the event horizon and settle down in the region between the Schwarzschild singularity and event horizon asymptotically. Only asymptotically these coordinates become rigidly connected to the Schwarzschild geometry. Nonetheless, moving punctures in numerical relativity are closely related with evolving Brill–Lindquist black holes.

3.5 PN Expansion of the Routh Functional

In case of the full Einstein theory, a formal PN expansion of the Routh functional in powers of $1/c^2$ is feasible. Using the definition $h_{ij}^{\text{TT}} = \frac{16\pi G}{c^4} \hat{h}_{ij}^{\text{TT}}$, we may write

$$R[x_a^i, p_{ai}, h_{ij}^{\text{TT}}, \partial_t h_{ij}^{\text{TT}}] - c^2 \sum_a m_a = \sum_{n=0}^{\infty} \frac{1}{c^{2n}} R_n[x_a^i, p_{ai}, \hat{h}_{ij}^{\text{TT}}, \partial_t \hat{h}_{ij}^{\text{TT}}]. \quad (95)$$

Furthermore, also the field equation for h_{ij}^{TT} can be put into a PN series form,

$$\left(\Delta - \frac{\partial_t^2}{c^2}\right) \hat{h}_{ij}^{\text{TT}} = \sum_{n=0}^{\infty} \frac{1}{c^{2n}} D_{(n)ij}^{\text{TT}}[x^k, x_a^k, p_{ak}, \hat{h}_{kl}^{\text{TT}}, \partial_t \hat{h}_{kl}^{\text{TT}}]. \quad (96)$$

This equation has to be solved iteratively with the aid of retarded integrals, which themselves have to be expanded in powers of $1/c$. In higher orders, however, log-of- $1/c$ terms will show up [4].

3.6 Near-Zone Energy Loss Versus Far-Zone Energy Flux

The change in time of the matter Hamiltonian (it is minus the Lagrangian for the gravitational field) reads, assuming \mathcal{R} to be *local* in the gravitational field,

$$\frac{dR}{dt} = \frac{\partial R}{\partial t} = \int \frac{\partial \mathcal{R}}{\partial h} \dot{h} + \int \frac{\partial \mathcal{R}}{\partial \nabla h} \nabla \dot{h} + \int \frac{\partial \mathcal{R}}{\partial \dot{h}} \ddot{h}, \quad (97)$$

where

$$R = R[x_a^i, p_{ai}, h, \dot{h}] = \int \mathcal{R}(x_a^i, p_{ai}, h, \nabla h, \dot{h}) \quad (98)$$

with abbreviations

$$\int \equiv \int d^3 \mathbf{x}, \quad h \equiv h_{ij}^{\text{TT}}, \quad \nabla h \equiv \partial_k h_{ij}^{\text{TT}}, \quad \dot{h} \equiv \partial_t h_{ij}^{\text{TT}}. \quad (99)$$

Above, the equation for dR/dt is valid provided the equations of motion

$$\dot{p}_{ai} = -\frac{\partial \mathcal{R}}{\partial x_a^i}, \quad \dot{x}_a^i = \frac{\partial \mathcal{R}}{\partial p_{ai}} \quad (100)$$

hold. Furthermore, we have

$$\begin{aligned} \int \frac{\partial \mathcal{R}}{\partial \nabla h} \nabla \dot{h} + \int \frac{\partial \mathcal{R}}{\partial \dot{h}} \ddot{h} &= \int \nabla \left(\frac{\partial \mathcal{R}}{\partial \nabla h} \dot{h} \right) + \frac{d}{dt} \int \left(\frac{\partial \mathcal{R}}{\partial \dot{h}} \dot{h} \right) \\ &\quad - \int \nabla \left(\frac{\partial \mathcal{R}}{\partial \nabla h} \right) \dot{h} - \int \frac{d}{dt} \left(\frac{\partial \mathcal{R}}{\partial \dot{h}} \right) \dot{h}. \end{aligned} \quad (101)$$

Introducing the canonical field momentum

$$\frac{c^3}{16\pi G} \pi = -\frac{\partial \mathcal{R}}{\partial \dot{h}}, \quad (102)$$

with abbreviation $\pi \equiv \pi_{\text{TT}}^{ij}$, and the Legendre transform

$$H = R + \frac{c^3}{16\pi G} \int \pi \dot{h}, \quad \text{or} \quad R = H - \frac{c^3}{16\pi G} \int \pi \dot{h}, \quad (103)$$

the energy loss equation takes the form

$$\frac{dH}{dt} = \int \nabla \left(\frac{\partial \mathcal{R}}{\partial \nabla h} \dot{h} \right) + \int \frac{\partial \mathcal{R}}{\partial \dot{h}} \ddot{h} - \int \nabla \left(\frac{\partial \mathcal{R}}{\partial \nabla h} \right) \dot{h} - \int \frac{d}{dt} \left(\frac{\partial \mathcal{R}}{\partial \dot{h}} \right) \dot{h}. \quad (104)$$

The application of the field equations

$$\frac{\partial \mathcal{R}}{\partial h} - \nabla \left(\frac{\partial \mathcal{R}}{\partial \nabla h} \right) - \frac{d}{dt} \left(\frac{\partial \mathcal{R}}{\partial \dot{h}} \right) = 0 \quad (105)$$

results in, employing the leading order quadratic field structure of \mathcal{R} ,

$$\begin{aligned} \frac{dH}{dt} &= \int \nabla \left(\frac{\partial \mathcal{R}}{\partial \nabla h} \dot{h} \right) = \oint_{\text{fz}} ds \frac{\partial \mathcal{R}}{\partial \nabla h} \dot{h} = \frac{c^4}{32\pi G} \oint_{\text{fz}} ds (\nabla h) \dot{h} \\ &= -\frac{c^3}{32\pi G} \oint_{\text{fz}} d\Omega r^2 \dot{h}^2, \end{aligned} \quad (106)$$

where ‘‘fz’’ denotes the far zone (see, e.g., Section 6.1), $d\Omega$ the solid-angle element, and r the radial coordinate of the two-surface of integration with surface-area element $ds = \mathbf{n} r^2 d\Omega$. Here the further *assumption* has been made that the volume integrals in Eq. 97 may have the outer-most region of the far zone as outer boundary. The expression

$$\mathcal{L} = \frac{c^3}{32\pi G} \oint_{\hat{t}_z} d\Omega r^2 (\dot{h}_{ij}^{\text{TT}})^2 \quad (107)$$

is the well known total energy flux (luminosity \mathcal{L}) of gravitational waves. The Newtonian and 1 PN wave generation fit into the above scheme of local Routhian density and far-zone as outer boundary, which can be inferred from [45].

4 Binary Point Masses to Higher PN Order

Most compact representations of dynamical systems are with Hamiltonians. Up to the 3.5 PN order, the Hamiltonian of binary point-mass systems is explicitly known, reading

$$\begin{aligned} H(t) = & m_1 c^2 + m_2 c^2 + H_N + \frac{1}{c^2} H_{[1PN]} + \frac{1}{c^4} H_{[2PN]} \\ & + \frac{1}{c^5} H_{[2.5PN]}(t) + \frac{1}{c^6} H_{[3PN]} + \frac{1}{c^7} H_{[3.5PN]}(t). \end{aligned} \quad (108)$$

The nonautonomous dissipative Hamiltonians $H_{[2.5PN]}(t)$ and $H_{[3.5PN]}(t)$ are written as explicitly depending on time because they depend on the gravitational field variables or, in case those are reduced to matter variables, on primed matter variables, see Section 4.4.

To simplify expressions like in Section 3.3, we go over to the center-of-mass frame $\mathbf{p}_1 + \mathbf{p}_2 = 0$ and also define

$$\begin{aligned} \tilde{H} = & (H - mc^2)/\mu, \quad \mu = m_1 m_2 / m, \quad m = m_1 + m_2, \quad \nu = \mu / m, \\ & \mathbf{p} = \mathbf{p}_1 / \mu, \quad p_r = (\mathbf{n} \cdot \mathbf{p}), \quad \mathbf{q} = (\mathbf{x}_1 - \mathbf{x}_2) / Gm, \quad \mathbf{n} = \mathbf{q} / |\mathbf{q}| \end{aligned} \quad (109)$$

with $0 \leq \nu \leq 1/4$ ($\nu = 0$ test-body case, $\nu = 1/4$ equal-mass case).

4.1 Conservative Hamiltonians

The conservative binary Hamiltonians read in reduced variables (the dissipative Hamiltonians will be treated in Section 4.4), see [20],

$$\tilde{H}_N = \frac{p^2}{2} - \frac{1}{q}, \quad (110)$$

$$\tilde{H}_{[1PN]} = \frac{1}{8}(3\nu - 1)p^4 - \frac{1}{2}[(3 + \nu)p^2 + \nu p_r^2] \frac{1}{q} + \frac{1}{2q^2}, \quad (111)$$

$$\begin{aligned}
\tilde{H}_{[2PN]} &= \frac{1}{16}(1 - 5\nu + 5\nu^2)p^6 \\
&\quad + \frac{1}{8}\left[(5 - 20\nu - 3\nu^2)p^4 - 2\nu^2 p_r^2 p^2 - 3\nu^2 p_r^4\right] \frac{1}{q} \\
&\quad + \frac{1}{2}\left[(5 + 8\nu)p^2 + 3\nu p_r^2\right] \frac{1}{q^2} - \frac{1}{4}(1 + 3\nu) \frac{1}{q^3}, \tag{112}
\end{aligned}$$

$$\begin{aligned}
\tilde{H}_{[3PN]} &= \frac{1}{128}(-5 + 35\nu - 70\nu^2 + 35\nu^3)p^8 \\
&\quad + \frac{1}{16}\left[(-7 + 42\nu - 53\nu^2 - 5\nu^3)p^6 + (2 - 3\nu)\nu^2 p_r^2 p^4 \right. \\
&\quad \quad \left. + 3(1 - \nu)\nu^2 p_r^4 p^2 - 5\nu^3 p_r^6\right] \frac{1}{q} \\
&\quad + \left[\frac{1}{16}(-27 + 136\nu + 109\nu^2)p^4 + \frac{1}{16}(17 + 30\nu)\nu p_r^2 p^2 \right. \\
&\quad \quad \left. + \frac{1}{12}(5 + 43\nu)\nu p_r^4\right] \frac{1}{q^2} \\
&\quad + \left[\left(-\frac{25}{8} + \left(\frac{1}{64}\pi^2 - \frac{335}{48}\right)\nu - \frac{23}{8}\nu^2\right)p^2 \right. \\
&\quad \quad \left. + \left(-\frac{85}{16} - \frac{3}{64}\pi^2 - \frac{7}{4}\nu\right)\nu p_r^2\right] \frac{1}{q^3} \\
&\quad + \left[\frac{1}{8} + \left(\frac{109}{12} - \frac{21}{32}\pi^2\right)\nu\right] \frac{1}{q^4}. \tag{113}
\end{aligned}$$

These Hamiltonians constitute an important element in the construction of templates for gravitational waves emitted from compact binaries. They serve also as basis of the effective one-body (EOB) approach, where with the aid of a canonical transformation the dynamics is put into test-body form of a deformed Schwarzschild metric [17]. From the reduced Hamiltonians, where a factor of $1/\nu$ is factorized out, the standard test-body dynamics is very easily obtained, simply by putting $\nu = 0$.

4.2 Dynamical Invariants

Dynamical invariants related to our previous dynamics are easily calculated within a Hamiltonian framework [19]. Let us denote the radial action by $i_r(E, j)$ with $E = \tilde{H}$ and $p^2 = p_r^2 + j^2/r^2$ ($\mathbf{p} = p_r \mathbf{e}_r + p_\varphi \mathbf{e}_\varphi$ with orthonormal basis $\mathbf{e}_r, \mathbf{e}_\varphi$ in the orbital plane). Then it holds

$$i_r(E, j) = \frac{1}{2\pi} \oint dr p_r, \tag{114}$$

where the integration is originally defined from minimum to minimum radial distance. Thus all expressions derived hereof relate to orbits completed in this sense. From analytical mechanics it is known that the phase of the completed orbit revolution Φ is given by

$$\frac{\Phi}{2\pi} = 1 + k = -\frac{\partial}{\partial j} i_r(E, j) \quad (115)$$

and the orbital period P reads

$$\frac{P}{2\pi Gm} = \frac{\partial}{\partial E} i_r(E, j). \quad (116)$$

Explicitly we get, for the periastron advance parameter k ,

$$k = \frac{1}{c^2} \frac{3}{j^2} \left\{ 1 + \frac{1}{c^2} \left[\frac{5}{4}(7-2\nu) \frac{1}{j^2} + \frac{1}{2}(5-2\nu) E \right] + \frac{1}{c^4} \left[a_1(\nu) \frac{1}{j^4} + a_2(\nu) \frac{E}{j^2} + a_3(\nu) E^2 \right] \right\}, \quad (117)$$

and for the orbital period,

$$\begin{aligned} \frac{P}{2\pi Gm} = \frac{1}{(-2E)^{3/2}} & \left\{ 1 - \frac{1}{c^2} \frac{1}{4}(15-\nu)E \right. \\ & + \frac{1}{c^4} \left[\frac{3}{2}(5-2\nu) \frac{(-2E)^{3/2}}{j} - \frac{3}{32}(35+30\nu+3\nu^2) E^2 \right] \\ & \left. + \frac{1}{c^6} \left[a_2(\nu) \frac{(-2E)^{3/2}}{j^3} - 3a_3(\nu) \frac{(-2E)^{5/2}}{j} + a_4(\nu) E^3 \right] \right\}, \quad (118) \end{aligned}$$

where

$$a_1(\nu) = \frac{5}{2} \left(\frac{77}{2} + \left(\frac{41}{64} \pi^2 - \frac{125}{3} \right) \nu + \frac{7}{4} \nu^2 \right), \quad (119)$$

$$a_2(\nu) = \frac{105}{2} + \left(\frac{41}{64} \pi^2 - \frac{218}{3} \right) \nu + \frac{45}{6} \nu^2, \quad (120)$$

$$a_3(\nu) = \frac{1}{4}(5-5\nu+4\nu^2), \quad (121)$$

$$a_4(\nu) = \frac{5}{128}(21-105\nu+15\nu^2+5\nu^3). \quad (122)$$

These expressions have direct applications to binary pulsars [28]. Explicit analytical orbit solutions of the conservative dynamics through 3 PN order are given in [54].

4.3 ISCO and the PN Framework

The motion of a test-body in the Schwarzschild metric is known to have its innermost stable circular orbit (ISCO) at $6MG/c^2$, in Schwarzschild coordinates. For test-bodies in rotating black holes (Kerr black holes) the ISCO lowers down up to MG/c^2 in case of direct motion and goes up to $9MG/c^2$ for retrograde motion, both motions in the equatorial plane. The ISCO of MG/c^2 is just the corotating case, where the Kerr black hole rotates as fast as the test-body is orbiting. In both limiting cases of test-body motion MG/c^2 and $9MG/c^2$, the black holes have maximal spins.

Within a Hamiltonian formalism the determination of the ISCO can be done straightforwardly. Just the following two equations have to be satisfied in the center-of-mass frame within the class of orbital circles ($p_r = 0$),

$$\frac{\partial H(r, J)}{\partial r} = 0 \quad (\text{dynamical circles}), \quad \frac{\partial^2 H(r, J)}{\partial r^2} = 0, \quad (123)$$

where r is the relative radial coordinate and J the orbital angular momentum. With the aid of the relation for the orbital frequency

$$\omega = \frac{dH(J)}{dJ}, \quad (124)$$

which holds for circular motion, the condition for the ISCO can also be put into the form

$$\frac{dH(\omega)}{d\omega} = 0. \quad (125)$$

Of course, one also could work with the Legendre transform $F(\omega) = H(J) - J\omega$ and put $\frac{d^2 F(\omega)}{d\omega^2} = 0$ but the outcome would be the same (recall, $dH = \omega dJ$, $dF = -Jd\omega$). Stability in the present approach with circular orbits means $\frac{d^2 F}{d\omega^2} > 0$ or, $\frac{dJ}{d\omega} < 0$ or $\frac{dH}{d\omega} < 0$.

In the context of approximation calculations, the main difference between the $H(r, J)$ and $H(\omega)$ approaches is that r is not a coordinate invariant variable; in the case of approximately known $H(r, J)$, this results in different values for the ISCO depending on the chosen coordinates. This is not the case with an approximately known $H(\omega)$ because it is coordinate invariant. Hereof, however, it does not follow that the ISCO calculated with $H(\omega)$ is more realistic than the ones calculated via $H(r, J)$ rather the varieties of ISCOs obtained via approximate $H(r, J)$ show up the uncertainty in their true location.

A test particle in the Schwarzschild spacetime on circular orbits has the reduced Hamiltonian

$$\begin{aligned} \hat{H}(x) &= \frac{1-2x}{(1-3x)^{1/2}} - 1 \\ &= -\frac{1}{2}x + \frac{3}{8}x^2 + \frac{27}{16}x^3 + \frac{675}{128}x^4 + \frac{3969}{256}x^5 + \dots, \end{aligned} \quad (126)$$

where

$$\hat{H}(x) \equiv \frac{H(x) - mc^2}{mc^2}, \quad x = \left(\frac{GM\omega}{c^3}\right)^{2/3}. \quad (127)$$

The condition $\frac{d\hat{H}(x)}{dx} = 0$ yields $x = 1/6 \approx 0.167$ or, in Schwarzschild coordinates $R = 6GM/c^2$. Evidently, ISCOs are located close to the reliability limit of PN expansions [20].

Using the dynamical invariants, the angular frequency of circular motion can be written as

$$\omega_{\text{circ}} = \omega_{\text{radial}} + \omega_{\text{periastron}} = 2\pi \frac{1+k}{P}. \quad (128)$$

With the aid of the definition $x = \left(\frac{GM\omega_{\text{circ}}}{c^3}\right)^{2/3}$, the binary dynamics yield, through 3PN order, that is, $c^2 \hat{H}_{3PN} = \tilde{H}_N + \frac{1}{c^2} \tilde{H}_{[1PN]} + \frac{1}{c^4} \tilde{H}_{[2PN]} + \frac{1}{c^6} \tilde{H}_{[3PN]}$,

$$\begin{aligned} \hat{H}_{3PN}(x) &= -\frac{x}{2} + \left(\frac{3}{8} + \frac{1}{24}\nu\right)x^2 + \left(\frac{27}{16} - \frac{19}{16}\nu + \frac{1}{48}\nu^2\right)x^3 \\ &\quad + \left(\frac{675}{128} + \left(-\frac{34445}{1152} + \frac{205}{192}\pi^2\right)\nu + \frac{155}{192}\nu^2 + \frac{35}{10368}\nu^3\right)x^4. \end{aligned} \quad (129)$$

The ISCO, calculated with the aid of Eq. 125, turns out to be $x \approx 0.255$ [3, 32].

To likely improve the PN truncation a representation of the energy is helpful, which in the test-particle limit is known to be a ratio of two simple polynomials for example,

$$e \equiv (1 + \hat{H})^2 - 1 = -x \frac{1-4x}{1-3x} \quad (130)$$

For the binary 3 PN Hamiltonian, e turns out to be

$$\begin{aligned} e(x) &= -x \left[1 - \left(1 - \frac{1}{3}\nu\right)x - \left(3 - \frac{35}{12}\nu\right)x^2 \right. \\ &\quad \left. - \left(9 + \frac{5}{24} \left(\frac{41}{4}\pi^2 - \frac{4309}{15}\nu\right) + \frac{103}{36}\nu^2 - \frac{1}{81}\nu^3\right)x^3 + \dots \right]. \end{aligned} \quad (131)$$

Applying to this expression the technique of padding, that is, putting it on a ratio-of-polynomials footing, the ISCO turned out to be, for equal-mass binaries, $x \approx 0.198$ [20].

To make contact to a recent discussion about the existence or nonexistence of ISCOs for equal-mass binaries in a PN setting [12], we discuss the more general

stability conditions for noncircular orbits using the Hamiltonian $H(r, p_r, J)$, where $J = p_\phi$. The crucial conditions for stability are

$$(1) : \frac{\partial^2 H}{\partial r^2} > 0, \quad (2) : \frac{\partial^2 H}{\partial p_r^2} > 0, \quad (3) : \frac{\partial^2 H}{\partial r^2} \frac{\partial^2 H}{\partial p_r^2} > 0. \quad (132)$$

Instability occurs if one of the $>$ -signs turns to zero. This particularly means that the $>$ -sign in (3) has to be zero. On the other side, in approximation calculations, where truncated series occur, the expression (3) can be zero without one of the expressions (1) or (2) being zero because the product of two series of the order n PN, which is of order $2n$ PN, is again truncated at n PN and thus can be zero without one or both of the factors being zero. In Ref. [12] the condition (3) has been given priority because it turned out to be coordinate invariant through 3 PN order (notice in this regard, $p_r = \partial W / \partial r$, where W is an action).

4.4 PN Dissipative Binary Dynamics

The leading order 2.5 PN dissipative binary orbital dynamics is described by the nonautonomous Hamiltonian, [65],

$$H_{[2.5PN]}(t) = \frac{2G}{5c^5} \frac{d^3 Q_{ij}(t)}{dt^3} \left(\frac{p_{1i} p_{1j}}{m_1} + \frac{p_{2i} p_{2j}}{m_2} - \frac{G m_1 m_2}{r_{12}} \right), \quad (133)$$

where

$$Q_{ij}(t) = \sum_{a=1,2} m_a (x_a^i x_a^{\prime j} - \frac{1}{3} \mathbf{x}_a^2 \delta_{ij}) \quad (134)$$

is the Newtonian mass-quadrupole tensor. Evidently, only after the Hamilton equations of motion are calculated the primed position and momentum variables resulting via $Q_{ij}(t)$ from time differentiations and use of the equations of motion are allowed to be identified with the unprimed position and momentum variables. The 3.5 PN Hamiltonian is known too, but it will not be given here because of quite lengthy expressions [45]. Applications of the 2.5 PN Hamiltonian can be found in, for example, [17, 37, 50, 63], where in Ref. [37] a transformation to the Burke–Thorne gauge (coordinate conditions) is performed.

5 Toward Binary Spinning Black Holes

Within the ADM formalism the action functional (i.e., the integral of the Lagrangian) of rotating bodies must have the following structure as long as the lengths of the spins are preserved in time,

$$W = \int dt \left(\sum_a p_{ai} \dot{x}_a^i + \sum_a S_a^{(i)} \Omega_a^{(i)} + \frac{c^3}{16\pi G} \int d^3\mathbf{x} \pi_{TT}^{ij} \dot{h}_{ij}^{TT} - H_{ADM} \left[x_a^i, p_{ai}, S_a^{(j)}, h_{ij}^{TT}, \pi_{TT}^{ij} \right] \right). \quad (135)$$

Here, $\Omega_a^{(i)} = \Omega_{a(i)} = \frac{1}{2} \epsilon_{ijk} \Lambda_{a(l)(j)} \dot{\Lambda}_{a(l)(k)}$, $\Lambda_{a(i)(k)} \Lambda_{a(j)(k)} = \Lambda_{a(k)(i)} \Lambda_{a(k)(j)} = \delta_{ij}$, $\epsilon_{ijk} = (i-j)(j-k)(k-i)/2$; and p_{ai} , x_a^i , $S_a^{(i)} = S_{a(i)}$, and $\delta\Theta_a^{(i)} = \frac{1}{2} \epsilon_{ijk} \Lambda_{a(l)(j)} \delta\Lambda_{a(l)(k)}$ are the independent matter variables, where the index a again numerates the particles. Notice that $\Theta_a^{(i)}$ are anholonomic variables related to the angle-type variables $\Lambda_{a(i)(j)}$ of the proper rotations (spins). $\Omega_a^{(i)}$ is the spin precession angular frequency vector of the a th particle. The equations of motion for the particles read,

$$\dot{x}_a^i(t) = \frac{\delta \int dt' H_{ADM}}{\delta p_{ai}(t)}, \quad \dot{p}_{ai}(t) = -\frac{\delta \int dt' H_{ADM}}{\delta x_a^i(t)}, \quad (136)$$

$$\Omega_a^{(i)}(t) = \frac{\delta \int dt' H_{ADM}}{\delta S_a^{(i)}(t)}, \quad \dot{S}_a^{(i)}(t) = \epsilon_{ijk} \Omega_a^{(j)}(t) S_a^{(k)}(t), \quad (137)$$

where the last equation results from the action functional through variation with respect to $\Theta_a^{(i)}$; for more details, see, for example, [39]. The Hamiltonian that generates both the evolution equations for all dynamical variables as well as the constraint equations through variation with respect to the Lagrange multipliers, the lapse and shift functions N and N^i , respectively, is given by

$$H = \int d^3\mathbf{x} (N\mathcal{H} - N^i \mathcal{H}_i) + E[\gamma_{ij}], \quad (138)$$

where

$$E[\gamma_{ij}] = \frac{c^4}{16\pi G} \oint_{i^0} ds_i (\gamma_{ij,j} - \gamma_{jj,i}) \quad (139)$$

is a surface integral at spacelike infinity i^0 with ds_i the two-dimensional surface-area element, [61], and

$$\mathcal{H} = \mathcal{H}^{\text{field}} + \mathcal{H}^{\text{matter}}, \quad (140)$$

$$\mathcal{H}_i = \mathcal{H}_i^{\text{field}} + \mathcal{H}_i^{\text{matter}}, \quad (141)$$

with

$$\frac{16\pi G}{c^4} \mathcal{H}^{\text{field}} = -\gamma^{1/2} \mathbf{R} + \frac{1}{\gamma^{1/2}} \left(\pi_j^i \pi_i^j - \frac{1}{2} \pi_i^i \pi_j^j \right), \quad (142)$$

$$\frac{16\pi G}{c^3} \mathcal{H}_i^{\text{field}} = 2\partial_j \pi_i^j + \pi^{kl} \partial_i \gamma_{kl}, \quad (143)$$

are the total Hamilton and (linear) momentum densities. After imposing the constraint equations $\mathcal{H} = \mathcal{H}_i = 0$ and the coordinate conditions (32–34), the energy expression $E[\gamma_{ij}]$ turns into the ADM-Hamiltonian,

$$E[\gamma_{ij}] = H \left[x_a^i, p_{ai}, S_a^{(i)}, h_{ij}^{\text{TT}}, \pi_{\text{TT}}^{ij} \right]. \quad (144)$$

To linear order in the spin variables, the matter densities read (recall, $\delta_a = \delta(\mathbf{x} - \mathbf{x}_a)$); on simplicity reasons the index a will not show up in the following equations)

$$\mathcal{H}^{\text{matter}} = -(np)c\delta - \frac{c}{2} t_{ij}^k \gamma^{ij}{}_{,k} - \left[\frac{cPl}{mc - np} \gamma^{ij} \gamma^{kl} \hat{S}_{jk} \delta \right]_{,i}, \quad (145)$$

$$\begin{aligned} \mathcal{H}_i^{\text{matter}} &= p_i \delta + \frac{1}{2} \left[\gamma^{mk} \hat{S}_{ik} \delta \right]_{,m} \\ &\quad - \left[\frac{PlPk}{np(mc - np)} (\gamma^{mk} \delta_i^p + \gamma^{mp} \delta_i^k) \gamma^{ql} \hat{S}_{qp} \delta \right]_{,m}, \end{aligned} \quad (146)$$

with

$$-np \equiv -n^\mu p_\mu = (m^2 c^2 + \gamma^{ij} p_i p_j)^{1/2} \quad (147)$$

and

$$t_{ij}^k = \gamma^{kl} \frac{\hat{S}_l(i P_j)}{np} \delta + \gamma^{kl} \gamma^{mn} \frac{\hat{S}_m(i P_j) P_n P_l}{(np)^2 (mc - np)} \delta, \quad (148)$$

where n^μ is the future directed unit vector field orthogonal to the $t = \text{const}$ hypersurfaces, $n_\mu = (-N, 0, 0, 0)$. The above expressions were shown to be correct up to (and including) the orders S/c^4 and S/c^2 in $\mathcal{H}^{\text{matter}}$ and $\mathcal{H}_i^{\text{matter}}$, respectively.

Introducing a dreibein field $e_j^{(i)}$ with $e_j^{(i)} e_k^{(i)} = \gamma_{jk}$ and $e_{(i)k} e_{(j)}^k = \delta_{ij}$, a spin tensor $S_{(k)(l)} = e_{(k)}^i e_{(l)}^j \hat{S}_{ij}$ can be introduced that fulfills the relation

$$\gamma^{ik} \gamma^{jl} \hat{S}_{ij} \hat{S}_{kl} = 2S_{(i)} S_{(j)} = \text{const}, \quad (149)$$

where $2S_{(i)} = \epsilon_{ijk} S_{(j)(k)}$. Crucial for our canonical formalism is the constancy in time of $S_{(i)} S_{(i)}$. Because of the symmetry property of the metric coefficients γ_{ij} , the symmetric root of γ_{ij}

$$e_{il} e_{lj} = \gamma_{ij}, \quad e_{ij} = e_{ji} \quad (150)$$

can be taken for the dreibein field, that is, $e_{(j)k} = e_{jk}$.

To the order the formalism has been developed consistently, the following relations hold,

$$p_{ai} = \int_{V_a} d^3\mathbf{x} \mathcal{H}_i^{\text{matter}}, \quad (151)$$

$$J_{aij} = \int_{V_a} d^3\mathbf{x} (x^i \mathcal{H}_j^{\text{matter}} - x^j \mathcal{H}_i^{\text{matter}}) = x_a^i p_{aj} - x_a^j p_{ai} + S_{a(i)(j)}, \quad (152)$$

where V_a denotes the volume of particle a . Furthermore,

$$\{x_a^i, p_{aj}\} = \delta_{ij}, \quad \{S_{a(i)}, S_{a(j)}\} = \epsilon_{ijk} S_{a(k)}, \quad \text{zero otherwise}, \quad (153)$$

and the total linear and angular momenta, respectively, take the forms $P_i = \sum_a p_{ai}$ and $J_{ij} = \sum_a J_{aij}$.

The crucial consistency relation reads,

$$\frac{\delta H^{\text{matter}}}{\delta \gamma^{ij}} = \frac{c}{2} N \sqrt{\gamma} T_{ij} \quad (154)$$

with

$$H^{\text{matter}} = \int d^3\mathbf{x} (N \mathcal{H}^{\text{matter}} - N^i \mathcal{H}_i^{\text{matter}}). \quad (155)$$

It is fulfilled to the needed order,

$$\sqrt{\gamma} T_{ij} = -\frac{P_i P_j}{np} \delta + t_{ij,k}^k + \mathcal{O}(G). \quad (156)$$

In asymptotically flat spacetimes the Poincaré group is a global symmetry group. Its generators P^μ and $J^{\mu\nu}$ are conserved and fulfill the Poincaré algebra, see, for example, [61],

$$\{P^\mu, P^\nu\} = 0, \quad (157)$$

$$\{P^\mu, J^{\rho\sigma}\} = -\eta^{\mu\rho} P^\sigma + \eta^{\mu\sigma} P^\rho, \quad (158)$$

$$\{J^{\mu\nu}, J^{\rho\sigma}\} = -\eta^{\nu\rho} J^{\mu\sigma} + \eta^{\mu\rho} J^{\nu\sigma} + \eta^{\sigma\mu} J^{\rho\nu} - \eta^{\sigma\nu} J^{\rho\mu}. \quad (159)$$

The meaning of the components are energy $P^0 = H/c$, linear momentum $P^i = P_i$, angular momentum $J^{ij} = J_{ij}$, and Lorentz boost $J^{i0}/c \equiv K^i = G^i - t P^i$. A center-of-mass vector can be defined by $X^i = c^2 G^i/H$. This vector, however, is not a canonical position vector, see, for example, [39]. The energy H and the center-of-mass vector $G^i = G_i$ have the representations

$$H = -\frac{c^4}{16\pi G} \int d^3\mathbf{x} \Delta\phi = -\frac{c^4}{16\pi G} \oint_{i0} r^2 d\Omega \mathbf{n} \nabla\phi, \quad (160)$$

$$G^i = -\frac{c^2}{16\pi G} \int d^3\mathbf{x} x^i \Delta\phi = -\frac{c^2}{16\pi G} \oint_{i0} r^2 d\Omega n^j (x^i \partial_j - \delta_{ij})\phi, \quad (161)$$

where i^0 denotes spacelike infinity, $r^2 d\Omega \mathbf{n}$ is the two-dimensional surface-area element, and \mathbf{n} the radial unit vector. The two quantities, H , G^i , are the most involved ones of those entering the Poincaré algebra.

In terms of three-dimensional quantities, the Poincaré algebra reads, see, for example [21], with $J_{ij} = \varepsilon_{ijk} J_k$,

$$\{P_i, H\} = \{J_i, H\} = 0, \quad (162)$$

$$\{J_i, P_j\} = \varepsilon_{ijk} P_k, \quad (163)$$

$$\{J_i, J_j\} = \varepsilon_{ijk} J_k, \quad (164)$$

$$\{J_i, G_j\} = \varepsilon_{ijk} G_k, \quad (165)$$

$$\{G_i, H\} = P_i, \quad (166)$$

$$\{G_i, P_j\} = \frac{1}{c^2} H \delta_{ij}, \quad (167)$$

$$\{G_i, G_j\} = -\frac{1}{c^2} \varepsilon_{ijk} J_k. \quad (168)$$

The Poincaré algebra has been extensively used in the calculations of PN Hamiltonians for spinning binaries [40,41]. Hereby the most important equation was (166), which tells that the total linear momentum has to be a total time derivative. Once this equation has even fixed the kinetic ambiguity in nondimensional regularization calculations [21]. The kinetic ambiguity got also fixed by a Lorentzian version of the Hadamard regularization based on the Fock-de Donder approach [10].

5.1 Approximate Hamiltonians for Spinning Binaries

All the Hamiltonians, and the center-of-mass vectors too, given in this section have been derived or rederived in recent papers by the author and his collaborators employing general relativity in canonical form [25,40,41,67,69], adjusting it to the motion of binary black holes.

The Hamiltonian of leading-order (LO) spin-orbit coupling reads

$$H_{SO}^{LO} = \sum_a \sum_{b \neq a} \frac{G}{c^2 r_{ab}^2} (\mathbf{S}_a \times \mathbf{n}_{ab}) \cdot \left[\frac{3m_b}{2m_a} \mathbf{p}_a - 2\mathbf{p}_b \right], \quad (169)$$

and the one of leading-order spin(1)–spin(2) coupling is given by

$$H_{S_1 S_2}^{LO} = \sum_a \sum_{b \neq a} \frac{G}{2c^2 r_{ab}^3} [3(\mathbf{S}_a \cdot \mathbf{n}_{ab})(\mathbf{S}_b \cdot \mathbf{n}_{ab}) - (\mathbf{S}_a \cdot \mathbf{S}_b)], \quad (170)$$

where $r_{ab} \mathbf{n}_{ab} = \mathbf{x}_a - \mathbf{x}_b$, $a \neq b$, and $a, b = 1, 2$. The more complicated Hamiltonian is the one with spin-squared terms because it relates to the rotational deformation of

spinning black holes. To leading order, say for spin(1), it reads (details of derivation are given later)

$$H_{S_1^2}^{LO} = \frac{Gm_2}{2c^2m_1r_{12}^3} [3(\mathbf{S}_1 \cdot \mathbf{n}_{12})(\mathbf{S}_1 \cdot \mathbf{n}_{12}) - (\mathbf{S}_1 \cdot \mathbf{S}_1)]. \quad (171)$$

The LO spin-orbit and spin(a)-spin(b) center-of-mass vectors take the form

$$\mathbf{G}_{SO}^{LO} = \sum_a \frac{1}{2c^2m_a} (\mathbf{p}_a \times \mathbf{S}_a), \quad \mathbf{G}_{S_1S_2}^{LO} = 0, \quad \mathbf{G}_{S_1^2}^{LO} = 0. \quad (172)$$

Within the conservative 3 PN dynamics for spinless point masses, the center-of-mass vector has been calculated in [21]. Applications of the LO spin Hamiltonians can be found in, for example, [2, 18, 66]. Other references treating the LO spin dynamics are, for example, [29, 49, 60, 71]. For applications of the next-to-leading-order (NLO) spin dynamics, presented straight below, see, for example, [5, 26].

The Hamiltonian of the NLO spin-orbit coupling reads, $r = r_{12}$,

$$\begin{aligned} H_{SO}^{NLO} = & -G \frac{((\mathbf{p}_1 \times \mathbf{S}_1) \cdot \mathbf{n}_{12})}{c^4 r^2} \left[\frac{5m_2 \mathbf{p}_1^2}{8m_1^3} + \frac{3(\mathbf{p}_1 \cdot \mathbf{p}_2)}{4m_1^2} - \frac{3\mathbf{p}_2^2}{4m_1 m_2} \right. \\ & \left. + \frac{3(\mathbf{p}_1 \cdot \mathbf{n}_{12})(\mathbf{p}_2 \cdot \mathbf{n}_{12})}{4m_1^2} + \frac{3(\mathbf{p}_2 \cdot \mathbf{n}_{12})^2}{2m_1 m_2} \right] \\ & + G \frac{((\mathbf{p}_2 \times \mathbf{S}_1) \cdot \mathbf{n}_{12})}{c^4 r^2} \left[\frac{(\mathbf{p}_1 \cdot \mathbf{p}_2)}{m_1 m_2} + \frac{3(\mathbf{p}_1 \cdot \mathbf{n}_{12})(\mathbf{p}_2 \cdot \mathbf{n}_{12})}{m_1 m_2} \right] \\ & + G \frac{((\mathbf{p}_1 \times \mathbf{S}_1) \cdot \mathbf{p}_2)}{c^4 r^2} \left[\frac{2(\mathbf{p}_2 \cdot \mathbf{n}_{12})}{m_1 m_2} - \frac{3(\mathbf{p}_1 \cdot \mathbf{n}_{12})}{4m_1^2} \right] \\ & - G^2 \frac{((\mathbf{p}_1 \times \mathbf{S}_1) \cdot \mathbf{n}_{12})}{c^4 r^3} \left[\frac{11m_2}{2} + \frac{5m_2^2}{m_1} \right] \\ & + G^2 \frac{((\mathbf{p}_2 \times \mathbf{S}_1) \cdot \mathbf{n}_{12})}{c^4 r^3} \left[6m_1 + \frac{15m_2}{2} \right] + (1 \leftrightarrow 2), \quad (173) \end{aligned}$$

and the one of NLO spin(1)-spin(2) coupling is given by

$$\begin{aligned} H_{S_1S_2}^{NLO} = & \frac{G}{2m_1 m_2 c^4 r^3} [6((\mathbf{p}_2 \times \mathbf{S}_1) \cdot \mathbf{n}_{12})((\mathbf{p}_1 \times \mathbf{S}_2) \cdot \mathbf{n}_{12}) \\ & + \frac{3}{2}((\mathbf{p}_1 \times \mathbf{S}_1) \cdot \mathbf{n}_{12})((\mathbf{p}_2 \times \mathbf{S}_2) \cdot \mathbf{n}_{12}) \\ & - 15(\mathbf{S}_1 \cdot \mathbf{n}_{12})(\mathbf{S}_2 \cdot \mathbf{n}_{12})(\mathbf{p}_1 \cdot \mathbf{n}_{12})(\mathbf{p}_2 \cdot \mathbf{n}_{12}) \\ & - 3(\mathbf{S}_1 \cdot \mathbf{n}_{12})(\mathbf{S}_2 \cdot \mathbf{n}_{12})(\mathbf{p}_1 \cdot \mathbf{p}_2) + 3(\mathbf{S}_1 \cdot \mathbf{p}_2)(\mathbf{S}_2 \cdot \mathbf{n}_{12})(\mathbf{p}_1 \cdot \mathbf{n}_{12}) \end{aligned}$$

$$\begin{aligned}
& + 3(\mathbf{S}_2 \cdot \mathbf{p}_1)(\mathbf{S}_1 \cdot \mathbf{n}_{12})(\mathbf{p}_2 \cdot \mathbf{n}_{12}) + 3(\mathbf{S}_1 \cdot \mathbf{p}_1)(\mathbf{S}_2 \cdot \mathbf{n}_{12})(\mathbf{p}_2 \cdot \mathbf{n}_{12}) \\
& + 3(\mathbf{S}_2 \cdot \mathbf{p}_2)(\mathbf{S}_1 \cdot \mathbf{n}_{12})(\mathbf{p}_1 \cdot \mathbf{n}_{12}) - 3(\mathbf{S}_1 \cdot \mathbf{S}_2)(\mathbf{p}_1 \cdot \mathbf{n}_{12})(\mathbf{p}_2 \cdot \mathbf{n}_{12}) \\
& + (\mathbf{S}_1 \cdot \mathbf{p}_1)(\mathbf{S}_2 \cdot \mathbf{p}_2) - \frac{1}{2}(\mathbf{S}_1 \cdot \mathbf{p}_2)(\mathbf{S}_2 \cdot \mathbf{p}_1) + \frac{1}{2}(\mathbf{S}_1 \cdot \mathbf{S}_2)(\mathbf{p}_1 \cdot \mathbf{p}_2) \\
& + \frac{3}{2m_1^2 r^3} [-((\mathbf{p}_1 \times \mathbf{S}_1) \cdot \mathbf{n}_{12})(\mathbf{p}_1 \times \mathbf{S}_2) \cdot \mathbf{n}_{12}] \\
& + (\mathbf{S}_1 \cdot \mathbf{S}_2)(\mathbf{p}_1 \cdot \mathbf{n}_{12})^2 - (\mathbf{S}_1 \cdot \mathbf{n}_{12})(\mathbf{S}_2 \cdot \mathbf{p}_1)(\mathbf{p}_1 \cdot \mathbf{n}_{12}) \\
& + \frac{3}{2m_2^2 r^3} [-((\mathbf{p}_2 \times \mathbf{S}_2) \cdot \mathbf{n}_{12})(\mathbf{p}_2 \times \mathbf{S}_1) \cdot \mathbf{n}_{12}] \\
& + (\mathbf{S}_1 \cdot \mathbf{S}_2)(\mathbf{p}_2 \cdot \mathbf{n}_{12})^2 - (\mathbf{S}_2 \cdot \mathbf{n}_{12})(\mathbf{S}_1 \cdot \mathbf{p}_2)(\mathbf{p}_2 \cdot \mathbf{n}_{12}) \\
& + \frac{6(m_1 + m_2)G^2}{c^4 r^4} [(\mathbf{S}_1 \cdot \mathbf{S}_2) - 2(\mathbf{S}_1 \cdot \mathbf{n}_{12})(\mathbf{S}_2 \cdot \mathbf{n}_{12})]. \tag{174}
\end{aligned}$$

The calculation of the LO and NLO order S_1^2 -Hamiltonians needs more information about the source terms than given in Eqs. 145 and 146. To achieve the $LO + NLO$ S_1^2 -Hamiltonians, the following additional source in the Hamilton constraint is needed,

$$\begin{aligned}
\mathcal{H}_{S_1^2, \text{static}}^{\text{matter}} &= -\frac{1}{2m_1} \left(Q_1^{ij} \delta_1 \right)_{;ij} + \frac{1}{8m_1} \gamma_{mn} \gamma^{pj} \gamma^{ql} \gamma^{mi}_{;p} \gamma^{nk}_{;q} \hat{S}_{1ij} \hat{S}_{1kl} \delta_1 \\
&+ \frac{1}{4m_1} \left(\gamma^{ij} \gamma^{mn} \gamma^{kl}_{;m} \hat{S}_{1ln} \hat{S}_{1jk} \delta_1 \right)_{;i}, \tag{175}
\end{aligned}$$

where; i and, i denote three-dimensional covariant and partial derivatives, respectively, and where

$$Q_1^{ij} \equiv \gamma^{ik} \gamma^{jl} \gamma^{mn} \hat{S}_{1km} \hat{S}_{1nl} + \frac{2}{3} \mathbf{S}_1^2 \gamma^{ij}, \tag{176}$$

$$2\mathbf{S}_1^2 = \gamma^{ik} \gamma^{jl} \hat{S}_{1ij} \hat{S}_{1kl} = \text{const.} \tag{177}$$

Q_1^{ij} is the quadrupole tensor of the black hole with number 1 resulting from its rotational deformation. Herewith, beyond the previously shown LO Hamiltonian (171), the NLO Hamiltonian comes out in the form, employing the Poincaré algebra for unique fixation of all coefficients,

$$\begin{aligned}
H_{S_1^2}^{NLO} &= \frac{G}{c^4 r^3} \left[\frac{m_2}{4m_1^3} (\mathbf{p}_1 \cdot \mathbf{S}_1)^2 - \frac{3}{4m_1 m_2} \mathbf{p}_2^2 \mathbf{S}_1^2 \right. \\
&\quad \left. + \frac{3m_2}{8m_1^3} (\mathbf{p}_1 \cdot \mathbf{n})^2 \mathbf{S}_1^2 - \frac{3m_2}{8m_1^3} \mathbf{p}_1^2 (\mathbf{S}_1 \cdot \mathbf{n})^2 \right]
\end{aligned}$$

$$\begin{aligned}
 & -\frac{3m_2}{4m_1^3} (\mathbf{p}_1 \cdot \mathbf{n}) (\mathbf{S}_1 \cdot \mathbf{n}) (\mathbf{p}_1 \cdot \mathbf{S}_1) + \frac{9}{4m_1 m_2} \mathbf{p}_2^2 (\mathbf{S}_1 \cdot \mathbf{n})^2 \\
 & + \frac{3}{4m_1^2} (\mathbf{p}_1 \cdot \mathbf{p}_2) \mathbf{S}_1^2 - \frac{9}{4m_1^2} (\mathbf{p}_1 \cdot \mathbf{p}_2) (\mathbf{S}_1 \cdot \mathbf{n})^2 \\
 & - \frac{3}{2m_1^2} (\mathbf{p}_1 \cdot \mathbf{n}) (\mathbf{p}_2 \cdot \mathbf{S}_1) (\mathbf{S}_1 \cdot \mathbf{n}) + \frac{3}{m_1^2} (\mathbf{p}_2 \cdot \mathbf{n}) (\mathbf{p}_1 \cdot \mathbf{S}_1) (\mathbf{S}_1 \cdot \mathbf{n}) \\
 & + \frac{3}{4m_1^2} (\mathbf{p}_1 \cdot \mathbf{n}) (\mathbf{p}_2 \cdot \mathbf{n}) \mathbf{S}_1^2 - \frac{15}{4m_1^2} (\mathbf{p}_1 \cdot \mathbf{n}) (\mathbf{p}_2 \cdot \mathbf{n}) (\mathbf{S}_1 \cdot \mathbf{n})^2 \Big] \\
 & - \frac{G^2 m_2}{2c^4 r^4} \left[9(\mathbf{S}_1 \cdot \mathbf{n})^2 - 5\mathbf{S}_1^2 + \frac{14m_2}{m_1} (\mathbf{S}_1 \cdot \mathbf{n})^2 - \frac{6m_2}{m_1} \mathbf{S}_1^2 \right]. \quad (178)
 \end{aligned}$$

The spin precession equations of the Hamiltonians H_{SO}^{NLO} and $H_{S_1^2}^{NLO}$ have been calculated also in the papers [57] and [58], respectively, where the first paper [57] has benefited from paper [69]. The final spin precession equation of the second paper [58], Eq. 60, deviates from the corresponding one in [67]. A detailed inspection has shown that the last term in Eq. 62 of [58] has a wrong sign [68]. Using the correct sign, after redefinition of the spin variable, agreement with the Hamiltonian of Ref. [67] is achieved.

The NLO order spin-orbit and spin(a)-spin(b) center-of-mass vectors take the form

$$\begin{aligned}
 \mathbf{G}_{SO}^{NLO} = & -\sum_a \frac{\mathbf{p}_a^2}{8c^4 m_a^3} (\mathbf{p}_a \times \mathbf{S}_a) \\
 & + \sum_a \sum_{b \neq a} \frac{m_b G}{4c^4 m_a r_{ab}} \left[((\mathbf{p}_a \times \mathbf{S}_a) \cdot \mathbf{n}_{ab}) \frac{5\mathbf{x}_a + \mathbf{x}_b}{r_{ab}} - 5(\mathbf{p}_a \times \mathbf{S}_a) \right] \\
 & + \sum_a \sum_{b \neq a} \frac{G}{c^4 r_{ab}} \left[\frac{3}{2} (\mathbf{p}_b \times \mathbf{S}_a) - \frac{1}{2} (\mathbf{n}_{ab} \times \mathbf{S}_a) (\mathbf{p}_b \cdot \mathbf{n}_{ab}) \right. \\
 & \quad \left. - ((\mathbf{p}_b \times \mathbf{S}_a) \cdot \mathbf{n}_{ab}) \frac{\mathbf{x}_a + \mathbf{x}_b}{r_{ab}} \right], \quad (179)
 \end{aligned}$$

$$\mathbf{G}_{S_1 S_2}^{NLO} = \frac{G}{2c^4} \sum_a \sum_{b \neq a} \left\{ [3(\mathbf{S}_a \cdot \mathbf{n}_{ab})(\mathbf{S}_b \cdot \mathbf{n}_{ab}) - (\mathbf{S}_a \cdot \mathbf{S}_b)] \frac{\mathbf{x}_a}{r_{ab}^3} + (\mathbf{S}_b \cdot \mathbf{n}_{ab}) \frac{\mathbf{S}_a}{r_{ab}^2} \right\}, \quad (180)$$

$$\mathbf{G}_{S_1^2}^{NLO} = \frac{2m_2 G}{c^4 m_1} \left[\frac{3(\mathbf{S}_1 \cdot \mathbf{n}_{12})^2}{8r_{12}^3} (\mathbf{x}_1 + \mathbf{x}_2) + \frac{\mathbf{S}_1^2}{8r_{12}^3} (3\mathbf{x}_1 - 5\mathbf{x}_2) - \frac{(\mathbf{S}_1 \cdot \mathbf{n}_{12}) \mathbf{S}_1}{r_{12}^2} \right]; \quad (181)$$

summing up through 2 PN order in the spin parts, we obtain, also see [21],

$$\mathbf{G} = \mathbf{G}_N + \mathbf{G}_{1PN} + \mathbf{G}_{2PN} + \mathbf{G}_{3PN} + \mathbf{G}_{SO}^{LO} + \mathbf{G}_{SO}^{NLO} + \mathbf{G}_{S_1 S_2}^{NLO} + \mathbf{G}_{S_1^2}^{NLO} + \mathbf{G}_{S_2^2}^{NLO}. \quad (182)$$

Numerically, spins of black holes can be counted of order $1/c$ (maximum), thus the spinless parts are taken up to the 3 PN order.

The currently known conservative binary Hamiltonians for spinning black holes through order $1/c^4$ can be summarized as follows

$$\begin{aligned} H = & H_N + H_{1PN} + H_{2PN} + H_{3PN} \\ & + H_{SO}^{LO} + H_{S_1 S_2}^{LO} + H_{S_1^2}^{LO} + H_{S_2^2}^{LO} \\ & + H_{SO}^{NLO} + H_{S_1 S_2}^{NLO} + H_{S_1^2}^{NLO} + H_{S_2^2}^{NLO} \\ & + H_{p_1 S_2^3} + H_{p_2 S_1^3} + H_{p_1 S_1^3} + H_{p_2 S_2^3} \\ & + H_{p_1 S_1 S_2^2} + H_{p_2 S_2 S_1^2} + H_{p_1 S_2 S_1^2} + H_{p_2 S_1 S_2^2} \\ & + H_{S_1^2 S_2^2} + H_{S_1 S_2^3} + H_{S_2 S_1^3}. \end{aligned} \quad (183)$$

The Hamiltonians $H_{S_1^4}$ and $H_{S_2^4}$ in the approximation in question turned out to be zero.

6 Lorentz-Covariant Approach and PN Expansions

The Lorentz-covariant approach has found a quite thorough presentation in [4]. So we will not go into so many details as in the canonical approach presented in the previous sections.

The Einstein field equations are given by

$$G^{\mu\nu}(g_{\kappa\lambda}, \partial_\alpha g_{\kappa\lambda}, \partial_\alpha \partial_\beta g_{\kappa\lambda}) = \frac{8\pi G}{c^2} \frac{T^{\mu\nu}(g_{\kappa\lambda}; c^2)}{c^2}, \quad (184)$$

where $g_{\kappa\lambda}$ and $T^{\mu\nu}$ are the four-metric and the stress–energy tensor of the matter (e.g., fluid), respectively. The contracted Bianchi identities yield the four-dimensional equations of motion (EOM) for the matter,

$$\nabla_\nu G^{\mu\nu} \equiv 0 \quad \rightarrow \quad \nabla_\nu T^{\mu\nu} = 0 \quad (\text{EOM}), \quad (185)$$

where ∇_ν denotes the four-dimensional covariant derivative. The Landau–Lifshitz form of the Einstein field equations fits very well into Lorentz-covariant schemes. It takes the form [51],

$$\partial_\lambda \partial_\kappa U^{\mu\nu\lambda\kappa}(g_{\alpha\beta}) = \frac{16\pi G}{c^4} \tau_{\text{LL}}^{\mu\nu}(g_{\alpha\beta}, \partial_\gamma g_{\alpha\beta}), \quad (186)$$

with $U^{\mu\nu\lambda\kappa} = \mathbf{g}^{\mu\nu} \mathbf{g}^{\lambda\kappa} - \mathbf{g}^{\mu\lambda} \mathbf{g}^{\nu\kappa}$ and $\mathbf{g}^{\mu\nu} = \sqrt{-g} g^{\mu\nu}$, where g denotes the determinant of the metric tensor. $t_{\text{LL}}^{\mu\nu}$ is known as the Landau–Lifshitz stress–energy (or energy–momentum) pseudo-tensor of the gravitational field. It is unique in the sense of symmetry and dependence on the metric coefficients and its first derivatives only. The equations of motion now read

$$\partial_\nu \partial_\lambda \partial_\kappa U^{\mu\nu\lambda\kappa} \equiv 0 \quad \rightarrow \quad \partial_\nu \tau_{\text{LL}}^{\mu\nu} = 0 \quad (\text{EOM}), \quad (187)$$

with

$$\tau_{\text{LL}}^{\mu\nu} \equiv -g T^{\mu\nu} + \frac{c^4}{16\pi G} t_{\text{LL}}^{\mu\nu}(g_{\alpha\beta}, \partial_\gamma g_{\alpha\beta}). \quad (188)$$

Applying the condition of harmonic coordinates

$$\partial_\nu \mathbf{g}^{\mu\nu} = 0 \quad \text{or,} \quad \partial_\nu H^{\mu\nu} = 0 \quad \text{with} \quad H^{\mu\nu} = \sqrt{-g} g^{\mu\nu} - \eta^{\mu\nu}, \quad (189)$$

where $\eta^{\mu\nu}$ denotes the Minkowski metric, the Einstein field equations in Landau–Lifshitz form read (so-called relaxed field equations because they do not imply the equations of motion; rather the condition of harmonic coordinates implies the equations of motion),

$$\eta^{\alpha\beta} \partial_\alpha \partial_\beta H^{\mu\nu} = \frac{16\pi G}{c^4} \tau^{\mu\nu}, \quad (190)$$

where

$$\tau^{\mu\nu} = -g T^{\mu\nu} + \frac{c^4}{16\pi G} \Lambda^{\mu\nu} \quad (191)$$

and

$$\begin{aligned} \Lambda^{\mu\nu} = & -H^{\alpha\beta} \partial_\alpha \partial_\beta H^{\mu\nu} + \partial_\alpha H^{\mu\beta} \partial_\beta H^{\nu\alpha} + \frac{1}{2} g^{\mu\nu} g_{\alpha\beta} \partial_\lambda H^{\alpha\tau} \partial_\tau H^{\beta\lambda} \\ & + \frac{1}{8} (2g^{\mu\alpha} g^{\nu\beta} - g^{\mu\nu} g^{\alpha\beta}) (2g_{\lambda\tau} g_{\rho\sigma} - g_{\lambda\sigma} g_{\tau\rho}) \partial_\alpha H^{\lambda\sigma} \partial_\beta H^{\tau\rho} \\ & - g^{\mu\alpha} g_{\beta\tau} \partial_\lambda H^{\nu\tau} \partial_\alpha H^{\beta\lambda} - g^{\nu\alpha} g_{\beta\tau} \partial_\lambda H^{\mu\tau} \partial_\alpha H^{\beta\lambda} \\ & + g_{\alpha\beta} g^{\lambda\tau} \partial_\lambda H^{\mu\alpha} \partial_\tau H^{\beta\nu}. \end{aligned} \quad (192)$$

The $\Lambda^{\mu\nu}$ object starts with quadratic nonlinearities of the gravitational field. It is another stress–energy pseudo-tensor of the gravitational field.

6.1 PM and PN Expansions

The *formal* retarded solution (resulting under the condition of no incoming radiation) of the inhomogeneous wave equation (190) reads

$$H^{\mu\nu}(\mathbf{x}, t) = -\frac{4G}{c^4} \int d^3\mathbf{x}' \tau^{\mu\nu}(\mathbf{x}', t - \frac{|\mathbf{x} - \mathbf{x}'|}{c}; c^2) |\mathbf{x} - \mathbf{x}'|^{-1}. \quad (193)$$

A PM expansion in powers of G can now be introduced in the form,

$$H^{\mu\nu}(\mathbf{x}, t) = \sum_{n=1}^{\infty} G^n H_{[n]}^{\mu\nu}(\mathbf{x}, t; c). \quad (194)$$

If additionally the virial theorem holds,

$$\frac{GM}{Rc^2} \sim \frac{V^2}{c^2}, \quad (195)$$

where respectively M , R , and V are typical masses, radii, and velocities of the system in question, the PM expansion may be further expanded into a PN series in powers of $1/c$. Yet, because of the retardation structure of the solution, a PN expansion is achievable only in the near and far zones and this only in a generalized form with log- c terms showing up at higher orders starting from 4 PN, that is, $(1/c^2)^4 = 1/c^8$, on.

Let us assume now that the matter source is bounded by a sphere with radius R centered in the origin of the coordinate system and that for the typical gravitational wave length λ the relation $\lambda \gg R$ holds. The near zone is then defined by $|\mathbf{x}| \ll \lambda$. The *formal* PN expansion (near-zone PN expansion) is defined by

$$H_{\text{nz}}^{\mu\nu}(\mathbf{x}, t) = -\frac{4G}{c^4} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int d^3\mathbf{x}' \frac{\partial^n}{c^n \partial t^n} \tau^{\mu\nu}(\mathbf{x}', t; c^2) |\mathbf{x} - \mathbf{x}'|^{n-1}. \quad (196)$$

Additionally, the expansion

$$\tau^{\mu\nu}(\mathbf{x}', t; c^2) = \sum_{n=-1}^{\infty} \frac{1}{c^{2n}} \tau_{(n)}^{\mu\nu}(\mathbf{x}', t) \quad (197)$$

applies. In the far zone, where $r = |\mathbf{x}| \gg \lambda$ holds, a *formal* PN expansion (far-zone PN expansion) yields

$$H_{\text{fz}}^{\mu\nu}(\mathbf{x}, t) = -\frac{4G}{c^4 r} \sum_{n=0}^{\infty} \frac{1}{n!} \int d^3\mathbf{x}' \frac{\partial^n}{c^n \partial t^n} \tau^{\mu\nu}(\mathbf{x}', t - \frac{r}{c}; c^2) (\mathbf{x}' \cdot \mathbf{n})^n. \quad (198)$$

Here, the expansion

$$\tau^{\mu\nu}\left(\mathbf{x}', t - \frac{r}{c}; c^2\right) = \sum_{n=-1}^{\infty} \frac{1}{c^{2n}} \tau_{[n]}^{\mu\nu}\left(\mathbf{x}', t - \frac{r}{c}\right) \quad (199)$$

applies. The Eqs. 196 and 197 on the one side and (198) and (199) on the other are somewhat simplified in the sense that they do not show up log- c terms at higher orders in the expansions that result from the badly defined integrals of non-compact support. On the other side, as the expressions stand, they are mathematically not defined at all. Details can be found in the works by Blanchet, Damour, Will, and collaborators; particularly see the contribution by L. Blanchet in this volume; for tail terms, also see our Section 6.3.

6.2 PN Expansion in the Near Zone

Up to the 2 PN order the metric coefficients read

$$g_{00} = -1 + \frac{2}{c^2} V - \frac{2}{c^4} V^2 + \frac{8}{c^6} \left(\hat{X} + V_i V_i + \frac{V^3}{6} \right), \quad (200)$$

$$g_{0i} = -\frac{4}{c^3} V_i - \frac{8}{c^5} \hat{R}_i, \quad (201)$$

$$g_{ij} = \delta_{ij} \left[1 + \frac{2}{c^2} V + \frac{2}{c^4} V^2 \right] + \frac{4}{c^4} \hat{W}_{ij}. \quad (202)$$

With the following choice of the matter variables, respectively, mass, mass-current, and stress density,

$$\sigma = \frac{T^{00} + T^{ii}}{c^2}, \quad \sigma_i = \frac{T^{0i}}{c}, \quad \sigma_{ij} = T^{ij}, \quad (203)$$

the 2 PN potentials can be put into the form

$$V(\mathbf{x}, t) = G_{\text{ret}} \{-4\pi G \sigma\} \equiv G \int \frac{d^3 \mathbf{z}}{|\mathbf{x} - \mathbf{z}|} \sigma(\mathbf{z}, t - |\mathbf{x} - \mathbf{z}|/c), \quad (204)$$

$$V_i = G_{\text{ret}} \{-4\pi G \sigma_i\}, \quad (205)$$

$$\hat{W}_{ij} = G_{\text{ret}} \{-4\pi G (\sigma_{ij} - \delta_{ij} \sigma_{kk}) - \partial_i V \partial_j V\}, \quad (206)$$

$$\hat{R}_i = G_{\text{ret}} \left\{ -4\pi G (V \sigma_i - V_i \sigma) - 2 \partial_k V \partial_i V_k - \frac{3}{2} \partial_t V \partial_i V \right\}, \quad (207)$$

$$\hat{X} = G_{\text{ret}} \left\{ -4\pi G V \sigma_{ii} + 2V_i \partial_t \partial_i V + V \partial_t^2 V + \frac{3}{2} (\partial_t V)^2 - 2\partial_i V_j \partial_j V_i + \hat{W}_{ij} \partial_{ij}^2 V \right\}. \quad (208)$$

The potentials of the orders 2.5 PN and 3.5 PN are radiation–reaction potentials. They are most compactly given under Burke–Thorne coordinate conditions, reading

$$U^{\text{reac}}(\mathbf{x}, t) = -\frac{G}{5c^5} x^{ij} \hat{M}_{ij}^{[5]}(t) + \frac{G}{c^7} \left[\frac{1}{189} x^{ijk} \hat{M}_{ijk}^{[7]}(t) - \frac{1}{70} x^{kk} x^{ij} \hat{M}_{ij}^{[7]}(t) \right], \quad (209)$$

$$U_i^{\text{reac}}(\mathbf{x}, t) = \frac{G}{c^5} \left[\frac{1}{21} \hat{x}^{ijk} \hat{M}_{jk}^{[6]}(t) - \frac{4}{45} \epsilon_{ijk} x^{jm} \hat{S}_{km}^{[5]}(t) \right], \quad (210)$$

where the source multipole moments are given by

$$\hat{M}_{ij} = \int d^3\mathbf{y} \left(\hat{y}^{ij} \sigma + \frac{1}{14c^2} y^{kk} \hat{y}^{ij} \partial_t^2 \sigma - \frac{20}{21c^2} \hat{y}^{ijk} \partial_t \sigma_k \right), \quad (211)$$

$$\hat{M}_{ijk} = \int d^3\mathbf{y} \hat{y}^{ijk} \sigma, \quad (212)$$

$$\hat{S}_{ij} = \int d^3\mathbf{y} \epsilon_{km<i} \hat{y}^{j>k} \sigma_m. \quad (213)$$

The used definitions read $y^{ij} \equiv y^i y^j$, $y^{<ij>} \equiv \hat{y}^{ij} = \text{STF}(y^{ij})$, and, for example, $\hat{M}_{ijk}^{[7]}$ indicates the seventh time derivative of M_{ijk} . Explicitly, the 1 PN metric including the gravitational radiation reaction through 3.5 PN order is given by

$$g_{00} = -1 + \frac{2}{c^2} (U + U^{\text{reac}}) + \frac{1}{c^4} [\partial_t^2 \chi - 2U^2 - 4UU^{\text{reac}}], \quad (214)$$

$$g_{0i} = -\frac{4}{c^3} (U_i + U_i^{\text{reac}}), \quad (215)$$

$$g_{ij} = \delta_{ij} \left[1 + \frac{2}{c^2} (U + U^{\text{reac}}) \right], \quad (216)$$

where the potentials have the integral representations

$$U(\mathbf{x}, t) = G \int \frac{d^3\mathbf{y}}{|\mathbf{x} - \mathbf{y}|} \sigma(\mathbf{y}, t), \quad (217)$$

$$U_i(\mathbf{x}, t) = G \int \frac{d^3\mathbf{y}}{|\mathbf{x} - \mathbf{y}|} \sigma_i(\mathbf{y}, t), \quad (218)$$

$$\chi(\mathbf{x}, t) = G \int d^3\mathbf{y} |\mathbf{x} - \mathbf{y}| \sigma(\mathbf{y}, t). \tag{219}$$

These integrals are well defined. Evidently, multipole expansion and PN expansion nicely fit together; see also [59].

6.3 PN Expansion in the Far Zone

In the far zone, the multipole expansion of the transverse-traceless (TT) part of the gravitational field, obtained by algebraic projection with $P_{ijkl}(\mathbf{n})$, reads, for example, [70],

$$H_{\text{fzTT}}^{ij}(\mathbf{x}, t) = -\frac{G}{c^4} \frac{P_{ijkl}(\mathbf{n})}{r} \sum_{l=2}^{\infty} \left\{ \left(\frac{1}{c^2} \right)^{\frac{l-2}{2}} \frac{4}{l!} M_{km i_3 \dots i_l}^{[l]} \left(t - \frac{r_*}{c} \right) N_{i_3 \dots i_l} \right. \\ \left. + \left(\frac{1}{c^2} \right)^{\frac{l-1}{2}} \frac{8l}{(l+1)!} \epsilon_{pq(k} S_{m) p i_3 \dots i_l}^{[l]} \left(t - \frac{r_*}{c} \right) n_q N_{i_3 \dots i_l} \right\}, \tag{220}$$

where the leading mass-quadrupole tensor takes the form, for example, [15],

$$M_{ij} \left(t - \frac{r_*}{c} \right) \\ = \widehat{M}_{ij} \left(t - \frac{r_*}{c} \right) \\ + \frac{2Gm}{c^3} \int_0^{\infty} dv \left[\ln \left(\frac{v}{2b} \right) + \frac{11}{12} \right] \widehat{M}_{ij}^{[2]} \left(t - \frac{r_*}{c} - v \right) + \mathcal{O} \left(\frac{1}{c^4} \right) \tag{221}$$

with

$$r_* = r + \frac{2Gm}{c^2} \ln \left(\frac{r}{cb} \right) + \mathcal{O} \left(\frac{1}{c^3} \right)$$

showing a leading-order tail term. Notice the modification of the standard PN expansion through tail terms. The Eq. 220 nicely shows that multipole expansions in the far zone also do induce PN expansions.

The gravitational luminosity is generally given by ($H_{\text{fzTT}}^{ij} = -h_{ij}^{\text{TTfz}}$),

$$\mathcal{L}(t) = \frac{c^3}{32\pi G} \oint_{\text{fz}} \left(\partial_t H_{\text{fzTT}}^{ij} \right)^2 r^2 d\Omega. \tag{222}$$

Through 1.5 PN order, the luminosity explicitly reads,

$$\begin{aligned} \mathcal{L}(t) &= \frac{G}{5c^5} \sum_{n=0}^{\infty} \left(\frac{1}{c^2}\right)^n \hat{\mathcal{L}}_n(t) \\ &= \frac{G}{5c^5} \left\{ M_{ij}^{[3]} M_{ij}^{[3]} + \frac{1}{c^2} \left[\frac{5}{189} M_{ijk}^{[4]} M_{ijk}^{[4]} + \frac{16}{9} S_{ij}^{[3]} S_{ij}^{[3]} \right] \right\}. \end{aligned} \quad (223)$$

On reasons of energy balance, for any representation of the Einstein theory, the time-averaged energy loss has to fulfill a relation of the form

$$- \left\langle \frac{d\mathcal{E}(t - \frac{r_*}{c})}{dt} \right\rangle = \langle \mathcal{L}(t) \rangle, \quad (224)$$

where the time averaging procedure takes into account typical periods of the system. The derivation of this equation in Section 3.6 is known to be valid for the first two radiation emissions and reaction levels.

7 Energy Loss and Gravitational Wave Emission

The energy flux to n PN order in the far zone, denoted n PN(fz), implies energy loss to $(n + 5/2)$ PN order in the near zone, denoted $(n + 5/2)$ PN(nz). Hereof it follows that energy-loss calculations are quite efficient via energy-flux calculations. Because of this we will apply the balance property between emitted and lost energies to some PN orders to easily derive the energy loss from the energy flux. In general, only after averaging over orbital periods will both expressions coincide (see Eq. 224). In the case of circular orbits, however, this averaging procedure is not needed.

7.1 Orbital Decay to 4 PN Order

The binding energy of our binary system on circular orbits is given by μE_{circ} . Therefore, for the energy loss to 4 PN order, we get

$$- \mu \frac{dE_{\text{circ}}}{dt} = \mathcal{L} = \frac{32c^5}{5G} v^2 x^5 \left[1 - \left(\frac{1247}{336} + \frac{35}{12} v \right) x + 4\pi x^{3/2} \right], \quad (225)$$

where the 1.5 PN(fz) energy flux is taken from Ref. [8] where the 2 PN(fz) energy flux also can be found; for the 3.5 PN(fz) energy flux see [11, 13].

Taking into account the Eq. 129 we obtain a differential equation for x which is easily solved with accuracy $1/c^8$. In terms of the dimensionless time variable

$$\tau = \frac{\nu c^3}{5Gm}(t_c - t), \quad (226)$$

where t_c denotes the coalescence time, the solution reads [8],

$$x = \frac{1}{4}\tau^{-1/4} \left[1 + \left(\frac{743}{4032} + \frac{11}{48}\nu \right) \tau^{-1/4} - \frac{1}{5}\pi\tau^{-3/8} \right]. \quad (227)$$

Taking into account the relation between phase and frequency $\frac{d\phi}{dt} = \omega$, respectively, $\frac{d\phi}{d\tau} = -\frac{5}{\nu}x^{3/2}$, the phase evolution results in

$$\phi = \phi_c - \frac{1}{\nu}\tau^{5/8} \left[1 + \left(\frac{3715}{8064} + \frac{55}{96}\nu \right) \tau^{-1/4} - \frac{3}{4}\pi\tau^{-3/8} \right]. \quad (228)$$

7.2 Gravitational Waveform to 1.5 PN Order

The radiation field can be decomposed into two orthogonal polarization states. The polarization states h_+ and h_\times are defined by

$$h_+ = \frac{1}{2}(u_i u_j - v_i v_j) h_{ij}^{\text{TT}}, \quad (229)$$

$$h_\times = \frac{1}{2}(u_i v_j + v_i u_j) h_{ij}^{\text{TT}}, \quad (230)$$

where \mathbf{u} and \mathbf{v} denote two vectors in the polarization plane forming an orthogonal right-handed triad with the direction \mathbf{n} from the source to the detector. The detector is directly sensitive to a linear combination of the polarization waveforms h_+ and h_\times , namely,

$$h(t) = F_+ h_+(t) + F_\times h_\times(t), \quad (231)$$

where F_+ and F_\times are the so-called beam-pattern functions of the detector depending on two angles giving the direction $-\mathbf{n}$ of the source as seen from the detector and a polarization angle specifying the orientation of the vectors \mathbf{u} and \mathbf{v} around that direction.

For our binary system, the two polarizations h_+ and h_\times are chosen such that the polarization vectors \mathbf{u} and \mathbf{v} lie, respectively, along the major and minor axis of the projection onto the plane of the sky of the circular orbit, with \mathbf{u} oriented toward the ascending node, the point at which black hole 1 crosses the plane of the sky moving toward the observer. The result, to 1.5 PN(fz) order, reads [14] (the 2 PN(fz) wave form is given therein too)

$$h_{+,\times} = \frac{2G\mu x}{c^2 r} \left[H_{+,\times}^{[0]} + x^{1/2} H_{+,\times}^{[1/2]} + x H_{+,\times}^{[1]} + x^{3/2} H_{+,\times}^{[3/2]} \right], \quad (232)$$

where the plus polarization is given by

$$H_+^{[0]} = -(1 + c_i^2)\cos 2\psi, \quad (233)$$

$$H_+^{[1/2]} = -\frac{s_i}{8} \frac{\delta m}{m} [(5 + c_i^2)\cos\psi - 9(1 + c_i^2)\cos 3\psi], \quad (234)$$

$$H_+^{[1]} = \frac{1}{6} [19 + 19c_i^2 - 2c_i^4 - \nu(19 - 11c_i^2 - 6c_i^4)]\cos 2\psi - \frac{4}{3} s_i^2 (1 + c_i^2)(1 - 3\nu)\cos 4\psi, \quad (235)$$

$$H_+^{[3/2]} = \frac{s_i}{192} \frac{\delta m}{m} \{ [57 + 60c_i^2 - c_i^4 - 2\nu(49 - 12c_i^2 - c_i^4)]\cos\psi - \frac{27}{2} [73 + 40c_i^2 - 9c_i^4 - 2\nu(25 - 8c_i^2 - 9c_i^4)]\cos 3\psi + \frac{625}{2} (1 - 2\nu)s_i^2 (1 + c_i^2)\cos 5\psi \} - 2\pi(1 + c_i^2)\cos 2\psi, \quad (236)$$

and the cross polarization by

$$H_\times^{[0]} = -2c_i \sin 2\psi, \quad (237)$$

$$H_\times^{[1/2]} = -\frac{3}{4} s_i c_i \frac{\delta m}{m} [\sin\psi - 3\sin 3\psi], \quad (238)$$

$$H_\times^{[1]} = \frac{c_i}{3} [17 - 4c_i^2 - \nu(13 - 12c_i^2)]\sin 2\psi - \frac{8}{3} c_i s_i^2 (1 - 3\nu)\sin 4\psi, \quad (239)$$

$$H_\times^{[3/2]} = \frac{s_i c_i}{96} \frac{\delta m}{m} \{ [63 - 5c_i^2 - 2\nu(23 - c_i^2)]\sin\psi - \frac{27}{2} [67 - 15c_i^2 - 2\nu(19 - 15c_i^2)]\sin 3\psi + \frac{625}{2} (1 - 2\nu)s_i^2 \sin 5\psi \} - 4\pi c_i \sin 2\psi, \quad (240)$$

where $c_i = \cos i$ and $s_i = \sin i$ and i denotes the inclination angle between the direction of the detector, as seen from the binary's center-of-mass, and the normal to the orbital plane which is assumed to be right-handed with respect to the sense of motion so that $0 \leq i \leq \pi$. $\delta m = m_1 - m_2$, and the phase variable ψ is given by

$$\psi = \phi - 3x^{3/2} \ln \left(\frac{x}{x_0} \right), \quad (241)$$

where ϕ is the actual orbital phase of the binary, namely, the angle oriented in the sense of motion between the ascending node and the direction of black hole

1 ($\phi = 0 \bmod 2\pi$ when the two black holes lie along \mathbf{u} , with black hole 1 at the ascending node). The logarithmic phase modulation originates from the propagation of tails in the wave zone. The constant scale x_0 can be chosen arbitrarily; it relates to the arbitrary constant b in the Eq. 221. For details on higher order PN levels, see, for example, [4].

Acknowledgements I thank the organizers of the Orléans School on Mass for their kind invitation and Luc Blanchet for helpful remarks on the manuscript.

References

1. R. Arnowitt, S. Deser, C.W. Misner, in *Gravitation: an Introduction to Current Research*, ed. by L. Witten (Wiley, New York, 1962), pp. 227–265
2. B.M. Barker, R.F. O’Connell, *Gen. Rel. Grav.* **11**, 149 (1979)
3. L. Blanchet, *Phys. Rev. D* **65**, 124009 (2002)
4. L. Blanchet, *Living Rev. Rel.* **9**, URL: <http://www.livingreviews.org/lrr-2006-4>
5. L. Blanchet, A. Buonanno, G. Faye, *Phys. Rev. D* **74**, 104034 (2006); Erratum, *ibidem* **75**, 049903(E) (2007); Erratum, *ibidem* **81**, 089901(E) (2010)
6. L. Blanchet, T. Damour, G. Esposito-Farèse, *Phys. Rev. D* **69**, 124007 (2004)
7. L. Blanchet, T. Damour, G. Esposito-Farèse, B.R. Iyer, *Phys. Rev. Lett.* **93**, 091101 (2004)
8. L. Blanchet, T. Damour, B.R. Iyer, *Phys. Rev. D* **51**, 5360 (1995)
9. L. Blanchet, T. Damour, G. Schäfer, *Mon. Not. R. Astron. Soc.* **242**, 289 (1990)
10. L. Blanchet, G. Faye, *J. Math. Phys.* **42**, 4391 (2001)
11. L. Blanchet, G. Faye, B.R. Iyer, B. Joguet, *Phys. Rev. D* **65**, 061501(R) (2002); Erratum, *ibidem* **71**, 129902(E) (2005)
12. L. Blanchet, B.R. Iyer, *Class. Q. Grav.* **20**, 755 (2003)
13. L. Blanchet, B.R. Iyer, B. Joguet, *Phys. Rev. D* **65**, 064005 (2002); Erratum, *ibidem* **71**, 129903(E) (2005)
14. L. Blanchet, B.R. Iyer, C.M. Will, A.G. Wiseman, *Class. Q. Grav.* **13** (1996)
15. L. Blanchet, G. Schäfer, *Class. Q. Grav.* **10**, 2699 (1993)
16. D.R. Brill, R.W. Lindquist, *Phys. Rev.* **131**, 471 (1963)
17. A. Buonanno, T. Damour, *Phys. Rev. D* **59**, 084006 (1999)
18. T. Damour, *Phys. Rev. D* **64**, 124013 (2001)
19. T. Damour, P. Jaranowski, G. Schäfer, *Phys. Rev. D* **62**, 044024 (2000)
20. T. Damour, P. Jaranowski, G. Schäfer, *Phys. Rev. D* **62**, 084011 (2000)
21. T. Damour, P. Jaranowski, G. Schäfer, *Phys. Rev. D*, **62**, 021501(R) (2000); Erratum, *ibidem* **63**, 029903(E) (2001)
22. T. Damour, P. Jaranowski, G. Schäfer, *Phys. Lett. B* **513**, 147 (2001)
23. T. Damour, P. Jaranowski, G. Schäfer, *Phys. Rev. D* **63**, 044021 (2001); Erratum, *ibidem* **66**, 029901(E) (2002)
24. T. Damour, P. Jaranowski, G. Schäfer, in *Proc. 11th Marcel Grossmann Mtg*, Berlin 23–29 July 2006, ed. by H. Kleinert, R.T. Jantzen, R. Ruffini (World Scientific, Singapore, 2008), pp. 2490–2492
25. T. Damour, P. Jaranowski, G. Schäfer, *Phys. Rev. D* **77**, 064032 (2008)
26. T. Damour, P. Jaranowski, G. Schäfer, *Phys. Rev. D* **78**, 024009 (2008)
27. T. Damour, G. Schäfer, *J. Math. Phys.* **32**, 127 (1991)
28. T. Damour, G. Schäfer, *Nuovo Cimento B* **101**, 127 (1988)
29. P.D. D’Eath, *Phys. Rev. D* **12**, 2183 (1975)
30. F. Estabrook, H. Wahlquist, S. Christensen, B. DeWitt, L. Smarr, E. Tsiang, *Phys. Rev. D* **7**, 2814 (1973)

31. G. Faye, L. Blanchet, A. Buonanno, Phys. Rev. D **74**, 104033 (2006)
32. G. Faye, P. Jaranowski, G. Schäfer, Phys. Rev. D **69**, 124029 (2004)
33. V. Fock, *The Theory of Space Time and Gravitation*, 2nd rev. edn. (Pergamon Press, New York, 1966)
34. T. Futamase, Y. Itoh, Living Rev. Rel. **10**, URL: <http://www.livingreviews.org/lrr-2007-2>
35. J.B. Gilmore, A. Ross, Phys. Rev. D **78**, 124021 (2008)
36. W. Goldberger, I.Z. Rothstein, Phys. Rev. D **73**, 104029 (2006)
37. A. Gopakumar, G. Schäfer, Phys. Rev. D **77**, 104023 (2008)
38. M. Hannam, S. Husa, D. Pollney, B. Brügmann, N. O’Murchadha, Phys. Rev. Lett. **99**, 241102 (2007)
39. A. Hanson, T. Regge, Ann. Phys. (N.Y.) **87**, 498 (1974)
40. S. Hergt, G. Schäfer, Phys. Rev. D **77**, 104001 (2008)
41. S. Hergt, G. Schäfer, Phys. Rev. D **78**, 124004 (2008)
42. D.D. Holm, Physica D **17**, 1 (1985)
43. L. Infeld, J. Plebański, *Motion and Relativity* (Pergamon, New York, 1960)
44. Y. Itoh, Phys. Rev. D **69**, 064018 (2004)
45. P. Jaranowski, G. Schäfer, Phys. Rev. D **55**, 4712 (1997); C. Königsdorffer, G. Faye, G. Schäfer, Phys. Rev. D **68**, 044004 (2003)
46. P. Jaranowski, G. Schäfer, Phys. Rev. D **57**, 7274 (1998); Erratum, ibidem **63**, 029902(E) (2001)
47. P. Jaranowski, G. Schäfer, Phys. Rev. D **60**, 124003 (1999) and **61**, 064008 (2000)
48. P. Jaranowski, G. Schäfer, Phys. Rev. D **65**, 127501 (2002)
49. L. Kidder, Phys. Rev. D **52**, 821 (1995)
50. K.D. Kokkotas, G. Schäfer, Mon. Not. R. Astron. Soc. **275**, 301 (1995)
51. L.D. Landau, E.M. Lifshitz, *The Classical Theory of Fields* (Pergamon Press, Oxford, 1985)
52. T. Ledvinka, G. Schäfer, J. Bičák, Phys. Rev. Lett. **100**, 251101 (2008)
53. R.W. Lindquist, J. Math. Phys. **4**, 938 (1963)
54. R.-M. Memmesheimer, A. Gopakumar, G. Schäfer, Phys. Rev. D **71**, 044021 (2005)
55. C.W. Misner, Ann. Phys. (N.Y.) **24**, 102 (1963)
56. M.E. Pati, C.M. Will, Phys. Rev. D **65**, 104008 (2002)
57. R.A. Porto, I.Z. Rothstein, Phys. Rev. D **78**, 044012 (2008)
58. R.A. Porto, I.Z. Rothstein, Phys. Rev. D **78**, 044013 (2008)
59. O. Poujade, L. Blanchet, Phys. Rev. D **65**, 124020 (2002)
60. E. Racine, A. Buonanno, L. Kidder, Phys. Rev. D **80**, 044010 (2009)
61. T. Regge, T. Teitelboim, Ann. Phys. (N.Y.) **88**, 286 (1974)
62. M. Riesz, Acta Math. **81**, 1 (1949)
63. M. Ruffert, H.-Th. Janka, G. Schäfer, Astron. Astrophys. **311**, 532 (1996)
64. G. Schäfer, Phys. Lett. A **100**, 128 (1984)
65. G. Schäfer, in *Symposia Gaussiana, Conf. A: Mathematical and Theoretical Physics*, ed. by M. Behara, R. Fritsch, R.G. Lintz (Walter de Gruyter, Berlin, 1995), pp. 667–679
66. G. Schäfer, Gen. Rel. Grav. **36**, 2223 (2004)
67. J. Steinhoff, S. Hergt, G. Schäfer, Phys. Rev. D **78**, 101503(R) (2008)
68. J. Steinhoff, G. Schäfer, Phys. Rev. D **80**, 088501 (2009)
69. J. Steinhoff, G. Schäfer, S. Hergt, Phys. Rev. D **77**, 104018 (2008)
70. K.S. Thorne, Rev. Mod. Phys. **52**, 299 (1980)
71. K.S. Thorne, J.B. Hartle, Phys. Rev. D **31**, 1815 (1985)
72. C.M. Will, Living Rev. Rel. **9**, URL: <http://www.livingreviews.org/lrr-2006-3>

The Effective One-Body Description of the Two-Body Problem

Thibault Damour and Alessandro Nagar

Abstract The effective one-body (EOB) formalism is an analytical approach which aims at providing an accurate description of the motion and radiation of coalescing binary black holes with arbitrary mass ratio. We review the basic elements of this formalism and discuss its aptitude at providing accurate template waveforms to be used for gravitational wave (GW) data analysis purposes.

1 Introduction

A network of ground-based interferometric gravitational wave (GW) detectors (LIGO/VIRGO/GEO/. . .) is currently taking data near its planned sensitivity [97]. Coalescing black-hole binaries are among the most promising, and most exciting, GW sources for these detectors. In order to successfully detect GWs from coalescing black-hole binaries, and to be able to reliably measure the physical parameters of the source (masses, spins, etc.), it is necessary to know in advance the shape of the GW signals emitted by inspiralling and merging black holes. Indeed, the detection and subsequent data analysis of GW signals is made by using a large bank of *templates* that accurately represent the GW waveforms emitted by the source.

Here, we shall introduce the reader to one promising strategy toward having an accurate analytical¹ description of the motion and radiation of binary black holes,

¹Here we use the adjective “analytical” for methods that solve explicit (analytically given) ordinary differential equations (ODE), even if one uses standard (Runge–Kutta-type) numerical tools to solve them. The important point is that, contrary to 3D numerical relativity (NR) simulations, numerically solving ODEs is extremely fast, and can therefore be done (possibly even in real time) for a dense sample of theoretical parameters, such as orbital ($v = m_1 m_2 / M, \dots$) or spin ($\hat{a}_1 = S_1 / Gm_1^2, \theta_1, \varphi_1, \dots$) parameters.

T. Damour and A. Nagar (✉)

Institut des Hautes Etudes Scientifiques, 35 Route de Chartres, F-91440 Bures-sur-Yvette, France
e-mail: nagar@ihes.fr

which covers all its stages (inspiral, plunge, merger, and ringdown): the *effective one-body* (EOB) approach [35, 36, 45, 55]. As early as 2000 [36] this method made several quantitative and qualitative predictions concerning the dynamics of the coalescence, and the corresponding GW radiation, notably: (i) a blurred transition from inspiral to a “plunge” that is just a smooth continuation of the inspiral; (ii) a sharp transition, around the merger of the black holes, between a continued inspiral and a ring-down signal; and (iii) estimates of the radiated energy and of the spin of the final black hole. In addition, the effects of the individual spins of the black holes were investigated within the EOB [33, 45] and were shown to lead to a larger energy release for spins parallel to the orbital angular momentum, and to a dimensionless rotation parameter J/E^2 always smaller than unity at the end of the inspiral (so that a Kerr black hole can form right after the inspiral phase). All those predictions have been broadly confirmed by the results of the recent numerical simulations performed by several independent groups [4–6, 8, 29–31, 39–42, 70, 71, 76, 82, 89–91, 94–96, 99, 101] (for a review of NR results see also [92]). Note that, in spite of the high computer power used in these simulations, the calculation of one sufficiently long waveform (corresponding to specific values of the many continuous parameters describing the two arbitrary masses, the initial spin vectors, and other initial data) takes of the order of 2 weeks. This is a very strong argument for developing analytical models of waveforms.

Those recent breakthroughs in NR open the possibility of comparing in detail the EOB description to NR results. This EOB/NR comparison has been initiated in several works [34, 37, 38, 59–62, 65, 66, 83, 85]. The level of analytical/numerical agreement is unprecedented, relative to what has been previously achieved when comparing other types of analytical waveforms to numerical ones. In particular, Refs. [38, 62] have compared two different kinds of analytical waveforms, computed within the EOB framework, to the most accurate GW form currently available from the Caltech–Cornell group, finding that the phase and amplitude differences are of the order of the numerical error.

If the reader wishes to put the EOB results in contrast with other (Post-Newtonian (PN) or hybrid) approaches he can consult, for example, [1, 2, 7, 29, 30, 72–74].

Before reviewing some of the technical aspects of the EOB method, let us indicate some of the historical roots of this method. First, we note that the EOB approach comprises three, rather separate, ingredients:

1. A description of the conservative (Hamiltonian) part of the dynamics of two black holes
2. An expression for the radiation–reaction part of the dynamics
3. A description of the GW waveform emitted by a coalescing binary system

For each one of these ingredients, the essential inputs that are used in EOB works are high-order PN expanded results that have been obtained by many years of work, by many researchers (see references below). However, one of the key ideas in the EOB philosophy is to avoid using PN results in their original “Taylor-expanded” form (i.e., $c_0 + c_1 v + c_2 v^2 + c_3 v^3 + \dots + c_n v^n$), but to use them instead in some *resummed* form (i.e., some non-polynomial function of v ,

defined so as to incorporate some of the expected non-perturbative features of the exact result). The basic ideas and techniques for resumming each ingredient of the EOB are different and have different historical roots. Concerning the first ingredient, that is, the EOB Hamiltonian, it was inspired by an approach to electromagnetically interacting quantum two-body systems introduced by Brézin, Itzykson, and Zinn–Justin [32].

The resummation of the second ingredient, that is, the EOB radiation–reaction force \mathcal{F} , was originally inspired by the Padé resummation of the flux function introduced by Damour, Iyer, and Sathyaprakash [53]. Recently, a new and more sophisticated resummation technique for the radiation reaction force \mathcal{F} has been introduced by Damour, Iyer, and Nagar [52] and further employed in EOB/NR comparisons [62]. It will be discussed in detail below.

As for the third ingredient, that is, the EOB description of the waveform emitted by a coalescing black-hole binary, it was mainly inspired by the work of Davis, Ruffini, and Tiomno [68], which discovered the transition between the plunge signal and a ringing tail when a particle falls into a black hole. Additional motivation for the EOB treatment of the transition from plunge to ringdown came from work on the so-called close limit approximation [93].

Let us finally note that the EOB approach has been recently improved [52, 61, 62] by following a methodology consisting of studying, element by element, the physics behind each feature of the waveform, and on systematically comparing various EOB-based waveforms with “exact” waveforms obtained by NR approaches. Among these “exact” NR waveforms, it has been useful to consider the small-mass-ratio limit² $\nu \equiv m_1 m_2 / (m_1 + m_2)^2 \ll 1$, in which one can use the well-controllable “laboratory” of numerical simulations of test particles (with an added radiation–reaction force) moving in black-hole backgrounds [60, 83].

2 Motion and Radiation of Binary Black Holes: PN Expanded Results

Before discussing the various resummation techniques used in the EOB approach, let us briefly recall the “Taylor-expanded” results that have been obtained by pushing to high accuracies the PN methods.

Concerning the orbital dynamics of compact binaries, we recall that the 2.5PN-accurate³ equations of motion have been derived in the 1980s [44, 46, 81, 98]. Pushing the accuracy of the equations of motion to the 3PN ($\sim(\nu/c)^6$) level proved to be a nontrivial task. At first, the representation of black holes by delta-function sources and the use of the (non-diffeomorphism invariant) Hadamard regularization

² Beware that the fonts used in this chapter make the greek letter ν (indicating the symmetric mass ratio) look very similar to the latin letter $\nu \neq \nu$ indicating the velocity.

³ As usual “ n -PN accuracy” means that a result has been derived up to (and including) terms which are $\sim(\nu/c)^{2n} \sim (GM/c^2 r)^n$ fractionally smaller than the leading contribution.

method led to ambiguities in the computation of the badly divergent integrals that enter the 3PN equations of motion [23, 78]. This problem was solved by using the (diffeomorphism invariant) *dimensional regularization* method (i.e., analytic continuation in the dimension of space d) which allowed one to complete the determination of the 3PN-level equations of motion [18, 56]. They have also been derived by an Einstein–Infeld–Hoffmann-type surface-integral approach [77]. The 3.5PN terms in the equations of motion are also known [80, 84, 87].

Concerning the emission of gravitational radiation, two different *gravitational-wave generation formalisms* have been developed up to a high PN accuracy: (i) the Blanchet–Damour–Iyer formalism [12, 13, 15–17, 49, 50] combines a multipolar post-Minkowskian (MPM) expansion in the exterior zone with a PN expansion in the near zone; while (ii) the Will–Wiseman–Pati formalism [86, 87, 104, 105] uses a direct integration of the relaxed Einstein equations. These formalisms were used to compute increasingly accurate estimates of the GW forms emitted by inspiralling binaries. These estimates include both normal, near-zone generated PN effects (at the 1PN [16], 2PN [21, 22, 105], and 3PN [26, 27] levels), and more subtle, wave-zone generated (linear and nonlinear) “tail effects” [13, 17, 28, 106]. However, technical problems arose at the 3PN level. Similarly to what happened with the equation of motion, the representation of black holes by “delta-function” sources causes the appearance of dangerously divergent integrals in the 3PN multipole moments. The use of Hadamard (*partie finie*) regularization did not allow one to unambiguously compute the needed 3PN-accurate quadrupole moment. Only the use of the (formally) diffeomorphism-invariant *dimensional regularization* method allowed one to complete the 3PN-level gravitational-radiation formalism [20].

The works mentioned in this section (see also Blanchet’s contribution in this volume, and [14] for a detailed account and more references) finally lead to PN-expanded results for the motion and radiation of binary black holes. For instance, the 3.5PN equations of motion are given in the form ($a = 1, 2; i = 1, 2, 3$)

$$\frac{d^2 z_a^i}{dt^2} = A_a^{i \text{ cons}} + A_a^{i RR}, \quad (1)$$

where

$$A^{\text{cons}} = A_0 + c^{-2} A_2 + c^{-4} A_4 + c^{-6} A_6, \quad (2)$$

denotes the “conservative” 3PN-accurate terms, while

$$A^{RR} = c^{-5} A_5 + c^{-7} A_7, \quad (3)$$

denotes the time-asymmetric contributions, linked to “radiation reaction.”

On the other hand, if we consider for simplicity the inspiralling motion of a quasi-circular binary system, the essential quantity describing the emitted GW form is the *phase* ϕ of the quadrupolar GW amplitude $h(t) \simeq a(t) \cos(\phi(t) + \delta)$. PN theory allows one to derive several different functional expressions for the

GW phase ϕ , as a function either of time or of the instantaneous frequency. For instance, as a function of time, ϕ admits the following explicit expansion in powers of $\theta \equiv \nu c^3(t_c - t)/5GM$ (where t_c denotes a formal “time of coalescence,” $M \equiv m_1 + m_2$ and $\nu \equiv m_1 m_2/M^2$)

$$\phi(t) = \phi_c - \nu^{-1} \theta^{5/8} \left(1 + \sum_{n=2}^7 (a_n + a'_n \ln \theta) \theta^{-n/8} \right), \quad (4)$$

with some numerical coefficients a_n, a'_n that depend only on the dimensionless (symmetric) mass ratio $\nu \equiv m_1 m_2/M^2$. The derivation of the 3.5PN-accurate expansion (4) uses both the 3PN-accurate conservative acceleration (2) and a 3.5PN extension of the (fractionally) 1PN-accurate radiation reaction acceleration (3) obtained by assuming a balance between the energy of the binary system and the GW energy flux at infinity (see, e.g., [14]).

Among the many other possible ways [54] of using PN-expanded results to predict the GW phase $\phi(t)$, let us mention the semi-analytic T4 approximant [7, 85]. The GW phase defined by the T4 approximant happens to agree well during the inspiral with the NR phase in the equal mass case [29]. However, this agreement seems to be coincidental because the T4 phase exhibits significant disagreement with NR results for other mass ratios [66] (as well as for spinning black holes [73]).

3 Conservative Dynamics of Binary Black Holes: the EOB Approach

The PN-expanded results briefly reviewed in the previous section are expected to yield accurate descriptions of the motion and radiation of binary black holes only during their *early inspiralling* stage, that is, as long as the PN expansion parameter $\gamma_e = GM/c^2 R$ (where R is the distance between the two black holes) stays significantly smaller than the value $\sim \frac{1}{6}$ where the orbital motion is expected to become dynamically unstable (“last stable circular orbit” and beginning of a “plunge” leading to the merger of the two black holes). One needs a better description of the motion and radiation to describe the *late inspiral* (say $\gamma_e \gtrsim \frac{1}{12}$), as well as the subsequent *plunge* and *merger*. One possible strategy for having a complete description of the motion and radiation of binary black holes, covering all the stages (inspiral, plunge, merger, ringdown), would then be to try to “stitch together” PN-expanded analytical results describing the early inspiral phase with 3-D numerical results describing the end of the inspiral, the plunge, the merger, and the ringdown of the final black hole, see, for example, Refs. [9, 85].

However, we wish to argue that the EOB approach makes a better use of all the analytical information contained in the PN-expanded results (1)–(3). The basic claim (first made in [35, 36]) is that the use of suitable *resummation methods* should allow one to describe, by analytical tools, a *sufficiently accurate* approximation of the *entire waveform*, from inspiral to ringdown, including the non-perturbative

plunge and merger phases. To reach such a goal, one needs to make use of several tools: (i) resummation methods, (ii) exploitation of the flexibility of analytical approaches, (iii) extraction of the non-perturbative information contained in various numerical simulations, (iv) qualitative understanding of the basic physical features which determine the waveform.

Let us start by discussing the first tool used in the EOB approach: the systematic use of resummation methods. Essentially two resummation methods have been employed (and combined) and some evidence has been given that they do significantly improve the convergence properties of PN expansions. The first method is the systematic use of *Padé approximants*. It has been shown in Ref. [53] that near-diagonal Padé approximants of the radiation reaction force⁴ \mathcal{F} seemed to provide a good representation of \mathcal{F} down to the last stable orbit (LSO) (which is expected to occur when $R \sim 6GM/c^2$, i.e. when $\gamma_e \simeq \frac{1}{6}$). In addition, a new route to the resummation of \mathcal{F} has been proposed very recently in Ref. [52]. This approach, that will be discussed in detail below, is based on a new multiplicative decomposition of the metric multipolar waveform (which is originally given as a standard PN series). In this case, Padé approximants prove to be useful to further improve the convergence properties of one particular factor of this multiplicative decomposition.

The second resummation method is a novel approach to the dynamics of compact binaries, which constitutes the core of the EOB method.

For simplicity of exposition, let us first explain the EOB method at the 2PN level. The starting point of the method is the 2PN-accurate Hamiltonian describing (in Arnowitt–Deser–Misner-type coordinates) the conservative, or time symmetric, part of the equations of motion (1) [i.e. the truncation $A^{\text{cons}} = A_0 + c^{-2}A_2 + c^{-4}A_4$ of Eq. 2] say $H_{2\text{PN}}(\mathbf{q}_1 - \mathbf{q}_2, \mathbf{p}_1, \mathbf{p}_2)$. By going to the center of mass of the system ($\mathbf{p}_1 + \mathbf{p}_2 = 0$), one obtains a PN-expanded Hamiltonian describing the *relative motion*, $\mathbf{q} = \mathbf{q}_1 - \mathbf{q}_2$, $\mathbf{p} = \mathbf{p}_1 = -\mathbf{p}_2$:

$$H_{2\text{PN}}^{\text{relative}}(\mathbf{q}, \mathbf{p}) = H_0(\mathbf{q}, \mathbf{p}) + \frac{1}{c^2} H_2(\mathbf{q}, \mathbf{p}) + \frac{1}{c^4} H_4(\mathbf{q}, \mathbf{p}), \quad (5)$$

where $H_0(\mathbf{q}, \mathbf{p}) = \frac{1}{2\mu} \mathbf{p}^2 + \frac{GM\mu}{|\mathbf{q}|}$ (with $M \equiv m_1 + m_2$ and $\mu = m_1 m_2 / M$) corresponds to the Newtonian approximation to the relative motion, while H_2 describes 1PN corrections and H_4 2PN ones. It is well known that, at the Newtonian approximation, $H_0(\mathbf{q}, \mathbf{p})$ can be thought of as describing a “test particle” of mass μ orbiting around an “external mass” GM . The EOB approach is a *general relativistic generalization* of this fact. It consists in looking for an “external spacetime geometry” $g_{\mu\nu}^{\text{ext}}(x^\lambda; GM)$ such that the geodesic dynamics of a “test particle” of mass μ within $g_{\mu\nu}^{\text{ext}}(x^\lambda, GM)$ is *equivalent* (when expanded in powers of $1/c^2$) to the original, relative PN-expanded dynamics (5).

⁴ We henceforth denote by \mathcal{F} the *Hamiltonian* version of the radiation reaction term A^{RR} , Eq. 3, in the (PN-expanded) equations of motion. It can be heuristically computed up to (absolute) 5.5PN [19, 20, 27] and even 6PN [24] order by assuming that the energy radiated in GW at infinity is balanced by a loss of the dynamical energy of the binary system.

Let us explain the idea, proposed in [35], for establishing a “dictionary” between the real relative-motion dynamics, (5), and the dynamics of an “effective” particle of mass μ moving in $g_{\mu\nu}^{\text{ext}}(x^\lambda, GM)$. The idea consists in “thinking quantum mechanically.”⁵ Instead of thinking in terms of a classical Hamiltonian, $H(\mathbf{q}, \mathbf{p})$ [such as $H_{2\text{PN}}^{\text{relative}}$, Eq. 5], and of its classical bound orbits, we can think in terms of the quantized energy levels $E(n, \ell)$ of the quantum bound states of the Hamiltonian operator $H(\hat{\mathbf{q}}, \hat{\mathbf{p}})$. These energy levels will depend on two (integer valued) quantum numbers n and ℓ . Here (for a spherically symmetric interaction, as appropriate to H^{relative}), ℓ parametrizes the total orbital angular momentum ($\mathbf{L}^2 = \ell(\ell + 1)\hbar^2$), while n represents the “principal quantum number” $n = \ell + n_r + 1$, where n_r (the “radial quantum number”) denotes the number of nodes in the radial wave function. The third “magnetic quantum number” m (with $-\ell \leq m \leq \ell$) does not enter the energy levels because of the spherical symmetry of the two-body interaction (in the center of mass frame). For instance, a nonrelativistic Coulomb (or Newton!) interaction

$$H_0 = \frac{1}{2\mu} \mathbf{p}^2 + \frac{GM\mu}{|\mathbf{q}|}, \tag{6}$$

gives rise to the well-known result

$$E_0(n, \ell) = -\frac{1}{2} \mu \left(\frac{GM\mu}{n\hbar} \right)^2, \tag{7}$$

which depends only on n (this is the famous Coulomb degeneracy). When considering the PN corrections to H_0 , as in Eq. 5, one gets a more complicated expression of the form

$$E_{2\text{PN}}^{\text{relative}}(n, \ell) = -\frac{1}{2} \mu \frac{\alpha^2}{n^2} \left[1 + \frac{\alpha^2}{c^2} \left(\frac{c_{11}}{n\ell} + \frac{c_{20}}{n^2} \right) + \frac{\alpha^4}{c^4} \left(\frac{c_{13}}{n\ell^3} + \frac{c_{22}}{n^2\ell^2} + \frac{c_{31}}{n^3\ell} + \frac{c_{40}}{n^4} \right) \right], \tag{8}$$

where we have set $\alpha \equiv GM\mu/\hbar = G m_1 m_2/\hbar$, and where we consider, for simplicity, the (quasi-classical) limit where n and ℓ are large numbers. The 2PN-accurate result (8) had been derived by Damour and Schäfer [67] as early as 1988. The dimensionless coefficients c_{pq} are functions of the symmetric mass ratio $\nu \equiv \mu/M$, for instance $c_{40} = \frac{1}{8}(145 - 15\nu + \nu^2)$. In classical mechanics (i.e., for large n and ℓ), it is called the “Delaunay Hamiltonian”, that is, the Hamiltonian expressed in terms of the *action variables*⁶ $J = \ell\hbar = \frac{1}{2\pi} \oint p_\varphi d\varphi$, and $N = n\hbar = I_r + J$, with $I_r = \frac{1}{2\pi} \oint p_r dr$.

⁵ This is related to an idea emphasized many times by John Archibald Wheeler: quantum mechanics can often help us in going to the essence of classical mechanics.

⁶ We consider, for simplicity, “equatorial” motions with $m = \ell$, that is, classically, $\theta = \frac{\pi}{2}$.

The energy levels (8) encode, in a gauge-invariant way, the 2PN-accurate relative dynamics of a “real” binary. Let us now consider an auxiliary problem: the “effective” dynamics of one body, of mass μ , following a geodesic in some “external” (spherically symmetric) metric⁷

$$g_{\mu\nu}^{\text{ext}} dx^\mu dx^\nu = -A(R) c^2 dT^2 + B(R) dR^2 + R^2(d\theta^2 + \sin^2\theta d\varphi^2). \quad (9)$$

Here, the *a priori unknown* metric functions $A(R)$ and $B(R)$ will be constructed in the form of expansions in $GM/c^2 R$:

$$\begin{aligned} A(R) &= 1 + a_1 \frac{GM}{c^2 R} + a_2 \left(\frac{GM}{c^2 R}\right)^2 + a_3 \left(\frac{GM}{c^2 R}\right)^3 + \dots; \\ B(R) &= 1 + b_1 \frac{GM}{c^2 R} + b_2 \left(\frac{GM}{c^2 R}\right)^2 + \dots, \end{aligned} \quad (10)$$

where the dimensionless coefficients a_n, b_n depend on ν . From the Newtonian limit, it is clear that we should set $a_1 = -2$. By solving (by separation of variables) the “effective” Hamiltonian–Jacobi equation

$$g_{\text{eff}}^{\mu\nu} \frac{\partial S_{\text{eff}}}{\partial x^\mu} \frac{\partial S_{\text{eff}}}{\partial x^\nu} + \mu^2 c^2 = 0,$$

$$S_{\text{eff}} = -\mathcal{E}_{\text{eff}} t + J_{\text{eff}} \varphi + S_{\text{eff}}(R), \quad (11)$$

one can straightforwardly compute (in the quasi-classical, large quantum numbers limit) the Delaunay Hamiltonian $\mathcal{E}_{\text{eff}}(N_{\text{eff}}, J_{\text{eff}})$, with $N_{\text{eff}} = n_{\text{eff}} \hbar$, $J_{\text{eff}} = \ell_{\text{eff}} \hbar$ (where $N_{\text{eff}} = J_{\text{eff}} + I_R^{\text{eff}}$, with $I_R^{\text{eff}} = \frac{1}{2\pi} \oint p_R^{\text{eff}} dR$, $P_R^{\text{eff}} = \partial S_{\text{eff}}(R)/dR$). This yields a result of the form

$$\begin{aligned} \mathcal{E}_{\text{eff}}(n_{\text{eff}}, \ell_{\text{eff}}) &= \mu c^2 - \frac{1}{2} \mu \frac{\alpha^2}{n_{\text{eff}}^2} \left[1 + \frac{\alpha^2}{c^2} \left(\frac{c_{11}^{\text{eff}}}{n_{\text{eff}} \ell_{\text{eff}}} + \frac{c_{20}^{\text{eff}}}{n_{\text{eff}}^2} \right) \right. \\ &\quad \left. + \frac{\alpha^4}{c^4} \left(\frac{c_{13}^{\text{eff}}}{n_{\text{eff}} \ell_{\text{eff}}^3} + \frac{c_{22}^{\text{eff}}}{n_{\text{eff}}^2 \ell_{\text{eff}}^2} + \frac{c_{31}^{\text{eff}}}{n_{\text{eff}}^3 \ell_{\text{eff}}} + \frac{c_{40}^{\text{eff}}}{n_{\text{eff}}^4} \right) \right], \end{aligned} \quad (12)$$

where the dimensionless coefficients c_{pq}^{eff} are now functions of the unknown coefficients a_n, b_n entering the looked for “external” metric coefficients (10).

At this stage, one needs (as in the famous AdS/CFT correspondence) to define a “dictionary” between the real (relative) two-body dynamics, summarized in Eq. 8, and the effective one-body one, summarized in Eq. 12. As, on both sides, quantum

⁷ It is convenient to write the “external metric” in Schwarzschild-like coordinates. Note that the external radial coordinate R differs from the two-body ADM-coordinate relative distance $R^{\text{ADM}} = |\mathbf{q}|$. The transformation between the two coordinate systems has been determined in Refs. [35,55].

mechanics tells us that the action variables are quantized in integers ($N_{\text{real}} = n\hbar$, $N_{\text{eff}} = n_{\text{eff}}\hbar$, etc.) it is most natural to identify $n = n_{\text{eff}}$ and $\ell = \ell_{\text{eff}}$. One then still needs a rule for relating the two different energies $E_{\text{real}}^{\text{relative}}$ and \mathcal{E}_{eff} . Ref. [35] proposed to look for a general map between the real energy levels and the effective ones (which, as seen when comparing (8) and (12), cannot be directly identified because they do not include the same rest-mass contribution⁸), namely

$$\frac{\mathcal{E}_{\text{eff}}}{\mu c^2} - 1 = f\left(\frac{E_{\text{real}}^{\text{relative}}}{\mu c^2}\right) = \frac{E_{\text{real}}^{\text{relative}}}{\mu c^2} \left(1 + \alpha_1 \frac{E_{\text{real}}^{\text{relative}}}{\mu c^2} + \alpha_2 \left(\frac{E_{\text{real}}^{\text{relative}}}{\mu c^2}\right)^2 + \dots\right). \tag{13}$$

The ‘‘correspondence’’ between the real and effective energy levels is illustrated in Fig. 1.

Finally, identifying $\mathcal{E}_{\text{eff}}(n, \ell)/\mu c^2$ to $f(E_{\text{real}}^{\text{relative}}/\mu c^2)$ yields six equations, relating the six coefficients $c_{pq}^{\text{eff}}(a_2, a_3; b_1, b_2)$ to the six $c_{pq}(v)$ and to the two energy coefficients α_1 and α_2 . It is natural to set $b_1 = +2$ (so that the linearized

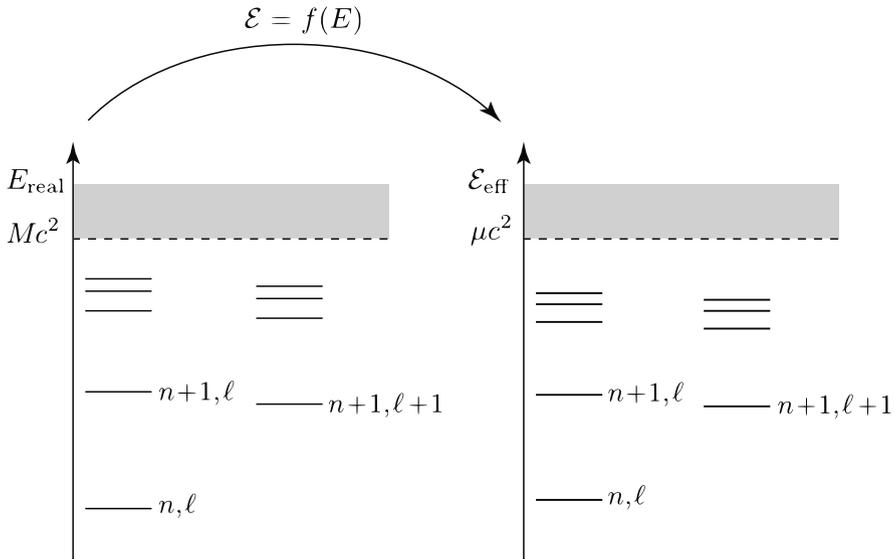


Fig. 1 Sketch of the correspondence between the quantized energy levels of the real and effective conservative dynamics. n denotes the ‘‘principal quantum number’’ ($n = n_r + \ell + 1$, with $n_r = 0, 1, \dots$ denoting the number of nodes in the radial function), while ℓ denotes the (relative) orbital angular momentum ($L^2 = \ell(\ell + 1)\hbar^2$). Though the EOB method is purely classical, it is conceptually useful to think in terms of the underlying (Bohr–Sommerfeld) quantization conditions of the action variables I_R and J to motivate the identification between n and ℓ in the two dynamics

⁸ Indeed $E_{\text{real}}^{\text{total}} = Mc^2 + E_{\text{real}}^{\text{relative}} = Mc^2 + \text{Newtonian terms} + \text{IPN}/c^2 + \dots$, while $\mathcal{E}_{\text{effective}} = \mu c^2 + N + \text{IPN}/c^2 + \dots$.

effective metric coincides with the linearized Schwarzschild metric with mass $M = m_1 + m_2$). One then finds that there exists a *unique* solution for the remaining five unknown coefficients a_2, a_3, b_2, α_1 , and α_2 . This solution is very simple:

$$a_2 = 0, \quad a_3 = 2\nu, \quad b_2 = 4 - 6\nu, \quad \alpha_1 = \frac{\nu}{2}, \quad \alpha_2 = 0. \quad (14)$$

Note, in particular, that the map between the two energies is simply

$$\frac{\mathcal{E}_{\text{eff}}}{\mu c^2} = 1 + \frac{E_{\text{real}}^{\text{relative}}}{\mu c^2} \left(1 + \frac{\nu}{2} \frac{E_{\text{real}}^{\text{relative}}}{\mu c^2} \right) = \frac{s - m_1^2 c^4 - m_2^2 c^4}{2 m_1 m_2 c^4} \quad (15)$$

where $s = (\mathcal{E}_{\text{real}}^{\text{tot}})^2 \equiv (Mc^2 + E_{\text{real}}^{\text{relative}})^2$ is Mandelstam's invariant $= -(p_1 + p_2)^2$. Note also that, at 2PN accuracy, the crucial “ g_{00}^{ext} ” metric coefficient $A(R)$ (which fully encodes the energetics of circular orbits) is given by the remarkably simple PN expansion

$$A_{2\text{PN}}(R) = 1 - 2u + 2\nu u^3, \quad (16)$$

where $u \equiv GM/(c^2 R)$ and $\nu \equiv \mu/M \equiv m_1 m_2/(m_1 + m_2)^2$.

The dimensionless parameter $\nu \equiv \mu/M$ varies between 0 (in the test mass limit $m_1 \ll m_2$) and $\frac{1}{4}$ (in the equal-mass case $m_1 = m_2$). When $\nu \rightarrow 0$, Eq. 16 yields back, as expected, the well-known Schwarzschild time–time metric coefficient $-g_{00}^{\text{Schw}} = 1 - 2u = 1 - 2GM/c^2 R$. One therefore sees in Eq. 16 the role of ν as a *deformation parameter* connecting a well-known test-mass result to a nontrivial and new 2PN result. It is also to be noted that the 1PN EOB result $A_{1\text{PN}}(R) = 1 - 2u$ happens to be ν -independent, and therefore identical to $A^{\text{Schw}} = 1 - 2u$. This is remarkable in view of the many nontrivial ν -dependent terms in the 1PN relative dynamics. The physically real 1PN ν -dependence happens to be fully encoded in the function $f(E)$ mapping the two energy spectra given in Eq. 15 above.

Let us emphasize the remarkable simplicity of the 2PN result (16). The 2PN Hamiltonian (5) contains eleven rather complicated ν -dependent terms. After transformation to the EOB format, the dynamical information contained in these eleven coefficients gets *condensed* into the very simple additional contribution $+2\nu u^3$ in $A(R)$, together with an equally simple contribution in the radial metric coefficient: $(A(R)B(R))_{2\text{PN}} = 1 - 6\nu u^2$. This condensation process is even more drastic when one goes to the next (conservative) PN order: the 3PN level, that is, additional terms of order $\mathcal{O}(1/c^6)$ in the Hamiltonian (5). As mentioned above, the complete obtention of the 3PN dynamics has represented quite a theoretical challenge and the final, resulting Hamiltonian is quite complicated. Even after going to the center of mass frame, the 3PN additional contribution $\frac{1}{c^6} H_6(\mathbf{q}, \mathbf{p})$ to Eq. 5 introduces eleven new complicated ν -dependent coefficients. After transformation to the EOB format [55], these eleven new coefficients get “condensed” into only *three* additional terms: (i) an additional contribution to $A(R)$, (ii) an additional contribution to $B(R)$, and (iii) a $\mathcal{O}(\mathbf{p}^4)$ modification of the “external” geodesic Hamiltonian. For instance, the crucial 3PN g_{00}^{ext} metric coefficient becomes

$$A_{3\text{PN}}(R) = 1 - 2u + 2\nu u^3 + a_4 \nu u^4, \quad (17)$$

where $u = GM/(c^2 R)$,

$$a_4 = \frac{94}{3} - \frac{41}{32} \pi^2 \simeq 18.6879027, \quad (18)$$

while the additional contribution to $B(R)$ gives

$$D_{3\text{PN}}(R) \equiv (A(R)B(R))_{3\text{PN}} = 1 - 6\nu u^2 + 2(3\nu - 26)\nu u^3. \quad (19)$$

Remarkably, it is found that the very simple 2PN energy map Eq. 15 does not need to be modified at the 3PN level.

The fact that the 3PN coefficient a_4 in the crucial “effective radial potential” $A_{3\text{PN}}(R)$, Eq. 17, is rather large and positive indicates that the ν -dependent nonlinear gravitational effects lead, for comparable masses ($\nu \sim \frac{1}{4}$), to a last stable (circular) orbit (LSO) which has a higher frequency and a larger binding energy than what a naive scaling from the test-particle limit ($\nu \rightarrow 0$) would suggest. Actually, the PN-expanded form (17) of $A_{3\text{PN}}(R)$ does not seem to be a good representation of the (unknown) exact function $A_{\text{EOB}}(R)$ when the (Schwarzschild-like) relative coordinate R becomes smaller than about $6GM/c^2$ (which is the radius of the LSO in the test-mass limit). In fact, by continuity with the test-mass case, one a priori expects that $A_{3\text{PN}}(R)$ always exhibits a simple zero defining an EOB “effective horizon” that is smoothly connected to the Schwarzschild event horizon at $R = 2GM/c^2$ when $\nu \rightarrow 0$. However, the large value of the a_4 coefficient does actually prevent $A_{3\text{PN}}$ to have this property when ν is too large, and in particular when $\nu = 1/4$, as it is visually explained in Fig. 2. The black curves in the figure represent the A function at 1PN (solid line), 2PN (dashed line) and 3PN (dash-dot line) approximation: while the 2PN curve still has a simple zero, the 3PN does not, due to the large value of a_4 . It was therefore suggested [55] to further resum⁹ $A_{3\text{PN}}(R)$ by replacing it by a suitable Padé (P) approximant. For instance, the replacement of $A_{3\text{PN}}(R)$ by

$$A_3^1(R) \equiv P_3^1[A_{3\text{PN}}(R)] = \frac{1 + n_1 u}{1 + d_1 u + d_2 u^2 + d_3 u^3}, \quad (20)$$

ensures that the $\nu = \frac{1}{4}$ case is smoothly connected with the $\nu = 0$ limit, as Fig. 2 clearly shows.¹⁰

The use of Eq. 20 was suggested before one had any (reliable) non-perturbative information on the binding of close black-hole binaries. Later, a comparison with

⁹ The PN-expanded EOB building blocks $A(R)$, $B(R)$, ... already represent a *resummation* of the PN dynamics in the sense that they have “condensed” the many terms of the original PN-expanded Hamiltonian within a very concise format. But one should not refrain to further resum the EOB building blocks themselves, if this is physically motivated.

¹⁰ We recall that the coefficient n_1 and (d_1, d_2, d_3) of the Padé approximant are determined by the condition that the first four terms of the Taylor expansion of A_3^1 in powers of $u = GM/(c^2 R)$ coincide with $A_{3\text{PN}}$.

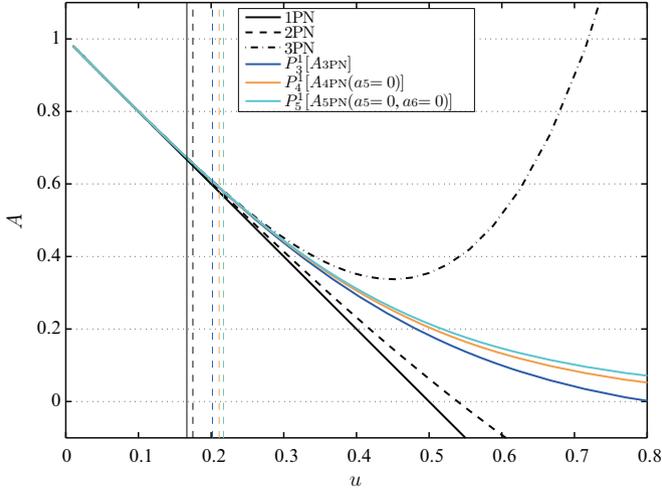


Fig. 2 Various approximations and Padé resummation of the EOB radial potential $A(u)$, where $u = GM/(c^2 R)$, for the equal-mass case $\nu = 1/4$. The vertical dashed lines indicate the corresponding (adiabatic) LSO location [35] defined by the condition $d^2 \mathcal{E}_{\text{eff}}^0/dR^2 = d\mathcal{E}_{\text{eff}}^0/dR = 0$, where $\mathcal{E}_{\text{eff}}^0$ is the effective energy along the sequence of circular orbits (i.e., when $P_R^{\text{eff}} = 0$)

some “waveless” numerical simulations of circular black-hole binaries [48] has given some evidence that Eq. 20 is physically adequate. In Refs. [45, 48] it was also emphasized that, in principle, the comparison between numerical data and EOB-based predictions should allow one to determine the effect of the unknown higher PN contributions to Eq. 17. For instance, one can add a 4PN-like term $+a_5 \nu u^5$ or a 5PN-like term $+a_6 \nu u^6$ in Eq. 17, and then Padé the resulting radial function. The new *resummed* A potential will exhibit an explicit dependence on a_5 (at 4PN) or (a_5, a_6) (at 5PN), that is

$$A_4^1(R; a_5, \nu) = P_4^1 [A_{3\text{PN}}(R) + \nu a_5 u^5], \quad (21)$$

or

$$A_5^1(R; a_5, a_6, \nu) = P_5^1 [A_{3\text{PN}}(R) + \nu a_5 u^5 + \nu a_6 u^6]. \quad (22)$$

Comparing the predictions of $A_4^1(R; a_5, \nu)$ or $A_5^1(R; a_5, a_6, \nu)$ to numerical data might then determine what is the physically preferred “effective” value of the unknown coefficient a_5 (if working at 4PN effective accuracy) or of the doublet (a_5, a_6) (when including also 5PN corrections). For illustrative purposes, Fig. 2 shows the effect of the Padé resummation with $a_5 = a_6 = 0$ and $\nu = 1/4$. Note that the Padé resummation procedure is injecting some “information” beyond that contained in the numerical values of the PN expansion coefficients a_n ’s of $A(R)$. As a consequence, the operation of Padéing and of restricting a_5 and a_6 to the (3PN-compatible) values $a_5 = 0 = a_6$ do not commute: $A_4^1(R; 0, 1/4) \neq A_5^1(R; 0, 0, 1/4) \neq A_3^1(R, 1/4)$. In this respect, let us also mention that the 4PN

a_5 -dependent Padé approximant $A_4^1(R; a_5, \nu)$ exactly reduces to the 3PN Padé approximant $A_3^1(R; \nu)$ when a_5 is replaced by the following function of ν ,

$$a_5^{3\text{PN}}(\nu) \equiv \frac{\nu(3392 - 123\pi^2)^2}{18432(\nu - 4)}. \quad (23)$$

Note that the value of the A_3^1 -reproducing effective 4PN coefficient $a_5^{3\text{PN}}(\nu)$ in the equal mass case is $a_5^{3\text{PN}}(1/4) \simeq -17.158031$. This is numerically compatible with the value $a_5 = -17.16$ quoted in Ref. [30] (but note that the correct A_3^1 -reproducing 4PN coefficient depends on the symmetric mass ratio ν). Similarly, when working at the 5PN level, $A_5^1(R; a_5, a_6, \nu)$ exactly reduces to the 4PN Padé approximant $A_4^1(R; a_5, \nu)$ when a_6 is replaced by the following function of both ν and a_5 :

$$\begin{aligned} a_6^{4\text{PN}}(\nu, a_5) \\ \equiv \frac{\nu(2304a_5^2 + 96(3392 - 123\pi^2)a_5 + (3776 - 123\pi^2)(32(3\nu + 94) - 123\pi^2))}{24[(3776 - 123\pi^2)\nu - 1536]}. \end{aligned} \quad (24)$$

The use of NR data to constrain the values of the higher PN parameters (a_5, a_6) is an example of the useful *flexibility* [51] of analytical approaches: the fact that one can tap numerically based, non-perturbative information to improve the EOB approach. The flexibility of the EOB approach related to the use of the a_5 -dependent radial potential $A_4^1(R; a_5, \nu)$ has been exploited in several recent works [30, 37, 38, 61, 65, 66] focusing on the comparison of EOB-based waveforms with waveforms computed via NR simulations. Collectively, all these studies have shown that it is possible to constrain a_5 (together with other flexibility parameters related to the resummation of radiation reaction, see below) so as to yield an excellent agreement (at the level of the published numerical errors) between EOB and NR waveforms. The result, however, cannot be summarized by stating that a_5 is constrained to be in the vicinity of a special numerical value. Rather, one finds a strong correlation between a_5 and other parameters, notably the radiation reaction parameter ν_{pole} introduced below. More recently, Ref. [62] could get rid of the flexibility parameters (such as ν_{pole}) related to the resummation of radiation reaction, and has shown that one can get an excellent agreement with NR data by using *only* the flexibility in the doublet (a_5, a_6) (the other parameters being essentially fixed internally to the formalism). We shall discuss this result further in Section 5 below.

The same kind of ν -continuity argument discussed so far for the A function needs to be applied also to the $D(R)_{3\text{PN}}$ function defined in Eq. 19. A straightforward way to ensure that the D function stays positive when R decreases (since it is $D = 1$ when $\nu \rightarrow 0$) is to replace $D_{3\text{PN}}(R)$ by $D_3^0(R) \equiv P_3^0[D_{3\text{PN}}(R)]$, where P_3^0 indicates the (0, 3) Padé approximant and explicitly reads

$$D_3^0(R) = \frac{1}{1 + 6\nu u^2 - 2(3\nu - 26)\nu u^3}. \quad (25)$$

The resummation of A (via Padé approximants) is necessary for ensuring the existence and ν -continuity of a *LSO* (see vertical lines in Fig. 2), as well as the existence and ν -continuity of a *last unstable orbit*, that is, of a ν -deformed analog of the light ring $R = 3GM/c^2$ when $\nu \rightarrow 0$. We recall that, when $\nu = 0$, the light ring corresponds to the circular orbit of a massless particle, or of an extremely relativistic massive particle, and is technically defined by looking for the maximum of $A(R)/R^2$, that is, by solving $(d/dR)(A(R)/R^2) = 0$. When $\nu \neq 0$ and when considering the quasi-circular plunge following the crossing of the LSO, the “effective” meaning of the “ ν -deformed light ring” (technically defined by solving $(d/dR)(A(R : \nu)/R^2) = 0$) is to entail, in its vicinity, the existence of a maximum of the orbital frequency $\Omega = d\varphi/dt$ (the resummation of $D(R)$ plays a useful role in ensuring the ν -continuity of this plunge behavior).

4 Description of Radiation–Reaction Effects in the EOB Approach

In the previous section we have described how the EOB method encodes the conservative part of the relative orbital dynamics into the dynamics of an “effective” particle. Let us now briefly discuss how to complete the EOB dynamics by defining some *resummed* expressions describing radiation reaction effects. One is interested in circularized binaries, which have lost their initial eccentricity under the influence of radiation reaction. For such systems, it is enough (as shown in [36]) to include a radiation reaction force in the p_φ equation of motion only. More precisely, we are using phase space variables $r, p_r, \varphi, p_\varphi$ associated to polar coordinates (in the equatorial plane $\theta = \frac{\pi}{2}$). Actually it is convenient to replace the radial momentum p_r by the momentum conjugate to the “tortoise” radial coordinate $R_* = \int dR (B/A)^{1/2}$, that is, $P_{R_*} = (A/B)^{1/2} P_R$. The real EOB Hamiltonian is obtained by first solving Eq. 15 to get $E_{\text{real}}^{\text{total}} = \sqrt{s}$ in terms of \mathcal{E}_{eff} , and then by solving the effective Hamiltonian–Jacobi equation¹¹ to get \mathcal{E}_{eff} in terms of the effective phase space coordinates \mathbf{q}_{eff} and \mathbf{p}_{eff} . The result is given by two nested square roots (we henceforth set $c = 1$):

$$\hat{H}_{\text{EOB}}(r, p_{r_*}, \varphi) = \frac{H_{\text{EOB}}^{\text{real}}}{\mu} = \frac{1}{\nu} \sqrt{1 + 2\nu (\hat{H}_{\text{eff}} - 1)}, \quad (26)$$

where

$$\hat{H}_{\text{eff}} = \sqrt{p_{r_*}^2 + A(r) \left(1 + \frac{p_\varphi^2}{r^2} + z_3 \frac{p_{r_*}^4}{r^2} \right)}, \quad (27)$$

¹¹ Completed by the $\mathcal{O}(\mathbf{p}^4)$ terms that must be introduced at 3PN.

with $z_3 = 2\nu(4-3\nu)$. Here, we are using suitably rescaled dimensionless (effective) variables: $r = R/GM$, $p_{r_*} = P_{R_*}/\mu$, $p_\varphi = P_\varphi/\mu GM$, as well as a rescaled time $t = T/GM$. This leads to equations of motion $(r, \varphi, p_{r_*}, p_\varphi)$ of the form

$$\frac{d\varphi}{dt} = \frac{\partial \hat{H}_{\text{EOB}}}{\partial p_\varphi} \equiv \Omega, \quad (28)$$

$$\frac{dr}{dt} = \left(\frac{A}{B}\right)^{1/2} \frac{\partial \hat{H}_{\text{EOB}}}{\partial p_{r_*}}, \quad (29)$$

$$\frac{dp_\varphi}{dt} = \hat{\mathcal{F}}_\varphi, \quad (30)$$

$$\frac{dp_{r_*}}{dt} = -\left(\frac{A}{B}\right)^{1/2} \frac{\partial \hat{H}_{\text{EOB}}}{\partial r}, \quad (31)$$

which explicitly read

$$\frac{d\varphi}{dt} = \frac{Ap_\varphi}{\nu r^2 \hat{H} \hat{H}_{\text{eff}}} \equiv \Omega, \quad (32)$$

$$\frac{dr}{dt} = \left(\frac{A}{B}\right)^{1/2} \frac{1}{\nu \hat{H} \hat{H}_{\text{eff}}} \left(p_{r_*} + z_3 \frac{2A}{r^2} p_{r_*}^3 \right) \quad (33)$$

$$\frac{dp_\varphi}{dt} = \hat{\mathcal{F}}_\varphi \quad (34)$$

$$\frac{dp_{r_*}}{dt} = -\left(\frac{A}{B}\right)^{1/2} \frac{1}{2\nu \hat{H} \hat{H}_{\text{eff}}} \left\{ A' + \frac{p_\varphi^2}{r^2} \left(A' - \frac{2A}{r} \right) + z_3 \left(\frac{A'}{r^2} - \frac{2A}{r^3} \right) p_{r_*}^4 \right\} \quad (35)$$

where $A' = dA/dr$. As explained above the EOB metric function $A(r)$ is defined by Padé resumming the Taylor-expanded result (10) obtained from the matching between the real and effective energy levels [as we were mentioning, one uses a similar Padé resumming for $D(r) \equiv A(r) B(r)$]. One similarly needs to resum $\hat{\mathcal{F}}_\varphi$, that is, the φ component of the radiation reaction which has been introduced on the right-hand side (r.h.s.) of Eq. 30. During the quasi-circular inspiral $\hat{\mathcal{F}}_\varphi$ is known (from the PN work mentioned in Section 2 above) in the form of a Taylor expansion of the form

$$\hat{\mathcal{F}}_\varphi^{\text{Taylor}} = -\frac{32}{5} \nu \Omega^5 r_\omega^4 \hat{F}^{\text{Taylor}}(\nu_\varphi), \quad (36)$$

where $\nu_\varphi \equiv \Omega r_\omega$, and $r_\omega \equiv r[\psi(r, p_\varphi)]^{1/3}$ is a modified EOB radius, with ψ being defined as

$$\psi(r, p_\varphi) = \frac{2}{r^2} \left(\frac{dA(r)}{dr} \right)^{-1} \left[1 + 2\nu \left(\sqrt{A(r) \left(1 + \frac{p_\varphi^2}{r^2} \right)} - 1 \right) \right], \quad (37)$$

which generalizes the 2PN-accurate Eq. 22 of Ref. [47]. In Eq. 36 we have defined

$$\hat{F}^{\text{Taylor}}(\nu) = 1 + A_2(\nu) \nu^2 + A_3(\nu) \nu^3 + A_4(\nu) \nu^4 + A_5(\nu) \nu^5 + A_6(\nu, \log \nu) \nu^6 + A_7(\nu) \nu^7 + A_8(\nu = 0, \log \nu) \nu^8, \quad (38)$$

where we have added to the known 3.5PN-accurate comparable-mass result the small-mass-ratio 4PN contribution [102]. We recall that the small-mass contribution to the Newton-normalized flux is actually known up to 5.5PN order, that is, to ν^{11} included. The standard Taylor expansion of the flux, 38, has rather poor convergence properties when considered up to the LSO. This is illustrated in Fig. 3 in the small-mass limit $\nu = 0$. The convergence of the PN-expanded flux can be studied in detail in the $\nu = 0$ limit, because in this case one can compute an “exact” result numerically (using black hole perturbation theory [43, 107]). The “exact” energy flux shown in Fig. 3 is obtained as a sum over multipoles

$$F^{\ell_{\max}} = \sum_{\ell=2}^{\ell_{\max}} \sum_{m=1}^{\ell} F_{\ell m}, \quad (39)$$

where $F_{\ell m} = F_{|\ell| m}$ already denotes the sum of two equal contributions corresponding to $+m$ and $-m$ ($m \neq 0$ as $F_{\ell 0}$ vanishes for circular orbits). To be precise, the “exact” result exhibited in Fig. 3 is given by the rather accurate approximation $F^{(6)}$

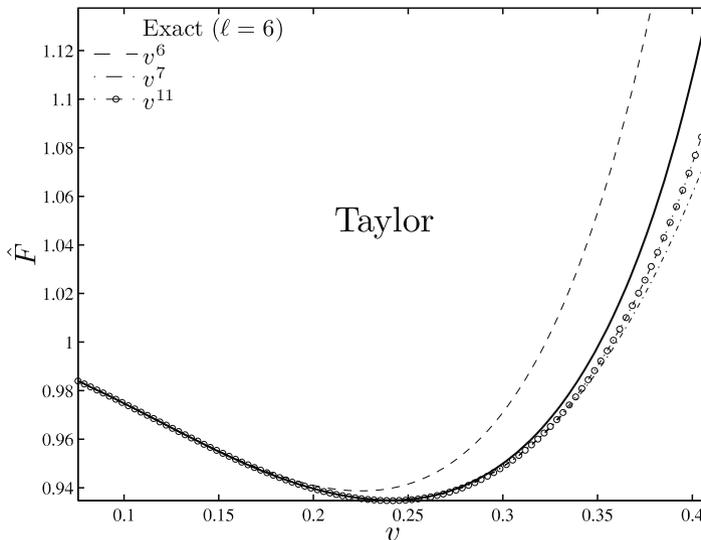


Fig. 3 The extreme-mass-ratio limit ($\nu = 0$): the Newton-normalized energy flux emitted by a particle on circular orbits. The figure illustrates the scattering of the standard Taylor expansion of the flux around the “exact” numerical result (computed up to $\ell = 6$) obtained via perturbation theory

obtained by choosing $\ell_{\max} = 6$; that is, by truncating the sum over ℓ in Eq. 39 beyond $\ell = 6$. In addition, one normalizes the result onto the “Newtonian” (i.e., quadrupolar) result $F_{22}^N = 32/5(\mu/M)^2 v^{10}$. In other words, the solid line in Fig. 3 represents the quantity $\hat{F} \equiv F^{(6)}/F_{22}^N$.

For clarity, we selected only three Taylor approximants: 3PN (v^6), 3.5PN (v^7), and 5.5PN (v^{11}). These three values suffice to illustrate the rather large scatter among Taylor approximants, and the fact that, near the LSO, the convergence toward the exact value (solid line) is rather slow, and non-monotonic. [See also Fig. 1 in Ref. [88] and Fig. 3 of Ref. [53] for fuller illustrations of the scattered and non-monotonic way in which successive Taylor expansions approach the numerical result.] The results shown in Fig. 3 elucidate that the Taylor series 38 is inadequate to give a reliable representation of the energy loss during the plunge. That is the reason why the EOB formalism advocates the use of a “resummed” version of \mathcal{F}_φ , that is, a non-polynomial function replacing Eq. 38 at the r.h.s. of the Hamilton’s equation (and coinciding with it in the $v/c \ll 1$ limit).

Two methods have been proposed to perform such a resummation. The first method, that strongly relies on the use of Padé approximants, was introduced by Damour, Iyer, and Sathyaprakash [53] and, with different degrees of sophistication, has been widely used in the literature dealing with the EOB formalism [30, 33, 34, 36–38, 59–61, 65, 66, 83, 85]. The second resummation method has been recently introduced by Damour, Iyer, and Nagar [52] and exploited to provide a self-consistent expression of the radiation reaction force in Ref. [62]. This latter resummation procedure is based on (i) a new multiplicative decomposition of the gravitational metric waveform which yields a (ii) resummation of each multipolar contribution to the energy flux. The use of Padé approximants is a useful tool (but not the only one) that proves helpful to further improve the convergence properties of each multipolar contribution to the flux. The following two sections are devoted to highlighting the main features of the two methods. For pedagogical reasons the calculation is first done in the small-mass limit ($\nu \rightarrow 0$) and then generalized to the comparable mass case.

4.1 Resummation of \hat{F}^{Taylor} Using a One-Parameter Family of Padé Approximants: Tuning v_{pole}

Following [53], one resums \hat{F}^{Taylor} by using the following Padé resummation approach. First, one chooses a certain number v_{pole} which is intended to represent the value of the orbital velocity v_φ at which the exact angular momentum flux would become infinite if one were to formally analytically continue $\hat{\mathcal{F}}_\varphi$ along *unstable* circular orbits below the LSO: then, given v_{pole} , one defines the resummed $\hat{F}(v_\varphi)$ as

$$\hat{F}^{\text{resummed}}(v_\varphi) = \left(1 - \frac{v_\varphi}{v_{\text{pole}}}\right)^{-1} P_4^4 \left[\left(1 - \frac{v_\varphi}{v_{\text{pole}}}\right) \hat{F}^{\text{Taylor}}(v_\varphi; \nu = 0) \right], \quad (40)$$

where P_4^4 denotes a (4, 4) Padé approximant.

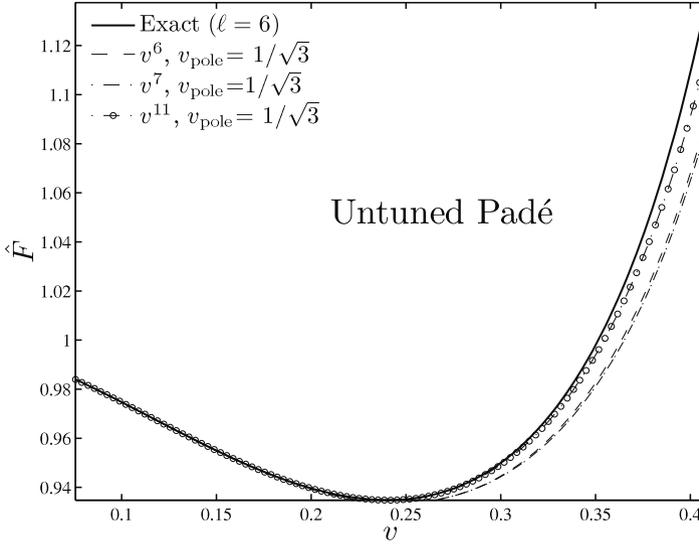


Fig. 4 The extreme-mass-ratio limit ($v = 0$). Padé resummation of the Taylor-expanded energy flux of Fig. 3 as proposed in Ref. [53] with $v_{\text{pole}} = 1/\sqrt{3}$. The sequence of Padé approximants is less scattered than the corresponding Taylor ones and closer to the exact result

If one first follows the reasoning line of [53], and fixes the location of the pole in the resummed flux at the standard Schwarzschild value $v_{\text{pole}}^{(v=0)} = 1/\sqrt{3}$, one gets the result in Fig. 4. By comparison to Fig. 3, one can appreciate the significantly better (and monotonic) way in which successive *Padé approximants* approach (in L_∞ norm on the full interval $0 < x < x_{\text{LSO}}$) the numerical result. Ref. [53] also showed that the observationally relevant overlaps (of both the “faithfulness” and the “effectualness” types) between analytical and numerical adiabatic signals were systematically better for Padé approximants than for Taylor ones. Note that this figure is slightly different from the corresponding results in panel (b) of Fig. 3 in [53] (in particular, the present result exhibits a better “convergence” of the v^{11} curve). This difference is due to the new treatment of the logarithmic terms $\propto \log x$. Instead of factoring them out in front as proposed in [53], we consider them here (following [61]) as being part of the “Taylor coefficients” $f_n(\log x)$ when Padéing the flux function.

A remarkable improvement in the (L_∞) closeness between $\hat{F}^{\text{Padé-resummed}}(v)$ and $\hat{F}^{\text{Exact}}(v)$ can be obtained, as suggested by Damour and Nagar [61] (following ideas originally introduced in Ref. [51]), by suitably flexing the value of v_{pole} . As proposed in Ref. [61], v_{pole} is tuned until the difference between the resummed and the exact flux at the LSO is zero (or at least smaller than 10^{-4}). The resulting closeness between the exact and tuned-resummed fluxes is illustrated in Fig. 5. It is so good (compared to the previous figures, where the differences were clearly visible) that we need to complement the figure with Table 1. This table compares in a quantitative way the result of the “untuned” Padé resummation ($v_{\text{pole}} = 1/\sqrt{3}$) of Ref. [53]

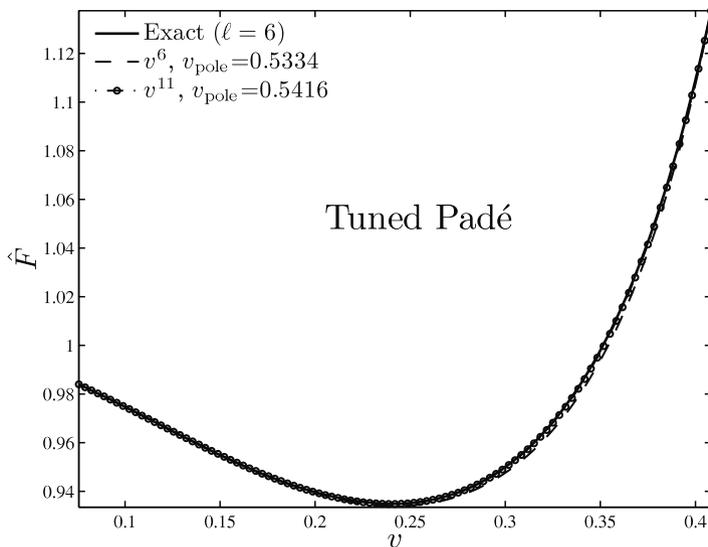


Fig. 5 The extreme-mass-ratio limit ($\nu = 0$). Same of Fig. 4 but *flexing* the value of the parameter ν_{pole} so to improve the agreement with the exact result

Table 1 Errors in the flux of the two (untuned or tuned) Padé resummation procedures. From left to right, the columns report: the PN-order; the difference between the resummed and the exact flux, $\Delta\hat{F} = \hat{F}^{\text{Resummed}} - \hat{F}^{\text{Exact}}$, at the LSO, and the L_∞ norm of $\Delta\hat{F}$, $\|\Delta\hat{F}\|_\infty$ (computed over the interval $0 < \nu < \nu_{\text{LSO}}$), for $\nu_{\text{pole}} = 1/\sqrt{3}$; the *flexed* value of ν_{pole} used here; \hat{F} at the LSO and the corresponding L_∞ norm (over the same interval) for the flexed value of ν_{pole}

PN-order	$\hat{F}_{\text{LSO}}^{1/\sqrt{3}}$	$\ \Delta\hat{F}\ _\infty^{1/\sqrt{3}}$	ν_{pole}	$\hat{F}_{\text{LSO}}^{\nu_{\text{pole}}}$	$\ \Delta\hat{F}\ _\infty^{\nu_{\text{pole}}}$
3 (ν^6)	-0.048	0.048	0.5334	7.06×10^{-5}	0.00426
3.5 (ν^7)	-0.051	0.051	0.5425	5.50×10^{-5}	0.00429
5.5 (ν^{11})	-0.022	0.022	0.5416	2.52×10^{-5}	0.000854

to the result of the “ ν_{pole} -tuned” Padé resummation described here. Defining the function $\Delta\hat{F}(\nu; \nu_{\text{pole}}) = \hat{F}^{\text{Resummed}}(\nu; \nu_{\text{pole}}) - \hat{F}^{\text{Exact}}(\nu)$ measuring the difference between a resummed and the exact energy flux, Table 1 lists both the values of $\Delta\hat{F}$ at $\nu = \nu_{\text{LSO}}$ and its L_∞ norm on the interval $0 < \nu < \nu_{\text{LSO}}$ for both the untuned and tuned cases. Note, in particular, how the ν_{pole} -flexing approach permits to reduce the L_∞ norm over this interval by more than an order of magnitude with respect to the untuned case. Note that the closeness between the tuned flux and the exact one is remarkably good (4.3×10^{-3}) already at the 3PN level.

It has recently been shown in several works [38, 61, 65, 66] that the *flexibility* in the choice of ν_{pole} could be advantageously used to get a close agreement with NR data (at the level of the numerical error). We will not comment here any further on this *parameter-dependent* resummation procedure of the energy flux and address the reader to the aforementioned references for further details.

4.2 Parameter-Free Resummation of Waveform and Energy Flux

In this section we shall introduce the reader to the new resummation technique for the multipolar waveform (and thus for the energy flux) introduced in Ref. [60, 61] and perfected in [52]. The aim is to summarize here the main ideas discussed in [52] as well as to collect most of the relevant equations that are useful for implementation in the EOB dynamics. To be precise, the new results discussed in Ref. [52] are twofold: on the one hand, that work generalized the $\ell = m = 2$ resummed waveform of [60, 61] to higher multipoles by using the most accurate currently known PN-expanded results [10, 25, 79] as well as the higher PN terms that are known in the test-mass limit [102, 103]; on the other hand, it introduced a *new resummation procedure* that consists in considering a new theoretical quantity, denoted as $\rho_{\ell m}(x)$, which enters the (ℓ, m) waveform (together with other building blocks, see below) only through its ℓ -th power: $h_{\ell m} \propto (\rho_{\ell m}(x))^\ell$. Here, and below, x denotes the invariant PN-ordering parameter $x \equiv (GM\Omega/c^3)^{2/3}$.

The main novelty introduced by Ref. [52] is to write the (ℓ, m) multipolar waveform emitted by a circular nonspinning compact binary as the *product* of several factors, namely

$$h_{\ell m}^{(\varepsilon)} = \frac{GMv}{c^2 R} n_{\ell m}^{(\varepsilon)} c_{\ell+\varepsilon}(v) x^{(\ell+\varepsilon)/2} Y^{\ell-\varepsilon, -m} \left(\frac{\pi}{2}, \Phi \right) \hat{S}_{\text{eff}}^{(\varepsilon)} T_{\ell m} e^{i\delta_{\ell m}} \rho_{\ell m}^\ell. \quad (41)$$

Here, ε denotes the parity of $\ell + m$ ($\varepsilon = \pi(\ell + m)$), that is, $\varepsilon = 0$ for “even-parity” (mass-generated) multipoles ($\ell + m$ even), and $\varepsilon = 1$ for “odd-parity” (current-generated) ones ($\ell + m$ odd); $n_{\ell m}^{(\varepsilon)}$ and $c_{\ell+\varepsilon}(v)$ are numerical coefficients; $\hat{S}_{\text{eff}}^{(\varepsilon)}$ is a μ -normalized effective source (whose definition comes from the EOB formalism); $T_{\ell m}$ is a resummed version [60, 61] of an infinite number of “leading logarithms” entering the *tail effects* [13, 17]; $\delta_{\ell m}$ is a supplementary phase (which corrects the phase effects not included in the *complex* tail factor $T_{\ell m}$), and, finally, $(\rho_{\ell m})^\ell$ denotes the ℓ -th power of the quantity $\rho_{\ell m}$ which is the new building block introduced in [52]. Note that in previous papers [60, 61] the quantity $(\rho_{\ell m})^\ell$ was denoted as $f_{\ell m}$ and we will mainly use this notation below. Before introducing explicitly the various elements entering the waveform 41 it is convenient to decompose $h_{\ell m}$ as

$$h_{\ell m} = h_{\ell m}^{(N, \varepsilon)} \hat{h}_{\ell m}^{(\varepsilon)}, \quad (42)$$

where $h_{\ell m}^{(N, \varepsilon)}$ is the Newtonian contribution and $\hat{h}_{\ell m}^{(\varepsilon)} \equiv \hat{S}_{\text{eff}}^{(\varepsilon)} T_{\ell m} e^{i\delta_{\ell m}} f_{\ell m}$ represents a resummed version of all the PN corrections. The PN correcting factor $\hat{h}_{\ell m}^{(\varepsilon)}$, as well as all its building blocks, has the structure $\hat{h}_{\ell m}^{(\varepsilon)} = 1 + \mathcal{O}(x)$.

Entering now in the discussion of the explicit form of the elements entering Eq. 41, we have that the v -independent numerical coefficients are given by

$$n_{\ell m}^{(0)} = (im)^\ell \frac{8\pi}{(2\ell + 1)!!} \sqrt{\frac{(\ell + 1)(\ell + 2)}{\ell(\ell - 1)}}, \quad (43)$$

$$n_{\ell m}^{(1)} = -(im)^\ell \frac{16\pi i}{(2\ell + 1)!!} \sqrt{\frac{(2\ell + 1)(\ell + 2)(\ell^2 - m^2)}{(2\ell - 1)(\ell + 1)\ell(\ell - 1)}}, \quad (44)$$

while the ν -dependent coefficients $c_{\ell+\varepsilon}(\nu)$ (such that $|c_{\ell+\varepsilon}(\nu = 0)| = 1$), can be expressed in terms of ν (as in Ref. [25, 79]), although they are more conveniently written in terms of the two mass ratios $X_1 = m_1/M$ and $X_2 = m_2/M$ in the form

$$\begin{aligned} c_{\ell+\varepsilon}(\nu) &= X_2^{\ell+\varepsilon-1} + (-)^{\ell+\varepsilon} X_1^{\ell+\varepsilon-1} \\ &= X_2^{\ell+\varepsilon-1} + (-)^m X_1^{\ell+\varepsilon-1}. \end{aligned} \quad (45)$$

In the second form of the equation we have used the fact that, as $\varepsilon = \pi(\ell + m)$, $\pi(\ell + \varepsilon) = \pi(m)$.

Let us turn now to discussing the structure of the $\hat{S}_{\text{eff}}^{(\varepsilon)}$ and $T_{\ell m}$ factors. To this aim, following Ref. [52], we recall that along the sequence of EOB circular orbits, which are determined by the condition $\partial_u \{A(u)[1 + j_0^2 u^2]\} = 0$, the effective EOB Hamiltonian (per unit μ mass) reads

$$\hat{H}_{\text{eff}} = \frac{H_{\text{eff}}}{\mu} = \sqrt{A(u)(1 + j_0^2 u^2)} \quad (\text{circular orbits}), \quad (46)$$

where the squared angular momentum is given by

$$j_0^2(u) = -\frac{A'(u)}{(u^2 A(u))'} \quad (\text{circular orbits}), \quad (47)$$

with the prime denoting d/du . Inserting this u -parametric representation of j^2 in Eq. 46 defines the u -parametric representation of the effective Hamiltonian $\hat{H}_{\text{eff}}(u)$. In the even-parity case (corresponding to mass moments), since the leading order source of gravitational radiation is given by the energy density, Ref. [52] defined the even-parity “source factor” as

$$\hat{S}_{\text{eff}}^{(0)}(x) = \hat{H}_{\text{eff}}(x) \quad \ell + m \quad \text{even}, \quad (48)$$

where $x = (GM\Omega/c^3)^{2/3}$. In the odd-parity case, they explored two, equally motivated, possibilities. The first one consists simply in still factoring $\hat{H}_{\text{eff}}(x)$; that is, in defining $\hat{S}_{\text{eff}}^{(1,H)} = \hat{H}_{\text{eff}}(x)$ also when $\ell + m$ is odd. The second one consists in factoring the angular momentum \mathcal{J} . Indeed, the angular momentum density $\varepsilon_{ijk} x^j \tau^{0k}$ enters as a factor in the (odd-parity) current moments, and \mathcal{J} occurs (in the small- ν limit) as a factor in the source of the Regge–Wheeler–Zerilli odd-parity multipoles. This leads us to define as second possibility

$$\hat{S}_{\text{eff}}^{(1,J)} = \hat{j}(x) \equiv x^{1/2} j(x) \quad \ell + m \quad \text{odd}, \quad (49)$$

where \hat{j} denotes what can be called the “Newton-normalized” angular momentum, namely the ratio $\hat{j}(x) = j(x)/j_N(x)$ with $j_N(x) = 1/\sqrt{x}$. In Ref. [52], the relative merits of the two possible choices were discussed. Although the analysis in the adiabatic $\nu = 0$ limit showed that they are equivalent from the practical point of view (because they both yield waveforms that are very close to the exact numerical result) we prefer to consider only the J -factorization in the following, that we will treat as our standard choice.

The second building block in our factorized decomposition is the “tail factor” $T_{\ell m}$ (introduced in Refs. [60, 61]). As mentioned above, $T_{\ell m}$ is a resummed version of an infinite number of “leading logarithms” entering the transfer function between the near-zone multipolar wave and the far-zone one, due to *tail effects* linked to its propagation in a Schwarzschild background of mass $M_{\text{ADM}} = H_{\text{EOB}}^{\text{real}}$. Its explicit expression reads

$$T_{\ell m} = \frac{\Gamma(\ell + 1 - 2i\hat{k})}{\Gamma(\ell + 1)} e^{\pi\hat{k}} e^{2i\hat{k} \log(2kr_0)}, \quad (50)$$

where $r_0 = 2GM$ and $\hat{k} \equiv GH_{\text{EOB}}^{\text{real}} m\Omega$ and $k \equiv m\Omega$. Note that \hat{k} differs from k by a rescaling involving the *real* (rather than the *effective*) EOB Hamiltonian, computed at this stage along the sequence of circular orbits.

The tail factor $T_{\ell m}$ is a complex number that already takes into account some of the dephasing of the partial waves as they propagate out from the near zone to infinity. However, as the tail factor only takes into account the leading logarithms, one needs to correct it by a complementary dephasing term, $e^{i\delta_{\ell m}}$, linked to subleading logarithms and other effects. This subleading phase correction can be computed as being the phase $\delta_{\ell m}$ of the complex ratio between the PN-expanded $\hat{h}_{\ell m}^{(e)}$ and the above-defined source and tail factors. In the comparable-mass case ($\nu \neq 0$), the 3PN δ_{22} phase correction to the leading quadrupolar wave was originally computed in Ref. [61] (see also Ref. [60] for the $\nu = 0$ limit). Full results for the subleading partial waves to the highest possible PN-accuracy by starting from the currently known 3PN-accurate ν -dependent waveform [25] have been obtained in [52].

The last factor in the multiplicative decomposition of the multipolar waveform can be computed as being the modulus $f_{\ell m}$ of the complex ratio between the PN-expanded $\hat{h}_{\ell m}^{(e)}$ and the above-defined source and tail factors. In the comparable mass case ($\nu \neq 0$), the f_{22} modulus correction to the leading quadrupolar wave was computed in Ref. [61] (see also Ref. [60] for the $\nu = 0$ limit). For the subleading partial waves, Ref. [52] explicitly computed the other $f_{\ell m}$ ’s to the highest possible PN-accuracy by starting from the currently known 3PN-accurate ν -dependent waveform [25]. In addition, as originally proposed in Ref. [61], to reach greater accuracy the $f_{\ell m}(x; \nu)$ ’s extracted from the 3PN-accurate $\nu \neq 0$ results are completed by adding higher order contributions coming from the $\nu = 0$ results [102, 103]. In the particular f_{22} case discussed in [61], this amounted to adding 4PN and 5PN $\nu = 0$ terms. This “hybridization” procedure was then systematically pursued for all the other multipoles, using the 5.5PN accurate calculation of the

multipolar decomposition of the GW energy flux of Refs. [102, 103]. Note that such hybridization procedure is *not* equivalent to the straightforward hybrid sum ansatz, $\tilde{h}_{\ell m} = \tilde{h}_{\ell m}^{\text{known}}(\nu) + \tilde{h}_{\ell m}^{\text{higher}}(\nu = 0)$ (where $\tilde{h}_{\ell m} \equiv h_{\ell m}/\nu$) that one may have thought to implement.

In the even-parity case, the determination of the modulus $f_{\ell m}$ is unique. In the odd-parity case, it depends on the choice of the source which, as explained above, can be connected either to the effective energy or to the angular momentum. We will consider both cases and distinguish them by adding either the label H or \mathcal{J} to the corresponding $f_{\ell m}$. Note, in passing, that, since in both cases the factorized effective source term (H_{eff} or \mathcal{J}) is a real quantity, the phases $\delta_{\ell m}$'s are the same.

The above-explained procedure defines the $f_{\ell m}$'s as Taylor-expanded PN series of the type

$$f_{\ell m}(x; \nu) = 1 + c_1^{f_{\ell m}}(\nu)x + c_2^{f_{\ell m}}(\nu)x^2 + c_3^{f_{\ell m}}(\nu, \log(x))x^3 + \dots \quad (51)$$

Note that one of the virtues of our factorization is to have separated the half-integer powers of x appearing in the usual PN-expansion of $h_{\ell m}^{(\epsilon)}$ from the integer powers, the tail factor, together with the complementary phase factor $e^{i\delta_{\ell m}}$, having absorbed all the half-integer powers. In Ref. [66] all the $f_{\ell m}$'s (both for the H and \mathcal{J} choices) have been computed up to the highest available (ν -dependent or not) PN accuracy. In the formulas for the $f_{\ell m}$'s given below, we “hybridize” them by adding to the known ν -dependent coefficients $c_n^{f_{\ell m}}(\nu)$ in Eq. 51 the $\nu = 0$ value of the higher order coefficients: $c_{n'}^{f_{\ell m}}(\nu = 0)$. The 1PN-accurate $f_{\ell m}$'s for $\ell + m$ even and also for $\ell + m$ odd can be written down for all ℓ . The complete result for the $f_{\ell m}$'s that are known with an accuracy higher than 1PN are listed in Appendix B of Ref. [66]. Here, for illustrative purposes, we quote only the lowest $f_{\ell m}^{\text{even}}$ and $f_{\ell m}^{\text{odd}, J}$ up to $\ell = 3$ included:

$$\begin{aligned} f_{22}(x; \nu) = & 1 + \frac{1}{42}(55\nu - 86)x + \frac{(2047\nu^2 - 6745\nu - 4288)}{1512}x^2 \\ & + \left(\frac{114635\nu^3}{99792} - \frac{227875\nu^2}{33264} + \frac{41}{96}\pi^2\nu - \frac{34625\nu}{3696} \right. \\ & \quad \left. - \frac{856}{105}\text{eulerlog}_2(x) + \frac{21428357}{727650} \right)x^3 \\ & + \left(\frac{36808}{2205}\text{eulerlog}_2(x) - \frac{5391582359}{198648450} \right)x^4 \\ & + \left(\frac{458816}{19845}\text{eulerlog}_2(x) - \frac{93684531406}{893918025} \right)x^5 + \mathcal{O}(x^6), \quad (52) \end{aligned}$$

$$\begin{aligned}
f_{21}^J(x; \nu) &= 1 + \left(\frac{23\nu}{42} - \frac{59}{28} \right) x + \left(\frac{85\nu^2}{252} - \frac{269\nu}{126} - \frac{5}{9} \right) x^2 \\
&\quad + \left(\frac{88404893}{11642400} - \frac{214}{105} \text{eulerlog}_1(x) \right) x^3 \\
&\quad + \left(\frac{6313}{1470} \text{eulerlog}_1(x) - \frac{33998136553}{4237833600} \right) x^4 + \mathcal{O}(x^5), \quad (53)
\end{aligned}$$

$$\begin{aligned}
f_{33}(x; \nu) &= 1 + \left(2\nu - \frac{7}{2} \right) x + \left(\frac{887\nu^2}{330} - \frac{3401\nu}{330} - \frac{443}{440} \right) x^2 \\
&\quad + \left(\frac{147471561}{2802800} - \frac{78}{7} \text{eulerlog}_3(x) \right) x^3 \\
&\quad + \left(39 \text{eulerlog}_3(x) - \frac{53641811}{457600} \right) x^4 + \mathcal{O}(x^5), \quad (54)
\end{aligned}$$

$$\begin{aligned}
f_{32}^J(x; \nu) &= 1 + \frac{320\nu^2 - 1115\nu + 328}{90(3\nu - 1)} x \\
&\quad + \frac{39544\nu^3 - 253768\nu^2 + 117215\nu - 20496}{11880(3\nu - 1)} x^2 \\
&\quad + \left(\frac{110842222}{4729725} - \frac{104}{21} \text{eulerlog}_2(x) \right) x^3 + \mathcal{O}(x^4), \quad (55)
\end{aligned}$$

$$\begin{aligned}
f_{31}(x; \nu) &= 1 + \left(-\frac{2\nu}{3} - \frac{13}{6} \right) x + \left(-\frac{247\nu^2}{198} - \frac{371\nu}{198} + \frac{1273}{792} \right) x^2 \\
&\quad + \left(\frac{400427563}{75675600} - \frac{26}{21} \text{eulerlog}_1(x) \right) x^3 \\
&\quad + \left(\frac{169}{63} \text{eulerlog}_1(x) - \frac{12064573043}{1816214400} \right) x^4 + \mathcal{O}(x^5). \quad (56)
\end{aligned}$$

For convenience and readability, we have introduced the following ‘‘eulerlog’’ functions: $\text{eulerlog}_m(x) = \gamma_E + \log 2 + \frac{1}{2} \log x + \log m$, where $\gamma_E = 0.57721\dots$ is Euler’s constant.

The decomposition of the total PN-correction factor $\hat{h}_{\ell m}^{(\varepsilon)}$ into several factors is in itself a resummation procedure which has already improved the convergence of the PN series one has to deal with: indeed, one can see that the coefficients entering increasing powers of x in the $f_{\ell m}$ ’s tend to be systematically smaller than the coefficients appearing in the usual PN expansion of $\hat{h}_{\ell m}^{(\varepsilon)}$. The reason for this is essentially twofold: (i) the factorization of $T_{\ell m}$ has absorbed powers of $m\pi$, which contributed to make large coefficients in $\hat{h}_{\ell m}^{(\varepsilon)}$, and (ii) the factorization of either \hat{H}_{eff} or \hat{j}

has (in the $\nu = 0$ case) removed the presence of an inverse square-root singularity located at $x = 1/3$, which caused the coefficient of x^n in any PN-expanded quantity to grow as 3^n as $n \rightarrow \infty$. To prevent some potential misunderstandings, let us emphasize that we are talking here about a singularity entering the analytic continuation (to larger values of x) of a mathematical function $h(x)$ defined (for small values of x) by considering the formal adiabatic circular limit. The point is that, in the $\nu \rightarrow 0$ limit, the radius of convergence and therefore the growth with n of the PN coefficients of $h(x)$ (Taylor-expanded at $x = 0$), are linked to the singularity of the analytically continued $h(x)$ which is nearest to $x = 0$ in the complex x -plane. In the $\nu \rightarrow 0$ case, the nearest singularity in the complex x -plane comes from the source factor $\hat{H}_{\text{eff}}(x)$ or $\hat{j}(x)$ in the waveform and is located at the light-ring $x_{\text{LR}}(\nu = 0) = 1/3$. In the $\nu \neq 0$ case, the EOB formalism transforms the latter (inverse square-root) singularity in a more complicated (“branching”) singularity where $d\hat{H}_{\text{eff}}/dx$ and $d\hat{j}/dx$ have inverse square-root singularities located at what is called [34, 36, 37, 61, 65] the (Effective)¹² “EOB-light-ring,” that is, the (adiabatic) maximum of Ω , $x_{\text{ELR}}^{\text{adiab}}(\nu) \equiv (M\Omega_{\text{max}}^{\text{adiab}})^{2/3} \gtrsim 1/3$.

Despite this improvement, the resulting “convergence” of the usual Taylor-expanded $f_{\ell m}(x)$ ’s quoted above does not seem to be good enough, especially near or below the LSO, in view of the high accuracy needed to define GW templates. For this reason, Refs. [60, 61] proposed to further resum the $f_{22}(x)$ function via a Padé (3,2) approximant, $P_2^3\{f_{22}(x; \nu)\}$, so as to improve its behavior in the strong-field-fast-motion regime. Such a resummation gave an excellent agreement with numerically computed waveforms, near the end of the inspiral and during the beginning of the plunge, for different mass ratios [60, 65, 66]. As we were mentioning above, a new route for resumming $f_{\ell m}$ was explored in Ref. [52]. It is based on replacing $f_{\ell m}$ by its ℓ -th root, say

$$\rho_{\ell m}(x; \nu) = [f_{\ell m}(x; \nu)]^{1/\ell}. \quad (57)$$

The basic motivation for replacing $f_{\ell m}$ by $\rho_{\ell m}$ is the following: the leading “Newtonian-level” contribution to the waveform $h_{\ell m}^{(\varepsilon)}$ contains a factor $\omega^\ell r_{\text{harm}}^\ell \nu^\varepsilon$ where r_{harm} is the harmonic radial coordinate used in the MPM formalism [16, 49]. When computing the PN expansion of this factor one has to insert the PN expansion of the (dimensionless) harmonic radial coordinate r_{harm} , $r_{\text{harm}} = x^{-1}(1 + c_1 x + \mathcal{O}(x^2))$, as a function of the gauge-independent frequency parameter x . The PN re-expansion of $[r_{\text{harm}}(x)]^\ell$ then generates terms of the type $x^{-\ell}(1 + \ell c_1 x + \dots)$. This is one (though not the only one) of the origins of 1PN corrections in $h_{\ell m}$ and $f_{\ell m}$ whose coefficients grow linearly with ℓ . The study of [52] has pointed out that these ℓ -growing terms are problematic for the accuracy of the PN-expansions. Our replacement of $f_{\ell m}$ by $\rho_{\ell m}$ is a cure for this problem.

¹² Beware that this “Effective EOB-light-ring” occurs for a circular-orbit radius slightly larger than the purely dynamical (circular) EOB-light-ring (where H_{eff} and \mathcal{J} would formally become infinite).

More explicitly, the investigation of 1PN corrections to GW amplitudes [16, 49, 79] has shown that, in the even-parity case (but see also Appendix A of Ref. [52] for the odd-parity case),

$$c_1^{f_{\ell m}}(\nu) = -\ell \left(1 - \frac{\nu}{3}\right) + \frac{1}{2} + \frac{3}{2} \frac{c_{\ell+2}(\nu)}{c_\ell(\nu)} - \frac{b_\ell(\nu)}{c_\ell(\nu)} - \frac{c_{\ell+2}(\nu)}{c_\ell(\nu)} \frac{m^2(\ell+9)}{2(\ell+1)(2\ell+3)}, \quad (58)$$

where $c_\ell(\nu)$ is defined in Eq. 45 and

$$b_\ell(\nu) \equiv X_2^\ell + (-)^\ell X_1^\ell. \quad (59)$$

Focusing on the $\nu = 0$ case for simplicity (since the ν dependence of $c_1^{f_{\ell m}}(\nu)$ is quite mild [52]), the above result shows that the PN expansion of $f_{\ell m}$ starts as

$$f_{\ell m}^{\text{even}}(x; 0) = 1 - \ell x \left(1 - \frac{1}{\ell} + \frac{m^2(\ell+9)}{2\ell(\ell+1)(2\ell+3)}\right) + \mathcal{O}(x^2). \quad (60)$$

The crucial thing to note in this result is that as ℓ gets large (keeping in mind that $|m| \leq \ell$), the coefficient of x will be negative and will approximately range between $-5\ell/4$ and $-\ell$. This means that when $\ell \geq 6$ the 1PN correction in $f_{\ell m}$ would by itself make $f_{\ell m}(x)$ vanish before the ($\nu = 0$) LSO $x_{\text{LSO}} = 1/6$. For example, for the $\ell = m = 6$ mode, one has $f_{66}^{\text{1PN}}(x; 0) = 1 - 6x(1 + 11/42) \approx 1 - 6x(1 + 0.26)$ which means a correction equal to -100% at $x = 1/7.57$ and larger than -100% at the LSO, namely $f_{66}^{\text{1PN}}(1/6; 0) \approx 1 - 1.26 = -0.26$. This value is totally incompatible with the ‘‘exact’’ value $f_{22}^{\text{exact}}(x_{\text{LSO}}) = 0.66314511$ computed from numerical data in Ref. [52].

Finally, one uses the newly resummed multipolar waveforms 41 to define a resummation of the *radiation reaction force* \mathcal{F}_φ defined as

$$\mathcal{F}_\varphi = -\frac{1}{\Omega} F^{(\ell_{\text{max}})}, \quad (61)$$

where the (instantaneous, circular) GW flux $F^{(\ell_{\text{max}})}$ is defined as

$$F^{(\ell_{\text{max}})} = \frac{2}{16\pi G} \sum_{\ell=2}^{\ell_{\text{max}}} \sum_{m=1}^{\ell} |R\dot{h}_{\ell m}|^2 = \frac{2}{16\pi G} \sum_{\ell=2}^{\ell_{\text{max}}} \sum_{m=1}^{\ell} (m\Omega)^2 |Rh_{\ell m}|^2. \quad (62)$$

As an example of the performance of the new resummation procedure based on the decomposition of $h_{\ell m}$ given by Eq. 41, let us focus, as before, on the computation of the GW energy flux emitted by a test particle on circular orbits on Schwarzschild spacetime. Figure 6 illustrates the remarkable improvement in the closeness between $\hat{F}^{\text{New-resummed}}$ and \hat{F}^{Exact} . The reader should compare this result with the previous Fig. 3 (the straightforward Taylor approximants to the flux),

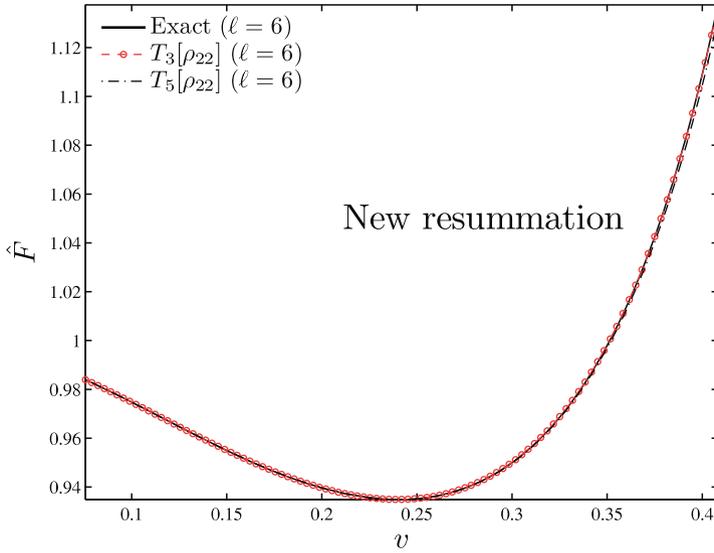


Fig. 6 Performance of the new resummation procedure described in Ref. [52]. The total GW flux \hat{F} (up to $\ell_{\max} = 6$) computed from inserting in Eq. 62 the factorized waveform 41 with the Taylor-expanded $\rho_{\ell m}$'s (with either 3PN or 5PN accuracy for ρ_{22}) is compared with the “exact” numerical data

Fig. 4 (the Padé resummation with $\nu_{\text{pole}} = 1/\sqrt{3}$) and Fig. 5 (the ν_{pole} -tuned Padé resummation). To be fully precise, Fig. 6 plots two examples of fluxes obtained from our new $\rho_{\ell m}$ -representation for the individual multipolar waveforms $h_{\ell m}$. These two examples differ in the choice of approximants for the $\ell = m = 2$ partial wave. One example uses for ρ_{22} its 3PN Taylor expansion, $T_3[\rho_{22}]$, while the other one uses its 5PN Taylor expansion, $T_5[\rho_{22}]$. All the other partial waves are given by their maximum known Taylor expansion.¹³ Note that the fact that we use here for the $\rho_{\ell m}$'s some straightforward Taylor expansions does not mean that this new procedure is not a resummation technique. Indeed, the defining resummation features of our procedure have four sources: (i) the factorization of the PN corrections to the waveforms into four different blocks, namely $\hat{S}_{\text{eff}}^{(\varepsilon)}$, $T_{\ell m}$, $e^{i\delta_{\ell m}}$ and $\rho_{\ell m}^\ell$ in Eq. 41; (ii) the fact that $\hat{S}_{\text{eff}}^{(\varepsilon)}$ is by itself a resummed source whose PN expansion would contain an infinite number of terms; (iii) the fact that the tail factor is a closed form expression (see Eq. 50 above) whose PN expansion also contains an infinite number of terms; and (iv) the fact that we have replaced the Taylor expansion of $f_{\ell m} \equiv \rho_{\ell m}^\ell$ by that of its ℓ -th root, namely $\rho_{\ell m}$.

In conclusion, Eqs. 41 and 62 introduce a new recipe to resum the (ν -dependent) GW energy flux that is alternative to the (ν_{pole} -tuned) one given by Eq. 40. The two

¹³ We recall that Ref. [52] has also shown that the agreement improves even more when the Taylor expansion of the function ρ_{22} is further suitably Padé resummed.

main advantages of the new resummation are: (i) it gives a better representation of the exact result in the $\nu \rightarrow 0$ limit (compare Fig. 6 to Fig. 5), and (ii) it is *parameter-free*: the only flexibility that one has in the definition of the waveform and flux is the choice of the analytical representation of the function f_{22} , like, for instance, $P_2^3\{f_{22}\}$, $(T_3[\rho_{22}])^2$, $(T_5[\rho_{22}])^2$, etc. (although Ref. [52] has pointed out the good consistency among all these choices). Note, that when $\nu \neq 0$, the GW energy flux will depend on the choice of resummation of the radial potential $A(R)$ through the Hamiltonian (for the even-parity modes) or the angular momentum (for the odd-parity modes). At the practical level, this means that the EOB model, implemented with the new resummation procedure of the energy flux (and waveform) described so far, will essentially only depend on the doublet of parameters (a_5, a_6) , that can in principle be constrained by comparison with (accurate) NR results. Contrary to the previous ν_{pole} -resummation of the radiation reaction, this route to resummation is free of radiation–reaction flexibility parameters. We will consider it as our “standard” route to the resummation of the energy flux in the following sections discussing in details the properties of the EOB dynamics and waveforms.

5 EOB Dynamics and Waveforms

In this section, we marry together all the EOB building blocks described in the previous sections and discuss the characteristic of the dynamics of the two black holes as provided by the EOB approach. In the following three subsections we discuss in some detail: (i) the setup of initial data for the EOB dynamics with negligible eccentricity (Section 5.1); (ii) the structure of the full EOB waveform, covering inspiral, plunge, merger, and ringdown, with the introduction of suitable Next-to-Quasi-Circular (NQC) effective corrections to it (and thus to the energy flux) (Section 5.2); (iii) the explicit structure of the EOB dynamics, discussing the solution of the dynamical equations.

5.1 Post–Post-Circular Initial Data

In this subsection we discuss in detail the so-called post–post-circular dynamical initial data (positions and momenta) as introduced in Section III B of [61]. This kind of (improved) construction is needed to have initial data with negligible eccentricity. Since the construction of the initial data is analytical, including the correction is useful to start the system relatively close and to avoid evolving the EOB equation of motion for a long time in order to make the system circularize itself.

To explain the improved construction of initial data let us introduce a formal bookkeeping parameter ε (to be set to 1 at the end) in front of the radiation reaction $\hat{\mathcal{F}}_\varphi$ in the EOB equations of motion. One can then show that the

quasi-circular inspiralling solution of the EOB equations of motion formally satisfies

$$p_\varphi = j_0(r) + \varepsilon^2 j_2(r) + O(\varepsilon^4), \quad (63)$$

$$p_{r_*} = \varepsilon \pi_1(r) + \varepsilon^3 \pi_3(r) + O(\varepsilon^5). \quad (64)$$

Here, $j_0(r)$ is the usual *circular* approximation to the inspiralling angular momentum as explicitly given by Eq. 47 above. The order ε (“post-circular”) term $\pi_1(r)$ is obtained by: (i) inserting the circular approximation $p_\varphi = j_0(r)$ on the left-hand side (l.h.s.) of Eq. 10 of [59], (ii) using the chain rule $dj_0(r)/dt = (dj_0(r)/dr)(dr/dt)$, (iii) replacing dr/dt by the right-hand side (r.h.s) of Eq. 9 of [59] and (iv) solving for p_{r_*} at the first order in ε . This leads to an explicit result of the form (using the notation defined in Ref. [59])

$$\varepsilon \pi_1(r) = \left[v \hat{H} \hat{H}_{\text{eff}} \left(\frac{B}{A} \right)^{1/2} \left(\frac{dj_0}{dr} \right)^{-1} \hat{\mathcal{F}}_\varphi \right]_0, \quad (65)$$

where the subscript 0 indicates that the r.h.s. is evaluated at the leading circular approximation $\varepsilon \rightarrow 0$. The post-circular EOB approximation (j_0, π_1) was introduced in Ref. [36] and then used in most of the subsequent EOB papers [33, 34, 37, 59, 83, 85]. The *post-post-circular* approximation (order ε^2), introduced in Ref. [61] and then used systematically in Ref. [62, 65, 66], consists of: (i) formally solving Eq. 35 with respect to the explicit p_φ^2 appearing on the r.h.s.; (ii) replacing p_{r_*} by its post-circular approximation, Eq. 65; (iii) using the chain rule $d\pi_1(r)/dt = (d\pi_1(r)/dr)(dr/dt)$; and (iv) replacing dr/dt in terms of π_1 (to leading order) by using Eq. 33. The result yields an explicit expression of the type $p_\varphi^2 \simeq j_0^2(r)[1 + \varepsilon^2 k_2(r)]$ of which one finally takes the square root. In principle, this procedure can be iterated to get initial data at any order in ε . As it will be shown below, the post-post-circular initial data ($j_0 \sqrt{1 + \varepsilon^2 k_2}, \pi_1$) are sufficient to lead to negligible eccentricity when starting the integration of the EOB equations of motion at radius $r \equiv R/(GM) = 15$.

5.2 EOB Waveforms

At this stage we have essentially discussed all the elements that are needed to compute the EOB dynamics obtained by solving the EOB equation of motion, Eqs. 32–35. The dynamics of the system yields $(\mathbf{q}(t), \mathbf{p}(t)) \equiv (\varphi(t), r(t), p_\varphi(t), p_{r_*}(t))$, namely, the trajectory in phase space. The (multipolar) metric waveform during the inspiral and plunge phase, up to the EOB “merger time” t_m (that is defined as the maximum of the orbital frequency Ω), is a function of this trajectory, that is, $h_{\ell m}^{\text{insplunge}} \equiv h_{\ell m}^{\text{insplunge}}(\mathbf{q}(t), \mathbf{p}(t))$. Focusing only on the dominant $\ell = m = 2$ waveform, the waveform that describes the full process of the binary

black-hole coalescence (i.e., inspiral, plunge, merger, and ringdown) can be split in to two parts:

- The *insplunge waveform*: $h_{22}^{\text{insplunge}}(t)$, computed along the EOB dynamics up to merger, which includes (i) the resummation of the “tail” terms described above and (ii) some effective parametrization of NQC effects. The $\ell = m = 2$ metric waveform explicitly reads

$$\left(\frac{Rc^2}{GM}\right) h_{22}^{\text{insplunge}}(t) = v n_{22}^{(0)} c_2(v) x \hat{h}_{22}(v; x) f_{22}^{\text{NQC}} Y^{2,-2}\left(\frac{\pi}{2}, \Phi\right), \quad (66)$$

where the argument x is taken to be (following [47]) $x = v_\phi^2 = (r_\omega \Omega)^2$ (where r_ω was introduced in Eq. 36 above). The resummed version of f_{22} entering in $\hat{h}_{22}(x)$ used here is given by the following *Padé-resummed* function $f_{22}^{\text{Pf}} \equiv P_2^3[f_{22}^{\text{Taylor}}(x; v)]$. In the waveform h_{22} above we have introduced (following [62]) a new ingredient, a “NQC” correction factor of the form¹⁴

$$f_{22}^{\text{NQC}}(a_1, a_2) = 1 + a_1 \frac{p_{r_*}^2}{(r\Omega)^2} + a_2 \frac{\ddot{r}}{r\Omega^2}, \quad (67)$$

where a_1 and a_2 are free parameters that have to be fixed. A crucial facet of the new EOB formalism presented here consists in trying to be as predictive as possible by reducing to an absolute minimum the number of “flexibility parameters” entering our theoretical framework. One can achieve this aim by “analytically” determining the two parameters a_1, a_2 entering [via the NQC factor Eq. 67] the (asymptotic) quadrupolar EOB waveform $\hat{R}h_{22}^{\text{EOB}}$ (where $\hat{R} = R/M$) by imposing: (a) that the modulus $|\hat{R}h_{22}^{\text{EOB}}|$ reaches, at the EOB-determined “merger time” t_m , a *local maximum*, and (b) that the value of this maximum EOB modulus is equal to a certain (dimensionless) function of v , $\varphi(v)$. In Ref. [62] we calibrated $\varphi(v)$ (independently of the EOB formalism) by extracting from the best current NR simulations the maximum value of the modulus of the NR quadrupolar *metric* waveform $|\hat{R}h_{22}^{\text{NR}}|$. Using the data reported in [99] and [66], and considering the “Zerilli-normalized” asymptotic metric waveform $\Psi_{22} = \hat{R}h_{22}/\sqrt{24}$, we found $\varphi(v) \simeq 0.3215v(1 - 0.131(1 - 4v))$. Our requirements (a) and (b) impose, for any given $A(u)$ potential, *two constraints* on the *two parameters* a_1, a_2 . We can solve these two constraints (by an iteration procedure) and thereby uniquely determine the values of a_1, a_2 corresponding to any given $A(u)$ potential. In particular, in the case considered here where $A(u) \equiv A(u; a_5, a_6, v)$ this uniquely determines a_1, a_2 in function of a_5, a_6 and v . Note that this is done while also consistently using the “improved” version of h_{22} given by Eq. 66 to compute the radiation reaction force via Eq. 62.

¹⁴ Note that one could also similarly improve the subleading higher-multipolar-order contributions to \mathcal{F}_ϕ . In addition, other (similar) expressions of the NQC factors can be found in the literature [38, 65, 66].

- A simplified representation of the transition between plunge and ringdown by smoothly *matching* (following Refs. [60]), on a $(2p + 1)$ -toothed “comb” $(t_m - p\delta, \dots, t_m - \delta, t_m, t_m + \delta, \dots, t_m + p\delta)$ centered around a matching time t_m , the inspiral-plus-plunge waveform to a ring-down waveform, made of the superposition of several¹⁵ quasi-normal-mode complex frequencies,

$$\left(\frac{Rc^2}{GM}\right) h_{22}^{\text{ringdown}}(t) = \sum_N C_N^+ e^{-\sigma_N^+(t-t_m)}, \quad (68)$$

with $\sigma_N^+ = \alpha_N + i\omega_N$, and where the label N refers to indices (ℓ, ℓ', m, n) , with $(\ell, m) = (2, 2)$ being the Schwarzschild-background multipolarity of the considered (metric) waveform $h_{\ell m}$, with $n = 0, 1, 2, \dots$ being the “overtone number” of the considered Kerr-background Quasi-Normal-Mode, and ℓ' the degree of its associated spheroidal harmonics $S_{\ell' m}(a\sigma, \theta)$. As discussed in [36] and [60], and already mentioned above, the physics of the transition between plunge and ringdown (which was first understood in the classic work of Davis et al. [68]) suggests to choose as matching time t_m , in the comparable-mass case, the EOB time when the EOB orbital frequency $\Omega(t)$ reaches its *maximum* value.

Finally, one defines a complete, quasi-analytical EOB waveform (covering the full process from inspiral to ringdown) as:

$$h_{22}^{\text{EOB}}(t) = \theta(t_m - t) h_{22}^{\text{insplunge}}(t) + \theta(t - t_m) h_{22}^{\text{ringdown}}(t), \quad (69)$$

where $\theta(t)$ denotes Heaviside’s step function. The final result is a waveform that only depends on the *two* parameters (a_5, a_6) which parametrize some flexibility on the Padé resummation of the basic radial potential $A(u)$, connected to the yet uncalculated (4PN, 5PN, and higher) PN contributions.

5.3 EOB Dynamics

We conclude this section by discussing the features of the typical EOB dynamics obtained by solving the EOB equation of motion 32–35 with post–post-circular initial data. The resummation of the radiation reaction force uses the multiplicative decomposition of $h_{\ell m}$ given by Eq. 41 with NQC correction to the $\ell = m = 2$ multipole given by Eq. 67. We fix the free parameters to the model to be $a_5 = 0$, $a_6 = -20$ (see below why) while a_1 and a_2 are obtained consistently according to the iteration procedure discussed above. The system is started at $r_0 = 15$ and $\varphi_0 = 0$. The post–post-circular initial data give $p_\varphi^0 = 4.31509298$ and $p_{r^*}^0 = -0.00109847$.

¹⁵ Refs. [60, 65] use $p = 2$, that is, a five-teethed comb, and, correspondingly, five positive-frequency Kerr Quasi-Normal Modes.

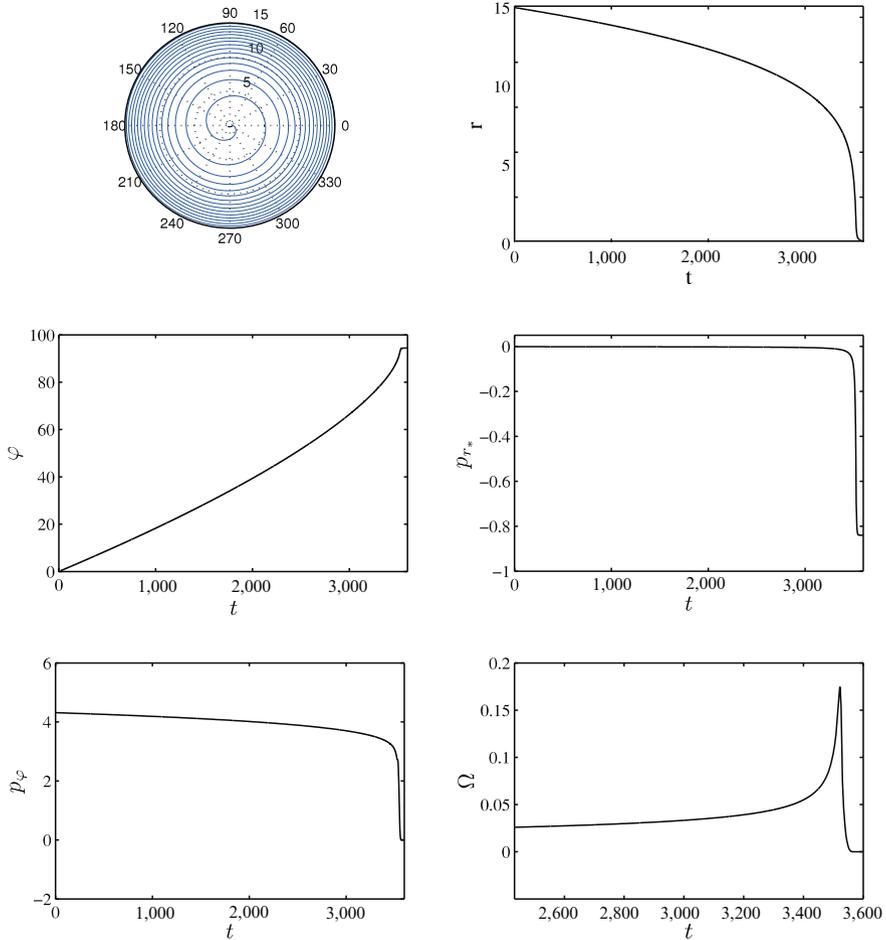


Fig. 7 EOB dynamics for $a_5 = 0$ and $a_6 = -20$. Clockwise from the top left panel, the panels report: the trajectory, the radial separation $r(t)$, the radial momentum p_{r^*} (conjugate to r_*), the orbital frequency $\Omega(t)$, the angular momentum $p_\varphi(t)$, and the orbital phase $\varphi(t)$

The result of the outcome of the integration of the EOB equation of motion is displayed in Fig. 7 together with the trajectory (top-left panel) and the orbital frequency (bottom-right panel). On this plot we remark two things. First, the fact that the orbital frequency has a maximum at time $t_m = 3,522$ that identifies, in EOB, the merger (and matching) time. Second, the fact that p_{r^*} tends to a finite value after the merger (contrary to p_r , that would diverge), yielding a more controllable numerical treatment of the late part of the EOB dynamics.

6 EOB and NR Waveforms

So far we have seen that (at least) two different EOB models (of dynamics and waveforms) are available. They differ, essentially, in the way the resummation of the GW energy flux yielding the radiation reaction force is performed. The first EOB model, that we will refer to as the “old” one, basically uses a Padé-resummation of the energy flux with an external parameter v_{pole} that must be fixed in some way. The second EOB model, that we will refer to as the “improved” one, uses a more sophisticated resummation procedure of the energy flux, multipole by multipole, in such a way that the final result depends explicitly only on the same parameters (a_5, a_6) that are used to parametrize higher PN contribution to the conservative part of the dynamics.

In the last 3 years, the power of the “old” EOB model has been exploited in various comparisons with NR data, aiming at constraining in some way the space of the EOB flexibility parameters (notably represented by a_5 and v_{pole}) by looking at regions in the parameter space where the agreement between the numerical and analytical waveforms is at the level of numerical error. For example, after a preliminary comparison done in Ref. [34], Buonanno et al. [37] compared *restricted* EOB waveforms¹⁶ to NR waveforms computed by the NASA-Goddard group, showing that it is possible to tune the value of a_5 so as to have a good agreement between the two set of data. In particular, for $a_5 = 60$ and v_{pole} given according to the (nowadays outdated) suggestion of Ref. [53], in the equal-mass case ($\nu = 1/4$), they found that the dephasing between (restricted) EOB and NR waveforms (covering late inspiral, merger and ringdown) stayed within ± 0.030 GW cycles over 14 GW cycles. In the case of a mass ratio 4 : 1 ($\nu = 0.16$), the dephasing stayed within ± 0.035 GW cycles over 9 GW cycles.

Later, the *resummed* factorized EOB waveform of Eq. 66 above within the “old” EOB model has been compared to several set of equal-mass and unequal-mass NR waveforms: (i) in the comparison with the very accurate inspiralling simulation of the Caltech–Cornell group [29] the dephasing stayed smaller than ± 0.001 GW cycles over 30 GW cycles (and the amplitudes agreed at the $\sim 10^{-3}$ level) [61]; (ii) in the comparison [65] with a late-inspiral–merger–ringdown NR waveform computed by the Albert Einstein Institute group, the dephasing stayed smaller than ± 0.005 GW cycles over 12 GW cycles; (iii) in the (joint) comparison [66] between EOB and very accurate equal-mass inspiralling simulation of the Caltech–Cornell group [29] and late-inspiral–merger–ringdown waveform for 1:1, 2:1 and 4:1 mass-ratio data computed by the Jena group it was possible to tune the EOB flexibility parameters (notably a_5 and v_{pole}) so that the dephasing stayed at the level of the numerical error. The same “old” model, with resummed factorized waveform, and the parameter-dependent (using v_{pole}) resummation of radiation reaction force, was recently extended by adding 6 more flexibility parameters to the

¹⁶ The terminology “restricted” refers to a waveform which uses only the leading *Newtonian* approximation, $h_{\ell m}^{(N,\varepsilon)}$, to the waveform.

ones already introduced in Refs. [61, 66], and was “calibrated” on the high-accuracy Caltech–Cornell equal-mass data [38]. This calibration showed that only 5 flexibility parameters (a_5 , v_{pole} , and three parameters related to non-quasi-circular corrections to the waveform amplitude) actually suffice to make the “old” EOB and NR waveform agree, both in amplitude and phase, at the level of the numerical error (this multi-flexed EOB model brings in an improvement with respect to the one of Refs. [61, 66] especially for what concerns the agreement between the waveform amplitude around the merger).

Recently, Ref. [62] has introduced and fully exploited the possibilities of the “improved” EOB formalism described above, taking advantage of: (i) the multiplicative decomposition of the (resummed) multipolar waveform advocated in Eq. 41 above; (ii) the effect of the NQC corrections to the waveform (and energy flux) given by Eq. 66, and, most importantly; (iii) the parameter-free resummation of radiation reaction \mathcal{F}_φ . In Ref. [62] the (a_5, a_6) -dependent predictions made by the “improved” formalism were compared to the high-accuracy waveform from an equal-mass BBH ($\nu = 1/4$) computed by the Caltech–Cornell group [99] (and now made available on the web). It was found that there is a strong degeneracy between a_5 and a_6 in the sense that there is an excellent EOB-NR agreement for an extended region in the (a_5, a_6) -plane. More precisely, the phase difference between the EOB (metric) waveform and the Caltech–Cornell one, considered between GW frequencies $M\omega_L = 0.047$ and $M\omega_R = 0.31$ (i.e., the last 16 GW cycles before merger), stays smaller than 0.02 radians within a long and thin banana-like region in the (a_5, a_6) -plane. This “good region” approximately extends between the points $(a_5, a_6) = (0, -20)$ and $(a_5, a_6) = (-36, +520)$. As an example (which actually lies on the boundary of the “good region”), we have followed [62] in considering here the specific values $a_5 = 0, a_6 = -20$ (to which correspond, when $\nu = 1/4, a_1 = -0.036347, a_2 = 1.2468$). We henceforth use M as time unit.

This result relies on the proper comparison between NR and EOB time series, which is a delicate subject. In fact, to compare the NR and EOB phase time series $\phi_{22}^{\text{NR}}(t_{\text{NR}})$ and $\phi_{22}^{\text{EOB}}(t_{\text{EOB}})$ one needs to shift, by additive constants, both one of the time variables, and one of the phases. In other words, we need to determine τ and α such that the “shifted” EOB quantities

$$t'_{\text{EOB}} = t_{\text{EOB}} + \tau \quad \phi'_{22}{}^{\text{EOB}} = \phi_{22}^{\text{EOB}} + \alpha, \quad (70)$$

“best fit” the NR ones. One convenient way to do so is first to “pinch” the EOB/NR phase difference at two different instants (corresponding to two different frequencies). More precisely, one can choose two NR times $t_1^{\text{NR}}, t_2^{\text{NR}}$, which determine two corresponding GW frequencies¹⁷ $\omega_1 = \omega_{22}^{\text{NR}}(t_1^{\text{NR}})$, $\omega_2 = \omega_{22}^{\text{NR}}(t_2^{\text{NR}})$, and then find the time shift $\tau(\omega_1, \omega_2)$ such that the shifted EOB phase difference, between ω_1 and ω_2 , $\Delta\phi^{\text{EOB}}(\tau) \equiv \phi'_{22}{}^{\text{EOB}}(t_2^{\text{EOB}}) - \phi'_{22}{}^{\text{EOB}}(t_1^{\text{EOB}}) = \phi_{22}^{\text{EOB}}(t_2^{\text{EOB}} + \tau) - \phi_{22}^{\text{EOB}}(t_1^{\text{EOB}} + \tau)$ is equal to the corresponding (unshifted) NR phase difference

¹⁷ Alternatively, one can start by giving oneself ω_1, ω_2 and determine the NR instants $t_1^{\text{NR}}, t_2^{\text{NR}}$ at which they are reached.

$\Delta\phi^{\text{NR}} \equiv \phi_{22}^{\text{NR}}(t_2^{\text{NR}}) - \phi_{22}^{\text{NR}}(t_1^{\text{NR}})$. This yields one equation for one unknown (τ), and (uniquely) determines a value $\tau(\omega_1, \omega_2)$ of τ . (Note that the $\omega_2 \rightarrow \omega_1 = \omega_m$ limit of this procedure yields the one-frequency matching procedure used in [29].) After having so determined τ , one can uniquely define a corresponding best-fit phase shift $\alpha(\omega_1, \omega_2)$ by requiring that, say, $\phi_{22}^{\text{EOB}}(t_1^{\text{EOB}}) \equiv \phi_{22}^{\text{EOB}}(t_1^{\text{EOB}}) + \alpha = \phi_{22}^{\text{NR}}(t_1^{\text{NR}})$.

Having so related the EOB time and phase variables to the NR ones we can straightforwardly compare all the EOB time series to their NR correspondants. In particular, we can compute the (shifted) EOB–NR phase difference

$$\Delta^{\omega_1, \omega_2} \phi_{22}^{\text{EOBNR}}(t_{\text{NR}}) \equiv \phi_{22}^{\text{EOB}}(t^{\text{EOB}}) - \phi_{22}^{\text{NR}}(t^{\text{NR}}). \tag{71}$$

Figure 8 compares¹⁸ (the real part of) our analytical *metric* quadrupolar waveform $\Psi_{22}^{\text{EOB}}/\nu$ to the corresponding (Caltech–Cornell) NR *metric* waveform $\Psi_{22}^{\text{NR}}/\nu$. This NR metric waveform has been obtained by a double time integration (following the procedure of Ref. [66]) from the original, publicly available, *curvature* waveform ψ_4^{22} . Such a curvature waveform has been extrapolated *both* in resolution and in extraction radius. The agreement between the analytical prediction and the NR result is striking, even around the merger. See Fig. 9 which closes up on the merger. The vertical line indicates the location of the EOB-merger time, that is, the location of the maximum of the orbital frequency.

The phasing agreement between the waveforms is excellent over the full time span of the simulation (which covers 32 cycles of inspiral and about 6 cycles of ringdown), while the modulus agreement is excellent over the full span, apart from

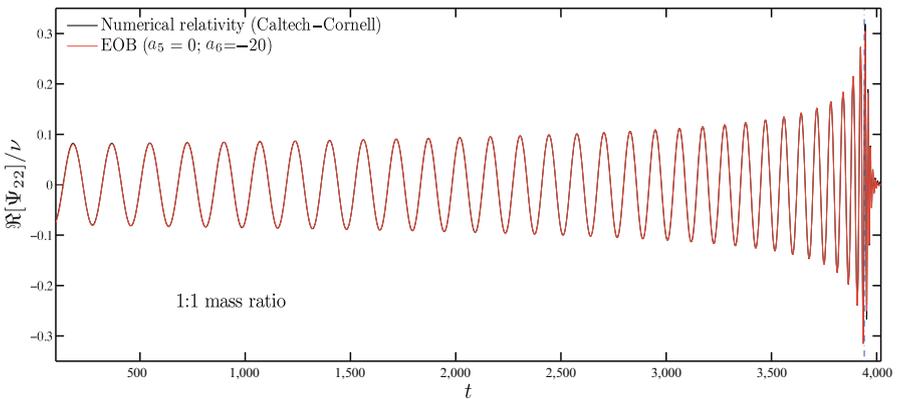


Fig. 8 This figure illustrates the comparison between the “improved” EOB waveform (quadrupolar ($\ell = m = 2$) metric waveform 66 with parameter-free radiation reaction 61 and with $a_5 = 0$, $a_6 = -20$) with the most accurate NR waveform (equal-mass case) nowadays available. The phase difference between the two is $\Delta\phi \leq \pm 0.01$ radians during the entire inspiral and plunge. Ref. [62] has shown that this agreement is at the level of the numerical error

¹⁸ The two frequencies used for this comparison, by means of the “two-frequency pinching technique” mentioned above, are $M\omega_1 = 0.047$ and $M\omega_2 = 0.31$.

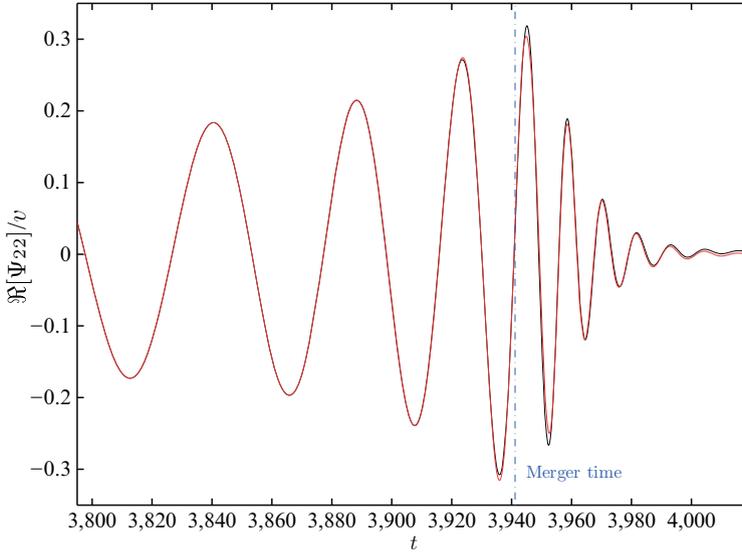


Fig. 9 Close up around merger of the waveforms of Fig. 8. Note the excellent agreement between *both* modulus and phasing also during the ringdown phase

two cycles after merger where one can notice a difference. More precisely, the phase difference, $\Delta\phi = \phi_{\text{metric}}^{\text{EOB}} - \phi_{\text{metric}}^{\text{NR}}$, remains remarkably small ($\sim \pm 0.02$ radians) during the entire inspiral and plunge ($\omega_2 = 0.31$ being quite near the merger). By comparison, the root-sum of the various numerical errors on the phase (numerical truncation, outer boundary, extrapolation to infinity) is about 0.023 radians during the inspiral [99]. At the merger, and during the ringdown, $\Delta\phi$ takes somewhat larger values ($\sim \pm 0.1$ radians), but it oscillates around zero, so that, on average, it stays very well in phase with the NR waveform (as is clear on Fig. 9). By comparison, we note that [99] mentions that the phase error linked to the extrapolation to infinity doubles during ringdown. We then note that the total “two-sigma” NR error level estimated in [99] rises to 0.05 radians during ringdown, which is comparable to the EOB/NR phase disagreement. In addition, Ref. [62] compared the “improved” EOB waveform to accurate NR data (obtained by the Jena group [66]) on the coalescence of *unequal mass-ratio* black-hole binaries. Figure 10 shows the result of the EOB/NR waveform comparison for a 2:1 mass ratio, corresponding to $\nu = 2/9$. When $a_5 = 0$, $a_6 = -20$, one finds $a_1 = -0.017017$ and $a_2 = 1.1906$. Again, the agreement is excellent, and within the numerical error bars.

Finally, Ref. [62] explored another aspect of the physical soundness of the EOB analytical formalism: the *triple* comparison between (i) the NR GW energy flux at infinity (computed in [30]); (ii) the corresponding analytically predicted GW energy flux at infinity (computed by summing $|\dot{h}_{\ell m}|^2$ over ℓ, m); and (iii) (minus) the *mechanical* energy loss of the system, as predicted by the general EOB formalism, that is, the “work” done by the radiation reaction $\dot{E}_{\text{mechanical}} = \Omega \mathcal{F}_\phi$. This comparison is shown in Fig. 11, which should be compared to Fig. 9 of [30].

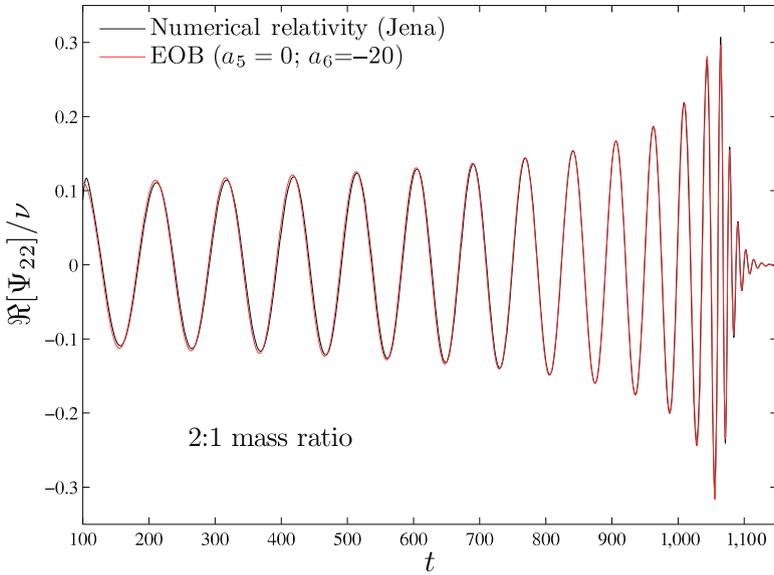


Fig. 10 Comparison between NR and EOB metric waveform for the 2:1 mass ratio

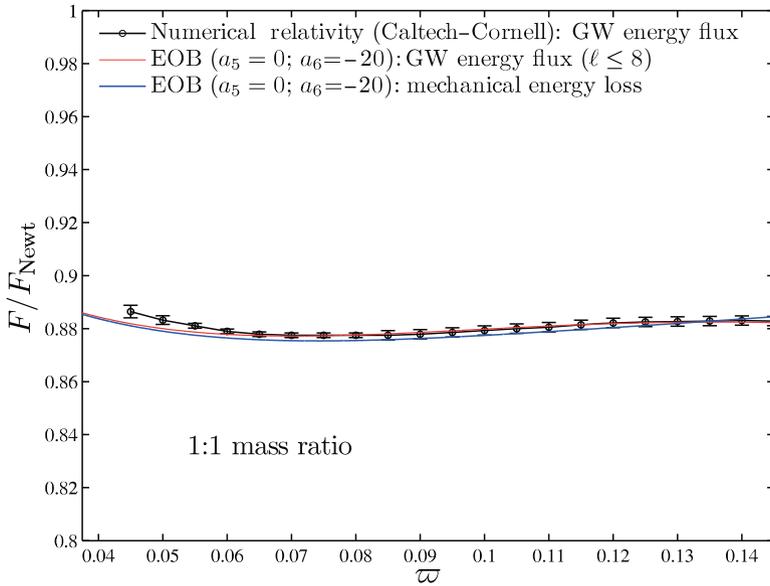


Fig. 11 The triple comparison between NR and EOB GW energy fluxes and the EOB mechanical energy loss

We kept here the same vertical scale as [30] which compared the NR flux to older versions of (resummed and non-resummed) analytical fluxes and needed such a $\pm 10\%$ vertical scale to accommodate all the models they considered. (The horizontal axis is the frequency ϖ of the differentiated metric waveform \dot{h}_{22} .) By contrast, we see again the striking closeness (at the $\sim 2 \times 10^{-3}$ level) between the EOB and NR GW fluxes. As both fluxes include higher multipoles than the (2, 2) one, this closeness is a further test of the agreement between the improved EOB formalism and NR results. (We think that the $\sim 2\sigma$ difference between the (coinciding) analytical curves and the NR one on the left of the figure is due to uncertainties in the flux computation of [30], possibly related to the method used there of computing \dot{h} .) Note that the rather close agreement between the analytical energy flux and the mechanical energy loss during late inspiral is not required by physics (because of the well-known ‘‘Schott term’’ [100]), but is rather an indication that $\dot{h}_{\ell m}$ can be well approximated by $-im\Omega h_{\ell m}$.

7 Conclusions

We have reviewed the basic elements of the EOB formalism. This formalism is still under development. The various existing versions of the EOB formalism have all shown their capability to reproduce within numerical errors the currently most accurate NR simulations of coalescing binary black holes. These versions differ in the number of free theoretical parameters. Recently, a new ‘‘improved’’ version of the formalism has been defined which contains essentially only *two* free theoretical parameters.

Among the successes of the EOB formalism let us mention:

1. An analytical understanding of the nonadiabatic late-inspiral dynamics and of its ‘‘blurred’’ transition to a quasi-circular plunge
2. The surprising possibility to analytically describe the merger of two black holes by a seemingly coarse approximation consisting of matching a continued inspiral to a ring-down signal
3. The capability, after using suitable resummation methods, to reproduce with exquisite accuracy *both* the phase and the amplitude of the GW signal emitted during the entire coalescence process, from early-inspiral to late-inspiral, plunge, merger, and ringdown
4. The GW energy flux predicted by the EOB formalism agrees, within numerical errors, with the most accurate numerical-relativity energy flux
5. The ability to correctly estimate (within a 2% error) the final spin and mass of nonspinning coalescing black-hole binaries (this issue has not been discussed in this review, but see Ref. [59])

We anticipate that the EOB formalism will also be able to provide an accurate description of more complicated systems than the nonspinning BBH discussed in this review. On the one hand, we think that the recently improved EOB framework can

be extended to the description of (nearly circularized) *spinning* black-hole systems by suitably incorporating both the PN-expanded knowledge of spin effects [3,57,69] and their possible EOB resummation [45,58]. On the other hand, the EOB formalism can also be extended to the description of binary neutron stars or mixed binary systems made of a black hole and a neutron star [64]. An important input for this extension is the use of the relativistic tidal properties of neutron stars [11,63,75].

Finally, we think that the EOB formalism has opened the realistic possibility of constructing (with minimal computational resources) a very accurate, large bank of GW templates, thereby helping in both detecting and analyzing the signals emitted by inspiralling and coalescing binary black holes. Though we have had in mind in this review essentially ground-based detectors, we think that the EOB method can also be applied to space-based ones, that is, to (possibly eccentric) large-mass-ratio systems.

Acknowledgement AN is grateful to Alessandro Spallicci, Bernard Whiting, and all the organizers of the “Ecole thématique du CNRS sur la masse (origine, mouvement, mesure).” Among the many colleagues whom we benefitted from, we would like to thank particularly Emanuele Berti, Bernd Brügmann, Alessandra Buonanno, Nils Dorband, Mark Hannam, Sascha Husa, Bala Iyer, Larry Kidder, Eric Poisson, Denis Pollney, Luciano Rezzolla, B.S. Sathyaprakash, Angelo Tartaglia, and Loic Villain, for fruitful collaborations and discussions. We are also grateful to Marie-Claude Vergne for help with Fig. 1.

References

1. P. Ajith, S. Babak, Y. Chen, M. Hewitson, B. Krishnan, A.M. Sintes, J.T. Whelan, J. Gonzalez, M. Hannam, S. Husa, M. Koppitz, D. Pollney, L. Rezzolla, L. Santamaria, U. Sperhake, J. Thornburg, *Class. Q. Grav.* **24**, S689 (2007)
2. P. Ajith, S. Babak, Y. Chen, M. Hewitson, B. Krishnan, A.M. Sintes, J.T. Whelan, B. Brügmann, P. Diener, N. Dorband, J. Gonzalez, M. Hannam, S. Husa, D. Pollney, L. Rezzolla, L. Santamaria, U. Sperhake, J. Thornburg, *Phys. Rev. D* **77**, 104017 (2008)
3. K.G. Arun, A. Buonanno, G. Faye, E. Ochsner, *Phys. Rev. D* **79**, 104023 (2009)
4. J.G. Baker, J. Centrella, D.I. Choi, M. Koppitz, J. van Meter, *Phys. Rev. Lett.* **96**, 111102 (2006)
5. J.G. Baker, J. Centrella, D.I. Choi, M. Koppitz, J.R. van Meter, M.C. Miller, *Astrophys. J.* **653**, L93 (2006)
6. J.G. Baker, J. Centrella, D.I. Choi, M. Koppitz, J. van Meter, *Phys. Rev. D* **73**, 104002 (2006)
7. J.G. Baker, J.R. van Meter, S.T. McWilliams, J. Centrella, B.J. Kelly, *Phys. Rev. Lett.* **99**, 181101 (2007)
8. J.G. Baker, M. Campanelli, F. Pretorius, Y. Zlochower, *Class. Q. Grav.* **24**, S25 (2007)
9. J.G. Baker, S.T. McWilliams, J.R. van Meter, J. Centrella, D.I. Choi, B.J. Kelly, M. Koppitz, *Phys. Rev. D* **75**, 124024 (2007)
10. E. Berti, V. Cardoso, J.A. Gonzalez, U. Sperhake, M. Hannam, S. Husa, B. Brügmann, *Phys. Rev. D* **76**, 064034 (2007)
11. T. Binnington, E. Poisson, *Phys. Rev. D* **80**, 084018 (2009)
12. L. Blanchet, *Phys. Rev. D* **51**, 2559 (1995)
13. L. Blanchet, *Class. Q. Grav.* **15**, 113 (1998); Erratum, *ibidem* **22**, 3381 (2005)
14. L. Blanchet, *Living Rev. Rel.* **5**, URL: <http://www.livingreviews.org/lrr-2002-3>
15. L. Blanchet, T. Damour, *Phil. Trans. R. Soc. Lond. A* **320**, 379 (1986)
16. L. Blanchet, T. Damour, *Ann. Inst. H. Poincaré (A) Phys. Théor.* **50**, 377 (1989)

17. L. Blanchet, T. Damour, Phys. Rev. D **46**, 4304 (1992)
18. L. Blanchet, T. Damour, G. Esposito-Farese, Phys. Rev. D **69**, 124007 (2004)
19. L. Blanchet, T. Damour, G. Esposito-Farese, B.R. Iyer, Phys. Rev. D **51**, 5360 (1995); Erratum, *ibidem* **54**, 1860 (1996)
20. L. Blanchet, T. Damour, G. Esposito-Farese, B.R. Iyer, Phys. Rev. D **71**, 124004 (2005)
21. L. Blanchet, T. Damour, B.R. Iyer, Phys. Rev. D **51**, 5360 (1995); Erratum, *ibidem* **54**, 1860 (1996)
22. L. Blanchet, T. Damour, B.R. Iyer, C.M. Will, A.G. Wiseman, Phys. Rev. Lett. **74**, 3515 (1995)
23. L. Blanchet, G. Faye, Phys. Rev. D **63**, 062005 (2001)
24. L. Blanchet, G. Faye, B.R. Iyer, B. Joguet, Phys. Rev. D **65**, 061501(R) (2002); Erratum, *ibidem* **71**, 129902(E) (2005)
25. L. Blanchet, G. Faye, B.R. Iyer, S. Sinha, Phys. Rev. D **80**, 124018 (2009)
26. L. Blanchet, B.R. Iyer, Phys. Rev. D **71**, 024004 (2005)
27. L. Blanchet, B.R. Iyer, B. Joguet, Phys. Rev. D **65**, 064005 (2002); Erratum, *ibidem* **71**, 129903(E) (2005)
28. L. Blanchet, G. Schäfer, Class. Q. Grav. **10**, 2699 (1993)
29. M. Boyle, D.A. Brown, L.E. Kidder, A.H. Mroue, H.P. Pfeiffer, M.A. Scheel, G.B. Cook, S.A. Teukolsky, Phys. Rev. D **76**, 124038 (2007)
30. M. Boyle, A. Buonanno, L.E. Kidder, A.H. Mroue, Y. Pan, H.P. Pfeiffer, M.A. Scheel, Phys. Rev. D **78**, 104020 (2008)
31. M. Boyle, L. Lindblom, H. Pfeiffer, M. Scheel, L.E. Kidder, Phys. Rev. D **75**, 024006 (2007)
32. E. Brezin, C. Itzykson, J. Zinn-Justin, Phys. Rev. D **1**, 2349 (1970)
33. A. Buonanno, Y. Chen, T. Damour, Phys. Rev. D **74**, 104005 (2006)
34. A. Buonanno, G.B. Cook, F. Pretorius, Phys. Rev. D **75**, 124018 (2007)
35. A. Buonanno, T. Damour, Phys. Rev. D **59**, 084006 (1999)
36. A. Buonanno, T. Damour, Phys. Rev. D **62**, 064015 (2000)
37. A. Buonanno, Y. Pan, J.G. Baker, J. Centrella, B.J. Kelly, S.T. McWilliams, J.R. van Meter, Phys. Rev. D **76**, 104049 (2007)
38. A. Buonanno, Y. Pan, H.P. Pfeiffer, M.A. Scheel, L.T. Buchman, L.E. Kidder, Phys. Rev. D **79**, 124028 (2009)
39. M. Campanelli, C.O. Lousto, P. Marronetti, Y. Zlochower, Phys. Rev. Lett. **96**, 111101 (2006)
40. M. Campanelli, C.O. Lousto, Y. Zlochower, Phys. Rev. D **73**, 061501 (2006)
41. M. Campanelli, C.O. Lousto, Y. Zlochower, Phys. Rev. D **74**, 041501 (2006)
42. M. Campanelli, C.O. Lousto, Y. Zlochower, D. Merritt, Phys. Rev. Lett. **98**, 231102 (2007)
43. C. Cutler, E. Poisson, G.J. Sussman, L.S. Finn, Phys. Rev. D **47**, 1511 (1993)
44. T. Damour, in *Gravitational Radiation*, ed. by N. Deruelle, T. Piran (North Holland, Amsterdam, 1983) pp. 59–144
45. T. Damour, Phys. Rev. D **64**, 124013 (2001)
46. T. Damour, N. Deruelle, Phys. Lett. A **87**, 81 (1981)
47. T. Damour, A. Gopakumar, Phys. Rev. D **73**, 124006 (2006)
48. T. Damour, E.ourgoulhon, P. Grandclement, Phys. Rev. D **66**, 024007 (2002)
49. T. Damour, B.R. Iyer, Ann. Inst. H. Poincaré (A) Phys. Théor. **54**, 115 (1991)
50. T. Damour, B.R. Iyer, Phys. Rev. D **43**, 3259 (1991)
51. T. Damour, B.R. Iyer, P. Jaranowski, B.S. Sathyaprakash, Phys. Rev. D **67**, 064028 (2003)
52. T. Damour, B.R. Iyer, A. Nagar, Phys. Rev. D **79**, 064004 (2009)
53. T. Damour, B.R. Iyer, B.S. Sathyaprakash, Phys. Rev. D **57**, 885 (1998)
54. T. Damour, B.R. Iyer, B.S. Sathyaprakash, Phys. Rev. D **63**, 044023 (2001); Erratum, *ibidem* **72**, 029902(E) (2005)
55. T. Damour, P. Jaranowski, G. Schäfer, Phys. Rev. D **62**, 084011 (2000)
56. T. Damour, P. Jaranowski, G. Schäfer, Phys. Lett. B **513**, 147 (2001)
57. T. Damour, P. Jaranowski, G. Schäfer, Phys. Rev. D **77**, 064032 (2008)
58. T. Damour, P. Jaranowski, G. Schäfer, Phys. Rev. D **78**, 024009 (2008)
59. T. Damour, A. Nagar, Phys. Rev. D **76**, 044003 (2007)
60. T. Damour, A. Nagar, Phys. Rev. D **76**, 064028 (2007)

61. T. Damour, A. Nagar, Phys. Rev. D **77**, 024043 (2008)
62. T. Damour, A. Nagar, Phys. Rev. D **79**, 081503 (2009)
63. T. Damour, A. Nagar, Phys. Rev. D **80**, 084035 (2009)
64. T. Damour, A. Nagar, Phys. Rev. D **81**, 084016 (2010)
65. T. Damour, A. Nagar, E.N. Dorband, D. Pollney, L. Rezzolla, Phys. Rev. D **77**, 084017 (2008)
66. T. Damour, A. Nagar, M. Hannam, S. Husa, B. Bruegmann, Phys. Rev. D **78**, 044039 (2008)
67. T. Damour, G. Schäfer, Nuovo Cimento B **101**, 127 (1988)
68. M. Davis, R. Ruffini, J. Tiomno, Phys. Rev. D **5**, 2932 (1972)
69. G. Faye, L. Blanchet, A. Buonanno, Phys. Rev. D **74**, 104033 (2006)
70. J.A. Gonzalez, M.D. Hannam, U. Sperhake, B. Bruegmann, S. Husa, Phys. Rev. Lett. **98**, 231101 (2007)
71. J.A. Gonzalez, U. Sperhake, B. Bruegmann, M. Hannam, S. Husa, Phys. Rev. Lett. **98**, 091101 (2007)
72. A. Gopakumar, M. Hannam, S. Husa, B. Bruegmann, Phys. Rev. D **78**, 064026 (2008)
73. M. Hannam, S. Husa, B. Bruegmann, A. Gopakumar, Phys. Rev. D **78**, 104007 (2008)
74. M. Hannam, S. Husa, U. Sperhake, B. Bruegmann, J.A. Gonzalez, Phys. Rev. D **77**, 044020 (2008)
75. T. Hinderer, Astrophys. J. **677**, 1216 (2008)
76. S. Husa, M. Hannam, J.A. Gonzalez, U. Sperhake, B. Bruegmann, Phys. Rev. D **77**, 044037 (2008)
77. Y. Itoh, T. Futamase, Phys. Rev. D **68**, 121501 (2003)
78. P. Jaranowski, G. Schäfer, Phys. Rev. D **57**, 7274 (1998); Erratum, *ibidem* **63**, 029902(E) (2001)
79. L.E. Kidder, Phys. Rev. D **77**, 044016 (2008)
80. C. Königsdorfer, G. Faye, G. Schäfer, Phys. Rev. D **68**, 044004 (2003)
81. S.M. Kopejkin, Astron. Zh. **62**, 889 (1985)
82. M. Koppitz, D. Pollney, C. Reisswig, L. Rezzolla, J. Thornburg, P. Diener, E. Schnetter, Phys. Rev. Lett. **99**, 041102 (2007)
83. A. Nagar, T. Damour, A. Tartaglia, Class. Q. Grav. **24**, S109 (2007)
84. S. Nissanke, L. Blanchet, Class. Q. Grav. **22**, 1007 (2005)
85. Y. Pan, A. Buonanno, J.G. Baker, J. Centrella, B.J. Kelly, S.T. McWilliams, F. Pretorius, J.R. van Meter, Phys. Rev. D **77**, 024014 (2008)
86. M.E. Pati, C.M. Will, Phys. Rev. D **62**, 124015 (2000)
87. M.E. Pati, C.M. Will, Phys. Rev. D **65**, 104008 (2002)
88. E. Poisson, Phys. Rev. D **52**, 5719 (1995); Erratum and Addendum, *ibidem* **55**, 7980 (1997)
89. D. Pollney, C. Reisswig, L. Rezzolla, B. Szilagyi, M. Ansorg, B. Deris, P. Diener, E.N. Dorband, M. Koppitz, A. Nagar, E. Schnetter, Phys. Rev. D **76**, 124002 (2007)
90. F. Pretorius, Phys. Rev. Lett. **95**, 121101 (2005)
91. F. Pretorius, Class. Q. Grav. **23**, S529 (2006)
92. F. Pretorius, in *Physics of Relativistic Objects in Compact Binaries: from Birth to Coalescence*, ed. by M. Colpi, P. Casella, V. Gorini, U. Moschella, A. Possenti, Astrophysics and Space Science Library **359** (Springer, Dordrecht and Canopus Publishing Lim., Bristol, 2009), p. 305
93. R.H. Price, J. Pullin, Phys. Rev. Lett. **72**, 3297 (1994)
94. L. Rezzolla, E. Barausse, E.N. Dorband, D. Pollney, C. Reisswig, J. Seiler, S. Husa, Phys. Rev. D **78**, 044002 (2008)
95. L. Rezzolla, P. Diener, E.N. Dorband, D. Pollney, C. Reisswig, E. Schnetter, J. Seiler, Astrophys. J. **674**, L29 (2008)
96. L. Rezzolla, E.N. Dorband, C. Reisswig, P. Diener, D. Pollney, E. Schnetter, B. Szilagyi, Astrophysics **J679**, 1422 (2008)
97. B.S. Sathyaprakash, B.F. Schutz, Living Rev. Rel. **12**, URL: <http://www.livingreviews.org/lrr-2009-2>
98. G. Schäfer, Ann. Phys. (N.Y.) **161**, 81 (1985)
99. M.A. Scheel, M. Boyle, T. Chu, L.E. Kidder, K.D. Matthews, H.P. Pfeiffer, Phys. Rev. D **79**, 024003 (2009)

100. G.A. Schott, *Phil. Mag.* **29**, 49 (1915)
101. U. Sperhake, V. Cardoso, F. Pretorius, E. Berti, J.A. Gonzalez, *Phys. Rev. Lett.* **101**, 161101 (2008)
102. H. Tagoshi, M. Sasaki, *Prog. Theor. Phys.* **92**, 745 (1994)
103. T. Tanaka, H. Tagoshi, M. Sasaki, *Prog. Theor. Phys.* **96**, 1087 (1996)
104. C.M. Will, *Prog. Theor. Phys. Suppl.* **136**, 158 (1999)
105. C.M. Will, A.G. Wiseman, *Phys. Rev. D* **54**, 4813 (1996)
106. A.G. Wiseman, *Phys. Rev. D* **48**, 4757 (1993)
107. N. Yunes, E. Berti, *Phys. Rev. D* **77**, 124006 (2008)

Introduction to Gravitational Self-Force

Robert M. Wald

Abstract The motion of a sufficiently small body in general relativity should be accurately described by a geodesic. However, there should be “gravitational self-force” corrections to geodesic motion, analogous to the “radiation reaction forces” that occur in electrodynamics. It is of considerable importance to be able to calculate these self-force corrections in order to be able to determine such effects as inspiral motion in the extreme mass ratio limit. However, severe difficulties arise if one attempts to consider point particles in the context of general relativity. This article describes these difficulties and how they have been dealt with.

1 Motion of Bodies in General Relativity

General relativity with suitable forms of matter has a well posed initial value formulation. In principle, therefore, to determine the motion of bodies in general relativity – such as binary neutron stars or black holes – one simply needs to provide appropriate initial data (satisfying the constraint equations) on a spacelike slice and then evolve this data via Einstein’s equation. It would be highly desirable to obtain simple analytic descriptions of motion. However, it is clear that, in general, the motion of a body of finite size will depend on the details of its composition as well as the details of its internal states of motion. Therefore, one can hope to get a simple description of motion only in some kind of “point particle limit.” Such a limit encompasses many cases of physical interest, such as “extreme mass ratio” inspiral. Of particular interest are the “radiation reaction” or “self-force” effects occurring during inspiral – the radiation reaction being, of course, the cause of the inspiral.

By definition, a “point particle” is an object whose stress–energy tensor is given by a delta-function with support on a worldline. A delta-function makes perfectly good mathematical sense as a distribution. Now, if a “source term” in an

R.M. Wald (✉)
Enrico Fermi Institute and Department of Physics, University of Chicago,
5640 S. Ellis Avenue, Chicago, IL 60637, USA
e-mail: rmwa@uchicago.edu

equation is distributional in nature, then the solution to this equation can, at best, be distributional in nature. Thus, if one wishes to consider distributional sources, one must generalize the notion of partial differential equations to apply to distributions. In the case of linear equations, this can be done straightforwardly: The notion of differentiation of distributions is well defined, so it makes perfectly good mathematical sense to consider distributional solutions to linear partial differential equations with distributional sources. Indeed, it is very useful to do so, and, for example, for Maxwell's equations even if the notion of a "point charge" did not arise from physical considerations, it would be very convenient for purely mathematical reasons to consider solutions with a delta-function charge-current source.

However, the situation is different in the case of nonlinear partial differential equations. Products of distributions normally can only be defined under special circumstances,¹ so it does not usually even make mathematical sense to say that a distribution satisfies a nonlinear equation. Thus, for example, although Maxwell's equations are linear, the coupled system of Maxwell's equations together with the equations of motion of the charged matter sources are nonlinear. Consequently, the complete, "self-consistent" Maxwell/motion equations are nonlinear. Hence, a priori, these equations are mathematically ill defined for point-charge sources. During the past century, there has been considerable discussion and debate as to how to make sense of these equations.

2 Point Particles in General Relativity

Einstein's equation is nonlinear, so a priori it does not make sense to consider this equation with a distributional source. Nevertheless, it has been understood for the past 40 years that it does make mathematical sense to consider Einstein's equation with a shell of matter [10], that is, an object whose stress-energy tensor is given by a delta-function with support on a timelike hypersurface (the "shell"). Solutions to Einstein's equation with a shell of matter correspond to patching together two smooth solutions along a timelike hypersurface in such a way that the metrics induced by the solutions on the two sides of the shell agree, but the extrinsic curvature is discontinuous. Such a solution corresponds to having a metric that is continuous, but whose first derivative has a jump discontinuity across the shell, and whose second derivative thereby has a delta-function character on the shell. Since the curvature tensor is linear in the second derivatives of the metric and there are no terms containing products of first and second derivatives of the metric, there is no difficulty in making sense of the curvature tensor of such a metric as a distribution.

¹ The product of two distributions can be defined if the decay properties of their Fourier transforms are such that the Fourier convolution integral defining their product converges. This will be the case when the wavefront sets of the distributions satisfy an appropriate condition (see [8]).

Unfortunately, however, the situation is much worse for Einstein’s equation with a point particle source.² An analysis by Geroch and Traschen [6] shows that it does not make mathematical sense to consider solutions of Einstein’s equation with a distributional stress–energy tensor supported on a worldline. Mathematically, the expected behavior of the metric near a “point particle” is too singular to make sense of the nonlinear terms in Einstein’s equation, even as distributions. Physically, if one tried to compress a body to make it into a point particle, it should collapse to a black hole before a “point particle limit” can be reached.

3 Point Particles in Linearized Gravity

Since point particles do not make sense in general relativity, it might appear that no simplifications to the description of motion can be achieved. However, the situation is not quite this bad because it does make mathematical sense to consider solutions, h_{ab} , to the linearized Einstein equation off of an arbitrary background solution, g_{ab} , with a distributional stress–energy tensor supported on a worldline. Therefore, one might begin a treatment of gravitational self-force by considering solutions to

$$G_{ab}^{(1)}[h](t, x^i) = 8\pi M u_a(t) u_b(t) \frac{\delta^{(3)}(x^i - z^i(t))}{\sqrt{-g}} \frac{d\tau}{dt}. \quad (1)$$

Here u^a is the unit tangent (i.e., 4-velocity) of the worldline γ defined by $x^i = z^i(t)$, τ is the proper time along γ , and $\delta^{(3)}(x^i - z^i(t))$ denotes the coordinate delta-function, that is, $\int \delta^{(3)}(x^i - z^i(t)) d^3x^i = 1$. (The right side of this equation could also be written in a manifestly covariant form as $8\pi M \int \delta_4(x, z(\tau)) u_a(\tau) u_b(\tau) d\tau$ where δ_4 denotes the covariant 4-dimensional delta-function.) However, two major difficulties arise in any approach that seeks to derive self-force effects starting with the linearized Einstein equation:

- The linearized Bianchi identity implies that the right side of Eq. 1 must be conserved in the background spacetime. For the case of a point-particle stress–energy tensor as occurs here, conservation requires that the worldline γ of the particle is a geodesic of the background spacetime. Therefore, there are no solutions for non-geodesic source curves, making it hopeless to use the linearized Einstein equation to derive corrections to geodesic motion.
- Even if the first problem were solved, solutions to this equation are singular on the worldline of the particle. Therefore, naive attempts to compute corrections to the motion due to h_{ab} – such as demanding that the particle move on a geodesic of $g_{ab} + h_{ab}$ – are virtually certain to encounter severe mathematical difficulties, analogous to the difficulties encountered in treatments of the electromagnetic self-force problem.

² “Strings” – that is, objects with a distributional stress–energy tensor corresponding to a delta-function with support on a timelike surface of co-dimension two – are a borderline case; see [5].

4 Lorenz Gauge Relaxation

The first difficulty has been circumvented by a number of researchers by modifying the linearized Einstein equation as follows: Choose the Lorenz gauge condition, so that the linearized Einstein equation takes the form

$$\nabla^c \nabla_c \tilde{h}_{ab} - 2R^c{}_{ab}{}^d \tilde{h}_{cd} = -16\pi M u_a(t) u_b(t) \frac{\delta^{(3)}(x^i - z^i(t))}{\sqrt{-g}} \frac{d\tau}{dt} \quad (2)$$

$$\nabla^b \tilde{h}_{ab} = 0 \quad (3)$$

where $\tilde{h}_{ab} \equiv h_{ab} - \frac{1}{2} h g_{ab}$ with $h = h_{ab} g^{ab}$. The first equation, by itself, has solutions for any source curve γ ; it is only when the Lorenz gauge condition is adjoined that the equations are equivalent to the linearized Einstein equation and geodesic motion is enforced. We will refer to the Lorenz gauge form of the linearized Einstein equation (2) with the Lorenz gauge condition (3) *not* imposed – as the *relaxed* linearized Einstein equation. If one solves the relaxed linearized Einstein equation while simply *ignoring* the Lorenz gauge condition that was used to derive this equation, one allows for the possibility of non-geodesic motion. Of course, the relaxed linearized Einstein equation is not equivalent to the original linearized Einstein equation. However, because deviations from geodesic motion are expected to be small, the Lorenz gauge violation should likewise be small, and it thus has been argued that solutions to the two systems should agree to sufficient accuracy.

5 Hadamard Expansions

In order to overcome the second difficulty, it is essential to understand the nature of the singular behavior of solutions to the relaxed linearized Einstein equation on the worldline of the particle. In order to do this, we would like to have a short distance expansion for the (retarded) Green's function for a general system of linear wave equations like (2). A formalism for doing this was developed by Hadamard in the 1920s. It is easiest to explain the basic idea of the Hadamard expansion in the Riemannian case rather than the Lorentzian case, that is, for Laplace equations rather than wave equations. For simplicity, we consider a single equation of the form

$$g^{ab} \nabla_a \nabla_b \phi + A^a \nabla_a \phi + B\phi = 0, \quad (4)$$

where g_{ab} is a Riemannian metric, A^a is a smooth vector field, and B is a smooth function. In the Riemannian case, Green's functions – that is, distributional solutions to Eq. 4 with a delta-function source on the right side – are unique up to the addition of a smooth solution,³ so all Green's functions have the same singular

³ This follows immediately from “elliptic regularity,” since the difference between two Green's functions satisfies the source free Laplace equation (4).

behavior.⁴ In 4-dimensions, in the case where g_{ab} is flat and both $A^a = 0$ and $B = 0$, a Green's function with source at x' is given explicitly by

$$G(x, x') = \frac{1}{\sigma(x, x')} \quad (5)$$

where $\sigma(x, x')$ denotes the squared geodesic distance between x and x' . This suggests that we seek a solution to the generalized Laplace equation (4) of the form

$$G(x, x') = \frac{U(x, x')}{\sigma(x, x')} + V(x, x') \ln \sigma(x, x') + W(x, x') \quad (6)$$

where V and W are, in turn, expanded as

$$V(x, x') = \sum_{j=0}^{\infty} v_j(x, x') \sigma^j, \quad W(x, x') = \sum_{j=0}^{\infty} w_j(x, x') \sigma^j. \quad (7)$$

One now proceeds by substituting these expansions into the generalized Laplace equation (4), using the identity $g^{ab} \nabla_a \sigma \nabla_b \sigma = 4\sigma$, and then formally setting the coefficient of each power of σ to 0 (see, e.g., [4]). The leading order equation yields a first order ordinary differential equation for U that holds along each geodesic through x' . This equation has a unique solution – the square root of the van Vleck–Morette determinant – that is regular at x' . In a similar manner, setting the coefficient of the higher powers of σ to 0, we get a sequence of “recursion relations” for the quantities v_j and w_j , which uniquely determine them – except for w_0 , which can be chosen arbitrarily. In the analytic case, one can then show that the resulting series for V and W have a finite radius of convergence and that the above expansion provides a Green's function. In the C^∞ but nonanalytic case, there is no reason to expect the series to converge, but truncated or otherwise suitably modified versions of the series can be used to construct a *parametrix* for Eq. 4, that is, a solution to Eq. 4 with a source that differs from a δ -function by at most a C^n function. Even in the analytic case, the Hadamard series defines a Green's function only in a sufficiently small neighborhood of x' . Clearly, this neighborhood must be contained within a normal neighborhood of x' in order that σ even be defined.

A similar construction works in the Lorentzian case, that is, for an equation of the form (4) with g_{ab} of Lorentz signature. The corresponding Hadamard expansion for the retarded Green's function is

$$G_+(x, x') = U(x, x') \delta(\sigma) \Theta(t - t') + V(x, x') \Theta(-\sigma) \Theta(t - t') \quad (8)$$

⁴This is not true in the Lorentzian case; the singular behavior of, for example, the retarded, advanced, and Feynman propagators are different from each other.

where V again is given by a series whose coefficients v_j are uniquely determined by recursion relations. The following points should be noted:

- In both the Riemannian and Lorentzian cases, $V(x, x')$ satisfies Eq. 4 in x . For a self-adjoint equation (as will be the case for Eq. 4 if $A^a = 0$), we also have $V(x, x') = V(x', x)$, so – where defined – V is a smooth solution of the homogeneous equation (4) in each variable.
- As already noted above, $v_0(x, x')$ (i.e., the first term in the series (7) for V) is uniquely determined by a recursion relation that can be solved by integrating an ordinary differential equation along geodesics through x' . In particular, in the Lorentzian case, $v_0(x, x')$ can thereby be obtained on the portion, \mathcal{N} , of the future lightcone of x' lying within a normal neighborhood of x' . Since $\sigma(x, x') = 0$ for $x \in \mathcal{N}$, we have $V(x, x') = v_0(x, x')$ on \mathcal{N} . However, since – as just noted – $V(x, x')$ satisfies the wave equation in x , it is uniquely determined in the domain of dependence of \mathcal{N} . Thus, in the Lorentzian case, one obtains the form (8) for the retarded Green's function in a sufficiently small neighborhood of x' in a way that bypasses any convergence issues for the Hadamard series (7).
- For a globally hyperbolic spacetime, the retarded Green's function $G_+(x, x')$ is globally well defined. By contrast, as already emphasized, the Hadamard form (8) of $G_+(x, x')$ can be valid at best within a normal neighborhood of x' . One occasionally sees in the literature Hadamard formulas that are purported to be valid when multiple geodesics connect x and x' , wherein a summation is made of contributions of the form (6) or (8) for each geodesic. I do not believe that there is any mathematical justification for these formulas.
- As follows from the “propagation of singularities” theorem [9], it is rigorously true that, globally, $G_+(x, x')$ is singular if and only if there is a future directed null geodesic from x' to x (whether or not this geodesic lies within a normal neighborhood of x').

5.1 Hadamard Expansions for a Point Particle Source

Using the generalization of the Hadamard form (8) to the retarded Green's function for the relaxed linearized Einstein equation (2), one finds (after a very lengthy calculation) that the solution to Eq. 2 in a sufficiently small neighborhood of the point particle source in Fermi normal coordinates is

$$h_{\alpha\beta} = \frac{2M}{r} \delta_{\alpha\beta} - 8M a_{(\alpha} u_{\beta)} (1 - a_i x^i) + h_{\alpha\beta}^{\text{tail}} + M \mathcal{R}_{\alpha\beta} + O(r^2). \quad (9)$$

Here r denotes the distance to the worldline, a^α is the acceleration of the worldline, $\mathcal{R}_{\alpha\beta}$ denotes a term of order r constructed from the curvature of the background spacetime, and

$$h_{\alpha\beta}^{\text{tail}} \equiv M \int_{-\infty}^{\tau^-} \left(G_{+\alpha\beta\alpha'\beta'} - \frac{1}{2} g_{\alpha\beta} G_{+\gamma\alpha'\beta'}^\gamma \right) u^{\alpha'} u^{\beta'} d\tau'. \quad (10)$$

The symbol τ^- means that this integration is to be cut short of $\tau' = \tau$ to avoid the singular behavior of the Green's function there; this instruction is equivalent to using only the “tail” (i.e., interior of the light cone) portion of the Green's function. In the region where the Hadamard form (8) holds (i.e., for x sufficiently close to x'), this corresponds to the contribution to the Green's function arising from $V(x, x')$.

6 Equations of Motion Including Self-Force

With the above formula (9) for $h_{\alpha\beta}$ as a starting point, the equations of motion of a point particle – accurate enough to take account of self-force corrections – have been obtained by the following three approaches:

- One can proceed in parallel with the derivations of Dirac [3] and DeWitt and Brehme [2] for the electromagnetic case and derive the motion from conservation of total stress–energy [11]. This requires an (ad hoc) regularization of the “effective stress energy” associated to $h_{\alpha\beta}$.
- One can derive equations of motion from some simple axioms [14], specifically that (i) the difference in “gravitational force” between different curves of the same acceleration (in possibly different spacetimes) is given by the (angle average of the) difference in $-\Gamma^\mu_{\alpha\beta} u^\alpha u^\beta$ where $\Gamma^\mu_{\alpha\beta}$ is the Christoffel symbol associated with $h_{\alpha\beta}$ and (ii) the gravitational self-force vanishes for a uniformly accelerating worldline in Minkowski spacetime. This provides a mathematically clean and simple way of obtaining equations of motion, but it is not a true “derivation” since the motion should follow from the assumptions of general relativity without having to make additional postulates.
- One can derive equations of motion via matched asymptotic expansions [11, 13]. The idea here is to postulate a suitable metric form (namely, Schwarzschild plus small perturbations) near the “particle,” and then “match” this “near zone” expression to the “far zone” formula (9) for $h_{\alpha\beta}$. Equations of motion then arise from the matching after imposition of a gauge condition. This approach is the closest of the three to a true derivation, but a number of ad hoc and/or not fully justified assumptions have been made, most notably Lorenz gauge relaxation.

6.1 The MiSaTaQuWa Equations

All three approaches have led to the following system of equations (in the case where there is no “incoming radiation,” i.e., h_{ab} vanishes in the asymptotic past):

$$\nabla^c \nabla_c \tilde{h}_{ab} - 2R^c{}_{ab}{}^d \tilde{h}_{cd} = -16\pi M u_a(t) u_b(t) \frac{\delta^{(3)}(x^i - z^i(t))}{\sqrt{-g}} \frac{d\tau}{dt} \quad (11)$$

$$u^b \nabla_b u^a = - \left(g^{ab} + u^a u^b \right) \left(\nabla_d h_{bc}^{\text{tail}} - \frac{1}{2} \nabla_b h_{cd}^{\text{tail}} \right) u^c u^d \quad (12)$$

where it is understood that the retarded solution to the equation for \tilde{h}_{ab} is to be chosen. Equations 11 and 12 are known as the MiSaTaQuWa equations. Note that the equation of motion (12) for the particle formally corresponds to the perturbed geodesic equation in the metric $g_{ab} + h_{ab}^{\text{tail}}$. However, it should be emphasized that h_{ab}^{tail} fails to be differentiable on the worldline of the particle and fails to be a (homogeneous) solution to the relaxed linearized Einstein equation. (This lack of differentiability affects only the spatial derivatives of the spatial components of h_{ab}^{tail} , so the right side of Eq. 12 is well defined.) Thus, one cannot interpret h_{ab}^{tail} as an effective, regularized, perturbed metric.

6.2 The Detweiler–Whiting Reformulation

An equivalent reformulation of Eq. 12 that does admit an interpretation as perturbed geodesic motion in an effective, regularized, perturbed metric has been given by Detweiler and Whiting [1], who proceed as follows. The symmetric Green’s function is defined by $G_{\text{sym}} = (G_+ + G_-)/2$ where G_- is the advanced Green’s function. For the case of a self-adjoint wave equation of the form (4) – that is, with $A^a = 0$ and g_{ab} Lorentzian – the Hadamard expansion of G_{sym} is given by

$$G_{\text{sym}}(x, x') = \frac{1}{2} [U(x, x') \delta(\sigma) + V(x, x') \Theta(-\sigma)]. \quad (13)$$

As previously noted $V(x, x')$ is symmetric (i.e., $V(x, x') = V(x', x)$) and is a homogeneous solution of Eq. 4 in each variable. However, $V(x, x')$ is, at best, defined only when x lies in a normal neighborhood of x' . In the region where V is defined, Detweiler and Whiting define a new Green’s function by

$$G_{\text{DW}}(x, x') = \frac{1}{2} [U(x, x') \delta(\sigma) - V(x, x') \Theta(\sigma)]. \quad (14)$$

The Detweiler–Whiting Green’s function has the very unusual property of having no support in the interior of the future or past light cones. Detweiler and Whiting show that Eq. 12 is equivalent to perturbed geodesic motion in the metric $g_{ab} + h_{ab}^R$ where h_{ab}^R is obtained by applying $G_+ - G_{\text{DW}}$ for the relaxed linearized Einstein equation to the worldline source. Since h_{ab}^R is a smooth, homogeneous solution to the relaxed linearized Einstein equation, it can be given an interpretation as an effective, regularized, perturbed metric. Of course, an observer making spacetime measurements near the particle would see the metric $g_{ab} + h_{ab}$, not $g_{ab} + h_{ab}^R$.

7 How Should Gravitational Self-Force Be Derived?

Although there is a general consensus that Eqs. 11 and 12 (or the Detweiler–Whiting version of Eq. 12) should provide a good description of the self-force corrections to the motion of a sufficiently small body, it is important that these equations be put on a firmer foundation both to clarify their range of validity and to potentially enable the systematic calculation of higher order corrections. It is clear that in order to obtain a precise and rigorous derivation of gravitational self-force, it will be necessary to take some kind of “point particle limit,” wherein the size, R , of the body goes to 0. However, to avoid difficulties associated with the nonexistence of point particles in general relativity, it is essential that one lets M go to 0 as well. If M goes to 0 more slowly than R , the body should collapse to a black hole before the limit $R \rightarrow 0$ is achieved. On the other hand, one could consider limits where M goes to 0 more rapidly than R , but finite size effects would then dominate over self-force effects as $R \rightarrow 0$. This suggests that we consider a one-parameter family of solutions to Einstein’s equation, $g_{ab}(\lambda)$, for which the body scales to 0 size and mass in an asymptotically self-similar way as $\lambda \rightarrow 0$, so that the ratio R/M approaches a constant in the limit.

Recently, Gralla and I [7] have considered such one-parameter families of bodies (or black holes). In the limit as $\lambda \rightarrow 0$ – where the body shrinks down to a worldline γ and “disappears” – we proved that γ must be a geodesic. Self-force and finite size effects then arise as perturbative corrections to γ . To first order in λ , these corrections are described by a deviation vector Z^i along γ . In [7], Gralla and I proved that, in the Lorenz gauge, this deviation vector satisfies

$$\frac{d^2 Z^i}{dt^2} = \frac{1}{2M} S^{kl} R_{kl0}{}^i - R_{0j0}{}^i Z^j - \left(h^{\text{tail}i}{}_{0,0} - \frac{1}{2} h_{00}{}^{,i} \right). \quad (15)$$

The first term in this equation corresponds to the usual “spin force” [12], that is, the leading order finite size correction to the motion. The second term is the usual right side of the geodesic deviation equation. (This term must appear since the corrections to motion must allow for the possibility of a perturbation to a nearby geodesic.) The last term corresponds to the self-force term appearing on right side of Eq. 12. It should be emphasized that Eq. 15 arises as the perturbative correction to geodesic motion for any one-parameter family satisfying our assumptions, and holds for black holes as well as ordinary bodies.

Although the self-force term in Eq. 15 corresponds to the right side of Eq. 12, these equations have different meanings. Equation 15 is a first order perturbative correction to geodesic motion, and no Lorenz gauge relaxation is involved in this equation since h_{ab} is sourced by a geodesic γ . By contrast, Eq. 12 is supposed to hold even when the cumulative deviations from geodesic motion are large, and Lorenz gauge relaxation is thereby essential. Given that Eq. 15 holds rigorously as a perturbative result, what is the status of the MiSaTaQuWa equations (11), (12)? In [7], we argued that the MiSaTaQuWa equations arise as “self-consistent

perturbative equations” associated with the perturbative result (15). If the deviations from geodesic motion are *locally* small – even though cumulative effects may yield large deviations from any individual geodesic over long periods of time – then the MiSaTaQuWa equations should provide an accurate description of motion.

Acknowledgements This research was supported in part by NSF grant PHY04-56619 to the University of Chicago.

References

1. S. Detweiler, B.F. Whiting, *Phys. Rev. D* **67**, 024025 (2003)
2. B.S. DeWitt, R.W. Brehme, *Ann. Phys. (N.Y.)* **9**, 220 (1960)
3. P.A.M. Dirac, *Proc. R. Soc. Lond. A* **167**, 148 (1938)
4. P.R. Garabedian, *Partial Differential Equations* (Wiley, New York, 1964)
5. D. Garfinkle, *Class. Q. Grav.* **16**, 4101 (1999)
6. R. Geroch, J. Traschen, *Phys. Rev. D* **36**, 1017 (1987)
7. S.E. Gralla, R.M. Wald, *Class. Q. Grav.* **25**, 205009 (2008)
8. L. Hormander, *The Analysis of Linear Partial Differential Operators, I* (Springer, Berlin, 1985)
9. L. Hormander, *The Analysis of Linear Partial Differential Operators, IV* (Springer, Berlin, 1985)
10. W. Israel, *Nuovo Cimento B* **44**, 1 (1966); *Lettere alla Redazione, ibidem* **48**, 463 (1967)
11. Y. Mino, M. Sasaki, T. Tanaka, *Phys. Rev. D* **55**, 3457 (1997)
12. A. Papapetrou, *Proc. R. Soc. Lond. A* **209**, 248 (1951)
13. E. Poisson, *Living Rev. Rel.* **7**, URL (cited on 23 December 2008): <http://www.livingreviews.org/lrr-2004-6>
14. T.C. Quinn, R.M. Wald, *Phys. Rev. D* **56**, 3381 (1997)

Derivation of Gravitational Self-Force

Samuel E. Gralla and Robert M. Wald

Abstract We analyze the issue of “particle motion” in general relativity in a systematic and rigorous way by considering a one-parameter family of metrics corresponding to having a body (or black hole) that is “scaled down” to 0 size and mass in an appropriate manner. We prove that the limiting worldline of such a one-parameter family must be a geodesic of the background metric and obtain the leading order perturbative corrections, which include gravitational self-force, spin force, and geodesic deviation effects. The status of the MiSaTaQuWa equation is explained as a candidate “self-consistent perturbative equation” associated with our rigorous perturbative result.

1 Difficulties with Usual Derivations

It is of considerable interest to determine the motion of a body in general relativity in the limit of small size, taking into account the deviations from geodesic motion arising from gravitational self-force effects. There is a general consensus that the gravitational self-force is given by the “MiSaTaQuWa equations”: In the absence of incoming radiation, the motion of a particle of mass M is given by

$$u^\nu \nabla_\nu u^\mu = -\frac{1}{2}(g^{\mu\nu} + u^\mu u^\nu)(2\nabla_\sigma h_{\nu\rho}^{\text{tail}} - \nabla_\nu h_{\rho\sigma}^{\text{tail}})|_{z(\tau)} u^\rho u^\sigma, \quad (1)$$

where $g_{\mu\nu}$ is the background metric, u^μ is the four-velocity of the worldline z^μ , and

$$h_{\mu\nu}^{\text{tail}}(x) = M \int_{-\infty}^{\tau^-} \left(G_{\mu\nu\mu'\nu'}^+ - \frac{1}{2} g_{\mu\nu} G_{\rho\mu'\nu'}^+ \right) (x, z(\tau')) u^{\mu'} u^{\nu'} d\tau', \quad (2)$$

S.E. Gralla (✉) and R.M. Wald (✉)

Enrico Fermi Institute and Department of Physics, University of Chicago,
5640 S. Ellis Avenue, Chicago, IL 60637, USA

e-mail: sgralla@uchicago.edu; rmwa@uchicago.edu

in which G^+ is the retarded Green's function for the wave operator $\nabla^\alpha \nabla_\alpha \tilde{h}_{\mu\nu} - 2R^\alpha{}_{\mu\nu} \tilde{h}^{\alpha\beta}$. (Note that the τ^- limit of integration indicates that only the part of G^+ interior to the light cone contributes to $h_{\mu\nu}^{\text{tail}}$.) However, all derivations contain some unsatisfactory features. This is not surprising in view of the fact that, as noted in [6], “point particles” do not make sense in nonlinear theories like general relativity!

- Derivations that treat the body as a point particle require unjustified “regularizations.”
- Derivations using matched asymptotic expansions [3, 4] make a number of ad hoc and/or unjustified assumptions.
- The axioms of the Quinn–Wald axiomatic approach [5] have not been shown to follow from Einstein's equation.
- All of the above derivations employ at some stage a “phony” version of the linearized Einstein equation with a point-particle source, wherein the Lorenz gauge version of the linearized Einstein equation is written down, but the Lorenz gauge condition is not imposed.

2 Rigorous Derivation Requirements

How should gravitational self-force be rigorously derived? A precise formula for gravitational self-force can hold only in a limit where the size, R , of the body goes to 0. Since “point-particles” do not make sense in general relativity – collapse to a black hole would occur before a point-particle limit could be taken – the mass, M , of the body must also go to 0 as $R \rightarrow 0$. In the limit as $R, M \rightarrow 0$, the worldtube of the body should approach a curve, γ , which should be a geodesic of the “background metric.” The self-force should arise as the lowest order in M correction to γ . In the following, we shall describe an approach that we have recently taken to derive gravitational self-force in this manner. Details can be found in [1].

The discussion above suggests that we consider a one-parameter family of solutions to Einstein's equation, $(g_{\mu\nu}(\lambda), T_{\mu\nu}(\lambda))$, with $R(\lambda) \rightarrow 0$ and $M(\lambda) \rightarrow 0$ as $\lambda \rightarrow 0$. But, what conditions should be imposed on $(g_{\mu\nu}(\lambda), T_{\mu\nu}(\lambda))$ to ensure that it corresponds to a body that is shrinking down to 0 size, without undergoing wild oscillations, drastically changing its shape, or doing other crazy things as it does so?

3 Limits of Spacetimes

As a very simple, explicit example of the kind of one-parameter family we seek, consider the Schwarzschild–de Sitter metrics with $M = \lambda$,

$$ds^2(\lambda) = - \left(1 - \frac{2\lambda}{r} - Cr^2\right) dt^2 + \left(1 - \frac{2\lambda}{r} - Cr^2\right)^{-1} dr^2 + r^2 d\Omega^2. \quad (3)$$

If we take the limit as $\lambda \rightarrow 0$ at fixed coordinates (t, r, θ, ϕ) with $r > 0$, it is easily seen that we obtain the de Sitter metric – with the de Sitter spacetime worldline γ defined by $r = 0$ corresponding to the location of the black hole “before it disappeared.” However, there is also another limit that can be taken. At each time t_0 , introduce rescaled coordinates $\bar{r} = r/\lambda$, $\bar{t} = (t - t_0)/\lambda$. We then correspondingly “blow up” the metric $g_{\mu\nu}(\lambda)$ by multiplying it by λ^{-2} , that is, we define

$$\bar{g}_{\mu\nu}(\lambda) \equiv \lambda^{-2} g_{\mu\nu}(\lambda). \tag{4}$$

We then have

$$d\bar{s}^2(\lambda) = -(1 - 2/\bar{r} - \lambda^2 C \bar{r}^2) d\bar{t}^2 + (1 - 2/\bar{r} - \lambda^2 C \bar{r}^2)^{-1} d\bar{r}^2 + \bar{r}^2 d\Omega^2. \tag{5}$$

In the limit as $\lambda \rightarrow 0$ (at fixed $(\bar{t}, \bar{r}, \theta, \phi)$) the “de Sitter background” becomes irrelevant. The limiting metric is simply the Schwarzschild metric of unit mass. The fact that the limit as $\lambda \rightarrow 0$ exists can be attributed to the fact that the Schwarzschild black hole is shrinking to zero in a manner where, in essence, nothing changes except the overall scale.

4 Our Basic Assumptions

The simultaneous existence of both of the above types of limits characterizes the type of one-parameter family of spacetimes $g_{\mu\nu}(\lambda)$ that we wish to consider. More precisely, we wish to consider a one parameter family of solutions $g_{\mu\nu}(\lambda)$ satisfying the following properties:

- (i) Existence of the “ordinary limit”: There exist coordinates x^α such that the metric $g_{\mu\nu}(\lambda, x^\alpha)$ is jointly smooth in (λ, x^α) , at least for $r > \bar{R}\lambda$ for some constant \bar{R} , where $r \equiv \sqrt{\sum (x^i)^2}$ ($i = 1, 2, 3$). For all λ and for $r > \bar{R}\lambda$, $g_{\mu\nu}(\lambda)$ is a vacuum solution of Einstein’s equation. Furthermore, $g_{\mu\nu}(\lambda = 0, x^\alpha)$ is smooth in x^α , including at $r = 0$, and, for $\lambda = 0$, the curve γ defined by $r = 0$ is timelike.
- (ii) Existence of the “scaled limit”: For each t_0 , we define $\bar{t} \equiv (t - t_0)/\lambda$, $\bar{x}^i \equiv x^i/\lambda$. Then the metric $\bar{g}_{\bar{\mu}\bar{\nu}}(\lambda; t_0; \bar{x}^\alpha) \equiv \lambda^{-2} g_{\bar{\mu}\bar{\nu}}(\lambda; t_0; \bar{x}^\alpha)$ is jointly smooth in $(\lambda, t_0; \bar{x}^\alpha)$ for $\bar{r} \equiv r/\lambda > \bar{R}$.

4.1 Additional Uniformity Requirement

The above two conditions must be supplemented by an additional “uniformity requirement,” which can be explained as follows. From the definitions of $\bar{g}_{\bar{\mu}\bar{\nu}}$ and \bar{x}^μ , we can relate coordinate components of the barred metric in barred

coordinates to coordinate components of the unbarred metric in corresponding unbarred coordinates,

$$\bar{g}_{\bar{\mu}\bar{\nu}}(\lambda; t_0; \bar{t}, \bar{x}^i) = g_{\mu\nu}(\lambda; t_0 + \lambda\bar{t}, \lambda\bar{x}^i). \quad (6)$$

Now introduce new variables $\alpha \equiv r$ and $\beta \equiv \lambda/r = 1/\bar{r}$, and view the metric components $g_{\mu\nu}(\lambda)$ as functions of $(\alpha, \beta, t, \theta, \phi)$, where θ and ϕ are defined in terms of x^i by the usual formula for spherical polar angles. We have

$$\bar{g}_{\bar{\mu}\bar{\nu}}(\alpha\beta, t_0; \bar{t}, 1/\beta, \theta, \phi) = g_{\mu\nu}(\alpha\beta, t = t_0 + \lambda\bar{t}; \alpha, \theta, \phi). \quad (7)$$

Then, by assumption (ii) we see that for $0 < \beta < 1/\bar{R}$, $g_{\mu\nu}$ is smooth in (α, β) for all α including $\alpha = 0$. By assumption (i), we see that for all $\alpha > 0$, $g_{\mu\nu}$ is smooth in (α, β) for $\beta < 1/\bar{R}$, including $\beta = 0$. Furthermore, for $\beta = 0$, $g_{\mu\nu}$ is smooth in α , including $\alpha = 0$.

We now impose the additional uniformity requirement on our one-parameter family of spacetimes:

- (iii) $g_{\mu\nu}$ is jointly smooth in (α, β) at $(0, 0)$.

We already know from our previous assumptions that $g_{\mu\nu}(\lambda; t_0, r, \theta, \phi)$ and its derivatives with respect to x^α approach a limit if we let $\lambda \rightarrow 0$ at fixed r and then let $r \rightarrow 0$. The uniformity requirement implies that the same limits are attained whenever λ and r both go to 0 in any way such that λ/r goes to 0.

It has recently been proven in [2] that an analog of the uniformity requirement holds for electromagnetism in Minkowski spacetime in the following sense: Consider a one-parameter family of charge-current sources of the form $J^\mu(\lambda, t, x^i) = \tilde{J}^\mu(\lambda, t, x^i/\lambda)$ where \tilde{J}^μ is a smooth function of its arguments and $x^i = 0$ defines a timelike worldline. Then the retarded solution, $F_{\mu\nu}(\lambda, x^\mu)$, is a smooth function of the variables $(\alpha, \beta, t, \theta, \phi)$ in a neighborhood of $(\alpha, \beta) = (0, 0)$. In the gravitational case, we do not have a simple relationship between the metric and the stress–energy source, and in the nonlinear regime, it would not make sense to formulate the uniformity condition in terms of the behavior of the stress–energy. Consequently, we have formulated this condition in terms of the behavior of the metric itself.

5 Geodesic Motion

The uniformity requirement implies that the metric components can be approximated near $(\alpha, \beta) = (0, 0)$ with a finite Taylor series in α and β ,

$$g_{\mu\nu}(\lambda; t, r, \theta, \phi) = \sum_{n=0}^N \sum_{m=0}^M r^n \left(\frac{\lambda}{r}\right)^m (a_{\mu\nu})_{nm}(t, \theta, \phi), \quad (8)$$

where remainder terms have been dropped. This gives a *far zone expansion*. Equivalently, we have

$$\bar{g}_{\bar{\mu}\bar{\nu}}(\lambda; t_0; \bar{t}, \bar{r}, \theta, \phi) = \sum_{n=0}^N \sum_{m=0}^M (\lambda \bar{r})^n \left(\frac{1}{\bar{r}}\right)^m (a_{\mu\nu})_{nm}(t_0 + \lambda \bar{t}, \theta, \phi). \quad (9)$$

Further Taylor expanding this formula with respect to the time variable yields a *near zone expansion*. Note that since we can express $\bar{g}_{\bar{\mu}\bar{\nu}}$ at $\lambda = 0$ as a series in $1/\bar{r}$ as $\bar{r} \rightarrow \infty$ and since $\bar{g}_{\bar{\mu}\bar{\nu}}$ at $\lambda = 0$ does not depend on \bar{t} , we see that $\bar{g}_{\bar{\mu}\bar{\nu}}(\lambda = 0)$ is a stationary, asymptotically flat spacetime.

The curve γ to which our body shrinks as $\lambda \rightarrow 0$ (see condition (i) above) can now be proven to be a geodesic of the metric $g_{\mu\nu}(\lambda = 0)$ as follows: Choose the coordinates x^α so that at $\lambda = 0$ they correspond to Fermi normal coordinates about the worldline γ . In particular, we have $g_{\mu\nu} = \eta_{\mu\nu}$ on γ at $\lambda = 0$. It follows from (8) that near γ (i.e., for small r) the metric $g_{\mu\nu}$ must take the form

$$g_{\mu\nu} = \eta_{\mu\nu} + O(r) + \lambda \left(\frac{C_{\mu\nu}(t, \theta, \phi)}{r} + O(1) \right) + O(\lambda^2). \quad (10)$$

Now, for $r > 0$, the coefficient of λ , namely,

$$h_{\mu\nu} = \frac{C_{\mu\nu}}{r} + O(1), \quad (11)$$

must satisfy the vacuum linearized Einstein equation off the background spacetime $g_{\mu\nu}(\lambda = 0)$. However, since each component of $h_{\mu\nu}$ is a locally L^1 function, it follows immediately that $h_{\mu\nu}$ is well defined as a distribution. It is not difficult to show that, as a distribution, $h_{\mu\nu}$ satisfies the linearized Einstein equation with source of the form $N_{\mu\nu}(t)\delta^{(3)}(x^i)$, where $N_{\mu\nu}$ is given by a formula involving the limit as $r \rightarrow 0$ of the angular average of $C_{\mu\nu}$ and its first derivative. The linearized Bianchi identity then immediately implies that $N_{\mu\nu}$ is of the form $Mu_\mu u_\nu$ with M constant, and that γ is a geodesic for $M \neq 0$.

6 Corrections to Motion

Our main interest, however, is not to rederive geodesic motion but to find the leading order corrections to geodesic motion that arise from finite mass and finite size effects. To define these corrections, we need to have a notion of the “location” of the body to first order in λ . This can be defined as follows: Since $\bar{g}_{\bar{\mu}\bar{\nu}}(\lambda = 0)$ is an asymptotically flat spacetime, its mass dipole moment can be set to 0 (at all t_0) as a gauge condition on the coordinates \bar{x}^i . These coordinates then have the interpretation of being “center of mass coordinates” for the spacetime $\bar{g}_{\bar{\mu}\bar{\nu}}(\lambda = 0)$. In terms of our original coordinates x^α , the transformation to center of mass coordinates at all t_0 corresponds to a coordinate transformation $x^\alpha \rightarrow \hat{x}^\alpha$ of the form

$$\hat{x}^\alpha(t) = x^\alpha - \lambda A^\alpha(t, x^i) + O(\lambda^2). \quad (12)$$

To first order in λ , the worldline defined by $\hat{x}^i = 0$ should correspond to the “position” of the body. The first-order displacement from γ in the original coordinates is then given simply by

$$Z^i(t) \equiv A^i(t, x^j = 0). \quad (13)$$

The quantity Z^i is most naturally interpreted as a “deviation vector field” defined on γ . Our goal is to derive relations (if any) that hold for Z^i that are independent of the choice of one-parameter family satisfying our assumptions.

6.1 Calculation of the Perturbed Motion

We now choose the x^α coordinates – previously chosen to agree with Fermi normal coordinates on γ at $\lambda = 0$ – to correspond to the Lorenz/harmonic gauge to first order in λ . To order λ^2 , the leading order in r terms in $g_{\alpha\beta}$ are

$$\begin{aligned} g_{\alpha\beta}(\lambda; t, x^i) &= \eta_{\alpha\beta} + B_{\alpha i \beta j}(t) x^i x^j + O(r^3) \\ &+ \lambda \left(\frac{2M}{r} \delta_{\alpha\beta} + h_{\alpha\beta}^{\text{tail}}(t, 0) + h_{\alpha\beta i}^{\text{tail}}(t, 0) x^i + M \mathcal{R}_{\alpha\beta}(t) + O(r^2) \right) \\ &+ \lambda^2 \left(\frac{M^2}{r^2} (-2t_\alpha t_\beta + 3n_\alpha n_\beta) + \frac{2}{r^2} P_i(t) n^i \delta_{\alpha\beta} + \frac{1}{r^2} t_{(\alpha} S_{\beta)j}(t) n^j \right. \\ &\quad \left. + \frac{1}{r} K_{\alpha\beta}(t, \theta, \phi) + H_{\alpha\beta}(t, \theta, \phi) + O(r) \right) + O(\lambda^3). \quad (14) \end{aligned}$$

Here B and \mathcal{R} are expressions involving the curvature of $g_{\mu\nu}(\lambda = 0)$ and we have introduced the “unknown” tensors K and H . The quantities P^i and $S_{\alpha\beta}$ turn out to be the mass dipole and spin of the “near-zone” background spacetime $\bar{g}_{\bar{\mu}\bar{\nu}}(\lambda = 0)$. For simplicity, we have assumed no “incoming radiation.” Hadamard expansion techniques and the second-order perturbation theory were used to derive this expression.

Using the coordinate shift $x^\mu \rightarrow x^\mu - \lambda A^\mu$ to cancel the mass dipole term, the above expression translates into the following expression for the scaled metric

$$\begin{aligned} \bar{g}_{\bar{\alpha}\bar{\beta}}(\hat{t}_0) &= \eta_{\alpha\beta} + \frac{2M}{\bar{r}} \delta_{\alpha\beta} + \frac{M^2}{\bar{r}^2} (-2t_\alpha t_\beta + 3n_\alpha n_\beta) + \frac{1}{\bar{r}^2} t_{(\alpha} S_{\beta)j} n^j + O\left(\frac{1}{\bar{r}^3}\right) \\ &+ \lambda \left[h_{\alpha\beta}^{\text{tail}} + 2A_{(\alpha\beta)} + \frac{1}{r} K_{\alpha\beta} + \frac{\bar{t}}{\bar{r}^2} t_{(\alpha} \dot{S}_{\beta)j} n^j + O\left(\frac{1}{\bar{r}^2}\right) + \bar{t} O\left(\frac{1}{\bar{r}^3}\right) \right] \\ &+ \lambda^2 \left[B_{\alpha i \beta j} \bar{x}^i \bar{x}^j + h_{\alpha\beta, \gamma}^{\text{tail}} \bar{x}^\gamma + M \mathcal{R}_{\alpha\beta}(\bar{x}^i) + 2B_{\alpha i \beta j} A^i \bar{x}^j + 2A_{(\alpha, \beta)\gamma} \bar{x}^\gamma + H_{\alpha\beta} \right. \\ &\quad \left. + \frac{\bar{t}}{\bar{r}} \dot{K}_{\alpha\beta} + \frac{\bar{t}^2}{\bar{r}^2} t_{(\alpha} \ddot{S}_{\beta)j} n^j + O\left(\frac{1}{\bar{r}}\right) + \bar{t} O\left(\frac{1}{\bar{r}^2}\right) + \bar{t}^2 O\left(\frac{1}{\bar{r}^3}\right) \right] + O(\lambda^3), \quad (15) \end{aligned}$$

where quantities on the right side are evaluated at $t = t_0$, and the overdots denote derivatives with respect to t .

The terms that are first order in λ in this equation satisfy the linearized vacuum Einstein equation about the background “near zone” metric (i.e., the terms that are 0th order in λ). From this equation, we find that $dS_{ij}/dt = 0$, that is, to lowest order, spin is parallelly propagated along γ .

The terms that are second order in λ in this equation satisfy the linearized Einstein equation about the background “near zone” metric with source given by the second-order Einstein tensor of the first order terms. Extracting the $\ell = 1$, electric parity, even-under-time-reversal part of this equation at $O(1/\bar{r}^2)$ and $O(\bar{t}/\bar{r}^3)$, we obtain (after considerable algebra!)

$$Z_{i,00} = \frac{1}{2M} S^{kl} R_{k l 0 i} - R_{0 j 0 i} Z^j - \left(h_{i 0,0}^{\text{tail}} - \frac{1}{2} h_{00,i}^{\text{tail}} \right). \quad (16)$$

In other words, in the Lorenz gauge, the deviation vector field, Z^a , on γ that describes the first order perturbation to the motion satisfies

$$u^c \nabla_c (u^b \nabla_b Z^a) = \frac{1}{2M} R_{bcd}{}^a S^{bc} u^d - R_{bcd}{}^a u^b u^d Z^c - (g^{ab} + u^a u^b) \left(\nabla_d h_{bc}^{\text{tail}} - \frac{1}{2} \nabla_b h_{cd}^{\text{tail}} \right) u^c u^d. \quad (17)$$

7 Interpretation of Results

Equation 17 gives the desired leading order corrections to motion along the geodesic γ . The first term on the right side of this equation is the Papapetrou “spin force,” which is the leading order “finite size” correction. The second term is just the usual right hand side of the geodesic deviation equation; it is not a correction to geodesic motion but rather allows for the possibility that the perturbation may make the body move along a different geodesic. Finally, the last term describes the gravitational self-force that we had sought to obtain, that is, the corrections to the motion caused by the body’s self-field. Equation 17 gives the correct description of motion when the metric perturbation is in the Lorenz gauge. When the metric perturbation is expressed in a different gauge, the force may be different [1].

8 Self-Consistent Equations

Although we have now obtained the perturbative correction to geodesic motion due to spin and self-force effects, at late times the small corrections due to self-force effects should accumulate (e.g., during an inspiral), and eventually the orbit should deviate significantly from the original, unperturbed geodesic γ . When this happens,

it is clear that our perturbative description in terms of a deviation vector defined on γ will not be accurate. Clearly, going to any (finite) higher order in perturbation theory will not help (much). However, if the mass and size of the body are sufficiently small, we expect that its motion is well described *locally* as a small perturbation of *some* geodesic. *Therefore, one should obtain a good description of the motion by making up (!) a “self-consistent perturbative equation”* that satisfies the following criteria: (1) It has a well-posed initial value formulation. (2) It has the same “number of degrees of freedom” as the original system. (3) Its solutions correspond closely to the solutions of the original perturbation equation over a time interval where the perturbation remains small. In some sense, such a self-consistent perturbative equation would take into account the important (“secular”) higher order perturbative effects (to all orders), but ignore other higher order corrections. Such equations are commonly considered in physics. The MiSaTaQuWa equations appear to be a good candidate for a self-consistent perturbative equation associated with our perturbative result.

9 Summary

In summary, we have analyzed the motion of a small body or black hole in general relativity, assuming only the existence of a one-parameter family of solutions satisfying assumptions (i), (ii), and (iii) above. We showed that at lowest (“0th”) order, the motion of a “small” body is described by a geodesic, γ , of the “background” spacetime. We then derived a formula for the first order deviation of the “center of mass” worldline of the body from γ . The MiSaTaQuWa equations then arise as (candidate) “self-consistent perturbative equations” based on our first order perturbative result. Note that it is only at this stage that “phoney” linearized Einstein equations come into play.

We have recently applied this basic approach to the derivation of self-force in electromagnetism [2], and have argued that the reduced order form of the Abraham–Lorentz–Dirac equation provides an appropriate self-consistent perturbative equation associated with our first order perturbative result (whereas the original Abraham–Lorentz–Dirac equation is excluded). It should be possible to use this formalism to take higher order corrections to the motion into account in a systematic way in both the gravitational and electromagnetic cases.

References

1. S. Gralla, R. Wald, *Class. Q. Grav.* **25**, 205009 (2008)
2. S. Gralla, A. Harte, R. Wald, *Phys. Rev. D* **80**, 024031 (2009)
3. Y. Mino, M. Sasaki, T. Tanaka, *Phys. Rev. D* **55**, 3457 (1997)
4. E. Poisson, *Living Rev. Rel.* **7**, URL: <http://www.livingreviews.org/lrr-2004-6>
5. T.C. Quinn, R.M. Wald, *Phys. Rev. D* **56**, 3381 (1997)
6. R.M. Wald, *Introduction to Gravitational Self-Force* (contribution to this volume)

Elementary Development of the Gravitational Self-Force

Steven Detweiler

Abstract The gravitational field of a particle of small mass m moving through curved spacetime, with metric g_{ab} , is naturally and easily decomposed into two parts each of which satisfies the perturbed Einstein equations through $O(m)$. One part is an inhomogeneous field h_{ab}^S , which, near the particle, looks like the Coulomb m/r field with tidal distortion from the local Riemann tensor. This singular field is defined in a neighborhood of the small particle and does not depend upon boundary conditions or upon the behavior of the source in either the past or the future. The other part is a homogeneous field h_{ab}^R . In a perturbative analysis, the motion of the particle is then best described as being a geodesic in the metric $g_{ab} + h_{ab}^R$. This geodesic motion includes all of the effects that might be called radiation reaction and conservative effects as well.

1 Introduction

Newton's apple hangs in a tree. The force of gravity is balanced by the force from a branch, and the apple is at rest. Later, the apple falls and accelerates downward until it hits the ground.

Einstein's insight elevates the lowly force of gravity to exalted status as a servant of geometry. Einstein's apple, being sentient and hanging in a tree, explains its own non-geodesic, non free-fall, accelerated motion as being caused by the force it feels from the branch. When the apple is released by the branch, its subsequent free fall motion is geodesic and not accelerated. The apple is freed from all forces and does not accelerate until it hits the ground.

These two perspectives have differing explanations and differing descriptions of the motion, but the actual paths through the events of spacetime are the same. Newton's understanding that the gravitational mass is identical to the inertial mass

S. Detweiler (✉)
Institute of Fundamental Theory, Department of Physics, University of Florida,
Gainesville, FL 32611, USA
e-mail: det@ufl.edu

implies that an object of negligible mass in free-fall moves along a trajectory which is independent of the object's mass. Einstein's Equivalence Principle requires that such an object in free-fall moves along a geodesic of spacetime, a trajectory which is independent of the object's mass. Newton's free-fall motion and Einstein's geodesic motion describe such an object as moving along one and the same sequence of events in spacetime.

Thorne and Hartle [52] give a clear and careful description of the motion of a small nearly-Newtonian object through the geometry of spacetime. They conclude that such motion might have a small acceleration, consistent with Newtonian analysis,¹ from the coupling of the internal mass multipole moments of the object with the multipole moments of the external spacetime geometry, which are related to the components of the Riemann tensor in the vicinity of the object [cf. Eqs. 41–44]. If the object orbits a large black hole, then the analysis implies that the motion is geodesic as long as any asphericity of the object, perhaps caused by rotation or tidal distortion, can be ignored. An acceleration larger than allowed by the coupling of the multipole moments is inexplicable in the context of either General Relativity or of Newtonian gravity and must necessarily result from some non-gravitational force.

How does the Thorne-Hartle description meld with the notion that Einstein's apple orbits a black hole, emits gravitational waves, radiates away energy and angular momentum, and cannot then move along a geodesic of the black hole geometry? Radiation reaction is not a consequence of any asphericity of the apple. Does the apple move along a geodesic? Would the apple, being sentient, describe its own motion as free-fall?

For the moment consider the familiar electromagnetic radiation-reaction force on an accelerating charge q as given below in Eq. 14. A notable feature is that the force is proportional to q^2 . Consequently this force is often described as resulting from the charge q interacting with its own electromagnetic field, and the force is called the electromagnetic *self-force*.

Similar language is used with gravitation, but in that case the force is proportional to m^2 and the resulting acceleration is proportional to m . In general terms, the gravitational self-force is said to be responsible for any aspect of motion that is proportional to the mass m of the object at hand. Yet, with either Newtonian gravity or General Relativity, the motion of an object of small mass m is independent of m . *Gravitational self-force* appears to be an oxymoron.

But, even Newtonian gravity contains a gravitational self-force. One might describe the motion of the Moon about the Earth as free-fall in the Earth's gravitational field and conclude that

$$ma = m \left(\frac{2\pi}{T} \right)^2 r = \frac{GMm}{r^2} \quad (1)$$

¹ If the acceleration of gravity \mathbf{g} differs significantly across a large object, then the center of mass moves responding to some average, over the object, of \mathbf{g} which does not necessarily match a free-fall trajectory.

where r is the radius of the Moon's orbit, so that the orbital period is

$$T = \sqrt{4\pi^2 r^3 / GM} \quad (2)$$

A more accurate description of the motion includes the influence of the Moon back on the Earth. Then the Moon is in free-fall in the Earth's gravitational field while the Earth orbits their common center of mass. And the conclusion becomes

$$\begin{aligned} m \left(\frac{2\pi}{T} \right)^2 r &= \frac{GMm}{r^2(1 + m/M)^2} \\ T &= \sqrt{4\pi^2 r^3 / GM} [1 + m/M]. \end{aligned} \quad (3)$$

The mass of the Moon has an influence on its own motion in Eq. 3, and this influence could be (although it rarely is) described as a consequence of the Newtonian gravitational self-force. Nevertheless, Newton's law of gravity still implies that the Moon does not exert a net gravitational force on itself. The acceleration of the Moon is still properly lined up with the gradient of the Earth's gravitational potential, and the Moon's motion is described as free-fall or geodesic, depending upon whether one is Newton or Einstein.

To me it seems inappropriate to describe the presence of the m/M term in Eq. 3 as resulting from the interaction of the Moon with its own gravitational field. Rather, the m/M term arises because the Earth orbits the common center of mass of the Earth–Moon system.

The conundrum of radiation reaction as being consistent with geodesic motion can now be resolved. Einstein's apple orbiting a black hole must move along a geodesic, but the geometry through which it moves is the black hole metric disturbed by the presence of the apple. Nevertheless, this disturbed metric is a vacuum solution of the Einstein equations in the neighborhood of the apple. If the motion were not geodesic, then the apple could not explain its own motion as being free-fall in a vacuum gravitational field. Such motion would violate Newton's laws as well as Einstein's Equivalence Principle.

Throughout this manuscript we focus on the self-force acting on small objects which are otherwise in unconstrained, free-fall motion – this includes the most interesting case of the two body problem in general relativity. This specifically excludes forced motion of, for example, a mass bouncing on the end of a spring. This restricted interest allows us in a general way to avoid the mathematical complications of Green's functions in curved spacetime and to rely instead on a strongly intuitive perspective which may be backed up with detailed analysis.

1.1 Outline

The Newtonian self-force problem in this Introduction is expanded upon in Section 2, where it becomes clear that careful definitions of coordinates are difficult to come by, and that physics is best described in terms of precisely defined and physically measurable quantities.

In Section 3 we describe Dirac's [31] classical treatment of radiation reaction in the context of electricity and magnetism in a language which mimics our approach to the gravitational self-force and to an illustrative toy problem in Section 4.

Perturbation theory in General Relativity is described in Section 5.1, applied to locally inertial coordinates in Section 5.2, applied to a neighborhood around a point mass in Section 5.3, and used to describe a small object moving through spacetime in Section 5.4.

The gravitational self-force is described in Section 6, which includes discussions of the conservative and dissipative effects and of some different possible implementations of self-force analyses.

The important and yet very confusing issue of gauge freedom in perturbation theory is raised in Section 7. And an example of gauge confusion in action is given in Section 8.

An outline of the necessary steps in a self-force calculation is given in Section 9, and some recent examples of actual gravitational self-force results are in Section 10 and 10.1. Section 10.2 describes a possible future approach to self-force calculations which is amenable to a 3+1 numerical implementation in the style of numerical relativity.

Concluding remarks are in Section 11.

1.2 Notation

The notation matches that used in an earlier review by the author [25] and is described here and again later in context.

Spacetime tensor indices are taken from the first third of the alphabet a, b, \dots, h , indices which are purely spatial in character are taken from the middle third, i, j, \dots, q and indices from the last third r, s, \dots, z are associated with particular coordinate components. The operator ∇_a is the covariant derivative operator compatible with the metric at hand. We often use $x^i = (x, y, z)$ for the spatial coordinates, and t for a timelike coordinate. An overdot, as in $\dot{\mathcal{E}}_{ij}$, denotes a time derivative along a timelike worldline. The tensor η_{ab} is the flat Minkowski metric $(-1, 1, 1, 1)$, down the diagonal. The tensor f_{kl} is the flat, spatial Cartesian metric $(1, 1, 1)$, down the diagonal. The projection operator onto the two dimensional surface of a constant r two sphere is $\sigma_i^j = f_i^j - x_i x^j / r^2$. A capitalized index, A, B, \dots emphasizes that the index is spatial and tangent to such a two sphere. Thus when written as σ_{AB} the projection operator is exhibiting its alternative role as the metric of the two-sphere. The tensor ϵ_{ijk} is the spatial Levi-Civita tensor, which takes on values of ± 1 depending upon whether the permutation of the indices are even or odd in comparison to x, y, z . A representative length scale \mathcal{R} of the geometry in the region of interest in spacetime is the smallest of the radius of curvature, the scale of inhomogeneities, and the time scale for changes along a geodesic. Typically, if the region of interest is a distance r away from a massive object M , then $\mathcal{R}^{-2} \sim M/r^3$ provides a measure of tidal effects, and $\mathcal{R} \sim$ an orbital period.

2 Newtonian Examples of Self-Force and Gauge Issues

Newtonian gravity self-force effects appeared in the introduction. Why don't we discuss these effects in undergraduate classical mechanics? The primary reason is that the Newtonian two-body problem can be solved easily and analytically without mention of the self-force. But in addition, a description of the Newtonian self-force introduces substantial, unavoidable ambiguities which are similar to the relativistic choice of gauge. Only because gauge confusion haunts all of perturbation theory in General Relativity do we now examine the Newtonian self-force using an elementary example made unavoidably confusing.

Consider a smaller mass m_1 and a larger mass m_2 in circular orbits of radii r_1 and r_2 about their common center of mass, so

$$m_1 r_1 = m_2 r_2. \quad (4)$$

And their separation is

$$R = r_1 + r_2 = r_1(1 + m_1/m_2). \quad (5)$$

Newton's law of gravity gives

$$\frac{m_1 v_1^2}{r_1} = \frac{G m_1 m_2}{(r_1 + r_2)^2}. \quad (6)$$

The velocity v_1 of the small object could be measurable by a redshift experiment. For this Newtonian system

$$\begin{aligned} v_1^2 &= \frac{G m_2 r_1}{(r_1 + r_2)^2} \\ &= \frac{G m_2}{r_1(1 + m_1/m_2)^2}, \\ &= \frac{G m_2}{r_1} (1 - 2m_1/m_2 + \dots). \end{aligned} \quad (7)$$

Thus we could state that in the limit that $m_1 \rightarrow 0$, the gravitational self-force decreases the orbital speed v_1 by a fractional amount $-m_1/m_2$. But, as an alternative, it is also true that

$$\begin{aligned} v_1^2 &= \frac{G m_2}{R(1 + m_1/m_2)} \\ &= \frac{G m_2}{R} (1 - m_1/m_2 + \dots). \end{aligned} \quad (8)$$

Thus we could equally well state that in the limit that $m_1 \rightarrow 0$, the gravitational self-force decreases the orbital speed v_1 by a fractional amount $-m_1/2m_2$. Which would be correct?

How does the ambiguity arise? In the first treatment, near by the orbit the radius r_1 was implicitly held fixed while we took the limit $m_1 \rightarrow 0$, and in that limit R approaches r_1 from above. In the second treatment the separation R was implicitly

held fixed in the limit, and in that case r_1 approaches R from below. Which of these is the “correct” way to take the limit? When viewed near by, which is a better description of the size of the orbit r_1 or R ?

In this Newtonian situation there might be some specific reason to make one choice rather than the other and the confusion could be resolved by including the detail of which quantity is being held fixed during the limiting process. But, in General Relativity for a small mass m_1 orbiting a much more massive black hole m_2 the ambiguity persists. After including self-force effects on the motion of m_1 , it would be tempting to state that the Schwarzschild coordinate r of m_1 's location should be held fixed while $m_1 \rightarrow 0$ to reveal the true consequences of the gravitational self-force. However, only the spherical symmetry of the exact Schwarzschild geometry allows for the unambiguous definition of r . Whereas the actual perturbed geometry is not spherically symmetric and has no natural r coordinate.

A clear statement of a perturbative gauge choice (*cf* Section 7) that fixes the gauge freedom can provide a mathematically well-defined quantity r on the manifold. But physics has no preferred gauge and has no preferred choice for r , just as neither r_1 nor R is preferred in this Newtonian example.

Rather than arguing the benefits of one gauge choice over another, it is far better to discard the focus on the radius r_1 or the separation R of the orbit, and to consider only quantities that could be determined with clear, unambiguous physical measurements. The orbital frequency Ω could be determined from the periodicity of the system, and the speed of the less massive component v_1 could be measured via a Doppler shift. We now look for a relationship between these two physically measurable quantities.

From the Newtonian analysis above,

$$\Omega^2 = \frac{Gm_2}{r_1(r_1 + r_2)^2} = \frac{Gm_2}{r_1^3(1 + m_1/m_2)^2} \quad (9)$$

so that

$$r_1 = \left[\frac{Gm_2}{(1 + m_1/m_2)^2} \right]^{1/3} \Omega^{-2/3} \quad (10)$$

and

$$\begin{aligned} v_1^2 &= \frac{Gm_2 r_1}{(r_1 + r_2)^2} \\ &= \frac{Gm_2}{r_1(1 + m_1/m_2)^2} \\ &= \frac{(Gm_2 \Omega)^{2/3}}{(1 + m_1/m_2)^{4/3}} = (Gm_2 \Omega)^{2/3} \left(1 - \frac{4}{3} \frac{m_1}{m_2} + \dots \right) \end{aligned} \quad (11)$$

Next, it seems appropriate to define a quantity with units of length in terms of the physically measurable Ω ,

$$R_\Omega^3 = Gm_2/\Omega^2. \quad (12)$$

Now the velocity v_1 of the orbit and the orbital frequency Ω are related by

$$v_1^2 = \frac{Gm_2}{R\Omega} \left(1 - \frac{4}{3} \frac{m_1}{m_2} + \dots \right), \quad (13)$$

and in terms of these measurable quantities it is unambiguous to state that the gravitational self-force changes v_1 , for a fixed Ω by a fractional amount $-2m_1/3m_2$.

This describes the effect of the self-force on two physically measurable observables and thus qualifies as a true, unambiguous self-force effect.

3 Classical Electromagnetic Self-Force

The standard expression [32] for the electromagnetic radiation reaction force on a charged particle q is

$$\mathcal{F}_{\text{rad}} = \frac{2}{3} \frac{q^2}{c^3} \ddot{\mathbf{v}}. \quad (14)$$

Equation (14) has issues of interpretation, but it does indeed describe the radiation reaction force when applied with care.

Dirac's [31] derivation of Eq. 14 is my favorite and can be described in a way that blends rather well with my preferred description of the self-force and the toy problem described in the next section.

First, Dirac considers the causally interesting retarded electromagnetic field F_{ab}^{ret} of an accelerating charge. But, he also considers the advanced field F_{ab}^{adv} and then describes what I call the electromagnetic *singular source* S field in flat spacetime

$$F_{ab}^{\text{S}} = \frac{1}{2} (F_{ab}^{\text{ret}} + F_{ab}^{\text{adv}}). \quad (15)$$

The field F_{ab}^{S} might also be called the *symmetric* field, as in “symmetric under reversal of causal structure.” F_{ab}^{S} has unphysical causal features, but it is an exact solution to Maxwell's equations with a source. In curved spacetime the definition of the singular source S field is more complicated than in the flat-space version of Eq. 15.

Dirac next allows the charge q to be of finite size. Then he presents a subtle analysis using the conservation of the electromagnetic stress–energy tensor in a neighborhood of the charge to show that F_{ab}^{S} exerts no net force on the charge in the limit that the size of the charge is vanishingly small.

Now let F_{ab}^{act} be the actual, measurable electromagnetic field. Then F_{ab}^{act} may be separated into two parts

$$F_{ab}^{\text{act}} = F_{ab}^{\text{S}} + F_{ab}^{\text{R}} \quad (16)$$

where the remainder R-field is *defined* by

$$F_{ab}^{\text{R}} \equiv F_{ab}^{\text{act}} - F_{ab}^{\text{S}}. \quad (17)$$

Both F_{ab}^{act} and F_{ab}^{S} are solutions to Maxwell's equations, in the neighborhood of q , with identical sources. Thus F_{ab}^{R} is necessarily a vacuum solution of the electromagnetic field equations and is therefore regular in the neighborhood of the particle.

Dirac then states that the radiation reaction force on the charge q moving with four-velocity u^a is

$$\mathcal{F}_b^{\text{rad}} = qu^a F_{ab}^{\text{R}} \quad (18)$$

and later shows that this is consistent with Eq. 14. In this context F_{ab}^{R} might be called the *radiation reaction* field, in view of the force it exerts on the charge.

Imagine the situation as viewed by a local observer who moves with the particle and is able to measure and analyze the actual electromagnetic field only in a neighborhood which includes the particle but is substantially smaller than the wavelength of any radiation. The observer is therefore not privy to any information whatsoever about distant boundary conditions, or about the possible existence of electromagnetically active material outside the neighborhood or even about the possibility of electromagnetic radiation either ingoing or outgoing at a great distance.

After considering the motion of the charge, the observer could calculate F_{ab}^{S} and then subtract it from the measured F_{ab}^{act} to yield F_{ab}^{R} . Finally the observer could apply Eq. 18 and conclude that the Lorentz force law correctly describes the electromagnetic contribution to the acceleration of the charge, even though the observer might be completely unaware of the presence of the radiation.

Thus F_{ab}^{act} is decomposed into two parts [28]. One part F_{ab}^{S} is singular at the point charge, can be identified as the particle's own electromagnetic field, and exerts no force on the particle itself. The other part F_{ab}^{R} does exert a force on the particle, is a locally source-free solution of Maxwell's equations and can be locally identified only as an externally generated field of indeterminate origin. A local observer would have no direct information about the source of F_{ab}^{R} and, in particular, could not distinguish the effects of radiation reaction from the effects of boundary conditions.

4 A Toy Problem with Two Length Scales That Creates a Challenge for Numerical Analysis

Binary inspiral of a small black hole into a much larger one presents substantial difficulties to the numerical relativity community. Perhaps the primary difficulty results from having two very different length scales. On the one hand, a very coarse grid size would allow easy resolution of the metric of the large black hole as well as coverage out to the wavezone resulting in the efficient production of gravitational waveforms. On the other hand, a very fine grid size would provide the detailed information about the metric in a neighborhood of the small black hole necessary for tracking the evolution of the binary system and for providing accurate gravitational waveforms.

The following toy problem shares the two length-scale difficulty of binary inspiral. But it is elementary, not complicated by curved spacetime or subtle

dynamics, and yet leads to some insight on how the binary inspiral problem might be approached. In addition, its resolution involves some aspects of Dirac's analysis of electromagnetic radiation reaction as presented in the previous section.

Consider this flat space numerical analysis problem in electrostatics: An object of small radius r_o has a spherically symmetric electric charge density $\rho(r)$ with an associated electrostatic potential φ . The object is inside an odd shaped grounded, conducting box which is much larger than r_o . The boundary condition on the potential is that $\varphi = 0$ on the box. For simplicity assume that the small object is at rest at the origin of coordinates. Thus, there is no radiation and the field equation for φ is elliptic. Then

$$\nabla^2 \varphi = -4\pi\rho \quad (19)$$

where ∇ is the usual three-dimensional flat space gradient operator, and ∇^2 the Laplace operator. Let \mathbf{r} refer to the displacement from the center of the object at the origin to a general point in the domain within the box.

Here is the goal: Given $\rho(r)$, numerically determine φ as a function of \mathbf{r} everywhere inside the box, subject to the field equation (19) and to the boundary condition that $\varphi = 0$ on the boundary of the box. Then find the total force on the small object which results from its interaction with φ .

Here is the difficulty: If the object is much smaller than the box, then the difference in length scales complicates calculating φ . The object is very small so an accurate analysis would require a very fine grid size. However, the distance from the object to the boundary of the box is large compared to the size of object. Thus a relatively coarse grid size would be desired to speed up the numerical evaluation. The difficulty is exacerbated if we are also interested in the force from φ acting back on the object; this requires accurately knowing the value of φ inside the small object precisely where φ has substantial variability.

We will shortly introduce a variety of versions of the potential under consideration. For clarity, the *actual* electrostatic potential φ^{act} actually satisfies both the field equation (19) with the actual source and also the relevant boundary conditions. Thus, φ^{act} is the potential which an observer would actually measure for the problem at hand.

4.1 An Approach Which Avoids the Small Length Scale

To remove the two-length-scale numerical difficulty we take the following approach: In a neighborhood of the object the potential ought to be approximated by the function φ^S defined as the usual electrostatic potential of a spherical distribution of charge which for a constant charge density $\rho(r)$ and total charge q is

$$\begin{aligned} \text{for } r < r_o : \quad \varphi^S(r) &= \frac{q}{2r_o^3}(3r_o^2 - r^2) \\ \text{for } r > r_o : \quad \varphi^S(r) &= q/r. \end{aligned} \quad (20)$$

The *source* field $\varphi^S(r)$ is completely determined by local considerations in the neighborhood of the object, and it is chosen carefully to be an elementary solution of

$$\nabla^2 \varphi^S = -4\pi\rho. \quad (21)$$

Sometimes φ^S is called the *singular* field to emphasize the q/r behavior outside but near a small source. Viewed from near by, the actual field φ^{act} is approximately φ^S .

Given φ^S , the numerical problem may be reformulated in terms of the field

$$\varphi^R \equiv \varphi^{\text{act}} - \varphi^S \quad (22)$$

which is then a solution of

$$\nabla^2 \varphi^R = -\nabla^2 \varphi^S - 4\pi\rho = 0, \quad (23)$$

where the second equality follows from Eq. 21. The *regular* field φ^R is thus a source free solution of the field equation, and is sometimes called the *remainder* when the *subtrahend* φ^S is removed from the actual field φ^{act} in Eq. 22.

Viewed from afar, the boundary condition that $\varphi^{\text{act}} = 0$ on the box plays an important role and determines the boundary condition that $\varphi^R = -\varphi^S$ on the box. Thus, rewriting the problem in terms of the analytically known φ^S and the “to be determined numerically” φ^R leaves us with the boundary value problem

$$\nabla^2 \varphi^R = 0 \quad \text{with the boundary condition that } \varphi^R = -\varphi^S \text{ on the box.} \quad (24)$$

It is important to note that φ^R is a regular, source-free solution of the field equation.

In this formulation based upon Eq. 24 φ^R scales as the charge q but has no structure with the length scale of the source r_o . The small length scale has been completely removed from the problem. The removal is at the expense of introducing a complicated boundary condition – but at least the boundary condition does not have an associated small length scale. Once φ^R has been determined, the actual field $\varphi^{\text{act}} = \varphi^R + \varphi^S$ is easily constructed.²

But that’s not all: This formulation has the bonus that it simplifies the calculation of the force on the object from the field. The force is an integral over the volume of the object,

$$\mathbf{F} = - \int \rho(r) \nabla \varphi^{\text{act}} d^3x. \quad (25)$$

² Following Dirac’s [31] usage, I prefer to use the word “actual” to refer to the complete, and total field that might be measured at some location. Often in self-force treatises the “retarded field” plays this central role. But, this obscures the fact that, viewed from near by, a local observer unaware of boundary conditions could make no measurement which would reveal just what part of the field is the retarded field. This confusion is increased if the spacetime is not flat, so that the retarded field could be determined only if the entire spacetime geometry were known.

In the original formulation using Eq. 19, the actual field φ^{act} in the integral would be dominated by φ^{S} which changes dramatically over the length scale of the object, and φ^{R} could be easily lost in the noise of the computation. The spherical symmetry of φ^{S} and ρ imply that

$$\int \rho(r) \nabla \varphi^{\text{S}} d^3x = 0. \quad (26)$$

Then the substitution $\varphi^{\text{act}} \rightarrow \varphi^{\text{S}} + \varphi^{\text{R}}$ in the integral of Eq. 25 leads to the conclusion that

$$\mathbf{F} = - \int \rho(r) \nabla \varphi^{\text{R}} d^3x. \quad (27)$$

Thus the force acting on the object may be written in terms of only φ^{R} .

But that's not all: The field φ^{R} does not change significantly over a small length scale, so if the object is extremely small (Think: an approximation to a δ -function.) then an accurate approximation to the force is

$$\mathbf{F} = -q \nabla \varphi^{\text{R}}|_{r=0}. \quad (28)$$

when viewed from near by.

Standard jargon calls the force in Eq. 28 the “self-force” because it is necessarily proportional to q^2 and apparently results from the object interacting with “its own field.” But, it is important to note that this force clearly depends upon the shape of the box, *i.e.* the details of the boundary conditions. In my opinion the physics appears more intuitive to have “the object’s own field,” refer only to φ^{S} whose local behavior is defined uniquely and independently of any boundary conditions. And φ^{S} is also guaranteed to exert no force back on the charge. Then φ^{R} is a regular source-free solution to the field equation in the neighborhood of the object and is solely responsible for the force acting on the object. An observer local to the object would know $\rho(r)$, could calculate φ^{S} and measure φ^{act} . Subtracting φ^{S} from the actual field φ^{act} then results in the regular remainder $\varphi^{\text{R}} = \varphi^{\text{act}} - \varphi^{\text{S}}$. While the force described in Eq. 28 is indeed proportional to q^2 , it still seems sensible to refer to this as simply “the force” on the object.

4.2 An Alternative That Resolves Boundary Condition Issues

The previous resolution of the difficulty of the two length scales caused a change and complication of the boundary conditions. With a slight variation, the problem can be reformulated in a way that brings back the original, natural boundary conditions.

The alternative approach deals with the boundary condition complication by introducing a window function $W(r)$ [54] which has three properties:

- (A) $W(r) = 1$ in a region which includes at least the entire source $\rho(r)$, that is all $r \leq r_0$.

- (B) $W(r) = 0$ for $r > r_W$ where r_W is generally much larger than r_o but is restricted so that the entire region $r < r_W$ is inside the box.
 (C) $W(r)$ is C^∞ and changes only over a long length scale comparable to r_W .

For this alternative approach the field defined by

$$\Phi^R \equiv \varphi^{\text{act}} - W\varphi^S \quad (29)$$

is a solution of

$$\begin{aligned} \nabla^2 \Phi^R &= -\nabla^2(W\varphi^S) - 4\pi\rho \\ &= -\varphi^S \nabla^2 W - 2\nabla W \cdot \nabla \varphi^S - W \nabla^2 \varphi^S - 4\pi\rho \\ &= -\varphi^S \nabla^2 W - 2\nabla W \cdot \nabla \varphi^S \equiv S_{\text{eff}}, \end{aligned} \quad (30)$$

where S_{eff} is the *effective source* and the third equality follows from Eq. 21 and property (A). The boundary condition is now that $\Phi^R = 0$ on the box, which is the natural boundary condition. Thus, rewriting the problem in terms of the analytically known φ^S and the to-be-determined-numerically Φ^R leaves us with the field equation

$$\nabla^2 \Phi^R = S_{\text{eff}} \quad (31)$$

and the *natural* boundary condition that $\Phi^R = 0$ on the box.

It is important to note that the effective source S_{eff} defined in Eq. 30 is zero inside the small object where $W(r) = 1$ and changes only over a long length scale r_W . Thus the field Φ^R is a regular, source-free solution of the field equation inside the object, and outside the object Φ^R only changes over a long length scale r_W . And Eqs. 27 and 28 provide the force acting on the object, after φ^R is replaced with Φ^R .

This alternative approach completely removes the small length scale from the problem and leaves the natural boundary condition $\Phi^R = 0$ on the box intact.

In applications of this approach to problems in curved spacetime, the singular field φ^S is rarely known exactly. In fact, for a δ -function source often only a finite number of terms in an asymptotic expansion are available. This limits the differentiability of the source of Eq. 31 which, in turn, limits the differentiability of Φ^R at the particle. But the procedure remains quite adequate for solving self-force problems.

This approach to the self-force, which introduces a window function, has now been implemented for a scalar charge in a circular orbit of the Schwarzschild geometry and is discussed below in Section 10.2.

5 Perturbation Theory

Perturbation theory has had some great successes in General Relativity particularly in the realm of black holes [46, 49, 51, 60] by proving stability [46, 56, 58, 60], analyzing the quasi-normal modes, [18, 22, 34, 35, 44] and calculating the gravitational waves from objects falling in and around black holes [19–21, 23, 60] to highlight just a few of the earlier accomplishments.

In preparation for the era of gravitational wave astronomy, relativists are now turning their attention to second and higher order perturbation analysis. However, we focus on linear order and give a brief description of this theory.

In Section 5.1 we begin with an overview that emphasizes the Bianchi identity's implication that a perturbing stress–energy tensor T_{ab} must be conserved $\nabla^a T_{ab} = 0$ to have a well formulated perturbation problem. This requires that an object of small size and mass must move along a geodesic.

We use perturbation theory in Section 5.2 to describe the geometry in the vicinity of a timelike geodesic Γ of a vacuum spacetime. We specifically use a locally inertial and harmonic coordinate system, THZ coordinates introduced by Thorne and Hartle [52], to represent the metric as a perturbation of flat spacetime $g_{ab} = \eta_{ab} + H_{ab}$ in a particularly convenient manner within a neighborhood of the geodesic.

In Section 5.3 we put a small mass m down on this same geodesic Γ and treat its gravitational field h_{ab}^S as a perturbation of g_{ab} .

Finally, in Section 5.4 we identify h_{ab}^S as the S-field of m , the analogue of F_{ab}^S in Section 3 and of ϕ^S in Section 4. In particular h_{ab}^S is a metric perturbation which is singular at the location of m , is a solution of the field equation for a δ -function point mass and exerts no force back on the mass m itself.

5.1 Standard Perturbation Theory in General Relativity

We start with a spacetime metric g_{ab} which is a vacuum solution of the Einstein equations $G_{ab}(g) = 0$. Then we ask, “What is the slight perturbation h_{ab} of the metric created by a small object moving through the spacetime along some world-line Γ ?”

Let \mathcal{R} be a representative length scale of the geometry near the object which is the smallest of the radius of curvature, the scale of inhomogeneities, and the time scale for changes in curvature along the world line of the object. When we say “small object” we imply that the size d of the object is much less than \mathcal{R} and that the mass m is much smaller than d .

As a notational convenience, the Einstein tensor $G_{ab}(g + h)$ for a perturbed metric may be expanded in powers of h as

$$G(g + h) = G(g) + G^{(1)}(g, h) + G^{(2)}(g, h) + \dots \quad (32)$$

where $G^{(n)}(g, h) = O(h^n)$. The zeroth order term $G(g)$ is zero if g_{ab} is a vacuum solution of the Einstein equations. The first order part is $G_{ab}^{(1)}(g, h)$, which resembles a linear wave operator on h_{ab} and is equivalent to the operator $-E_{ab}(h)$ given below in Eq. 35. The second order part $G^{(2)}(g, h)$ consists of terms such as “ $\nabla h \nabla h$ ” or “ $h \nabla \nabla h$,” similar to the Landau–Lifshitz pseudo tensor [33]. The third and higher order terms in the expansion (32) are less familiar.

Next, we assume that the stress–energy tensor of the object T_{ab} is $O(m)$, and that the perturbation in the metric h_{ab} is also $O(m)$. At first perturbative order,

$$G_{ab}(g + h) = 8\pi T_{ab} + O(h^2). \quad (33)$$

We expand $G_{ab}(g + h)$ through first order in h via the symbolic operation

$$G_{ab}^{(1)}(g, h) = \frac{\delta G_{ab}}{\delta g_{cd}} h_{cd} \quad (34)$$

and define the wave operator mentioned above by $E_{ab}(h) \equiv -G_{ab}^{(1)}(g, h)$, so that

$$\begin{aligned} 2E_{ab}(h) = & \nabla^2 h_{ab} + \nabla_a \nabla_b h - 2\nabla_{(a} \nabla^c h_{b)c} \\ & + 2R_a{}^c{}_{b^d} h_{cd} + g_{ab}(\nabla^c \nabla^d h_{cd} - \nabla^2 h), \end{aligned} \quad (35)$$

with $h \equiv h_{ab}g^{ab}$. Also ∇_a and $R_a{}^c{}_{b^d}$ are the derivative operator and Riemann tensor of g_{ab} . If h_{ab} solves

$$E_{ab}(h) = -8\pi T_{ab}. \quad (36)$$

then Eq. 33 is satisfied.

In an actual project, the biggest technical task is usually solving Eq. 36. As an example, the study of gravitational radiation from an object orbiting a Schwarzschild black hole typically invokes the Regge–Wheeler–Zerilli formalism [46, 60].

With a vacuum-spacetime metric g_{ab} and *any* symmetric tensor k_{ab} , the Bianchi identity implies that

$$\nabla^a E_{ab}(k) = 0. \quad (37)$$

This is easily demonstrated by direct analysis, after starting with Eq. 35. Thus, for a solution of Eq. 36 to exist, it is necessary that the integrability condition

$$\nabla^a T_{ab} = 0 \quad (38)$$

for the stress–energy tensor be satisfied.

If the stress–energy tensor is only approximately conserved $\nabla^a T_{ab} = O(m^2)$ then the solution for h_{ab} might be in error at $O(m^2)$. In some circumstances this might be acceptable, in which case if T_{ab} represents the stress–energy tensor for a particle of small size, then the particle must move along an approximate geodesic of g_{ab} [40] with an acceleration no larger than $O(m)$. Then the integrability condition is nearly satisfied and h_{ab} can be determined from Eq. 36.

Next, one might be inclined to attempt the analysis of the Einstein equations through second order in the perturbation h_{ab} . But, this requires that T_{ab} be conserved, not in the metric g_{ab} , but rather in the first order perturbed metric $g_{ab} + h_{ab}$. Thus the worldline of a particle is not geodesic in g_{ab} and its acceleration as measured in g_{ab} is often said to result from the *gravitational self-force*. After the

self-force problem is solved for the $O(m)$ adjustment to the motion of the particle, then the second order field equation from Eq. 32 determines h_{ab} through $O(m^2)$.

As described by Thorne and Kovács [53], this process continues: With the improved metric, solve the dynamical equations for a more accurate worldline and stress–energy tensor. With the improved stress–energy tensor solve the field equations for a more accurate metric perturbation. Repeat.

This alternation of focus between the dynamical equations and the field equations is quite similar to that used in post-Newtonian analyses.

5.2 An Application of Perturbation Theory: Locally Inertial Coordinates

Before dealing with perturbing masses, we first consider vacuum perturbations of a vacuum spacetime and focus on a neighborhood of a timelike geodesic Γ where the metric appears as a perturbation H_{ab} of the flat Minkowskii metric η_{ab} .

This application is simplified by use of a convenient coordinate system described by Thorne and Hartle [52]. It is well known in General Relativity [57], that for a timelike geodesic Γ in spacetime there is a class of *locally inertial* coordinate systems $x^a = (t, x, y, z)$, with $r^2 = x^2 + y^2 + z^2$, which satisfies the following conditions:

- (A) The geodesic Γ is identified with $x = y = z = r = 0$ and t measures the proper time along the worldline.
- (B) On Γ , the metric takes the Minkowskii form $g_{ab} = \eta_{ab}$.
- (C) All first derivatives of g_{ab} vanish on Γ so that the Christoffel symbols also vanish on Γ .

Fermi-normal coordinates [39] provide an example which meets all of these locally inertial criteria.

With a locally inertial coordinate system in hand, it is natural to Taylor expand g_{ab} about Γ with

$$g_{ab} = \eta_{ab} + H_{ab} + \dots \quad (39)$$

where

$$\begin{aligned} H_{ab} &= {}_2H_{ab} + {}_3H_{ab}, \\ {}_2H_{ab} &= \frac{1}{2}x^i x^j \partial_i \partial_j g_{ab}, \\ {}_3H_{ab} &= \frac{1}{6}x^i x^j x^k \partial_i \partial_j \partial_k g_{ab}, \end{aligned} \quad (40)$$

and the partial derivatives are evaluated on Γ .

The quantities ${}_2H_{ab}$ and ${}_3H_{ab}$ scale as $O(r^2/\mathcal{R}^2)$ and $O(r^3/\mathcal{R}^3)$ in a small neighborhood of Γ , and these may be treated as perturbations of flat spacetime with r/\mathcal{R} being the small parameter. Recall that \mathcal{R} is a length scale of the background geometry. First order perturbation theory is applicable here because H_{ab} has no

$O(r/\mathcal{R})$ term but starts at $O(r^2/\mathcal{R}^2)$. Thus ${}_2H_{ab}$ and ${}_3H_{ab}$ may be treated as independent perturbations and the first nonlinear term appears at $O(r^4/\mathcal{R}^4)$. Thus, H_{ab} is a perturbation which must satisfy the source-free perturbed Einstein equations $E_{ab}(H) = 0$.

Thorne and Hartle [52] and Zhang [62] show that a particular choice of locally inertial coordinates leads to a relatively simple expansion of the metric. Initially they introduce spatial, symmetric, trace-free *multipole moments* of the external spacetime \mathcal{E}_{ij} , \mathcal{B}_{ij} , \mathcal{E}_{ijk} , and \mathcal{B}_{ijk} which are functions only of t and are directly related to the Riemann tensor evaluated on Γ by

$$\mathcal{E}_{ij} = R_{titj}, \quad (41)$$

$$\mathcal{B}_{ij} = \varepsilon_i{}^{pq} R_{pqjt}/2, \quad (42)$$

$$\mathcal{E}_{ijk} = [\partial_k R_{titj}]^{\text{STF}} \quad (43)$$

and

$$\mathcal{B}_{ijk} = \frac{3}{8} [\varepsilon_i{}^{pq} \partial_k R_{pqjt}]^{\text{STF}}. \quad (44)$$

Here $^{\text{STF}}$ means to take the symmetric, tracefree part with respect to the spatial indices, and ε_{ijk} is the flat, spatial Levi-Civita tensor, which takes on values of ± 1 depending upon whether the permutation of the indices are even or odd in comparison to x, y, z . Also, \mathcal{E}_{ij} and \mathcal{B}_{ij} are $O(1/\mathcal{R}^2)$, while \mathcal{E}_{ijk} and \mathcal{B}_{ijk} are $O(1/\mathcal{R}^3)$. All of the above multipole moments are tracefree because the external background geometry is assumed to be a vacuum solution of the Einstein equations.

Spatial STF tensors are closely related to linear combinations of spherical harmonics. For example the STF tensor \mathcal{E}_{ij} with two spatial indices is related to the $\ell = 2$ spherical harmonics $Y_{2,m}$ by

$$\mathcal{E}_{ij} x^i x^j = r^2 \sum_{m=-2}^2 E_{2,m} Y_{2,m}, \quad (45)$$

with the five independent components of \mathcal{E}_{ij} being determined by the five independent coefficients $E_{2,m}$.

Next an infinitesimal coordinate transformation (a perturbative gauge transformation, Section 7) changes the description of H_{ab} to a form where the partial derivatives in the Taylor expansion are equivalent to the components of the Riemann tensor and represented by the multipole moments. The result is

$$\begin{aligned} {}_2H_{ab} dx^a dx^b &= -\mathcal{E}_{ij} x^i x^j (dt^2 + f_{kl} dx^k dx^l) + \frac{4}{3} \varepsilon_{kpq} \mathcal{B}^q{}_i x^p x^i dt dx^k \\ &\quad - \frac{20}{21} \left[\dot{\mathcal{E}}_{ij} x^i x^j x_k - \frac{2}{5} r^2 \dot{\mathcal{E}}_k x^i \right] dt dx^k \\ &\quad + \frac{5}{21} \left[x_i \varepsilon_{j pq} \dot{\mathcal{B}}^q{}_k x^p x^k - \frac{1}{5} r^2 \varepsilon_{pqi} \dot{\mathcal{B}}_j{}^q x^p \right] dx^i dx^j + O(r^4/\mathcal{R}^4) \end{aligned} \quad (46)$$

and

$$\begin{aligned} {}_3H_{ab} dx^a dx^b &= -\frac{1}{3} \dot{\mathcal{E}}_{ijk} x^i x^j x^k (dt^2 + f_{lm} dx^l dx^m) \\ &\quad + \frac{2}{3} \varepsilon_{kpq} \mathcal{B}^q_{ij} x^p x^i x^j dt dx^k + O(r^4/\mathcal{R}^4), \end{aligned} \quad (47)$$

where f_{kl} is the flat, spatial Cartesian metric $(1, 1, 1)$, down the diagonal. The overdot represents a time derivative along Γ of, say, $\mathcal{E}_{ij} = O(\mathcal{R}^{-2})$, and then $\dot{\mathcal{E}}_{ij} = O(\mathcal{R}^{-3})$ because \mathcal{R} bounds the time scale for variation along Γ .

A straightforward evaluation of the Riemann tensor for the metric $\eta_{ab} + {}_2H_{ab} + {}_3H_{ab}$ confirms that the STF multipole moments are related to the Riemann tensor as claimed in Eqs. 41–44.

We call the locally inertial coordinates of Thorne, Hartle and Zhang used in Eqs. 46 and 47 THZ coordinates.

If interest is focused only on the lower orders $O(r^2/\mathcal{R}^2)$ and $O(r^3/\mathcal{R}^3)$, then THZ coordinates are not unique and freedom is allowed in their construction away from the worldline Γ . Given one set of THZ coordinates x^a , a new set defined from $x^a_{\text{new}} = x^a + \lambda^a_{ijklm} x^i x^j x^k x^l x^m$, where $\lambda^a_{ijklm} = O(1/\mathcal{R}^4)$ is an arbitrary function of proper time on Γ , preserves the defining form of the expansion given in Eqs. 46 and 47.

Work in preparation describes a direct, constructive procedure for finding a THZ coordinate system associated with any geodesic of a vacuum solution of the Einstein equations.

5.3 Metric Perturbations in the Neighborhood of a Point Mass

We are now prepared to use perturbation theory to determine h^S_{ab} , the gravitational analogue of F^S_{ab} in Section 3 and of φ^S in Section 4.

We consider the perturbative change h_{ab} in the metric g_{ab} caused by a point mass m traveling through spacetime. We look for the solution h_{ab} to Eq. 36 with the stress–energy tensor T^{ab} of a point mass

$$T^{ab} = m \int_{-\infty}^{\infty} \frac{u^a u^b}{\sqrt{-g}} \delta^4(x^a - X^a(s)) ds, \quad (48)$$

where $X^a(s)$ describes the worldline of m in an arbitrary coordinate system as a function of the proper time s along the worldline.

The integrability condition for Eq. 36 requires the conservation of T^{ab} , and we put m down on the geodesic Γ of the previous section and limit interest to a neighborhood of Γ where r^4/\mathcal{R}^4 is considered negligible although r^3/\mathcal{R}^3 is not. And we use THZ coordinates. The perturbed metric of Section 5.2 is now viewed as the “background” metric, $g_{ab} = \eta_{ab} + H_{ab}$, with H_{ab} given in Eqs. 46 and 47. The

stress–energy tensor T_{ab} for a point mass is particularly simple in THZ coordinates and has only one nonzero component,

$$T_{tt} = m\delta^3(x^i). \quad (49)$$

For this stress–energy tensor and this background metric, we call the solution to Eq. 36, h_{ab}^S , for reasons explained below, and its derivation is given elsewhere [24, 25]. Here we present the results:

$$h_{ab}^S = {}_0h_{ab}^S + {}_2h_{ab}^S + {}_3h_{ab}^S, \quad (50)$$

where

$${}_0h_{ab}^S dx^a dx^b = 2\frac{m}{r}(dt^2 + dr^2) \quad (51)$$

is the Coulomb m/r part of the Schwarzschild metric, and

$$\begin{aligned} {}_2h_{ab}^S dx^a dx^b &= \frac{4m}{r} \mathcal{E}_{ij} x^i x^j dt^2 - 2\frac{4m}{3r} \varepsilon_{kpq} \mathcal{B}^q_{ij} x^p x^i dt dx^k \\ &\quad + \dot{\mathcal{E}}_{ij} \text{ and } \dot{\mathcal{B}}_{ij} \text{ terms} \end{aligned} \quad (52)$$

are the quadrupole tidal distortions of the Coulomb part. The terms involving $\dot{\mathcal{E}}_{ij}$ and $\dot{\mathcal{B}}_{ij}$ are more complicated and are not given here. The octupole tidal distortions of the Coulomb field are

$$\begin{aligned} {}_3h_{ab}^S dx^a dx^b &= \frac{m}{3r} \mathcal{E}_{ijk} x^i x^j x^k \left[5 dt^2 + dr^2 + 2\sigma_{AB} dx^A dx^B \right] \\ &\quad - 2\frac{10m}{9r} \varepsilon_{kpq} \mathcal{B}^q_{ij} x^p x^i x^j dt dx^k. \end{aligned} \quad (53)$$

Recall that σ_{AB} is the two dimensional metric on the surface of a constant r two sphere.

The perturbation h_{ab}^S is a solution to Eq. 36 only in a neighborhood of Γ . The next perturbative-order terms that are not included in h_{ab}^S scale as mr^3/\mathcal{R}^4 . The operator E_{ab} involves second derivatives, and it follows that for h_{ab}^S given above

$$E_{ab}(h^S) = -8\pi T_{ab} + \mathcal{O}(mr/\mathcal{R}^4). \quad (54)$$

In some circumstances we might wish to introduce a window function W similar to that described in Section 4.2, which would multiply all of the terms on the right hand side of Eq. 50. If so, the window function near by m must be restricted by the condition that

$$W = 1 + \mathcal{O}(r^4/\mathcal{R}^4) \quad (55)$$

in order to preserve the delicate features of h_{ab}^S in a neighborhood of m , especially the property revealed in Eq. 54. Away from m , it is only necessary that W vanish in some smooth manner.

The perturbations ${}_2h_{ab}^S$ and ${}_3h_{ab}^S$ should not be confused with a consequence of Newtonian tides. When a small Newtonian object moves through spacetime, its mass distribution is tidally distorted by the external gravitational field. The extent of this distortion depends upon the size d of the object itself. For a self-gravitating, non-rotating incompressible fluid,³ the quadrupole distortion of the matter leads to a change in the Newtonian gravitational potential outside the object which scales as $\delta U \sim \mathcal{I}_{ij} x^i x^j / r^3 \sim d^5 / r^3 \mathcal{R}^2$, where \mathcal{I}_{ij} is the mass quadrupole moment tensor. Such behavior is not at all similar to that of ${}_2h_{tt}^S = O(mr/\mathcal{R}^2)$, and ${}_3h_{tt}^S = O(mr^2/\mathcal{R}^3)$.

The quadrupole distortion revealed in ${}_2h_{tt}^S$ is not a consequence of a distortion of the object m itself, but rather results from the curvature of spacetime acting on the monopole field of m and has no Newtonian counterpart.

5.4 A Small Object Moving Through Spacetime

As a concrete example we now focus on a small Newtonian object of mass m and characteristic size d moving through some given external vacuum spacetime with metric g_{ab} . Naturally, m is approximately moving along a geodesic Γ , and g_{ab} has a characteristic length and time scale \mathcal{R} associated with Γ . We assume that m and d are both much smaller than \mathcal{R} .

In a region comparable to d , the object appears Newtonian, and its gravitational potential can be determined. The structure of the object depends upon details like the density, type of matter, amount of rotation and whether it is stationary or oscillating.

The Newtonian object might have a mass quadrupole moment $\mathcal{I}_{ij} = O(md^2)$ perhaps sustained by internal stresses in the matter itself. Independent of the cause of the quadrupole moment, the external Newtonian gravitational potential would have a quadrupole part $\mathcal{I}_{ij} x^i x^j / r^5$.

The coupling between a mass quadrupole moment of the small object and an external octupole gravitational field $\mathcal{E}_{ijk} x^i x^j x^k$ results in the small acceleration of the center of mass, away from free-fall, given by [52, 61]

$$a^i = -\frac{1}{2m} \mathcal{E}^{ijk} \mathcal{I}_{jk} \quad (56)$$

in either the context of Newtonian physics or of General Relativity. This tidal acceleration scales as

$$a = O(d^2/\mathcal{R}^3). \quad (57)$$

If our small Newtonian object is actually a nonrotating fluid body then it would naturally be spherically symmetric except for distortion caused by an external tidal

³ A terse but adequate description of perturbative tidal effects on a Newtonian, self-gravitating, non-rotating, incompressible fluid is given on p. 467 of [16].

field such as $\mathcal{E}_{ij}x^i x^j$. In that case $\mathcal{I}_{ij} = \mathcal{O}(d^5/\mathcal{R}^2)$ as discussed at the end of Section 5.3 and in [16], and the tidal acceleration then scales as

$$a = \mathcal{O}(d^5/m\mathcal{R}^5). \quad (58)$$

We conclude that a Newtonian object in free motion is only allowed an acceleration away from free-fall which is limited as in Eq. 57 or 58. Any larger acceleration must involve some non-gravitational force.

It is also possible to analyze the situation if we replace the Newtonian object with a small Schwarzschild black hole of mass m . In that case it is easiest to turn the perturbation problem inside-out and to consider the Schwarzschild metric as the background with the metric perturbation being caused by H_{ab} given in Eqs. 46 and 47. One boundary condition is that h_{ab} approach H_{ab} for $m \ll r \ll \mathcal{R}$. The boundary condition at the event horizon is that h_{ab} be an ingoing wave, or well-behaved in the time independent limit. The time independent problem is well studied; historically in Refs. [46, 60], more recently in the present context in Ref. [24], and with slow time dependence in Refs. [41, 42].

In the time independent limit, the generic quadrupole perturbation of the metric of the Schwarzschild spacetime results in

$$\begin{aligned} (g_{ab}^{\text{Schw}} + h_{ab}^{\text{Schw}}) dx^a dx^b = & -\left(1 - \frac{2m}{r}\right) \left[1 - \mathcal{E}_{ij}x^i x^j \left(1 - \frac{2m}{r}\right)\right] dt^2 \\ & + \frac{4}{3} \varepsilon_{kpq} \mathcal{B}^q{}_i x^p x^i \left(1 - \frac{2m}{r}\right) dt dx^k \\ & + \left(\frac{1}{1 - 2m/r} - \mathcal{E}_{ij}x^i x^j\right) dr^2 \\ & + \left[r^2 - (r^2 - m^2)\mathcal{E}_{ij}x^i x^j\right] (d\theta^2 + \sin^2 \theta d\phi^2). \quad (59) \end{aligned}$$

In this expression x^i represents x, y and z which are related to r, θ and ϕ in the usual way in Cartesian space.

It is elementary to check that if $m = 0$ then this reduces to the time independent limit of Eq. 46. If \mathcal{E}_{ij} and $\mathcal{B}_{ij} = 0$ then this reduces to the Schwarzschild metric. And the terms which are bilinear in m and either \mathcal{E}_{ij} or \mathcal{B}_{ij} are equivalent to the time independent limit of Eq. 52. An expression with similar features holds for the octupole perturbations.

The metric of Eq. 59 represents a Schwarzschild black hole at rest on the geodesic Γ in a time-independent external spacetime. And note that there is no black hole quadrupole moment induced by the external quadrupole field as there are no quadrupole $1/r^3$ terms in this metric in the region where $m \ll r \ll \mathcal{R}$. The Schwarzschild black hole equivalent of \mathcal{I}_{ij} vanishes. It follows that, in this situation, the black hole has no acceleration away from Γ .

Time dependence in \mathcal{E}_{ij} slightly changes this situation. In [24], it is argued that with slow time dependence, with a time-scale $\mathcal{O}(\mathcal{R})$, the induced quadrupole field of the Schwarzschild metric in fact scales as $\sim m^5/r^3\mathcal{R}^2$, and that the acceleration from coupling with an external octupole field, $\mathcal{E}_{ijk}x^i x^j x^k \sim r^3/\mathcal{R}^3$, gives an acceleration

$$a = \mathcal{O}(m^4/\mathcal{R}^5). \quad (60)$$

This result is consistent with the Newtonian result in Eq. 58 if the size d of the Newtonian object is replaced with the mass m of the black hole.

An elementary approach using dimensional analysis arrives at this same result. Acceleration is a three-vector with a unit of 1/length. The only quantities in play are m , \mathcal{E}_{ij} and \mathcal{E}_{ijk} . The only combination of these which yields a vector with the units of acceleration is $m^4 \mathcal{E}^{ijk} \mathcal{E}_{jk} = \mathcal{O}(m^4/\mathcal{R}^5)$.

The field h_{ab}^S is now seen to satisfy the requirements desired for a ‘‘Singular field:’’

- (A) h_{ab}^S is a solution of the field equation in the vicinity of a δ -function mass source on a geodesic Γ .
- (B) h_{ab}^S exerts no force back on its δ -function source as evidenced by the facts that h_{ab}^S is the part of the perturbed Schwarzschild geometry that is linear in m , and that the small black hole has acceleration no larger than $\mathcal{O}(m^3/\mathcal{R}^4)$, while all that is required is that the acceleration be no larger than $\mathcal{O}(m^2/\mathcal{R}^3)$.

6 Self-Force from Gravitational Perturbation Theory

For an overview of the general approach to gravitational self-force problems about to be described, we refer back to the treatment of the electromagnetic self-force in Section 3, the toy-problem of Section 4, and particularly to the introduction of h_{ab}^S in Sections 5.3 and 5.4.

At a formal level, we begin with a metric g_{ab} which is a vacuum solution of the Einstein equation and look for an approximate solution for h_{ab}^{act} from

$$G(g + h^{\text{act}}) = 8\pi T + \mathcal{O}(h^2), \quad (61)$$

with appropriate boundary conditions, where $T_{ab} = \mathcal{O}(m)$ is the stress–energy tensor of a point particle m .

Initially we assume that m is moving along a geodesic Γ . In a neighborhood of Γ , h_{ab} is well approximated by h_{ab}^S . Thus we define h_{ab}^R via the replacement

$$h_{ab}^{\text{act}} = h_{ab}^S + h_{ab}^R, \quad (62)$$

and use the expansion in Eq. 32 and the definition in Eq. 35 to write

$$\begin{aligned} G_{ab}(g + h^{\text{act}}) &= G_{ab}(g) - E_{ab}(h^{\text{act}}) + \mathcal{O}(h^2) \\ &= -E_{ab}(h^R) - E_{ab}(h^S) + \mathcal{O}(h^2) \end{aligned} \quad (63)$$

where we use the assumption that $G_{ab}(g) = 0$ and the linearity of the operator $E_{ab}(h)$.

In Section 5.3 the properties of h_{ab}^S were chosen carefully so that

$$E_{ab}(h^S) = -8\pi T_{ab} + O(mr/\mathcal{R}^4) \text{ in a neighborhood of } m. \quad (64)$$

We can demonstrate this result by letting $4h_{ab}^S = O(mr^3/\mathcal{R}^4)$ be the next term not included in the expansion (50). The operator E_{ab} has second order spatial derivatives, and every time derivative brings in an extra factor of $1/\mathcal{R}$. Thus $E_{ab}(4h_{ab}^S) = O(mr/\mathcal{R}^4)$, and Eq. 64 follows.

Now we define the *effective source*

$$\begin{aligned} 8\pi S_{ab} &\equiv 8\pi T_{ab} + E_{ab}(h^S), \\ &= O(mr/\mathcal{R}^4). \end{aligned} \quad (65)$$

Thus S_{ab} is zero at $r = 0$, where it is continuous but not necessarily differentiable. Everywhere else S_{ab} is C^∞ .

The first perturbative order problem Eq. 61 is now reduced to solving

$$E_{ab}(h^R) = -8\pi S_{ab}, \quad (66)$$

and then Eq. 62 reconstructs h_{ab}^{act} . The limited differentiability of S_{ab} causes no fundamental difficulty for determining h_{ab}^R , and introduces no small length scale either. The resulting h_{ab}^R will be C^2 at the location of the point mass, and C^∞ elsewhere.

At this order of approximation Section 5.4 showed that the mass m moves along a geodesic of the *actual* metric g_{ab}^{act} with h_{ab}^S removed, i.e. along a geodesic of $g_{ab} + h_{ab}^{\text{act}} - h_{ab}^S = g_{ab} + h_{ab}^R$. Thus, the gravitational self-force results in geodesic motion not in g_{ab} but rather in $g_{ab} + h_{ab}^R$.

Admittedly, $g_{ab} + h_{ab}^R$ is not truly a vacuum solution of the Einstein equation. But, by construction it is clear that

$$G_{ab}(g + h^R) = O(mr/\mathcal{R}^4). \quad (67)$$

More terms of higher order in r/\mathcal{R} in the expression for h_{ab}^S would result in a remainder with more powers of r/\mathcal{R} on the right hand side of Eq. 67. But these would not change the first derivatives of h_{ab}^R on Γ which are all that would appear in the geodesic equation for m . So the expansion for h_{ab}^S as given in Section 5.4 is adequate for our purposes.

6.1 Dissipative and Conservative Parts

When viewed from near by, the effect of the gravitational self-force on a small mass m arises as a consequence of the purely local phenomenon of geodesic motion. In the neighborhood of m , it is impossible then to distinguish the dissipative part of the self-force from the conservative part.

Viewed from afar with the usually appropriate boundary conditions, the metric perturbation h_{ab}^{act} is actually the retarded field h_{ab}^{ret} and it is often useful then to distinguish the dissipative effects which remove energy and angular momentum from the conservative effects which might affect, say, the orbital frequency.

In the case that $h_{ab}^{\text{act}} = h_{ab}^{\text{ret}}$, it is natural to define the dissipative part of the regular field as

$$h_{ab}^{\text{dis}} = \frac{1}{2}(h_{ab}^{\text{ret}} - h_{ab}^{\text{adv}}) \quad (68)$$

The advanced and the retarded fields are each solutions of the same wave equation with the same δ -function source. Thus their difference is a solution of the homogeneous wave equation and is therefore regular at the point mass. And the dissipative effects of the self-force are revealed as geodesic motion in the metric $g_{ab} + h_{ab}^{\text{dis}}$.

In a complementary fashion, the conservative part of the regular field is naturally defined as

$$\begin{aligned} h_{ab}^{\text{con}} &= h_{ab}^{\text{R}} - \frac{1}{2}(h_{ab}^{\text{ret}} - h_{ab}^{\text{adv}}) \\ &= h_{ab}^{\text{ret}} - h_{ab}^{\text{S}} - \frac{1}{2}(h_{ab}^{\text{ret}} - h_{ab}^{\text{adv}}) \\ &= \frac{1}{2}(h_{ab}^{\text{ret}} + h_{ab}^{\text{adv}}) - h_{ab}^{\text{S}} \end{aligned} \quad (69)$$

And the conservative effects of the self-force are revealed as geodesic motion in the metric $g_{ab} + h_{ab}^{\text{con}}$.

With these definitions it is natural that

$$h_{ab}^{\text{R}} = h_{ab}^{\text{con}} + h_{ab}^{\text{dis}}. \quad (70)$$

This decomposition into conservative and dissipative parts follows an aspect of the procedure that Mino describes [36] as a possible method for computing the dissipative effects of gravitational radiation reaction on the Carter constant [14, 15] for a small mass orbiting a Kerr black hole.

6.2 Gravitational Self-Force Implementations

When it is actually time to search for some self-force consequences there are a number of different choices to be made.

6.2.1 Field Regularization Via the Effective Source

The majority of this review has been leading toward a natural implementation of self-force analysis using the standard 3 + 1 techniques of numerical relativity.

Assume that h_{ab}^R and its first derivatives, and also the position and four-velocity of m are known at one moment of time.

1. Use the position and four-velocity of m to analytically determine h_{ab}^S .
2. Obtain the effective source S_{ab} via Eq. 65.
3. Evolve Eq. 66 for h_{ab}^R one step forward in time.
4. Move the particle a step forward in time using the geodesic equation for $g_{ab} + h_{ab}^R$.
5. Repeat.

Section 10.2 describes the application of this approach to a scalar field problem and includes figures which reveal some generic characteristics of the source function.

6.2.2 Mode-Sum Regularization

Mode-sum regularization [1, 3] avoids the singularity of h_{ab}^{act} and its derivatives on Γ by an initial multipole-moment decomposition, say, into spherical harmonic components $h_{ab}^{\text{act}\ell m}$. With the assumption that h_{ab}^S is carefully defined away from m in a fashion that also allows for a decomposition in terms of spherical harmonics $h_{ab}^{S\ell m}$, then $h_{ab}^{R\ell m} = h_{ab}^{\text{act}\ell m} - h_{ab}^{S\ell m}$ would be the decomposition of h_{ab}^R . The collection of the multipole moments $h_{ab}^{S\ell m}$, their derivatives and various of their linear combinations are, together, known as “regularization parameters.” This essentially leads to the mode-sum regularization procedure of Barack and Ori [1, 3] which has been used in nearly all of the self-force calculations to date.

6.2.3 The Gravitational Self-Force Actually Resulting in Acceleration

We have strongly pushed our agenda of treating the gravitational self-force in local terms as geodesic motion through a vacuum spacetime $g_{ab} + h_{ab}^R$. However, when viewed from afar the worldline Γ of m is indeed accelerated and not a geodesic of the background geometry g_{ab} . This acceleration can be described as a consequence of m interacting with a spin-2 field h_{ab}^R which leads to the resulting acceleration

$$u^b \nabla_b u^a = - \left(g^{ab} + u^a u^b \right) u^c u^d \left(\nabla_c h_{db}^R - \frac{1}{2} \nabla_b h_{cd}^R \right) \quad (71)$$

away from the original worldline in the original metric g_{ab} .

Under some circumstances this might be a convenient interpretation. The resulting worldline would be identical to the geodesic of $g_{ab} + h_{ab}^R$ and would correctly incorporate all self-force effects, although the worldline would not be parameterized by the actual proper time. It is important to note that the acceleration of Eq. 71 cannot be measured with an accelerometer and, by itself, has no actual, direct physical consequence.

In the next section we describe some general consequences of gauge transformations in perturbation theory. Be warned that if Eq. 71 is used to calculate the deviation ζ^a of the worldline away from a geodesic in the background metric g_{ab} , then any gauge transformation whose gauge vector $\xi^a = -\zeta^a$, on the world line, would automatically set the right hand side of Eq. 71 to zero and leave m on its original geodesic. This possibility certainly confuses the interpretation of the right hand side of Eq. 71. Such a removal of the self-force only works as long as the deviation vector $\zeta^a \sim O(h)$. If self-force effects accumulate in time, such as from dissipation or orbital precession, then after a long enough time the effects of the self-force will be revealed.

7 Perturbative Gauge Transformations

In General Relativity, the phrase “choice of gauge” has different possible interpretations depending upon whether one is interested in perturbation theory or, say, numerical relativity. With numerical relativity, “choice of gauge” usually refers to the choice of a specific coordinate system, with the understanding that general covariance implies that the meaning of a calculated quantity might be as ambiguous as the coordinate system in use.

In perturbation theory the “choice of gauge” is more subtle. One considers the difference between the actual metric g_{ab}^{act} of a spacetime of interest and an abstract metric g_{ab} of a given, background spacetime. The difference

$$h_{ab} = g_{ab}^{\text{act}} - g_{ab} \quad (72)$$

is assumed to be small. The perturbed Einstein equations govern h_{ab} , and knowing h_{ab} might provide answers to questions concerning the propagation and emission of gravitational waves, for example.

In this perturbative context “choice of gauge” involves the choice of coordinates, but in a very precise sense [2, 9, 47, 50]. The subtraction in Eq. 72 is ambiguous. The two metrics reside on different manifolds, and there is no unique map from the events on one manifold to those of another. Usually the names of the coordinates are the same on the two manifolds, and this provides an implicit mapping between the manifolds. But this mapping is not unique. For example, the Schwarzschild geometry is spherically symmetric. This allows the Schwarzschild coordinate r to be defined in terms of the area $4\pi r^2$ of a spherically symmetric two-surface. The perturbed Schwarzschild geometry is not spherically symmetric, and to describe the coordinate r on the perturbed manifold as the “Schwarzschild r ” does not describe the meaning of r in any useful manner and is not a perturbative choice of gauge.

In perturbation theory a gauge transformation is an infinitesimal coordinate transformation of the perturbed spacetime

$$x_{\text{new}}^a = x_{\text{old}}^a + \xi^a, \quad \text{where } \xi^a = O(h), \quad (73)$$

and the coordinates x_{new}^a , x_{old}^a , and the coordinates on the abstract manifold are all described by the same names, for example (t, r, θ, ϕ) for perturbations of the Schwarzschild geometry. The transformation of Eq. 73 not only changes the components of a tensor by $O(h)$, in the usual way, but also changes the mapping between the two manifolds and hence changes the subtraction in Eq. 72. With the transformation (73),

$$h_{ab}^{\text{new}} = (g_{cd} + h_{cd}^{\text{old}}) \frac{\partial x_{\text{old}}^c}{\partial x_{\text{new}}^a} \frac{\partial x_{\text{old}}^d}{\partial x_{\text{new}}^b} - \left(g_{ab} + \xi^c \frac{\partial g_{ab}}{\partial x^c} \right). \quad (74)$$

The ξ^c in the last term accounts for the $O(h)$ change in the event of the background used in the subtraction. After an expansion, this provides a new description of h_{ab}

$$\begin{aligned} h_{ab}^{\text{new}} &= h_{ab}^{\text{old}} - g_{cb} \frac{\partial \xi^c}{\partial x^a} - g_{cb} \frac{\partial \xi^d}{\partial x^b} - \xi^c \frac{\partial g_{ab}}{\partial x^c} \\ &= h_{ab}^{\text{old}} - \mathcal{L}_\xi g_{ab} = h_{ab}^{\text{old}} - 2\nabla_{(a} \xi_{b)} \end{aligned} \quad (75)$$

through $O(h)$; the symbol \mathcal{L} represents the Lie derivative and ∇_a is the covariant derivative compatible with g_{ab} . A gauge transformation does not change the actual perturbed manifold, but it does change the coordinate description of the perturbed manifold.

A little clarity is revealed by noting that

$$E_{ab}(\nabla_{(c} \xi_{d)}) \equiv 0 \quad (76)$$

for any C^2 vector field ξ^a ; and if ξ^a has limited differentiability or is a distribution, then Eq. 76 holds in a distributional sense [25]. Thus $-2\nabla_{(a} \xi_{b)}$ is a homogeneous solution of the linear Eq. 35. It appears as though any $-2\nabla_{(a} \xi_{b)}$ may be added to an inhomogeneous solution of Eq. 35 to create a “new” inhomogeneous solution. In fact the new solution is physically indistinguishable from the old – they differ only by a gauge transformation with gauge vector ξ^a .

Generally, the four degrees of gauge freedom contained in the gauge vector ξ^a are used to impose four convenient conditions on h_{ab} . For perturbations of the Schwarzschild metric, it is common to use the Regge–Wheeler gauge which sets four independent parts of h_{ab} to zero; this results in some very convenient algebraic simplifications. The Lorenz gauge requires that $\nabla_a(h^{ab} - \frac{1}{2}g^{ab}h^c_c) = 0$ and is formally attractive but unwieldy in practice [1, 5, 27].

The Bianchi identity implies that there are four relations among the ten components of the Einstein equations. Choosing a gauge helps focus on a self-consistent method for solving a subset of these equations. A physicist might have a favorite for a gauge choice, but Nature has no preference whatsoever.

8 Gauge Confusion and the Gravitational Self-Force

If a particular physical consequence of the gravitational self-force requires a particular choice of gauge, then it is unlikely that this physical consequence has any useful interpretation. This was already demonstrated with the example presented in Section 2 where the magnitude of the effect of the Newtonian self-force on the period in an extreme-mass-ratio binary depended upon the definition of the variable r .

The quasi-circular orbits of the Schwarzschild geometry provide a fine example which reveals the insidious nature of gauge confusion in self-force analyses. Ref. [26] contains a thorough discussion of this subject and this section has two self-force examples which highlight the confusion that perturbative gauge freedom creates.

It is straightforward to determine the components of the geodesic equation for the metric $g_{ab}^{\text{Schw}} + h_{ab}^{\text{R}}$. A consequence of these is that the orbital frequency of m in a quasi-circular orbit about a Schwarzschild black hole of mass M is given by

$$\Omega^2 = \frac{M}{r^3} - \frac{r - 3M}{2r^2} u^a u^b \partial_r h_{ab}^{\text{R}} \quad (77)$$

which can be proven to be independent of the gauge choice. Clearly the self-force makes itself known to the orbital frequency through the last term. So we focus on the orbit at radius $r = 10M$, choose to work in the Lorenz gauge, work hard and successfully evaluate all of the components of the regularized field h_{ab}^{R} as well as its radial derivative. Then we calculate the second term in Eq. 77 and determine that Ω changes by a specific amount $\Delta\Omega_{\text{Lz}}$. We now know the gauge invariant change in the orbital frequency for m in the orbit at $10M$.

Or do we? To check this result we repeat the numerical work but this time use the Regge–Wheeler gauge, and find that the change in Ω is $\Delta\Omega_{\text{rw}}$ and

$$\Delta\Omega_{\text{rw}} \neq \Delta\Omega_{\text{Lz}} ! \quad (78)$$

What's going on? For a quasi circular orbit Ω can be *proven* to be independent of gauge, and yet with two different gauges we find two different orbital frequencies for the single orbit at $10M$.

When I first discovered this conundrum I was reminded of my experience trying to understand special relativity and believing that apparently paradoxical situations made special relativity logically inconsistent. Eventually the paradoxes vanished when I understood that coordinates named t , x , y and z are steeped in ambiguity and that only physical observables are worth calculating and discussing.

The resolution of this self-force confusion is similar. The two evaluations of Ω^2 are each correct. But, one is for the orbit at the Schwarzschild radial coordinate $r = 10M$ in the Lorenz gauge, while the other is at the Schwarzschild radial coordinate $r = 10M$ in the Regge–Wheeler gauge. These are two distinct orbits. In fact, the gauge vector ξ^a which transforms from the Lorenz gauge to the Regge–Wheeler gauge has a radial component ξ^r whose magnitude is just right to make the change in the first term in Eq. 77 balance the change in the second term.

The angular frequency of m orbiting a black hole is a physical observable and independent of any gauge choice. But the perturbed Schwarzschild geometry is not spherically symmetric and there is then no natural definition for a radial coordinate.

A second example of gauge confusion appears when one attempts to find the self-force effect on the rate of inspiral of a quasi-circular orbit of Schwarzschild. It is natural to find the energy E , Ω and dE/dt all as functions of the radius of a circular orbit and then to use

$$\frac{d\Omega}{dt} = \frac{dE}{dt} \times \frac{d\Omega/dr}{dE/dr} \quad (79)$$

to determine the rate of change of Ω . We can find the self force effect on each of these quantities so we can apparently find the self force effect on $d\Omega/dt$ which is a physical observable and must be gauge invariant.

This situation is subtle. Why do we believe Eq. 79? With some effort it can be shown that the geodesic equation for $g_{ab}^{\text{Schw}} + h_{ab}^{\text{R}}$ implies that Eq. 79 holds for a quasi-circular orbit [26]. Part of this proof depends upon the t -component of the geodesic equation which is

$$\frac{dE}{dt} = -\frac{1}{2u^t} u^a u^b \partial_t h_{ab}^{\text{R}}, \quad (80)$$

and this is a gravitational self-force effect. But, note that the right hand side of Eq. 79 is already first order in h_{ab}^{R} from the factor dE/dt . While self-force effects on $d\Omega/dr$ and dE/dr can be found, if these are included then second order self force effects on dE/dt must also be found for a consistent solution.

The end result is that you really can't see the effect of the conservative part of the self-force on the waveform for quasi-circular orbits using first order perturbation theory.

9 Steps in the Analysis of the Gravitational Self-Force

We now highlight the major steps involved in most gravitational self-force calculations.

First the metric perturbation h_{ab}^{act} is determined. For a problem in the geometry of the Schwarzschild metric, this involves solving the Regge–Wheeler [46] and the Zerilli [60] equations to determine the actual metric perturbations. The Kerr metric still presents some challenges. The Teukolsky [49, 51] formalism can provide the Weyl scalars but finding the metric perturbations [59] from these is difficult at best, and does not include the non-radiating monopole and dipole perturbations. One possibility for Kerr is to find the metric perturbations directly, perhaps in the Lorenz gauge, but this would likely require a 3+1 approach. Another possibility being

discussed [8] is to Fourier transform in ϕ , and then use a 2+1 formalism which results in an m -sum. Rotating black holes continue to be a challenge for self-force calculations.

Next, the singular field h_{ab}^S is identified for the appropriate geodesic in the background spacetime. A general expansion of the singular field is available [24], but it is not elementary to use.⁴ Work in progress provides a constructive procedure for the THZ coordinates in the neighborhood of a geodesic, and this would lead to explicit expressions for h_{ab}^S in the natural coordinates of the manifold. However, this procedure is not yet in print, and it is not yet clear how difficult it might be to implement.

Then the perturbation is regularized by subtracting the singular field from the actual field resulting in $h_{ab}^R = h_{ab}^{\text{act}} - h_{ab}^S$. Most applications have taken this step using the mode-sum regularization procedure of Barack and Ori [1, 3]. In this case, a mode-sum decomposition of the singular (or “direct,” cf. footnote 4) field is identified and then removed from the mode-sum decomposition of the actual field. The remainder is essentially the mode-sum decomposition of the regular field. Generally, this mode-sum converges slowly as a power law in the mode index, l or m . Although some techniques have been used to speed up this convergence [29]. More recently, “field regularization” (discussed in Section 10.2 and in [54]) has been used for scalar field self-force calculations. For this procedure in the gravitational case, Eq. 66 might be used to obtain the regular field h_{ab}^R directly via 3+1 analysis.

After the determination of h_{ab}^R , the effect of the gravitational self-force is then generically described as resulting in geodesic motion for m in the metric $g_{ab}^0 + h_{ab}^R$. This appears particularly straightforward to implement using field regularization. Alternatively, the motion might also be described as being accelerated by the gravitational self-force as described in Eq. 71.

At this point, one should be able to answer the original question – whatever that might have been! In fact, the original question should be given careful consideration *before* proceeding with the above steps. Formulating the question might be as difficult as answering it. It is useful to keep in mind that only physical observables and geometrical invariants can be defined in a manner independent of a choice of coordinates or a choice of perturbative gauge.

My prejudices about the above choices for each step are not well hidden. But, for whatever technique or framework is in use, a self-force calculation should have the focus trained upon a physical observable, not upon the method of analysis.

Self-force calculations unavoidably involve some subtlety. Experience leads me to be wary about putting trust in my own unconfirmed results. Good form requires independent means to check analyses. Comparisons with the previous work of others, with Newtonian and post-Newtonian analyses, or with other related analytic weak-field situations all lend credence to a result.

⁴ Expansions for the somewhat related “direct” field are also available [3, 4, 7, 37, 38, 43, 45], though their use is, similarly, not at all elementary.

10 Applications

Recently, the effect of the gravitational self-force on the orbital frequency of the innermost stable circular orbit of the Schwarzschild geometry has been reported by Barack and Sago [6]. They find that the self-force changes the orbital frequency of the ISCO by $0.4870(\pm 0.0006)m/M$. To date this result is by far the most interesting gravitational self-force problem that has been solved. But it is too recent a result to be described more fully herein.

10.1 Gravitational Self-Force Effects on Circular Orbits of the Schwarzschild Geometry

As an elementary example we consider a small mass m in a circular orbit about the Schwarzschild geometry. Details of this analysis may be found in [26]. The gravitational self-force affects both the orbital frequency Ω and also the Schwarzschild t -component of the four-velocity, u^t , which is related to a redshift measurement. The self-force effects on these quantities are known to be independent of the gauge choice for h_{ab} , as would be expected because they can each be determined by a physical measurement. However the radius of the orbit depends upon the gauge in use and has no meaning in terms of a physical measurement.

Notwithstanding the above, we define R_Ω via

$$\Omega^2 = M/R_\Omega^3 \quad (81)$$

as a natural radial measure of the orbit which inherits the property of gauge independence from Ω . The quantity u^t can be divided into two parts $u^t = {}_0u^t + {}_1u^t$, where each part is separately gauge independent. Further the functional relationships between Ω , ${}_0u^t$ and R_Ω are identical to their relationships in the geodesic limit,

$${}_0u^t = [1 - 3(\Omega M)^{2/3}] + O(m^2) \quad (82)$$

and shows no effect from the self-force. The remainder

$${}_1u^t = u^t - {}_0u^t \quad (83)$$

is a true consequence of the self-force, and we plot the numerically determined ${}_1u^t$ as a function of R_Ω in Fig. 1. The numerical data of Fig. 1 have also been carefully compared with and seen to be in agreement with the numerical results of Sago and Barack, as shown in [48], despite the fact that very different gauges were in use and different numerical methods were employed.

We have derived a post-Newtonian expansion for ${}_1u^t$ based upon the work of others [10, 11]. Our expansion is in powers of m/R_Ω , which is v^2/c^2 in the Newtonian limit, and we find

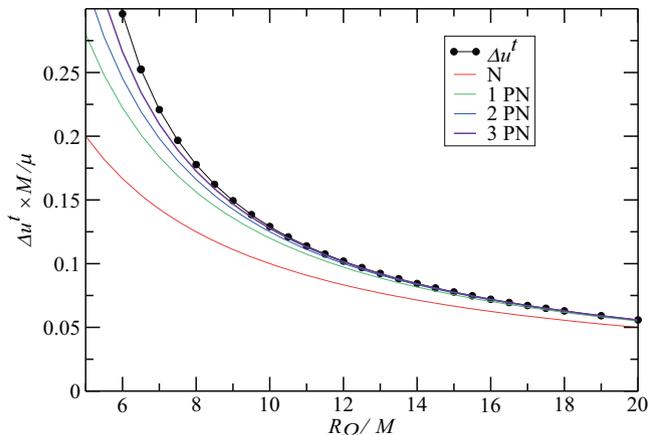


Fig. 1 From [26]. The quantity ${}_1u^t$, which is the gauge independent $O(m)$ part of u^t , is given as a function of R_Ω for circular orbits in the Schwarzschild geometry. Also shown are u^t_1 as calculated with Newtonian, 1PN and 2PN analyses in [26] based upon results in [11] and [10]. The 3PN line is based on a numerical determination of the 3PN coefficient in Eq. 84 in [26]

$${}_1u^t = \frac{m}{M} \left[-\left(\frac{M}{R_\Omega}\right) - 2\left(\frac{M}{R_\Omega}\right)^2 - 5\left(\frac{M}{R_\Omega}\right)^3 + \dots \right], \quad (84)$$

which includes terms of order v^6/c^6 . Further, with numerical analysis we have fit these results to determine a 3PN parameter of order v^8/c^8 and found that the coefficient of the $(M/R_\Omega)^4$ term is -27.61 ± 0.03 .

Recent work, with Blanchet, Le Tiec, and Whiting [12, 13], includes a full 3PN determination of the same 3PN coefficient as well as a more precise numerical determination via self-force analyses involving an increased range of data. The consistency of these two efforts has the possibility of giving greatly increased confidence in the self-force numerical analysis as well as in the post-Newtonian analysis, each of which involves substantial complications.

This self-force result is primarily only of academic interest. But it is consistent with a post-Newtonian expansion and includes an estimate of the previously unknown $O(v^8/c^8)$ coefficient in the expansion. Modest though it might be, this is a result.

10.2 Field Regularization Via the Effective Source

The ultimate goal of self-force analysis has become the generation of accurate gravitational waveforms from extreme mass-ratio inspiral (EMRI). It would be amusing to “see” numerically the waves emitted by a small black hole in a highly eccentric orbit about a much larger one and to see the changes in the orbit while the small hole loses energy and angular momentum.

Such a project appears to require a method to solve for the gravitational waves while simultaneously modifying the worldline of the small hole as it responds to the gravitational self-force. The toy problem in Section 4 shows how this might be done using the expertise of numerical relativity groups coupled with the self-force community.

Our group is in the early stages of development of infrastructure that any numerical relativity group could use to get gravitational self-force projects up and running with a minimum of effort. We intend to provide the software that will produce the regularized-field source S_{ab} , for a small mass m as a function of location and four-velocity. A numerical relativist could then evolve the linear field equation

$$E_{ab}(h^R) = -8\pi S_{ab} \quad (85)$$

for h_{ab}^R , while simultaneously adjusting the worldline according to Eq. 71.

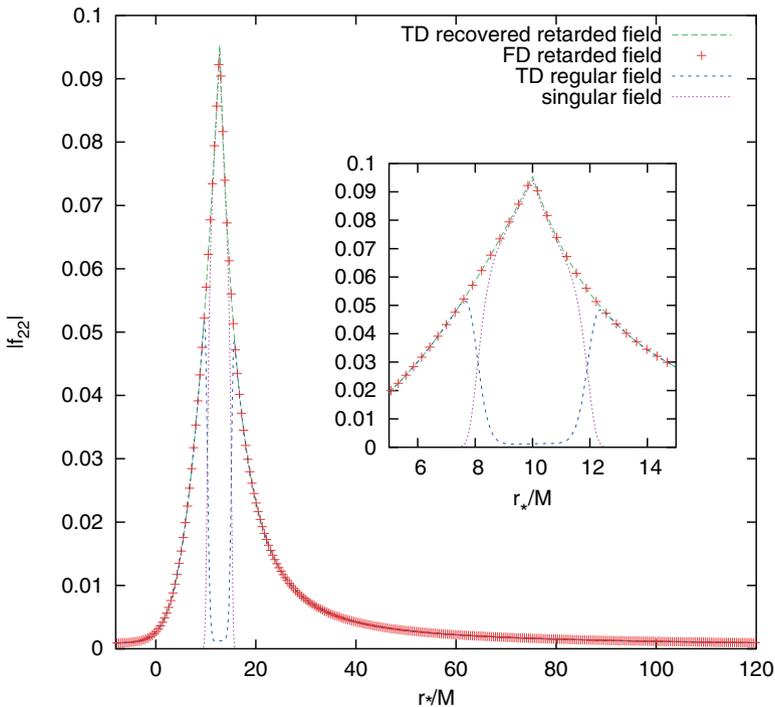


Fig. 2 From [54]. Comparison of time-domain (TD) and frequency-domain (FD) results for the $l = m = 2$ multipole moment of the scalar field. The regular field is represented by the blue dashed line. Adding this to the $l = m = 2$ multipole moment of the analytically known singular field, $W\psi^S$, results in the computed, actual field to good agreement. The inset shows near the point charge that ψ^R is very well behaved and that $\psi^R + W\psi^S$ is indistinguishable from the actual, retarded field ψ^{ret} , just as it should be

As described in Section 6 such a computation of h_{ab}^R would provide not only the effects of the gravitational self-force but also the gravitational wave itself.

Ian Vega [54] has led a first attempt at directly solving for the regularized field and self-force using a well tested problem involving a scalar charge in a circular orbit of the Schwarzschild geometry. This analysis used a multipole decomposition of the source and field. And Vega solved for the multipole components in the time domain using a 1+1 code. Figure 2 shows the $\ell = m = 2$ mode and compares the accurate frequency domain evaluation of the retarded field ψ^{ret} to the sum $\psi^S + \psi^R$ as determined using 1+1 methods with field-regularization as described in Section 4. Table 1 compares the numerical results of regularized fields and forces from the field-regularization approach of [54] with the mode-sum regularization procedure [3, 7] used in [29].

Figure 3 shows an example of the source-function used in a test of this approach with a scalar field. The “double bump” shape far from the charge is a characteristic of any function similar to $\nabla^2(W/|\mathbf{r}-\mathbf{r}_0|)$ with a window function W which satisfies the three window properties given in Section 4.2.

Table 1 From [54]. Summary of scalar field self-force results for a circular orbits at $R = 10M$ and $R = 12M$. The error is determined by a comparison with an accurate frequency-domain calculation [29]

	R	Time-domain	Frequency-domain	Error
$\partial_t \psi^R$	$10M$	3.750211×10^{-5}	3.750227×10^{-5}	0.000431%
$\partial_r \psi^R$	$10M$	1.380612×10^{-5}	1.378448×10^{-5}	0.157%
$\partial_t \psi^R$	$12M$	1.747278×10^{-5}	1.747254×10^{-5}	0.00139%
$\partial_r \psi^R$	$12M$	5.715982×10^{-6}	5.710205×10^{-6}	0.101%

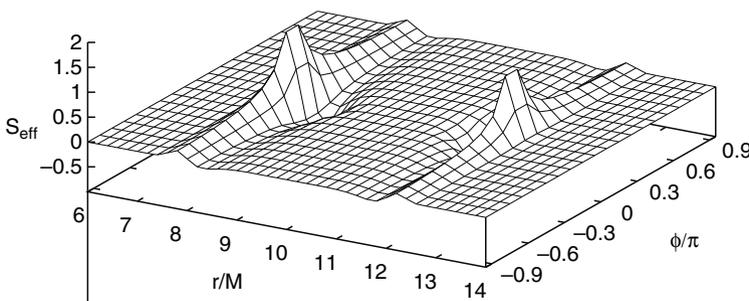


Fig. 3 From [54]. The effective source S_{eff} on the equatorial plane for a scalar charge in a circular orbit of the Schwarzschild metric. The particle is at $r/M = 10$, $\phi/\pi = 0$, where S_{eff} appears to have no structure on this scale. The spiky appearance is solely a consequence of the grid resolution of the figure. In fact the source is C^∞ everywhere except at the location of the scalar charge where S_{eff} appears quite calm on this scale

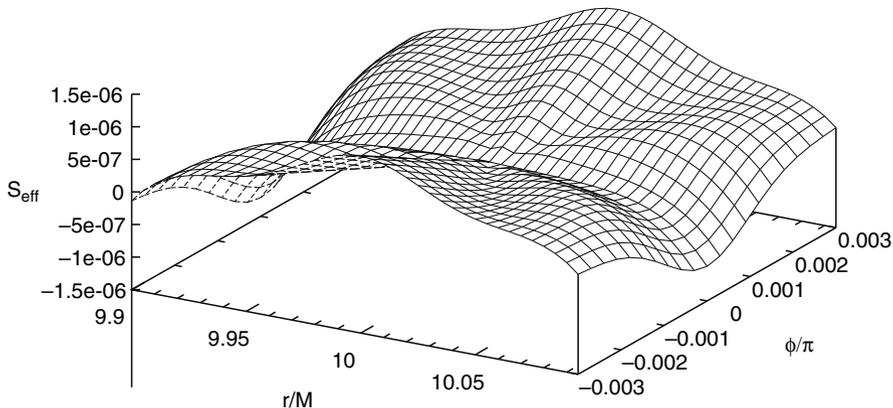


Fig. 4 From [54]. The effective source S_{eff} in the equatorial plane in the vicinity of the point source at $r/M = 10$, $\phi/\pi = 0$. Note the significant difference of scales with Fig. 3

Figure 4 reveals the C^0 nature of the effective source at the location of the particle on a dramatically different scale. It is important to note that limited differentiability of this sort does not introduce a small length scale into the numerical problem, and might be treated via a special stencil in the neighborhood of the charge.

A recent collaboration with Peter Diener, Wolfgang Tichy and Ian Vega [55] looks at the same test problem but involves two distinct 3+1 codes, which were developed completely independently. One uses pseudo-spectral methods, the other uses a multiblock code with high order matching across block boundaries. With a modest amount of effort these two codes, each developed for generic numerical relativity problems, were modified to accommodate the effective source of the scalar field and are able to determine all components of the effective source with errors less than 1%. The future of numerical 3+1 self-force analysis looks promising.

11 Concluding Remarks

Ptolemy was able to model accurately the motion of the planets in terms of epicycles and circles about the Earth. However, the precise choice of which circles and epicycles should be used was debated. Copernicus realized that a much cleaner description resulted from having the motion centered upon the sun. The two competing models were equally able to predict the positions of the planets for the important task of constructing horoscopes. But for understanding the laws of physics, Newton clearly favored the Copernican model.

There appear to be two rather distinct attitudes toward calculating the effects of the gravitational self-force for a mass m orbiting a black hole. Both lead to identical conclusions about physically measurable quantities. If the motion is to be described as accelerating in the black hole geometry, then the acceleration depends upon the perturbative gauge choice and is not related to any acceleration that an observer

local to m could actually measure. If the motion is described as geodesic in the spacetime geometry through which m moves, then it is immediately apparent that the only quantities worth calculating are those which are physically measurable, or at least independent of the gauge choice. With this second attitude, one is left with the rather satisfying perspective that the effects of the gravitational self-force are neither more nor less than the result of free-fall in a gravitational field.

In this review, I have eschewed mention of Green's functions. The asymptotic matching perspective promoted here seems more effective to me at getting to the physics of the gravitational self-force and less likely to lead to mathematical confusion.

The singular field h_{ab}^S , which plays a fundamental role, has a reasonably straightforward description in convenient locally inertial coordinates. And it appears nearly immediately in the DW [28] formulation of radiation reaction via the Green's function $G_{abc'd'}^S$.⁵ This Green's function has odd acausal structure with support on the past and future null cone of the field point and also in the spacelike related region outside these null cones. Such causal structure is consistent with the fact that h_{ab}^S exerts no self-force. Based upon personal conversations, this feature appears problematical to some. However, the integrability condition of the perturbed Einstein equation requires that the worldline of a point source be a geodesic. Geodesic motion is the General Relativistic equivalent of Newtonian no motion, and the singular field is the curved space equivalent of a Coulomb field. Not much is happening at the source or to the singular field. I cannot imagine that such behavior somehow leads to an effect that might be described as acausal.

The S-field h_{ab}^S is defined via an expansion in a neighborhood of the source and does not depend upon boundary conditions, and the restriction to geodesic motion precludes any unexpected behavior of the point mass in either the past or the future. The S-field is precisely the nearly-Newtonian monopole field with minor tidal distortions from the surrounding spacetime geometry.

While orbiting a black hole, Einstein's apple emits gravitational waves and spirals inward. However, the apple is in free fall and not accelerating. In fact, it is not moving in its locally inertial frame of reference, and is aware of neither its role as the source of any radiation nor of its role acting out the effects of radiation reaction.

S. Chandrasekhar was fond of describing a conversation with the sculptor Henry Moore. In his own words, Chandra "had the occasion to ask Henry Moore how one should view sculptures: from afar or from near by. Moore's response was that the greatest sculptures can be viewed – indeed should be viewed – from all distances since new aspects of beauty will be revealed at every scale" [17]. The self-force analysis in General Relativity also reveals different aspects when viewed from afar and when viewed from near by. From afar a small black hole dramatically emits gravitational waves while inspiralling toward a much larger black hole. From

⁵ In fact the singular field was discovered first [24] using matched asymptotic expansions. And the Green's function appeared only later during an attempt to show consistency with the usual DeWitt-Brehme [30] approach to radiation reaction.

near by the small hole reveals the quiet simplicity and grace of geodesic motion. Rather than “beauty,” a satisfying sense of physical consistency is “revealed at every scale.”

Acknowledgements My understanding of gravitational self-force effects has evolved over the past decade in large part in discussions with colleagues during the annual Capra meetings. I am deeply indebted to the organizers and participants of these fruitful meetings. And I am particularly pleased to have had recent collaborators Leor Barack, Peter Diener, Eric Poisson, Norichika Sago, Wolfgang Tichy, Ian Vega, and Bernard Whiting, who individually and as a group have kept me on track and moving forward. This work was supported in part by the National Science Foundation, through grant number PHY-0555484 with the University of Florida. Some of the numerical results described here were performed at the University of Florida High-Performance Computing Center (URL: <http://hpc.ufl.edu>).

References

1. L. Barack, Phys. Rev. D **64**, 084021 (2001)
2. L. Barack, A. Ori, Phys. Rev. D **64**, 124003 (2001)
3. L. Barack, A. Ori, Phys. Rev. D **66**, 084022 (2002)
4. L. Barack, A. Ori, Phys. Rev. D **67**, 024029 (2003)
5. L. Barack, N. Sago, Phys. Rev. D **75**, 064021 (2007)
6. L. Barack, N. Sago, Phys. Rev. Lett. **102**, 191101 (2009)
7. L. Barack, Y. Mino, H. Nakano, A. Ori, M. Sasaki, Phys. Rev. Lett. **88**, 091101 (2002)
8. L. Barack, D.A. Golbourn, N. Sago, Phys. Rev. D **76**, 124036 (2007)
9. J.M. Bardeen, Phys. Rev. D **22**, 1882 (1980)
10. L. Blanchet, Living Rev. Rel. **9**, URL: <http://www.livingreviews.org/lrr-2006-4>
11. L. Blanchet, G. Faye, B. Ponsot, Phys. Rev. D **58**, 124002 (1998)
12. L. Blanchet, S. Detweiler, A. Le Tiec, B.F. Whiting, Phys. Rev. D **81**, 064004 (2010)
13. L. Blanchet, S. Detweiler, A. Le Tiec, B.F. Whiting, Phys. Rev. D **81**, 084033 (2010)
14. B. Carter, Comm. Math. Phys. **10**, 280 (1968)
15. B. Carter, Phys. Rev. **174**, 1559 (1968)
16. S. Chandrasekhar, *Hydrodynamic and Hydromagnetic Stability*, International Series of Monographs on Physics, ed. by W. Marshall, D.H. Wilkinson (Oxford University at the Clarendon Press, Oxford, 1961; reprinted by Dover, New York, 1981)
17. S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Oxford University Press, Oxford, 1983)
18. S. Chandrasekhar, S. Detweiler, Proc. R. Soc. Lond. **344**, 441 (1975)
19. C.T. Cunningham, R.H. Price, V. Moncrief, Astrophys. J. **224**, 643 (1979)
20. C.T. Cunningham, R.H. Price, V. Moncrief, Astrophys. J. **230**, 870 (1979)
21. M. Davis, R. Ruffini, W.H. Press, R. Price, Phys. Rev. Lett. **27**, 1466 (1971)
22. S. Detweiler, in *Sources of Gravitational Radiation*, ed. by L. Smarr (Cambridge University Press, Cambridge, 1979), pp. 211–230
23. S. Detweiler, Astrophys. J. **239**, 292 (1980)
24. S. Detweiler, Phys. Rev. Lett. **86**, 1931 (2001)
25. S. Detweiler, Class. Q. Grav. **22**, S681 (2005)
26. S. Detweiler, Phys. Rev. D **77**, 124026 (2008)
27. S. Detweiler, E. Poisson, Phys. Rev. D **69**, 084019 (2004)
28. S. Detweiler, B.F. Whiting, Phys. Rev. D **67**, 024025 (2003)
29. S. Detweiler, E. Messaritaki, B.F. Whiting, Phys. Rev. D **67**, 104016 (2003)
30. B.S. DeWitt, R.W. Brehme, Ann. Phys. (N.Y.) **9**, 220 (1960)
31. P.A.M. Dirac, Proc. R. Soc. Lond. A **167**, 148 (1938)

32. J.D. Jackson, *Classical Electrodynamics*, 3rd edn. (Wiley, New York, 1998)
33. L.D. Landau, E.M. Lifshitz, *Classical Theory of Fields*, 4th edn. (Pergamon, Oxford, 1975)
34. E.W. Leaver, Proc. R. Soc. Lond. A **402**, 285 (1985)
35. E.W. Leaver, Phys. Rev. D **34**, 384 (1986)
36. Y. Mino, Phys. Rev. D **67**, 084027 (2003)
37. Y. Mino, H. Nakano, M. Sasaki, Prog. Theor. Phys. **108**, 1039 (2002)
38. Y. Mino, M. Sasaki, T. Tanaka, Phys. Rev. D **55**, 3457 (1997)
39. C.W. Misner, K.S. Thorne, J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973)
40. E. Poisson, Living Rev. Rel. **7**, URL: <http://www.livingreviews.org/lrr-2004-6>
41. E. Poisson, Phys. Rev. D **69**, 084007 (2004)
42. E. Poisson, Phys. Rev. D **70**, 084044 (2004)
43. E. Poisson, Phys. Rev. Lett. **94**, 161103 (2005)
44. W.H. Press, Astrophys. J. **170**, L105 (1971)
45. T.C. Quinn, R.M. Wald, Phys. Rev. D **56**, 3381 (1997)
46. T. Regge, J.A. Wheeler, Phys. Rev. **108**, 1063 (1957)
47. R. Sachs, in *Relativity, Groups and Topology*, ed. by B. DeWitt, C. DeWitt (Gordon and Breach, New York, 1964)
48. N. Sago, L. Barack, S. Detweiler, Phys. Rev. D **78**, 124024 (2008)
49. M. Sasaki, T. Nakamura, Prog. Theor. Phys. **67**, 1788 (1982)
50. J.M. Stewart, M. Walker, Proc. R. Soc. Lond. **341**, 49 (1974)
51. S. Teukolsky, Astrophys. J. **185**, 635 (1973)
52. K.S. Thorne, J.B. Hartle, Phys. Rev. D **31**, 1815 (1985)
53. K.S. Thorne, S.J. Kovács, Astrophys. J. **200**, 245 (1975)
54. I. Vega, S. Detweiler, Phys. Rev. D **77**, 084008 (2008)
55. I. Vega, P. Diener, W. Tichy, S. Detweiler, Phys. Rev. D **80**, 084021 (2009)
56. C.V. Vishveshwara, Phys. Rev. D **1**, 2870 (1970)
57. S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972)
58. B.F. Whiting, J. Math. Phys. **30**, 1301 (1989)
59. B.F. Whiting, L.R. Price, Class. Q. Grav. **22**, S589 (2005)
60. F.J. Zerilli, Phys. Rev. D **2**, 2141 (1970)
61. X.-H. Zhang, Phys. Rev. D **31**, 3130 (1985)
62. X.-H. Zhang, Phys. Rev. D **34**, 991 (1986)

Constructing the Self-Force

Eric Poisson

Abstract I present an overview of the methods involved in the computation of the scalar, electromagnetic, and gravitational self-forces acting on a point particle moving in a curved spacetime. For simplicity, the focus here will be on the scalar self-force. The lecture follows closely my review article on this subject [E. Poisson, *Living Rev. Relativ.* **7** (2004), <http://www.livingreviews.org/lrr-2004-6>]. I begin with a review of geometrical elements (Synge's world function, the parallel propagator). Next I introduce useful coordinate systems (Fermi normal coordinates and retarded light-cone coordinates) in a neighborhood of the particle's world line. I then present the wave equation for a scalar field in curved spacetime and the equations of motion for a particle endowed with a scalar charge. The wave equation is solved by means of a Green's function, and the self-force is constructed from the field gradient. Because the retarded field is singular on the world line, the self-force must involve a regularized version of the field gradient, and I describe how the regular piece of the self-field can be identified. In the penultimate section of the lecture I put the construction of the self-force on a sophisticated axiomatic basis, and in the concluding section I explain how one can do better by abandoning the dangerous fiction of a point particle.

1 Introduction

We consider a point particle moving on a world line γ in a curved spacetime with metric $g_{\alpha\beta}$. The particle is either endowed with a scalar charge q or an electric charge e , and we wish to calculate the effect of these charges on the motion of the particle. This motion is not geodesic, because the (scalar or electromagnetic) field created by the particle interacts with the particle and causes it to accelerate. In flat spacetime this effect is produced by a local distortion of the field lines associated

E. Poisson (✉)

Department of Physics, University of Guelph, Guelph, Ontario, N1G 2W1, Canada

e-mail: poisson@physics.uoguelph.ca

with the particle's acceleration. In curved spacetime there is no such local distortion when the particle moves freely; what happens instead is that the field interacts with the spacetime curvature and back-scatters toward the particle. As far as the particle is concerned, then, it interacts with an incoming wave, and the motion is not geodesic.

There is a gravitational analogue to these (scalar and electromagnetic) situations: Even in the absence of charges q and e , we may wish to go beyond the test-mass description and consider the effect of the particle's mass m on its motion. There are two ways of describing this effect. We might say that the particle moves on a *geodesic* in a *perturbed spacetime* with metric $g_{\alpha\beta} + h_{\alpha\beta}$. Or we might say that the particle moves on an *accelerated world line* in the *original spacetime* with metric $g_{\alpha\beta}$. It is useful to keep both points of view active, and most researchers working in this field go freely back and forth between these modes of description. In the second view, the particle's acceleration is associated with the perturbation $h_{\alpha\beta}$, which produces a gravitational self-force acting on the particle.

The scalar, electromagnetic, and gravitational self-force problems share many physical and mathematical features. In each case a moving charge is accompanied by a field: q produces a scalar field Φ , e produces an electromagnetic field A_α , and m produces a gravitational perturbation $h_{\alpha\beta}$. In each case the field satisfies a linear wave equation in the background spacetime: Φ satisfies a scalar wave equation, A_α a vectorial equation derived from Maxwell's equations, and $h_{\alpha\beta}$ a tensorial equation derived from the linearized Einstein equations. And in each case the self-force is equal to the gradient of the field evaluated on the particle's world line.

This last observation reveals the problematic nature of this investigation. It is straightforward enough to calculate the scalar field Φ , the vector potential A_α , and the gravitational perturbation $h_{\alpha\beta}$ at a distance from the particle. But because the particle is pointlike, these quantities diverge on the world line, and derivatives of these quantities are even more singular. How is one supposed to deal with these singular expressions and extract from them the finite pieces that produce a well-defined effect, namely, the self-force acting on the particle?

The purpose of this lecture is to offer some elements of answer to this question. My focus will be on the technical aspects of the problem, which I will try to describe without going overboard with derivations and mathematical precision. The preceding lecture by Bob Wald offers more on the conceptual aspects of the self-force, and other contributions describe ways of computing the self-force and its consequences. This lecture can be considered to be a light introduction to my massive review article published in *Living Reviews in Relativity* [9], to which I will frequently refer (as LRR); there the reader will find all the gory details of all the derivations omitted in the lecture.

For simplicity I will focus on the simplest exemplar of a self-force: the *scalar self-force* produced by a scalar field Φ on a scalar charge q . Understanding the details of this construction is a first step toward understanding the nature of the electromagnetic and gravitational self-forces; the computations involved are simpler, but the conceptual basis is essentially the same.

The self-force has a long history in theoretical physics, which is nicely summarized in a book by Herbert Spohn [12]. The standard reference for the elec-

tromagnetic self-force in flat spacetime is Dirac’s famous 1938 paper [4]. Dirac’s construction was generalized to curved spacetime in 1960 by DeWitt and Brehme [3]; a technical error in their work was corrected by Hobbs [7]. The gravitational self-force was first computed in 1997 by Mino, Sasaki, and Tanaka [8]; a more direct derivation (based on an axiomatic approach) was later provided by Quinn and Wald [11]. Finally, the scalar self-force – the main topic of this lecture – was first constructed in 2000 by Ted Quinn [10].

2 Geometric Elements

The construction of the self-force would be impossible without the introduction of geometric tools that were first fashioned by Synge [13] and independently by DeWitt and Brehme [3]. In this section I introduce the world function $\sigma(x, x')$ and the parallel propagator $g^\alpha_{\alpha'}(x, x')$.

Let x and x' be two points in spacetime, and let us assume that they are sufficiently close that there is a unique geodesic segment β linking them. The segment is described by the parametric relations $z^\mu(\lambda)$, in which λ is an affine parameter that runs from 0 to 1; we have that $z(0) = x'$ and $z(1) = x$. The vector $t^\mu := dz^\mu/d\lambda$ is tangent to β (Fig. 1).

The *world function* is defined by

$$\sigma(x, x') := \frac{1}{2} \int_0^1 g_{\mu\nu}(z) t^\mu t^\nu d\lambda. \tag{1}$$

It is numerically equal to half the squared geodesic distance between x and x' . When $\sigma < 0$ the separation between x and x' is timelike, and when $\sigma > 0$ it is spacelike. The equation $\sigma(x, x') = 0$ describes the light cones of each point; if x' is kept fixed then $\sigma(x) = 0$ describes the past and future light cones of x' ; if instead x is kept fixed then $\sigma(x') = 0$ describes the past and future light cones of x .

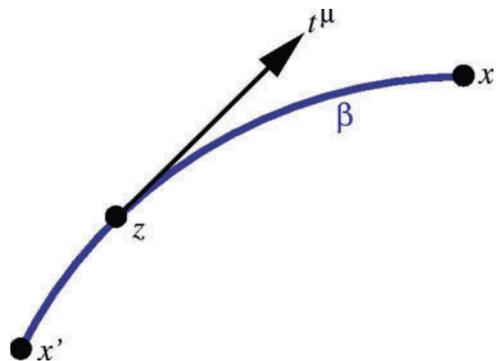


Fig. 1 Geodesic segment β between the spacetime points x' and x . The vector t^μ is tangent to this geodesic

The world function can be differentiated with respect to each argument. It can be shown (LRR Section 2.1.2) that

$$-\sigma^{\alpha'} := -\nabla^{\alpha'}\sigma(x, x') \quad (2)$$

is a vector at x' that is proportional to t^μ evaluated at that point. I use the convention that primed indices refer to the point x' , while unprimed indices refer to x . It can also be shown that its length is given by

$$g_{\alpha'\beta'}\sigma^{\alpha'}\sigma^{\beta'} = 2\sigma; \quad (3)$$

the length of $\sigma^{\alpha'}$ therefore measures the geodesic distance between the two points. Because the vector points from x' to x , we have a covariant notion of a displacement vector between the two points.

The *parallel propagator* takes a vector $A^{\alpha'}$ at x' and moves it to x by parallel transport on the geodesic segment β . We express this operation as

$$A^\alpha(x) = g_{\alpha'}^\alpha(x, x')A^{\alpha'}(x'), \quad (4)$$

in which A^α is the resulting vector at x . The operation is easily generalized to dual vectors and other types of tensors (LRR Section 2.3).

The world function and the parallel propagator can be employed in the construction of a Taylor expansion of a tensor about a reference point x' . Suppose that we have a tensor field $A^{\alpha\beta}(x)$ and that we wish to express it as an expansion in powers of the displacement away from x' . The role of the deviation vector is played by $-\sigma^{\alpha'}$, and the expansion coefficients will be ordinary tensors at x' . We might write something like

$$A^{\alpha'\beta'} + A^{\alpha'\beta'}_{\gamma'}(-\sigma^{\gamma'}) + \frac{1}{2}A^{\alpha'\beta'}_{\gamma'\delta'}(-\sigma^{\gamma'})(-\sigma^{\delta'}) + \dots,$$

but this defines a tensor at x' , not x . To get a proper expression for $A^{\alpha\beta}(x)$ we must also involve the parallel propagator, and we write

$$A^{\alpha\beta} = g_{\alpha'}^\alpha g_{\beta'}^\beta \left[A^{\alpha'\beta'} - A^{\alpha'\beta'}_{\gamma'}\sigma^{\gamma'} + \frac{1}{2}A^{\alpha'\beta'}_{\gamma'\delta'}\sigma^{\gamma'}\sigma^{\delta'} + \dots \right]. \quad (5)$$

Having postulated this form for the expansion, the expansion coefficients $A^{\alpha'\beta'}$, $A^{\alpha'\beta'}_{\gamma'}$, and so on can be computed by repeatedly differentiating the tensor field and evaluating the results in the limit $x \rightarrow x'$ (see LRR Section 2.4). For example, $A^{\alpha'\beta'} = \lim A^{\alpha\beta}$, as we might expect.

3 Coordinate Systems

Self-force computations are best carried out using covariant methods. It is convenient, however, to display the results in a coordinate system that is well suited to the description of a neighborhood of the world line γ . In this transcription

it is advantageous to keep the coordinates in a close correspondence with the geometric objects (such as $\sigma^{\alpha'}$) that appear in the covariant expressions. I find that two coordinate systems are particularly useful in this context: the *Fermi normal coordinates* ($t, x^a = s\omega^a$), and the *retarded null-cone coordinates* ($u, x^a = r\Omega^a$).

A third coordinate system, known as the *Thorne–Hartle–Zhang coordinates* [14, 15], has also appeared in the self-force literature – they are the favored choice of the Florida group led by Steve Detweiler and Bernard Whiting (see, e.g., Ref. [1]). The THZ coordinates are a variant of the Fermi coordinates, and they have some nice properties. But I find them less convenient to deal with than the Fermi or retarded coordinates, because they do not seem to possess a simple covariant definition. (The THZ coordinates may enjoy the mild Florida winters, but they are not robust enough to endure the tougher Canadian winters.) I shall not discuss the THZ coordinates here.

The Fermi and retarded coordinates share a basic geometrical construction on the world line. At each point on γ we erect a basis (u^μ, e_a^μ) of orthonormal vectors. The timelike vector u^μ is the particle’s velocity vector, and it is tangent to the world line. The spatial unit vectors e_a^μ are labeled with the index $a = 1, 2, 3$, and they are all orthogonal to u^μ ; they are also mutually orthogonal. The vectors are transported on γ so as to preserve their orthonormality properties. If the world line is a geodesic, then we might take the vectors e_a^μ to be parallel transported on γ . If instead the world line is accelerated, then we might take the spatial vectors to be Fermi–Walker transported on the world line (LRR Section 3.2.1). The tetrad of basis vectors satisfies the completeness relation

$$g^{\mu\nu} = -u^\mu u^\nu + \delta^{ab} e_a^\mu e_b^\nu, \tag{6}$$

which holds at any point on the world line.

Any tensor that is evaluated on γ can be decomposed in the basis (u^μ, e_a^μ) . For example, we might introduce the *frame components* of the Riemann tensor (Figs. 2 and 3),

$$R_{0a0b}(\tau) := R_{\mu\alpha\nu\beta} \Big|_\gamma u^\mu e_a^\alpha u^\nu e_b^\beta, \tag{7a}$$

$$R_{0abc}(\tau) := R_{\mu\alpha\beta\gamma} \Big|_\gamma u^\mu e_a^\alpha e_b^\beta e_c^\gamma, \tag{7b}$$

$$R_{abcd}(\tau) := R_{\alpha\beta\gamma\delta} \Big|_\gamma e_a^\alpha e_b^\beta e_c^\gamma e_d^\delta. \tag{7c}$$

They are functions of proper time τ on the world line.

The Fermi coordinates ($t, x^a = s\omega^a$) are constructed as follows (LRR Section 3.2). We select a point x in a neighborhood of the world line, and we locate the unique geodesic segment β that originates at x and intersects γ orthogonally. The intersection point is labeled \bar{x} , and t is the value of the proper-time parameter at this point: $\bar{x} = z(\tau = t)$. This defines the time coordinate t of the point x . The spatial coordinates x^a are defined by

$$x^a := -e_{\bar{a}}^a(\bar{x})\sigma^{\bar{a}}(x, \bar{x}); \tag{8}$$

Fig. 2 A tetrad of basis vectors on the world line γ . The unit timelike vector u^μ is tangent to the world line. The unit spatial vectors e_a^μ are mutually orthogonal and also orthogonal to u^μ

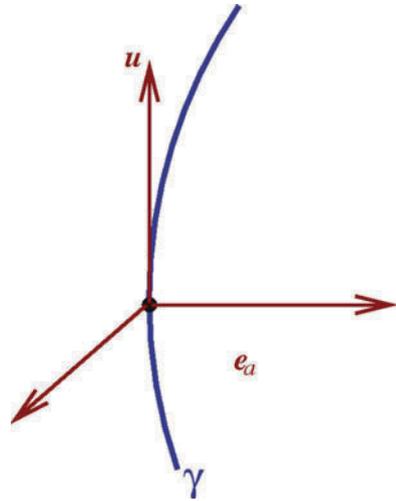
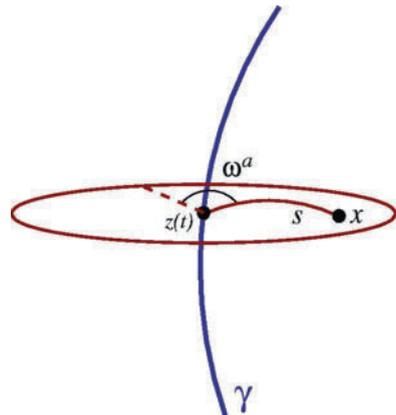


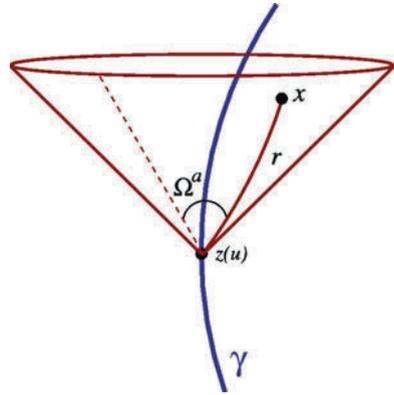
Fig. 3 Fermi coordinates



they are the projections in the basis $e_a^{\bar{\alpha}}$ of the deviation vector $-\sigma^{\bar{\alpha}}(x, \bar{x})$ between the points x and \bar{x} . The Fermi coordinates come with the condition $\sigma_{\bar{\alpha}}(x, \bar{x})u^{\bar{\alpha}}(\bar{x})=0$, which states that the deviation vector is orthogonal to the world line's tangent vector; it is this condition that identifies the intersection point $\bar{x} = z(t)$.

It is useful to introduce s as the proper distance between x and the world line. This is formally defined by $s^2 := 2\sigma(x, \bar{x})$, and it is easy to involve the completeness relation and show that $s^2 = \delta_{ab}x^ax^b$ (LRR Section 3.2.3); the Fermi distance s is therefore the usual Euclidean distance associated with the quasi-Cartesian coordinates x^a . It is also useful to introduce the direction cosines $\omega^a := x^a/s$; these quantities satisfy $\delta_{ab}\omega^a\omega^b = 1$, and they can be thought of as a radial unit vector that points away from the world line. Each hypersurface $t = \text{constant}$ is orthogonal to the world line (in the sense described above), and

Fig. 4 Retarded coordinates



each spacetime point x within the surface can be said to be simultaneous with $\bar{x} = z(t)$. The Fermi coordinates therefore provide a convenient notion of *rest frame* for the particle (Fig. 4).

The retarded coordinates $(u, x^a = r\Omega^a)$ are constructed as follows (LRR Section 3.3). Once more we select a point x in a neighborhood of the world line, but this time we locate the unique *null geodesic segment* β that originates at x and travels backward in time toward γ . The new intersection point is labeled x' , and u is the value of the proper-time parameter at this point: $x' = z(\tau = u)$. This defines the time coordinate u of the point x . The spatial coordinates x^a are defined exactly as before, by

$$x^a := -e_{\alpha'}^a(x')\sigma^{\alpha'}(x, x'). \tag{9}$$

The retarded coordinates come with the condition $\sigma(x, x') = 0$, which states that the points x and $x' = z(u)$ are linked by a null geodesic (which travels forward in time from x' to x).

It is useful to introduce r as a measure of light-cone distance between x and the world line. This is formally defined by $r := \sigma_{\alpha'}(x, x')u^{\alpha'}(x')$, and can be shown to be an affine parameter on the null geodesic that links x to x' (LRR Section 3.3.3). In addition, we have that $r^2 = \delta_{ab}x^ax^b$, and r is the usual Euclidean distance associated with the spatial coordinates x^a . It is also useful to introduce the direction cosines $\Omega^a := x^a/r$, which again play the role of a unit radial vector that points away from the world line.

Each hypersurface $u = \text{constant}$ is the future light cone of the point $z(u)$ on the world line. Any point x on this light cone is in direct causal contact with $z(u)$. For this reason the retarded coordinates give the simplest description of the scalar field Φ produced by a point charge q moving on the world line. The field satisfies a wave equation, and the radiation produced by the field essentially realizes the light cones that are so prominently featured in the construction of the coordinates.

4 Field Equation and Particle Motion

Let me recapitulate the problem that we wish to solve. We have a point particle of mass m and scalar charge q moving on a world line γ described by the parametric relations $z^\mu(\tau)$. The particle creates a scalar field $\Phi(x)$ and this field acts back on the particle and produces a force F_{self}^α . We wish to determine this self-force.

The scalar field obeys the linear wave equation

$$\square\Phi = -4\pi\mu, \quad (10)$$

where $\square := g^{\alpha\beta}\nabla_\alpha\nabla_\beta$ is the wave operator, and

$$\mu(x) = q \int_\gamma \delta_4(x, z(\tau)) d\tau \quad (11)$$

is the scalar-charge density, expressed as an integral over the world line. The four-dimensional delta function $\delta_4(x, z)$ is defined as a scalar quantity; it is normalized by $\int \delta_4(x, x') \sqrt{-g} d^4x = 1$.

The particle moves according to

$$m(\tau)a^\mu = q(g^{\mu\nu} + u^\mu u^\nu)\Phi_\nu, \quad (12)$$

where $a^\mu = Du^\mu/d\tau$ is the covariant acceleration and $\Phi_\nu := \nabla_\nu\Phi$ is the field gradient. The presence of the projector $g^{\mu\nu} + u^\mu u^\nu$ on the right-hand side ensures that the acceleration is orthogonal to the velocity. The field gradient, however, also has a component in the direction of the world line, and this produces a change in the particle's rest mass (LRR Section 5.1.1): $dm/d\tau = -qu^\mu\Phi_\mu$. We shall not be concerned with this effect here. Suffice it to say that the mass is not conserved because the scalar field can radiate monopole waves, which is impossible for electromagnetic and gravitational radiation.

The wave equation for Φ can be integrated, as we shall do in the following section, and the solution can be examined near the world line. Not surprisingly, the field is singular on the world line. This property makes the equation of motion meaningless as it stands, and we shall have to make sense of it in the course of our analysis.

5 Retarded Green's Function

The wave equation for Φ is solved by means of a *Green's function* $G(x, x')$ that satisfies

$$\square G(x, x') = -4\pi\delta_4(x, x'). \quad (13)$$

The solution is simply

$$\Phi(x) = \int G(x, x')\mu(x') \sqrt{-g'} d^4x' = q \int_\gamma G(x, z) d\tau, \quad (14)$$

and the difficulty of solving the wave equation has been transferred to the difficulty of computing the Green's function. We wish to construct the *retarded solution* to the wave equation, and this is accomplished by selecting the retarded Green's function $G_{\text{ret}}(x, x')$ among all the solutions to Green's equation. (Other choices will be considered below.) The retarded Green's function possesses the important property that it vanishes when the source point x' is in the future of the field point x . This ensures that $\Phi(x)$ depends on the past behavior of the source μ , but not on its future behavior.

The retarded Green's function is known to exist globally as a distribution if the spacetime is globally hyperbolic. But knowledge of the Green's function is required only in the immediate vicinity of the world line, so as to identify the behavior of Φ there; we shall not be concerned with the behavior of the Green's function when x and $z(\tau)$ are widely separated.

In this context the Green's function can be shown (LRR Section 4.3) to admit a *Hadamard decomposition* of the form

$$G_{\text{ret}}(x, x') = U(x, x')\delta_{\text{future}}(\sigma) + V(x, x')\Theta_{\text{future}}(-\sigma), \quad (15)$$

where $\sigma(x, x')$ is the world function introduced previously, and the two-point functions $U(x, x')$ and $V(x, x')$ are smooth when $x \rightarrow x'$. The retarded Green's function is not smooth in this limit, however, as we can see from the presence of the delta and theta functions. The first term involves $\delta_{\text{future}}(\sigma)$, the restriction of $\delta(\sigma)$ on the future light cone of the source point x' . The delta function is active when $\sigma(x, x') = 0$, and this describes (for fixed x') the future and past light cones of x' . We then eliminate the past branch of the light cone – for example, by multiplying $\delta(\sigma)$ by the step function $\Theta(t - t')$ – and this produces $\delta_{\text{future}}(\sigma)$. The second term involves $\Theta_{\text{future}}(-\sigma)$, a step function that is active when $\sigma < 0$, that is, when x and x' are timelike related; we also restrict the interior of the light cone to the future branch, so that x is necessarily in the future of x' .

The delta term in $G_{\text{ret}}(x, x')$ is sometimes called the *direct term*, and it corresponds to propagation from x' to x that takes place directly on the light cone. If the Green's function contained a direct term only (as it does in flat spacetime), the field at x would depend only on the conditions of the source μ at the corresponding retarded events x' , the intersection between the support of the source and x 's past light cone. In the case of a point particle this reduces to a single point $x' \equiv z(u)$. The theta term in $G_{\text{ret}}(x, x')$, which is sometimes called the *tail term*, corresponds to propagation within the light cone; this extra term (which is generically present in curved spacetime) brings a dependence from events x' that lie in the past of the retarded events. In the case of a point particle, the field at x depends on the particle's entire past history, from $\tau = -\infty$ to $\tau = u$.

There exists an algorithm to calculate $U(x, x')$ and $V(x, x')$ in the form of Taylor expansions in powers of $-\sigma^{\alpha'}$ (LRR Section 4.3.2). It returns

$$U(x, x') = 1 + \frac{1}{12}R_{\alpha'\beta'}\sigma^{\alpha'}\sigma^{\beta'} + \dots \quad (16)$$

and

$$V(x, x') = \frac{1}{12}R(x') + \dots, \quad (17)$$

where $R_{\alpha'\beta'}$ is the Ricci tensor at x' , and $R(x')$ the Ricci scalar at x' . In Ricci-flat spacetimes the expansions of $U - 1$ and V both begin at the fourth order in $\sigma^{\alpha'}$.

6 Alternate Green's Function

The *advanced* Green's function is given by (LRR Section 4.3)

$$G_{\text{adv}}(x, x') = U(x, x')\delta_{\text{past}}(\sigma) + V(x, x')\Theta_{\text{past}}(-\sigma), \quad (18)$$

in terms of the same two-point functions $U(x, x')$ and $V(x, x')$ that appear within the retarded Green's function. The difference is that the light cones are now restricted to the past branch, so that $G_{\text{adv}}(x, x')$ vanishes when x' is in the past of x . A solution to the wave equation constructed with the advanced Green's function would display anti-causal behavior: it would depend on the future history of the source.

Another useful choice of Green's function is the *Detweiler–Whiting singular* Green's function defined by (LRR Section 4.3.5)

$$G_{\text{S}}(x, x') = \frac{1}{2}U(x, x')\delta(\sigma) - \frac{1}{2}V(x, x')\Theta(\sigma). \quad (19)$$

Here the delta and theta functions are no longer restricted: both future and past branches contribute to the Green's function. In fact, the argument of the step function is now $+\sigma$, and this indicates that the second term is active when x and x' are spacelike related. For fixed x' , the singular Green's function is nonzero when x is either on, or outside, the (past and future) light cone of x' . Unlike the retarded and advanced Green's functions, $G_{\text{S}}(x, x')$ is not known to exist globally as a distribution (even for globally hyperbolic spacetimes); its local existence is not in doubt, however, and this suffices for our purposes.

All three Green's functions satisfy the same wave equation, $\square G = -4\pi\delta_4$, and all three give rise to fields $\Phi(x)$ that diverge on the particle's world line. A useful combination of Green's functions is the *Detweiler–Whiting regular* two-point function

$$G_{\text{R}}(x, x') = G_{\text{ret}}(x, x') - G_{\text{S}}(x, x'), \quad (20)$$

which satisfies the homogeneous wave equation $\square G_{\text{R}}(x, x') = 0$. This two-point function gives rise to a field

$$\Phi_{\text{R}} = \Phi_{\text{ret}} - \Phi_{\text{S}} \quad (21)$$

that also satisfies the homogeneous wave equation, $\square\Phi_{\text{R}} = 0$. This field is the difference between two singular fields, and since Φ_{ret} and Φ_{S} are equally singular

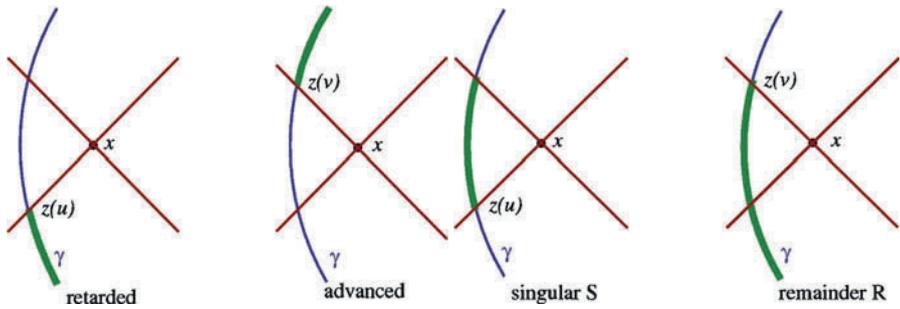


Fig. 5 Solutions to the wave equation. The retarded field is sourced by the past history of the point particle, up to the retarded position $z(u)$. The advanced field is sourced by the future history, starting from the advanced position $z(v)$. The singular field is sourced by the world-line segment that lies on and outside the light cone of the field-point x , starting from the retarded position $z(u)$ and ending at the advanced position $z(v)$. Finally, the regular field is sourced by the history of the particle up to the advanced position $z(v)$

near the world line, we find that the *regular remainder field* Φ_R stays bounded when $x \rightarrow z(\tau)$. And what’s more, the regular field is smooth, in the sense that it and any number of its derivatives possess a well-defined limit when $x \rightarrow z(\tau)$ (Fig. 5).

7 Fields Near the World Line

Using the ingredients presented in the preceding sections it is possible to show (LRR Sections 5.1.3–5.1.5) that close to the world line, the retarded, singular, and regular fields are given by

$$\Phi_{\text{ret}} = \frac{q}{r} + q \int_{-\infty}^u V(x, z) d\tau + O(r^2), \tag{22a}$$

$$\Phi_S = \frac{q}{r} - \frac{1}{3}q\dot{a}_a x^a + O(r^2), \tag{22b}$$

$$\Phi_R = \frac{1}{3}q\dot{a}_a x^a + q \int_{-\infty}^u V(x, z) d\tau + O(r^2), \tag{22c}$$

in the retarded coordinates $(u, x^a = r\Omega^a)$. These expressions involve the retarded distance r from x to $z(u)$ and the frame component $\dot{a}_a(u) := \dot{a}_\mu e_a^\mu$ of $\dot{a}^\mu := Da^\mu/d\tau$, the proper-time derivative of the acceleration vector. They involve also an integration over the past history of the particle. These expressions are valid for Ricci-flat spacetimes. We observe that the retarded and singular fields both diverge as q/r as we approach the world line, but that the regular remainder Φ_R is free of singularities.

In the Fermi coordinates ($t, x^a = s\omega^a$) we have the more complicated expressions:

$$\begin{aligned} \Phi_{\text{ret}} = & \frac{q}{s} - \frac{1}{2}qa_a\omega^a + qs\left(\frac{1}{8}\dot{a}_0 + \frac{1}{3}\dot{a}_a\omega^a - \frac{1}{6}R_{0a0b}\omega^a\omega^b\right) \\ & + q\int_{-\infty}^t V(x, z) d\tau + O(s^2), \end{aligned} \quad (23a)$$

$$\Phi_S = \frac{q}{s} - \frac{1}{2}qa_a\omega^a + qs\left(\frac{1}{8}\dot{a}_0 - \frac{1}{6}R_{0a0b}\omega^a\omega^b\right) + O(s^2), \quad (23b)$$

$$\Phi_R = \frac{1}{3}q\dot{a}_ax^a + q\int_{-\infty}^t V(x, z) d\tau + O(s^2). \quad (23c)$$

They involve the spatial distance s between x and $z(t)$, the frame components of the acceleration vector and its proper-time derivative, the frame components of the Riemann tensor, and the integration over the past history of the particle. Once more we see that the retarded and singular fields diverge in the limit $s \rightarrow 0$, but that the regular field is free of singularities. (The singular nature of Φ_S is also observed in a term such as $\frac{1}{2}qa_a\omega^a$ that stays bounded when $s \rightarrow 0$, but is directionally ambiguous on the world line.)

From these equations we may calculate the spatial derivatives of the singular and regular fields. We obtain

$$\begin{aligned} \nabla_a\Phi_S = & -\frac{q}{s^2}\omega_a - \frac{q}{2s}(\delta_a^b - \omega^b\omega_a)a_b + \frac{q}{8}\dot{a}_0\omega_a \\ & + \frac{q}{6}R_{0b0c}\omega_a\omega^b\omega^c - \frac{q}{3}R_{0a0b}\omega^b + O(s), \end{aligned} \quad (24a)$$

$$\nabla_a\Phi_R = \frac{1}{3}q\dot{a}_a + q\int_{-\infty}^t \nabla_a V(x, z) d\tau + O(s). \quad (24b)$$

As expected, the gradient of the singular field diverges as q/s^2 , but the gradient of the regular field is free of singularities. The gradient of the retarded field is $\nabla_a\Phi_{\text{ret}} = \nabla_a\Phi_S + \nabla_a\Phi_R$.

We notice that many terms in $\nabla_a\Phi_S$ are proportional to an odd number of radial vectors ω^a ; all such terms vanish when we average the field gradient over a spherical surface of constant s . This averaging leaves behind

$$\langle \nabla_a\Phi_S \rangle = -\left(\frac{q}{3s}\right)a_a + O(s), \quad (25)$$

which is still singular. To obtain this result we made use of the identity $\langle \omega^b\omega_a \rangle = \frac{1}{3}\delta_a^b$. Notice that because $\nabla_a\Phi_R$ is smooth at $s = 0$, an averaging simply returns the same expression: $\langle \nabla_a\Phi_R \rangle = \nabla_a\Phi_R$.

8 Self-Force

Let us now reflect on the results of the preceding section and try to make sense of Eq. 12 as an equation of motion for the particle. Our first attempt will be entirely heuristic; we shall add refinement to our treatment in the following section.

Given that Eq. 12 does not make sense as it stands when we insert $\Phi = \Phi_{\text{ret}}$ on its right-hand side, let us take the view that the equation is meant to apply to an extended body instead of a point particle, and let us average $\Phi_\mu := \nabla_\mu \Phi$ over the body's volume. This operation should be carried out in the body's rest frame, and for this purpose it is natural to adopt the Fermi coordinates. We aim, therefore, to average the spatial components Φ_a^{ret} of the field gradient. The simplest form of averaging was carried out already in the preceding section, and we obtained

$$\langle \Phi_a^{\text{ret}} \rangle = \langle \Phi_a^{\text{S}} \rangle + \langle \Phi_a^{\text{R}} \rangle = -\left(\frac{q}{3s}\right)a_a + \Phi_a^{\text{R}} + O(s), \tag{26}$$

in which Φ_a^{R} is evaluated (without obstacle) at $s = 0$. This expression corresponds to pretending that the body is a thin spherical shell of radius s .

Substitution into Eq. 12 and evaluation in the Fermi coordinates produces

$$(m + \delta m)a^a = q\Phi_{\text{R}}^a, \tag{27}$$

with $\delta m := q^2/(3s)$ denoting the contribution to the total body mass that comes from the field's energy. Absorbing this into a redefinition of the inertial mass m , the final tensorial expression for the equation of motion is

$$ma^\alpha = q(g^{\alpha\beta} + u^\alpha u^\beta)\nabla_\beta \Phi_{\text{R}}, \tag{28}$$

with

$$\nabla_\beta \Phi_{\text{R}} = \frac{1}{3}q\dot{a}_\beta + q \int_{-\infty}^\tau \nabla_\beta V(x, z) d\tau'. \tag{29}$$

This is Quinn's equation of motion [10] for a scalar charge q moving in a curved spacetime with metric $g_{\alpha\beta}$. The self-force involves an instantaneous term proportional to $\dot{a}^\alpha = Da^\alpha/d\tau$, as well as an integral over the particle's past history.

Equation 28 informs us that of the complete retarded field $\Phi_{\text{ret}} = \Phi_{\text{S}} + \Phi_{\text{R}}$, only the Detweiler–Whiting regular field Φ_{R} contributes to the self-force. The role of the singular field is merely to contribute to the particle's inertia, through a shift δm in its inertial mass. This contribution diverges in the limit $s \rightarrow 0$, but it would be finite for any extended body.

I must confess that this computation returns the wrong expression for the particle's self-energy. We obtained $\delta m = q^2/(3s)$, while the correct expression is $\delta m = q^2/(2s)$; we are wrong by a factor of 2/3. I believe that this discrepancy originates in an inconsistency between our assumed shape for the extended body – a spherical shell of radius s – and the field it produces, which we took to be equal to

the field produced by a point particle. I would conjecture that calculating the field actually produced by a spherical shell would give rise to the correct expression for δm , but leave unchanged the final result of Eq. 28 for the equation of motion.

9 Axiomatic Approach

The procedure outlined above is admittedly heuristic. It can, however, be formalized and put on an axiomatic basis that supplies Eq. 28 with a much improved pedigree. This is the approach that was first pursued by Ted Quinn and Bob Wald [10, 11]. They formulate two axioms that the scalar self-force F^α should satisfy:

Quinn–Wald Axiom 1 Two scalar particles move on world lines γ and $\tilde{\gamma}$ in two different spacetimes. At points z and \tilde{z} their acceleration vectors have equal lengths. The neighborhoods of z and \tilde{z} , as well as the acceleration vectors, are identified in Fermi coordinates. Then the difference in the self-forces is given by

$$F^\alpha - \tilde{F}^\alpha = q(g^{\alpha\beta} + u^\alpha u^\beta) \lim_{s \rightarrow 0} \langle \Phi_\beta - \tilde{\Phi}_\beta \rangle. \quad (30)$$

Here Φ and $\tilde{\Phi}$ are the retarded fields in each spacetime, and $\Phi_\beta := \nabla_\beta \Phi$ while $\tilde{\Phi}_\beta := \tilde{\nabla}_\beta \tilde{\Phi}$. The limit is well defined after the difference of field gradients is averaged over a sphere of radius s .

Quinn–Wald Axiom 2 $\tilde{F}^\alpha = 0$ in flat spacetime, for a particle with uniform acceleration.

The first axiom is essentially a statement that when two particles momentarily share the same acceleration, their fields are equally singular, and the difference (after averaging) possesses a well-defined limit when $s \rightarrow 0$. The second axiom is a scalar-charge analogue to a well-known result from flat-spacetime electrodynamics: a charged particle moving with a uniform acceleration does not undergo radiation reaction.

According to Eq. 24, the curved-spacetime expression for the gradient of the retarded field is

$$\begin{aligned} \Phi_a = & -\frac{q}{s^2} \omega_a - \frac{q}{2s} (\delta_a^b - \omega^b \omega_a) a_b + \frac{q}{8} \dot{a}_0 \omega_a \\ & + \frac{q}{6} R_{0b0c} \omega_a \omega^b \omega^c - \frac{q}{3} R_{0a0b} \omega^b + \nabla_a \Phi_R + O(s); \end{aligned} \quad (31)$$

this holds at time t in Fermi coordinates. The flat-spacetime expression is

$$\tilde{\Phi}_a = -\frac{q}{s^2} \omega_a - \frac{q}{2s} (\delta_a^b - \omega^b \omega_a) a_b + O(s), \quad (32)$$

and this also holds at time t in the same system of Fermi coordinates. Under the conditions of the Quinn–Wald axioms, the acceleration a_a that appears in Φ_a and $\tilde{\Phi}_a$ is one and the same. In the flat-spacetime expression we set \dot{a}_0 and \dot{a}_a to zero because

the acceleration is chosen to be uniform. In addition we eliminate the Riemann-tensor terms, as well as the integral over the particle’s past history – V necessarily vanishes in flat spacetime.

Subtraction yields

$$\Phi_a - \tilde{\Phi}_a = \frac{q}{8} \dot{a}_0 \omega_a + \frac{q}{6} R_{0b0c} \omega_a \omega^b \omega^c - \frac{q}{3} R_{0a0b} \omega^b + \nabla_a \Phi_R + O(s), \quad (33)$$

and we get

$$\langle \Phi_a - \tilde{\Phi}_a \rangle = \nabla_a \Phi_R + O(s) \quad (34)$$

after averaging over a sphere of constant s . The difference in the self-forces is therefore $F_a - \tilde{F}_a = q \nabla_a \Phi_R$. The second axiom finally returns $F_a = q \nabla_a \Phi_R$, which is equivalent to Eq. 28; we have reproduced Quinn’s expression for the scalar self-force. Notice that the second axiom eliminates the need to carry out an explicit renormalization of the mass.

Another axiomatic approach provides an even more immediate derivation of Quinn’s equation. This is the approach suggested by Steve Detweiler and Bernard Whiting [2], which is based on an observation and an alternate axiom:

Detweiler–Whiting Observation The retarded field Φ_{ret} can be decomposed uniquely into a singular piece Φ_S and a regular remainder Φ_R .

Detweiler–Whiting Axiom The singular field produces no force on the particle.

The immediate consequence of the axiom is that only the regular field participates in the self-force, and we once more arrive at Quinn’s equation.

The Detweiler–Whiting approach is very clean and provides a quick route to the final answer. The observation is not at all controversial, because the singular field is indeed uniquely defined by the prescription outlined in Sect. 6. The axiom, on the other hand, seems too good to be true. How can it just be asserted that the singular field produces no force?

A fairly compelling line of argument rests on the fact that according to its definition, the singular field is strongly time-symmetric, in the sense that the field at x does not depend on the future nor the past of the spacetime point; it instead depends on source points x' that are in a spacelike or lightlike relation with x . Since we would expect the self-force to be sensitive to the direction of time – an advanced field should produce a different force from a retarded field – it seems plausible that the singular field would not know whether to push or pull, and would therefore choose to do neither.

The argument is not water-tight. For example, an alternate singular field, defined by Dirac’s prescription $\frac{1}{2} \Phi_{\text{ret}} + \frac{1}{2} \Phi_{\text{adv}}$, would also be time-symmetric (though not strongly time-symmetric), and could also be asserted to produce no force. The resulting self-force, however, would be produced by $\frac{1}{2} \Phi_{\text{ret}} - \frac{1}{2} \Phi_{\text{adv}}$, and would depend on the entire history of the particle, both past and future. We would of course reject this candidate self-force on grounds of causality violation, but the argument nevertheless shows that there is more to the Detweiler–Whiting singular field than a time-symmetry property. Another hole in the argument lies in the link between

the time-symmetry of the singular field and the statement that it must exert no force: While the time-symmetry property clearly implies that the singular field cannot produce dissipative effects on the particle, there is no reason to rule out an eventual conservative contribution to the self-force.

The conclusion is that additional axioms are necessarily required to make sense of the equations of motion formulated for a point particle. The axioms may seem plausible and perhaps even self-evident, but they cannot be derived from first principles in the context of a classical field theory coupled to a point particle. Such a theory is inherently singular and ambiguous, and it necessarily requires external input in the form of additional axioms.

10 Conclusion

Can one do better than this? The answer is “no” if we insist in treating the point particle as a fundamental classical object. The answer, however, is “yes” if we properly understand that a point particle is merely a convenient substitute for what is fundamentally an extended body. In this view, the length scale of the moving body is ℓ , not 0. The body possesses a finite density of scalar charge, the scalar field is finite everywhere, and its motion traces a world tube in spacetime instead of a single world line. To determine this motion is a well-posed problem, but the description now involves a lot of additional details. Under usual circumstances, however, ℓ is much smaller than all other length scales present in the problem, such as the radius of curvature \mathcal{R} of the body’s trajectory. Under these circumstances the description of the motion can be simplified so as to involve a much smaller number of variables; in the limit $\ell/\mathcal{R} \rightarrow 0$ only the position of the center-of-mass matters, and all couplings between the body’s multipole moments and the external field become irrelevant. In this limit we recover a point-particle description, with the essential understanding that it is merely an approximate description that should not be considered to be fundamental.

To go through the details of this program is difficult, and it appears that very few authors have attempted it since the old days of Lorentz and Abraham. For a recent discussion, and a review of this literature, see the work by Harte [6]. Another important exception concerns the gravitational self-force acting on a small black hole (LRR Section 5.4), which is decidedly not treated as a point mass.

It is well known that in general relativity, the motion of gravitating bodies is determined, along with the spacetime metric, by the Einstein field equations; the equations of motion are not separately imposed. This observation provides a means of deriving the gravitational self-force without having to rely on the fiction of a point mass. In the powerful method of *matched asymptotic expansions*, the metric of the small black hole, perturbed by the tidal gravitational field of the external spacetime, is matched to the metric of the external spacetime, perturbed by the black hole. The equations of motion are then recovered by demanding that the metric be a valid solution to the vacuum field equations. In my opinion, this method (which was first

applied to the gravitational self-force problem by Mino et al. [8]) gives what is by far the most compelling derivation of the gravitational self-force. Indeed, the method is entirely free of conceptual and technical pitfalls – there are no singularities (except deep inside the black hole) and only retarded fields are employed.

In this assessment I respectfully disagree with my colleague Bob Wald, who finds that the method incorporates a number of unjustified assumptions. I would concede that expositions of the method – including my own in LRR – might not have sufficiently clarified some of its subtle aspects. But I see this as faulty exposition, not as an intrinsic difficulty with the method of matched asymptotic expansions. I refer the reader to the recent work by Sam Gralla and Bob Wald [5] for their views on this issue, and their own approach to the motion of an extended body in general relativity.

The introduction of a point particle in a classical field theory appears at first sight to be severely misguided. This is all the more true in a nonlinear theory such as general relativity. The lesson learned here is that surprisingly often, *one can get away with it*. The derivation of the gravitational self-force based on the method of matched asymptotic expansions does indeed show that the result obtained on the basis of a point-particle description can be reliable, in spite of all its questionable aspects. This is a remarkable observation, and one that carries a lot of convenience: It is indeed much easier to implement the point-mass description than to perform the matching of two metrics in two coordinate systems. The lesson, of course, carries over to the scalar and electromagnetic cases.

Acknowledgements I wish to thank the organizers of the school for their kind invitation to lecture; Orléans in the summer is a very nice place to be. I wish to thank the participants for many interesting discussions. And finally, I wish to thank Bernard Whiting for his patience. This work was supported by the Natural Sciences and Engineering Research Council of Canada.

References

1. S. Detweiler, *Class. Q. Grav.* **22**, S681 (2005)
2. S. Detweiler, B.F. Whiting, *Phys. Rev. D* **67**, 024025 (2003)
3. B.S. DeWitt, R.W. Brehme, *Ann. Phys. (N.Y.)* **9**, 220 (1960)
4. P.A.M. Dirac, *Proc. R. Soc. Lond. A* **167**, 148 (1938)
5. S.E. Gralla, R.M. Wald, *Class. Q. Grav.* **25**, 205009 (2008)
6. A.I. Harte, *Phys. Rev. D* **73**, 065006 (2006)
7. J.M. Hobbs, *Ann. Phys. (N.Y.)* **47**, 141 (1968)
8. Y. Mino, M. Sasaki, T. Tanaka, *Phys. Rev. D* **55**, 3457 (1997)
9. E. Poisson, *Living Rev. Rel.* **7**, URL (cited on 11 August 2010): <http://www.livingreviews.org/lrr-2004-6>
10. T.C. Quinn, *Phys. Rev. D* **62**, 064029 (2000)
11. T.C. Quinn, R.M. Wald, *Phys. Rev. D* **56**, 3381 (1997)
12. H. Spohn, *Dynamics of Charged Particles and Their Radiation Field* (Cambridge University Press, Cambridge, 2008)
13. J.L. Synge, *Relativity: the General Theory* (North-Holland, Amsterdam, 1960)
14. K.S. Thorne, J.B. Hartle, *Phys. Rev. D* **31**, 1815 (1985)
15. X.H. Zhang, *Phys. Rev. D* **34**, 991 (1986)

Computational Methods for the Self-Force in Black Hole Spacetimes

Leor Barack

Abstract We survey the set of computational methods devised for implementing the MiSaTaQuWa formulation in practice, for orbits around Kerr black holes. We focus on the *gravitational self-force* (SF) and review in detail two of these methods: (i) the standard *mode-sum method*, in which the perturbation field is decomposed into multipole harmonics and the MiSaTaQuWa regularization is performed, effectively, mode by mode; and (ii) *m-mode regularization*, whereby one regularizes individual azimuthal modes of the full perturbation. The implementation of these strategies involves the numerical integration of the relevant perturbation equations, and we discuss several practical issues that arise and ways to deal with them. These issues include the choice of gauge, the numerical representation of the particle singularity, and the handling of high-frequency contributions near the particle in frequency-domain calculations. As an example, we show results from an actual computation of the gravitational SF for an eccentric geodesic orbit around a Schwarzschild black hole, using direct numerical integration of the Lorenz-gauge perturbation equations in the time domain.

1 Introduction and Overview

The development of a robust formulation for the self-force (SF) in curved spacetime has been motivated strongly by the prospects of observing gravitational waves from extreme mass-ratio inspirals (EMRIs) with LISA (see Jennrich's contribution in this volume). It is estimated that LISA will observe hundreds such events [53], out to cosmological distances [52], allowing a powerful microscopy of the strong-field geometry outside astrophysical massive black holes. The rich science encapsulated in the EMRI waveforms [1, 9, 10, 39, 54] makes this class of sources a high priority for LISA. However, the full scientific promise of EMRI detections could only be

L. Barack (✉)
School of Mathematics, University of Southampton, Southampton, SO17 1BJ,
United Kingdom
e-mail: leor@soton.ac.uk

realized if accurate and faithful theoretical templates of inspiral waveforms were available by the time LISA flies. The underlying idealized physical problem is that of a structureless point particle of mass μ , set in a generic (eccentric, inclined) strong-field orbit around a Kerr black hole of a much larger mass, $M \gg \mu$. The goal is to model the gravitational waveforms emitted as radiation reaction drives the gradual inspiral up until the eventual plunge through the black hole’s event horizon. A prerequisite is to calculate the local gravitational SF acting on the inspiralling object, at least at leading order in the small mass [$O(\mu^2)$]. This defines the mission statement for the ongoing “SF program” that sets the context for this review.

The theoretical framework underpinning the SF program is reviewed elsewhere in this volume (see Wald’s, Detweiler’s, and Poisson’s contributions). It comprises the works by Mino, Sasaki, and Tanaka [86] (gravitational SF using matched asymptotic expansions), Quinn and Wald [103] (gravitational SF using an axiomatic approach), Detweiler and Whiting [44] (R-field reinterpretation of the perturbed motion), Gralla and Wald [57] (rigorous derivation of the SF using a one-parameter family of spacetimes), Barack and Ori [18] (gauge dependence of the gravitational SF), and Poisson [97] (a self-contained and pedagogical review, with an elegant reproduction of previous derivations). The main end product of this theoretical advance was a firmly established general formula for the SF in a class of spacetimes including Kerr. This formula (Eq. 2 below) will be the starting point of our review. We will refer to it, following Poisson [97], as the MiSaTaQuWa formula, an acronym based on the names of the authors of Refs. [86] and [103], where the formula was first derived.

Starting in the late 1990s, work began to recast the MiSaTaQuWa formula in a practical form and implement it in actual calculations of the SF. While the “holy grail” of this program remains the calculation of the gravitational SF for generic orbits in Kerr, much of the initial effort has concentrated on the toy problem of the scalar-field SF, and on simple classes of orbits (radial, circular) in Schwarzschild spacetime. The last few years, however, have seen first calculations of the gravitational (and electromagnetic) SFs for generic orbits in Schwarzschild – and work on Kerr is now under way. In our presentation we shall focus, for concreteness, on the gravitational problem.

The scope of our discussion will be restricted to work concerned with the direct evaluation of the MiSaTaQuWa SF along a given prespecified orbit (normally taken to be a geodesic of the background spacetime); we will not consider here the important question of how orbits evolve under the effect of the SF. There is a parallel research effort [49, 50, 55, 65, 67, 84, 100, 108] aimed to devise a faithful scheme for calculations of the slow (“adiabatic”) orbital evolution in LISA-relevant sources. This effort is largely based on a strategy proposed by Mino [81–83, 107], in which a time-average measure of the rate of change of the orbital “constants of motion” is calculated from a certain “radiative” Green’s function without resorting to the local SF, but neglecting its conservative effects. It is not inconceivable that this method would prove sufficiently accurate for LISA applications. However, ultimately, the performance and accuracy of this method could only be assessed against precise calculations of the full SF.

1.1 The MiSaTaQuWa Formula

The starting point for our discussion is the MiSaTaQuWa formula, which we now state. Consider a timelike geodesic Γ in Kerr spacetime and let τ be proper time along Γ . Let also $x^\alpha = z^\alpha(\tau)$ describe Γ in some smooth coordinate system, and $u^\alpha \equiv dz^\alpha/d\tau$ be the four velocity. Denote by $g_{\alpha\beta}$ the Kerr background metric, and by $h_{\alpha\beta}$ the retarded metric perturbation from a particle of mass μ whose worldline is Γ . Assume $h_{\alpha\beta}$ is given in the Lorenz gauge:

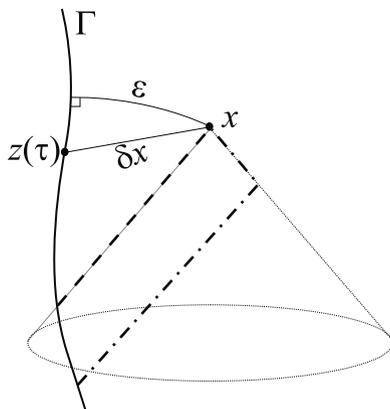
$$g^{\gamma\beta} \bar{h}_{\alpha\beta;\gamma} = 0 \quad \text{with} \quad \bar{h}_{\alpha\beta} \equiv h_{\alpha\beta} - \frac{1}{2} g_{\alpha\beta} h_{\gamma}{}^{\gamma}. \tag{1}$$

Throughout this article, and as usual in perturbation theory, indices are raised and lowered using the background metric $g_{\alpha\beta}$, and covariant derivatives (denoted by semicolons) are taken with respect to that metric. (We also use the conventional geometrized units where $G = c = 1$, and we adopt metric signature $-+++$.) At any spacetime point x the trace-reversed perturbation can be written as a sum of two pieces, $\bar{h}_{\alpha\beta} = \bar{h}_{\alpha\beta}^{\text{dir}} + \bar{h}_{\alpha\beta}^{\text{tail}}$, the former being the ‘‘direct’’ contribution coming from the intersection of the past light cone of x with Γ , and the latter being the ‘‘tail’’ contribution arising from the part of Γ inside this light cone (see Fig. 1). Both $\bar{h}_{\alpha\beta}$ and $\bar{h}_{\alpha\beta}^{\text{dir}}$ obviously diverge when evaluated on Γ ; however, $\bar{h}_{\alpha\beta}^{\text{tail}}$ is continuous and differentiable (though not smooth) there.

The MiSaTaQuWa formula states that the gravitational SF along Γ is given by¹

$$F_{\text{self}}^\alpha(z) = \lim_{x \rightarrow z} \mu k^{\alpha\beta\gamma\delta} \bar{h}_{\beta\gamma;\delta}^{\text{tail}} = \lim_{x \rightarrow z} \left(\mu k^{\alpha\beta\gamma\delta} \bar{h}_{\beta\gamma;\delta} - \mu k^{\alpha\beta\gamma\delta} \bar{h}_{\beta\gamma;\delta}^{\text{dir}} \right), \tag{2}$$

Fig. 1 An illustration of the setup described in the text. $z(\tau)$ is a point on the timelike worldline Γ (thick solid line) and x is a field point close to z , shown with a portion of its past light cone. ϵ is the spatial geodesic distance from x to Γ and $\delta x^\alpha \equiv x^\alpha - z^\alpha$. The metric perturbation at x consists of a *direct* and a *tail* contribution, illustrated by the thick dashed and dash-dot lines, respectively



¹ In the original MiSaTaQuWa formulation, the SF is not expressed directly in terms of the gradient of $\bar{h}_{\alpha\beta}^{\text{tail}}$, but rather as a worldline integral over the gradient of the relevant retarded Green’s function (cf. Eq. 1.9.6 of Poisson [97]). The commutation of the derivative operator and the worldline integral produces local terms at x , which, however, vanish (in the vacuum case which concerns us here) upon contraction with $k^{\alpha\beta\gamma\delta}$ at the limit $x \rightarrow z$.

where $k^{\alpha\beta\gamma\delta}(x)$ is any smooth off- Γ extension of the tensor $k_0^{\alpha\beta\gamma\delta}$ defined on Γ as²

$$k_0^{\alpha\beta\gamma\delta} = \frac{1}{2}g^{\alpha\delta}u^\beta u^\gamma - g^{\alpha\beta}u^\gamma u^\delta - \frac{1}{2}u^\alpha u^\beta u^\gamma u^\delta + \frac{1}{4}u^\alpha g^{\beta\gamma}u^\delta + \frac{1}{4}g^{\alpha\delta}g^{\beta\gamma}. \quad (3)$$

An alternative formulation, as compelling in its interpretation of the perturbed motion as it is useful for practical implementation, is due to Detweiler and Whiting [40, 44]. This formulation introduces the alternative splitting $\bar{h}_{\alpha\beta} = \bar{h}_{\alpha\beta}^S + \bar{h}_{\alpha\beta}^R$, where the “R” field (unlike the tail field) is a certain *smooth* solution of the perturbation equations, which, nonetheless, gives rise to the same physical SF as the tail field:

$$F_{\text{self}}^\alpha(z) = \lim_{x \rightarrow z} \mu k^{\alpha\beta\gamma\delta} \bar{h}_{\beta\gamma;\delta}^R = \lim_{x \rightarrow z} \left(\mu k^{\alpha\beta\gamma\delta} \bar{h}_{\beta\gamma;\delta} - \mu k^{\alpha\beta\gamma\delta} \bar{h}_{\beta\gamma;\delta}^S \right). \quad (4)$$

The gradient fields $\bar{h}_{\alpha\beta;\gamma}^R$ and $\bar{h}_{\alpha\beta;\gamma}^{\text{tail}}$ differ by terms which are continuous (yet not differentiable) on Γ (cf. Eq. 1.9.5 of [97]), and which vanish at $x \rightarrow z$ upon contraction with $k^{\alpha\beta\gamma\delta}$. The singular “S” field $\bar{h}_{\alpha\beta}^S$ and the direct field $\bar{h}_{\alpha\beta}^{\text{dir}}$ share the same leading-order singularity near Γ . More precisely, considering a particular point z on Γ and a nearby off- Γ field point x (see Fig. 1), we have [21, 87, 91]

$$\bar{h}_{\alpha\beta}^{\text{S,dir}}(x) = \frac{4\mu \hat{u}_\alpha(x) \hat{u}_\beta(x)}{\epsilon(x)} + \frac{\mu w_{\alpha\beta}^{\text{S,dir}}(x)}{\epsilon(x)} + c_{\alpha\beta}^{\text{S,dir}}, \quad (5)$$

where \hat{u}_β is the four-velocity vector parallelly propagated from z to x , ϵ is the spatial geodesic distance from x to Γ (i.e., the length of the short normal geodesic section connecting x to Γ), $w_{\alpha\beta}^{\text{S,dir}}$ are smooth functions of x (and z) which vanish at $x \rightarrow z$ at least *quadratically* in the coordinate differences $x^\alpha - z^\alpha$, and $c_{\alpha\beta}^{\text{S,dir}}$ are constants. The S and direct fields differ only in the explicit form of $w_{\alpha\beta}$ and (possibly) in the value of c , neither of which will be important in what follows.

Equations 2 and 4 prescribe the correct regularization of the gravitational SF and form the fundamental basis for all SF calculations. It is important to make the point that the calculation methods to be described below do not involve any further regularization: they are simply implementations of the MiSaTaQuWa formula. In particular, the mode-sum methods at the focus of our discussion (sometimes referred to, perhaps misleadingly, as “regularization” methods) simply recast the MiSaTaQuWa formula in a more practical – but otherwise equivalent – form.

1.2 Gauge Dependence

Another crucial point to have in mind is that the MiSaTaQuWa formula (2) is guaranteed to hold true only if $\bar{h}_{\alpha\beta}$ satisfies the Lorenz-gauge condition (1). Relatedly,

² The quantity $k^{\alpha\beta\gamma\delta} \bar{h}_{\beta\gamma;\delta}$ is the linear perturbation in the connection coefficients, $\delta\Gamma_{\beta\gamma}^\alpha(h_{\mu\nu})$, projected orthogonally to Γ ; that is, $k^{\alpha\beta\gamma\delta} \bar{h}_{\beta\gamma;\delta} = (g^{\beta\gamma} + u^\beta u^\gamma) \delta\Gamma_{\beta\gamma}^\alpha$.

the singular fields $\bar{h}_{\alpha\beta}^{S,\text{dir}}$ are guaranteed to have the form (5) only in the Lorenz gauge; other gauge choices may “distort” the local isotropy of the singularity (as demonstrated in Ref. [18] with a few examples).

Yet, it is important to understand the gauge dependence of the SF. A thorough analysis of this issue was presented in Ref. [18]. It was explained that the very definition of the gravitational SF through a mapping of the physical trajectory from a “perturbed” geometry onto a “background” spacetime automatically gives rise to a gauge ambiguity in the SF. From this recognition there readily follows a gauge transformation law for the SF: For a given physical metric perturbation $h_{\mu\nu}$ in a given gauge, consider an infinitesimal gauge displacement $x^\alpha \rightarrow x^\alpha - \xi^\alpha(\propto \mu)$, under which the perturbation transforms according to

$$h_{\mu\nu} \rightarrow h_{\mu\nu} + \xi_{\mu;\nu} + \xi_{\nu;\mu}. \quad (6)$$

The corresponding gravitational SF then transforms as [18]

$$F_{\text{self}}^\alpha \rightarrow F_{\text{self}}^\alpha - \mu \left[(g^{\alpha\lambda} + u^\alpha u^\lambda) \ddot{\xi}_\lambda + R^\alpha_{\mu\lambda\nu} u^\mu \xi^\lambda u^\nu \right], \quad (7)$$

where an overdot denotes covariant differentiation with respect to proper time, $R^\alpha_{\mu\lambda\nu}$ is the Riemann tensor associated with the background geometry, and the terms in square brackets are, of course, evaluated at the particle.

Strictly speaking, the notion of a gravitational SF in a non-Lorenz gauge only makes sense if, for ξ^α relating that gauge to the Lorenz gauge, the expression on the right-hand side of Eq. 7 is well defined. This is not at all an obvious condition; some simple counterexamples are analyzed in [18]. In gauges where the expression in square brackets in Eq. 7 does not have a well-defined (finite and direction independent) particle limit, one might still devise a useful notion of the SF by averaging over angular directions, or by taking a directional limit in a consistent fashion [18, 57].

1.3 Implementation Strategies

One faces several practical challenges in trying to implement the MiSaTaQuWa formula. Some analytic approximations (see under “quasi-local” and “weak field” below) can tackle the tail calculation directly, but an accurate treatment of the strong-field SF must involve a numerical computation, bringing about several technical difficulties. Foremost, how does one go about extracting the tail (or R) piece from the full retarded perturbation in practice? Equations 2 and 3 suggest using the subtraction $\bar{h}_{\alpha\beta} - \bar{h}_{\alpha\beta}^{\text{dir}}$ (or $\bar{h}_{\alpha\beta} - \bar{h}_{\alpha\beta}^S$). This, however, involves the removal of one divergent quantity from another, which is not easily tractable in actual numerical calculations.

Several strategies have been proposed for dealing with this “subtraction” problem, and in the main part of this review we shall describe a few of them in some detail. Here we proceed with a brief overview of the main implementation frameworks, including those based on analytic approximations.

1.3.1 Quasi-Local Calculations

This approach tackles the calculation of the tail contribution directly, by analytically evaluating the Hadamard expansion of the Green’s function [2–5, 93, 94]. Such calculations capture the “near” part of the tail, which, one might hope, represents the dominant contribution in problems of interest. Quasi-local calculations could be supplemented by a numerical computation of the “far” part of the tail, a strategy referred to as “matched expansions” (not to be confused with “matched asymptotic expansions”). The applicability of this idea was demonstrated very recently [37] with a full calculation in Nariai spacetime (a simple toy spacetime featuring many of the characteristics of Schwarzschild).

1.3.2 Weak-Field Analysis

The tail formula can be evaluated analytically for certain weak-field configurations [47, 96, 98–100], within a Newtonian or a post-Newtonian (PN) framework. Such work has provided important insight into the nature and properties of the SF. PN techniques have also been implemented in combination with the mode-sum method discussed below [63, 64, 89–91].

1.3.3 Radiation-Gauge Regularization

This approach, introduced in 2006 [70], proposes a reformulation of the MiSa-TaQuWa regularization in the radiation gauge, which enables a reconstruction of the R-part of the metric perturbation from a regularized Newman–Penrose scalar (ψ_0 or ψ_4). The main advantage of this method is that it reduces the numerical component of the calculation to a solution of a single scalar-like (Teukolsky’s) equation. However, some of the technical complexity is relegated to the metric reconstruction stage. This technique has been implemented so far only for a particle held static in Schwarzschild, but more interesting cases are currently being studied [69, 101]. See also the discussion in Section 3.1.

1.3.4 Mode-Sum Method

An approach whereby one evaluates the tail contribution mode by mode in a multipole expansion [7, 16, 17, 19, 21, 22, 45, 62–64, 75, 87]. The subtraction “full–direct” (or “full–S-field”) is performed mode by mode, avoiding the need to deal with divergent quantities. The method exploits the separability of the field equations in Kerr into multipole harmonics. The mode-sum method has provided the framework for the bulk of work on SF calculations over the last decade. We will discuss it in detail in Section 2.

1.3.5 “Puncture” Methods

A set of recently proposed methods custom-built for time-domain numerical implementation in $2 + 1$ or $3 + 1$ dimensions [11, 12, 69, 77, 116]. Common to these methods is the idea to utilize as a variable for the numerical time evolution a “punctured” field, constructed from the full field by removing a suitable singular piece, given analytically. The piece removed approximates the correct S-field sufficiently well that the resulting “residual” field is guaranteed to yield the correct MiSaTaQuWa SF. In the $2 + 1$ D version of this approach the regularization is done mode by mode in the azimuthal (m -mode) expansion of the full field. This procedure offers significant simplification; we shall review it in detail in Section 5.

SF Calculations to Date As we have mentioned already, the program to calculate the SF in black hole orbits has been progressing gradually, through the study of a set of simplified model problems. Some of the necessary computational techniques were first tested within the simpler framework of a scalar-field toy model before being applied to the electromagnetic (EM) and gravitational problems. Authors have considered special classes of orbits (static, radial, circular) before attempting more generic cases, and much of the work so far has focused on Schwarzschild orbits. The state of the art is that there now exist numerical codes for calculating the scalar, EM, and gravitational SFs for all (bound) geodesics in Schwarzschild spacetime. It is reasonable to predict that workers in the field will now increasingly be turning their attention to the Kerr problem.

The information in Tables 1–3 is meant to provide a quick reference to work done so far. It covers actual evaluations of the local SF that are based on the MiSaTaQuWa formulation (or the analogous scalar-field and EM formulations of Refs. [102] and [46, 66, 103], respectively), either directly or through one of the aforementioned implementation methods. We have included weak-field and PN implementations, but have not included quasi-local calculations and work based on the radiative Green’s function approach. The three tables list separately works on the scalar, EM, and gravitational SFs. In each table, works are listed roughly in chronological order. Some of the numerical techniques indicated under “calculation method” are discussed in Sections 4 and 5 of this review.

The rest of this review is structured as follows. Section 2 is a self-contained introduction to the mode-sum method. The basic idea is presented through an elementary example, followed by a formulation of the method as applied to generic orbits in Kerr. In Section 3, we discuss the practicalities of numerical calculations with point-like sources, and review some of the methods proposed to facilitate such calculations in both the frequency and time domains. Section 4 focuses on a particular implementation method, namely the direct time-domain integration of the Lorenz-gauge metric perturbation equations (in Schwarzschild). This method enabled the recent milestone calculation of the gravitational SF for eccentric orbits in Schwarzschild, and we present results from this calculation. In Section 5 we discuss puncture-type methods, proposed (with the Kerr problem in mind) as alternative to the standard mode-sum scheme. We focus on one particular variant: the m -mode regularization method. Section 6 reflects on recent advances and speculates on future directions.

Table 1 Calculations of the *scalar-field* SF within the MiSaTaQuWa framework. In this table, as well as in Tables 2 and 3, “direct” implies explicit evaluation of the MiSaTaQuWa tail term; “spectral” indicates numerical calculation through frequency-domain analysis; and “evolution” refers to numerical calculation via integration in the time domain

Case	Author(s)	Implementation strategy	Computation method
Schwarzschild: Static particle ^a	Burko [29]	Mode sum	Analytic
Newtonian potential: Generic motion	Pfenning and Poisson [96]	Direct	Analytic
Schwarzschild: Circular geodesics	Burko [30] Detweiler et al. [45, 48]	Mode sum	Spectral
Schwarzschild: Radial geodesics	Barack and Burko [8]	Mode sum	Evolution in 1+1D
Spherical mass shell: Static particle	Burko et al. [33]	Mode sum	Analytic
Schwarzschild: Circular geodesics	Nakano et al. [90] Hikida et al. [64]	Post-Newtonian	Analytic
Kerr–Newman: Static particle	Burko and Liu [32]	Mode sum	Semi-analytic
Isotropic cosmology: Static particle	Burko et al. [34]	Direct	Analytic
Isotropic cosmology: Slow motion	Haas and Poisson [61]	Direct	Analytic
Schwarzschild: Eccentric geodesics	Haas [59]	Mode sum	Evolution in 1+1D
Schwarzschild: Circular geodesics	Vega and Detweiler [116]	Puncture	Evolution in 1+1D
Kerr: circular Equatorial geodesics	Barack and Warburton [119]	Mode sum	Spectral

^aThis case was analyzed independently by Wiseman [120] using a different method.

Table 2 Calculations of the *electromagnetic* self-force based on MiSaTaQuWa formula

Case	Author(s)	Implementation strategy	Computation method
Schwarzschild: Static particle ^a	Burko [29]	Mode sum	Analytic
Newtonian potential: Generic motion	Pfenning and Poisson [96]	Direct	Analytic
Isotropic cosmology: Slow motion	Haas and Poisson [61]	Direct	Analytic
Schwarzschild: Static particle	Keidl et al. [70]	Radiation gauge Regularization	Analytic
Schwarzschild: Eccentric geodesics	Haas [60]	Mode sum	Evolution in 1+1D

^aThis calculation recovers results by Will and Smith [105] using a different method.

Table 3 Calculations of the *gravitational* self-force based on MiSaTaQuWa formula

Case	Author(s)	Implementation strategy	Computation method
Newtonian potential:	Pfenning and Poisson [96]	Direct	Analytic
Generic motion			
Schwarzschild:	Barack and Lousto [14]	Mode sum	1+1D evolution in Regge–Wheeler gauge
Radial geodesics			
Schwarzschild:	Keidl et al. [70]	Radiation gauge	Analytic
Static particle		Regularization	
Schwarzschild:	Barack and Sago [24]	Mode sum	1+1D evolution in Lorenz gauge
Circular geodesics			
Schwarzschild:	Detweiler [42]	Mode sum	Spectral numerics in Regge–Wheeler gauge
Circular geodesics			
Schwarzschild:	Barack and Sago [25, 26]	Mode sum	1+1D evolution in Lorenz gauge
Eccentric geodesics			

2 Mode-Sum Method

Let us write the MiSaTaQuWa formula (4) in the more compact form

$$F_{\text{self}}^\alpha(z) = \lim_{x \rightarrow z} [F^\alpha(x) - F_S^\alpha(x)], \tag{8}$$

where F^α and F_S^α are fields (the “full force” and “S force,” respectively) constructed through

$$F^\alpha(x) = \mu k^{\alpha\beta\gamma\delta} \bar{h}_{\beta\gamma;\delta}, \quad F_S^\alpha(x) = \mu k^{\alpha\beta\gamma\delta} \bar{h}_{\beta\gamma;\delta}^S. \tag{9}$$

Recall that $k^{\alpha\beta\gamma\delta}(x)$ are tensor fields defined through a (smooth) extension of $k_0^{\alpha\beta\gamma\delta}$ (given in Eq. 3) off the worldline. The fields $F^\alpha(x)$ and $F_S^\alpha(x)$ may be defined through any such extension, as long as the same extension is applied in both. Both F^α and F_S^α , of course, diverge at the particle, but their difference, $F_R^\alpha \equiv \mu k^{\alpha\beta\gamma\delta} \bar{h}_{\beta\gamma;\delta}^R$, is a *smooth* (analytic) function of x even at the particle.

In the mode-sum method we formally decompose the fields F^α and F_S^α into multipole (l, m) harmonics in the black hole spacetime. These harmonics are defined in the Kerr/Schwarzschild background based on the Boyer–Lindquist/Schwarzschild coordinates (t, r, θ, φ) in the standard way, that is, through a projection onto an orthogonal basis of angular functions defined on surfaces of constant t and r . Let us denote by $F^{\alpha l}(x)$ and $F_S^{\alpha l}(x)$ the l -mode contribution to F^α and F_S^α , respectively (summed over m). A key observation is that each of these l -mode fields is finite even at the particle. This suggests a natural regularization procedure, which, essentially, amounts to performing the subtraction $F^\alpha - F_S^\alpha$ *mode by mode*. The idea is best developed through an elementary example, as follows.

2.1 An Elementary Example

Consider a pointlike particle of mass μ at rest in flat space. The location of the particle is $\mathbf{x} = \mathbf{x}_p$ in a given Cartesian system. In this simple static configuration, the perturbed Einstein equations read

$$\nabla^2 \bar{h}_{tt} = -16\pi\mu \delta^3(\mathbf{x} - \mathbf{x}_p) \quad (10)$$

(with all other components vanishing), where $\bar{h}_{\alpha\beta}(\mathbf{x})$, recall, is the trace-reversed metric perturbation (Eq. 1), ∇^2 is the Laplacian operator, and we have represented the particle with a delta-function distribution. The static perturbation $\bar{h}_{\alpha\beta}$ automatically satisfies the Lorenz-gauge condition (1). Of course, in this simple case we can immediately write down the exact physical solution (it is the Coulomb-like solution $\bar{h}_{tt} = 4\mu/|\mathbf{x} - \mathbf{x}_p|$), and we simply have $\bar{h}_{tt}^S = \bar{h}_{tt}$ and $F_S^\alpha = F^\alpha$, so that Eq. 8 trivially gives $F_{\text{self}}^\alpha = 0$ as expected. However, for the sake of our discussion, let us instead proceed (indeed, rather artificially in our case) by considering the multipole expansion of the perturbation.

To this end, introduce polar coordinates (r, θ, φ) , such that our particle is located at $(r_0 \neq 0, \theta_0, \varphi_0)$, and expand h_{tt} in spherical harmonics on the spheres $r = \text{const}$, in the form

$$\bar{h}_{tt} = \sum_{l=0}^{\infty} \bar{h}_{tt}^l(r, \theta, \varphi), \quad \text{where} \quad \bar{h}_{tt}^l(r, \theta, \varphi) = \sum_{m=-l}^l \tilde{h}_{tt}^{lm}(r) Y_{lm}(\theta, \varphi). \quad (11)$$

This expansion separates Eq. 10 into radial and angular parts, the former reading (for each l, m)

$$\tilde{h}_{tt,rr}^{lm} + \frac{2}{r} \tilde{h}_{tt,r}^{lm} - \frac{l(l+1)}{r^2} \tilde{h}_{tt}^{lm} = \frac{-16\pi\mu}{r_0^2} Y_{lm}^*(\theta_0, \varphi_0) \delta(r - r_0), \quad (12)$$

where an asterisk denotes complex conjugation. The unique physical l, m -mode solution, continuous everywhere and regular at both $r = 0$ and $r \rightarrow \infty$, reads

$$\tilde{h}_{tt}^{lm}(r) = \frac{16\pi\mu}{(2l+1)r_0} Y_{lm}^*(\theta_0, \varphi_0) \times \begin{cases} (r/r_0)^{-l-1}, & r \geq r_0, \\ (r/r_0)^l, & r \leq r_0, \end{cases} \quad (13)$$

giving

$$\bar{h}_{tt}^l(r, \theta) = \frac{4\mu}{r_0} P_l(\cos \gamma) \times \begin{cases} (r/r_0)^{-l-1}, & r \geq r_0, \\ (r/r_0)^l, & r \leq r_0, \end{cases} \quad (14)$$

where P_l is the Legendre polynomial and γ is the angle subtended by the two radius vectors to \mathbf{x} and \mathbf{x}_p (see Fig. 2).

Now consider the multipole decomposition of the ‘‘full-force’’ field defined on the left-hand side of Eq. 9. This decomposition will, of course, depend on how the tensor $k^{\alpha\beta\gamma\delta}$ is extended off the particle. We choose here a simple ‘‘fixed k ’’ extension, defined through $k^{\alpha\beta\gamma\delta}(x) \equiv k_0^{\alpha\beta\gamma\delta}$ (in our Minkowskian coordinates (t, \mathbf{x})). Then

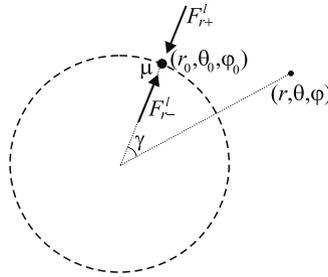


Fig. 2 An illustration of the simple setup described in the text: A particle of mass μ in flat space is at rest at $(r_0, \theta_0, \varphi_0)$. The gravitational field of the particle is decomposed into spherical harmonics, each contributing a finite amount $F_{r\pm}^l$ to the “one-sided” radial full force acting on the particle

$F_t = 0$ and $F_i = \mu k_i^{ttj} \bar{h}_{tt,j} = (\mu/4) \bar{h}_{tt,i}$. Focus now on the r component. The l mode of F_r is given in terms of the l mode of $\bar{h}_{tt,r}$ simply as $F_r^l = (\mu/4) \bar{h}_{tt,r}^l$. Using Eq. 14 and evaluating F_r at the particle (taking $\gamma \rightarrow 0$ followed by $r \rightarrow r_0^\pm$), we obtain

$$F_{r\pm}^l(\mathbf{x}_p) = \mp L \frac{\mu^2}{r_0^2} - \frac{\mu^2}{2r_0^2}, \quad \text{where } L \equiv l + \frac{1}{2}. \tag{15}$$

Here, the subscripts \pm indicate the two (different) values obtained by taking the particle limit from “outside” ($r \rightarrow r^+$) and “inside” ($r \rightarrow r^-$).

Let us note the following features manifest in the above simple analysis:

- The individual l modes of the metric perturbation, $\bar{h}_{\alpha\beta}^l$, are each continuous at the particle’s location, although their derivatives are discontinuous there.
- The individual l modes of the full force, F_α^l , have *finite* one-sided values at the particle.
- At large l , each of the one-sided values of the full force at the particle is dominated by a term $\propto l$. (The mode sum obviously diverges at the particle, reflecting the divergence of the full force F_α there.)

It turns out (as we shall see later) that all the above features are quite generic, and they carry over intact to the much more general problem of a particle moving in Kerr spacetime. Specifically, we find that, at any point along the particle’s trajectory, the (one-sided values of the) full-force l modes always have the large- l form

$$F_{\alpha\pm}^l = \pm LA_\alpha + B_\alpha + C_\alpha/L + O(L^{-2}). \tag{16}$$

In our elementary problem the power series in $1/L$ truncates at the L^0 term, but in general the series can be infinite. The l -independent coefficients A_α , B_α , and C_α , whose values depend on the geometry as well on the particle’s location and four-velocity, reflect the asymptotic structure of the particle singularity at large l . These coefficients, called *regularization parameters*, play a crucial role in the mode-sum regularization procedure, as we describe next.

2.2 The Mode-Sum Formula

Consider a mass particle moving on a geodesic trajectory in Kerr, and suppose we are interested in the value of the SF at a point z with Boyer–Lindquist coordinates $(t_0, r_0, \theta_0, \varphi_0)$ along the trajectory. Starting with Eq. 8, let us formally expand $F^\alpha(x)$ and $F_S^\alpha(x)$ in spherical harmonics³ on the surfaces $t, r = \text{const}$. Denoting the corresponding l -mode contributions (summed over m) by $F^{\alpha l}(x)$ and $F_S^{\alpha l}(x)$, we write⁴

$$F_{\text{self}}^\alpha(z) = \lim_{x \rightarrow z} \sum_{l=0}^{\infty} \left[F^{\alpha l}(x) - F_S^{\alpha l}(x) \right]. \quad (17)$$

Since $F^\alpha(x) - F_S^\alpha(x)$ is a smooth function for all x , the mode sum in Eq. 17 is guaranteed to converge exponentially for all x (this is a general mathematical property of the multipole expansion). In particular, the sum converges uniformly at $x = z$, and we are allowed to change the order of limit and summation. We expect, however, that the particle limit of the individual terms $F^{\alpha l}$ and $F_S^{\alpha l}$ is only defined in a one-sided sense. Hence, we write

$$F_{\text{self}}^\alpha(z) = \sum_{l=0}^{\infty} \left[F_{\pm}^{\alpha l}(z) - F_{S\pm}^{\alpha l}(z) \right], \quad (18)$$

where \pm indicates the values obtained by first taking the limits $t \rightarrow t_0, \theta \rightarrow \theta_0$ and $\varphi \rightarrow \varphi_0$, and then taking $r \rightarrow r_0^\pm$. Of course, the difference $F_{\pm}^{\alpha l}(z) - F_{S\pm}^{\alpha l}(z)$ does *not* depend on the direction from which the radial limit is taken: As the difference $F^{\alpha l}(x) - F_S^{\alpha l}(x)$ is the l mode of a smooth function, it is itself smooth for all x .

Furthermore, since the mode sum in Eq. 18 converges exponentially, we expect $F^{\alpha l}$ and $F_S^{\alpha l}$ to share the same large- l power expansion (16), with the same expansion coefficients. This motivates us to reexpress Eq. 18 in the form

$$F_{\text{self}}^\alpha(z) = \sum_{l=0}^{\infty} \left[F_{\pm}^{\alpha l}(z) \mp LA^\alpha - B^\alpha - C_\alpha/L \right] - \sum_{l=0}^{\infty} \left[F_{S\pm}^{\alpha l}(z) \mp LA^\alpha - B^\alpha - C_\alpha/L \right]. \quad (19)$$

Here, each of the terms in square brackets vanishes at least as $\sim l^{-2}$ as $l \rightarrow \infty$, and so each of the two sums converges at least as $\sim 1/l$. We hence arrive at the

³ Here we ignore the vectorial nature of F^α and F_S^α and treat each of their Boyer–Lindquist components as a scalar function. We do this for a mere mathematical convenience. See [62] for a more sophisticated, covariant treatment.

⁴ Note that the form of $F^{\alpha l}$ and $F_S^{\alpha l}$ will depend on the specific off-worldline extension chosen for the tensor k . However, this ambiguity disappears upon taking the limit $x \rightarrow z$, and the final SF is of course insensitive to the choice of extension.

following mode-sum reformulation of the MiSaTaQuWa equation:

$$F_{\text{self}}^\alpha(z) = \sum_{l=0}^{\infty} \left[F_{\pm}^{\alpha l}(z) \mp LA^\alpha - B^\alpha - C^\alpha \right] - D^\alpha \tag{20}$$

with

$$D^\alpha \equiv \sum_{l=0}^{\infty} \left[F_{S\pm}^{\alpha l}(z) \mp LA^\alpha - B^\alpha - C^\alpha \right]. \tag{21}$$

The *mode-sum formula* (20), first proposed in Refs. [17] (scalar-field case) and [7] (gravitational case) provides a practical way of calculating the SF, once the regularization parameters A^α , B^α , C^α , and D^α are known. The values of these parameters are obtained analytically via a local analysis of the singular (or direct) field near the particle. Early calculations of the regularization parameters [6, 7, 17], which were restricted to specific orbits, were based on a local analysis of the Green’s function near coincidence ($x \rightarrow z$) at large l . Later, after the local form of the direct field had been derived explicitly to sufficient accuracy [87], workers have been able to derive general expressions for the regularization parameters, valid for generic (geodesic) orbits – first in Schwarzschild [16, 19, 21], then in Kerr [20, 22]. Later derivations (in the scalar case) are presented in Ref. [45], where higher-order terms in the $1/L$ expansion were derived analytically in order to accelerate the convergence of the mode sum; and in Ref. [62], where the regularization parameters were redefined as scalar quantities using a covariant projection of the singular field onto a null tetrad based on the worldline.

In what follows, we sketch the derivation of the parameters (using the method of Refs. [16, 22]) for generic geodesics in Kerr.

In passing, we remind that the mode-sum formula (in the gravitational case) is formulated in the Lorenz gauge, just like the MiSaTaQuWa formula on which it relies. The values of the regularization parameters; the form of the large- l expansion of the singular force; or even the very definiteness of the SF – none of these is gauge independent. It has been shown, however, that the regularization parameters remain invariant under gauge transformations (from the Lorenz gauge) that are sufficiently regular [18].

2.3 Derivation of the Regularization Parameters

Consider a geodesic worldline Γ in Kerr, and refer back to Fig. 1. Let point z along Γ have Boyer–Lindquist coordinates $(t_0, r_0, \theta_0, \varphi_0)$. The singular force at x (a point in the immediate vicinity of z) is $F_S^\alpha(x) = \mu k^{\alpha\beta\gamma\delta} \bar{h}_{\beta\gamma;\delta}^S$ (Eq. 9), where $\bar{h}_{\beta\gamma}^S(x)$ is given in Eq. 5. We hereafter use the “fixed k” extension, defined through $k^{\alpha\beta\gamma\delta}(x) \equiv k_0^{\alpha\beta\gamma\delta}(z)$ (note this definition is coordinate dependent; here we refer specifically to contravariant Boyer–Lindquist components). We assume x^α are

smooth coordinates, and denote the coordinate difference between points x and z by $\delta x^\alpha \equiv x^\alpha - z^\alpha$. In Eq. 5 the quantities $\epsilon^2(x)$ and $\hat{u}_\alpha(x)$ are smooth functions of δx , and we may expand them in the form

$$\epsilon^2 = S_0 + S_1 + O(\delta x^4), \quad \hat{u}_\alpha = u_\alpha + \delta u_\alpha + O(\delta x^2), \quad (22)$$

where S_0 and S_1 are, respectively, quadratic and cubic in δx , and δu_α is linear in δx . Explicitly, these expansion coefficients read

$$S_0 = (g_{\alpha\beta} + u_\alpha u_\beta) \delta x^\alpha \delta x^\beta, \quad (23)$$

$$S_1 = \left(u_\lambda u_\gamma \Gamma_{\alpha\beta}^\lambda + g_{\alpha\beta, \gamma} / 2 \right) \delta x^\alpha \delta x^\beta \delta x^\gamma, \quad (24)$$

$$\delta u_\alpha = \Gamma_{\alpha\beta}^\lambda u_\lambda \delta x^\beta, \quad (25)$$

where the metric and its derivatives, as well as the background connections $\Gamma_{\alpha\beta}^\lambda$, are all evaluated at z (for a derivation of S_1 , see Appendix A of [19]).

We now substitute the above expansions in Eq. 5 and construct the field $F_S^\alpha(x)$. We may write down the resulting expression as a sum of terms sorted according to how they scale with δx :

$$F_S^\alpha(x) = \frac{P_{(1)}^\alpha(\delta x)}{\epsilon_0^3} + \frac{P_{(4)}^\alpha(\delta x)}{\epsilon_0^5} + \frac{P_{(7)}^\alpha(\delta x)}{\epsilon_0^7} + O(\delta x). \quad (26)$$

Here $\epsilon_0 \equiv S_0^{1/2}$, and $P_{(n)}^\alpha$ denotes a certain multilinear function of the coordinate differences δx^α , of homogeneous order n in δx . Note that the first term on the right-hand side scales as δx^{-2} , the second as δx^{-1} , and the third as δx^0 . The $O(\delta x)$ remainder disappears at the limit $x \rightarrow z$ and cannot affect the value of the final SF; it is therefore safe to ignore it. The explicit form of $P_{(7)}^\alpha$ will not be needed in our analysis. The two other coefficients read

$$P_{(1)}^\alpha = -\frac{1}{2} K_0^{\alpha\delta} S_{0,\delta}, \quad (27)$$

$$P_{(4)}^\alpha = -\frac{1}{2} K_0^{\alpha\delta} S_0 S_{1,\delta} + \frac{3}{4} K_0^{\alpha\delta} S_1 S_{0,\delta} - \frac{1}{2} K_1^{\alpha\delta} S_0 S_{0,\delta}, \quad (28)$$

where

$$K_0^{\alpha\delta} = 4k^{\alpha\beta\gamma\delta} u_\beta u_\gamma, \quad K_1^{\alpha\delta} = 4k^{\alpha\beta\gamma\delta} (\delta u_\beta u_\gamma + u_\beta \delta u_\gamma). \quad (29)$$

The l modes of the $F_S^\alpha(z)$ are constructed, by definition, through

$$\begin{aligned} F_{S_\pm}^{\alpha l} &= \lim_{x \rightarrow z_\pm} \sum_{m=-l}^l Y_{lm}(\theta, \varphi) \int d\Omega' Y_{lm}^*(\theta', \varphi') F_S^\alpha(r, t, \theta', \varphi'; z) \\ &= \lim_{\delta r \rightarrow 0_\pm} \frac{L}{2\pi} \int d\Omega' P_l(\cos \tilde{\theta}) F_S^\alpha(r, t_0, \theta', \varphi'; z), \end{aligned} \quad (30)$$

where $d\Omega' \equiv d \cos \theta' d\varphi'$, $\cos \tilde{\theta} \equiv \cos \theta' \cos \theta_0 + \sin \theta' \sin \theta_0 \cos(\varphi' - \varphi_0)$, and, recall, $L = l + 1/2$. In the second line we have already taken the limits $t \rightarrow t_0$, $\theta \rightarrow \theta_0$, and $\varphi \rightarrow \varphi_0$, thereby choosing to approach the particle along the radial direction, recalling that we expect two different one-side values (hence the designation “ \pm ”). It is now convenient to use $\tilde{\theta}$ as a new “polar” angular integration variable [noting $\tilde{\theta} = 0$ for $(\theta', \varphi') = (\theta_0, \varphi_0)$], along with a new “azimuthal” variable $\tilde{\varphi}$ based around the particle; then introduce a pair of (locally) Cartesian-like angular coordinates, through

$$x = \rho(\tilde{\theta}) \cos \tilde{\varphi}, \quad y = \rho(\tilde{\theta}) \sin \tilde{\varphi}. \quad (31)$$

Here $\rho \equiv 2 \sin(\tilde{\theta}/2) [= \tilde{\theta} + O(\tilde{\theta}^3)$ near the particle]. With no loss of generality, we choose $\tilde{\varphi}$ such that the four-velocity at z has no component in the y direction: $u^y = 0$. Note $x = y = 0$ at the particle, so $\delta x^x = x$ and $\delta x^y = y$.

To evaluate the integral in Eq. 30, let us consider separately the contributions from each of the three terms of F_S^α in Eq. 26. As explained in Ref. [19], the second and third terms (scaling near the particle as $\sim \delta x^{-1}$ and $\sim \delta x^0$, respectively) are sufficiently regular to allow interchanging the order of the limit and integral in Eq. 30. Also (relatedly), for these two terms the radial limit is well defined and the two-side ambiguity does not occur. The third term, in particular, contributes to $F_S^{\alpha l}$ an amount

$$\frac{L}{2\pi} \int d \cos \tilde{\theta} d\tilde{\varphi} P_l(\cos \tilde{\theta}) \frac{\hat{P}_{(7)}^\alpha(x, y)}{\hat{\epsilon}_0^7(x, y)}, \quad (32)$$

where hats denote evaluation at $\delta r = \delta t = 0$. We observe that $\hat{\epsilon}_0$ is an even function of both x and y , and so is $\cos \tilde{\theta}$. However, each of the terms in $\hat{P}_{(7)}^\alpha(x, y)$ (such as $\propto x^2 y^5$, $\propto x^4 y^3$, etc.) is odd in either x or y . It follows from symmetry that the integral in Eq. 32 vanishes. Hence, the term $\propto P_{(7)}^\alpha$ contributes nothing to the l mode $F_S^{\alpha l}$.

To evaluate the contribution of the term $\epsilon_0^{-5} P_{(4)}^\alpha$, first use Eq. 31 to factor ρ^4 out of $\hat{P}_{(4)}^\alpha$, writing $\hat{P}_{(4)}^\alpha(x, y) = \rho^4 \hat{P}_{(4)}^\alpha(\cos \tilde{\varphi}, \sin \tilde{\varphi})$. Similarly, write $\hat{\epsilon}_0(x, y) = \rho \hat{\epsilon}_0(\cos \tilde{\varphi}, \sin \tilde{\varphi})$. This allows us to separate the $\tilde{\theta}$ and $\tilde{\varphi}$ integrals for this contribution, in the form

$$\left[\frac{L}{2\pi} \int_{-1}^1 d \cos \tilde{\theta} \frac{P_l(\cos \tilde{\theta})}{\rho(\tilde{\theta})} \right] \left[\int_0^{2\pi} d\tilde{\varphi} \frac{\hat{P}_{(4)}^\alpha(\cos \tilde{\varphi}, \sin \tilde{\varphi})}{\hat{\epsilon}_0^5(\cos \tilde{\varphi}, \sin \tilde{\varphi})} \right]. \quad (33)$$

Here the $\tilde{\theta}$ integral is elementary: The entire term in the first set of square brackets reads simply $(2\pi)^{-1}$ (for any l). The $\tilde{\varphi}$ integral is l independent. It is not elementary, but it can be expressed explicitly in terms of elliptic integrals. We arrive at the important conclusion that the term $\epsilon_0^{-5} P_{(4)}^\alpha$ contributes to each l -mode $F_S^{\alpha l}$ a *constant amount*, independent of l . Let us denote this constant contribution B^α . To express B^α explicitly in terms of the Boyer–Lindquist coordinates and four-velocity

at z is straightforward but rather tedious [it involves, in particular, writing down the explicit coordinate transformation $(\theta, \varphi) \rightarrow (\tilde{\theta}, \tilde{\varphi})$]. This has been worked out in Ref. [22], and we give the result in the Appendix.

We have one more contribution to $F_S^{\alpha l}$ to consider: that of the first term in Eq. 26. This contribution reads

$$\lim_{\delta r \rightarrow 0^\pm} \frac{L}{2\pi} \int dx dy P_l(\cos \tilde{\theta}) \epsilon_0^{-3} P_{(1)}^\alpha, \tag{34}$$

where we have used the Jacobian $\partial(\tilde{\theta}, \tilde{\varphi})/\partial(x, y) = (\sin \theta)^{-1}$. Here ϵ_0 and $P_{(1)}^\alpha$ are already evaluated at $t = t_0$. The latter is simply a linear combination of δr , x , and y : $P_{(1)}^\alpha = c_r^\alpha \delta r + c_x^\alpha x + c_y^\alpha y$, where from Eqs. 27 and 23 the coefficients are $c_\beta^\alpha = -K_0^{\alpha\delta} (g_{\delta\beta} + u_\delta u_\beta)$. Since ϵ_0 and $\cos \tilde{\theta}$ are even functions of y , we find from symmetry that the term $\propto c_y^\alpha$ does not contribute to the integral in Eq. 34. Furthermore, expanding the Legendre Polynomial in the form $P_l(\cos \tilde{\theta}) = 1 + a_l \tilde{\theta}^2 + O(\tilde{\theta}^4)$ [where $a_l = -l(l + 1)/4$], we find that neither the $\propto a_l$ term nor the $O(\tilde{\theta}^4)$ remainder contribute to the integral: the former by virtue of the same symmetry argument which applied to the term $\epsilon_0^{-7} P_{(7)}^\alpha$, and the latter because it contributes to the integrand an amount of $O(\delta x^2)$, which we may absorb in the remainder term in Eq. 26 and thus ignore.

Thus, the contribution to $F_S^{\alpha l}$ from the term $\epsilon_0^{-3} P_{(1)}^\alpha$ reduces to

$$\lim_{\delta r \rightarrow 0^\pm} \frac{L}{2\pi} \int dx dy \epsilon_0^{-3} (c_r^\alpha \delta r + c_x^\alpha x) \tag{35}$$

(where, recall, ϵ_0^{-3} is evaluated at t_0). Notice this expression depends on l only through the prefactor L . The integral is a standard one, although care must be taken in evaluating the directional limit – how this is done is explained in detail in Ref. [19]. One finds that the term in Eq. 35 has the form⁵ $\pm LA^\alpha$, where the signs correspond to $\delta r \rightarrow 0^\pm$, and where the various Boyer–Lindquist components of A^α are given by [22]

$$\begin{aligned} A^r &= -\frac{\mu^2}{V} \left(\frac{\sin^2 \theta_0}{g_{rr} g_{\theta\theta} g_{\varphi\varphi}} \right)^{1/2} (V + u_r^2/g_{rr})^{1/2}, \\ A^t &= -(u_r/u_t) A^r, \quad A^\theta = A^\varphi = 0, \end{aligned} \tag{36}$$

with $V \equiv 1 + u_\theta^2/g_{\theta\theta} + u_\varphi^2/g_{\varphi\varphi}$.

⁵ It is easy to convince oneself that the two-sided limits $\delta r \rightarrow 0^\pm$ in Eq. 35 are equal in magnitude and different in sign: Notice $\epsilon_0(-\delta r, -x, y) = \epsilon_0(\delta r, x, y)$ (for $\delta t = 0$), which, together with the fact that the integration domain is symmetric in x , implies that the integral in Eq. 35 is an odd function of δr .

In summary, we find that the l modes $F_S^{\alpha l}$ are given *precisely* by

$$F_{S\pm}^{\alpha l} = \pm L A^\alpha + B^\alpha, \quad (37)$$

in which A^α and B^α are l -independent vectorial quantities whose values are given explicitly in Eqs. 36 and 65. As we mentioned before, we expect the l modes of the full field, $F_\pm^{\alpha l}$, to admit the same power expansion in L^{-1} as $F_{S\pm}^{\alpha l}$ [since the sum in Eq. 17 converges exponentially]. Hence, recalling Eq. 16, we find

$$C^\alpha = 0. \quad (38)$$

Furthermore, we find that each of the individual terms in the sum in Eq. 21 vanishes, giving

$$D^\alpha = 0. \quad (39)$$

We comment on a potentially confusing aspect of the above analysis: We have discarded in Eq. 26 those terms of $F_S^\alpha(x)$ that vanish at $x \rightarrow z$. Clearly, the multipole expansion of these neglected terms could contribute to $F_{S\pm}^{\alpha l}$ (e.g., they may well add a term $\propto L^{-2}$ in Eq. 37). Such terms, however, must add up to zero upon summation over l (when evaluated at the particle). They hence affect neither the value of D^α in Eq. 21, nor the value of the final SF in Eq. 20.

This completes the calculation of all necessary regularization parameters. The calculation of the SF through the mode-sum formula (20) now reduces to the problem of obtaining the full-force l -modes $F_\pm^{\alpha l}$ along the orbit. This is usually done numerically, and requires one to solve the appropriate perturbation equations. In the next section we discuss the multitude of numerical strategies proposed for this purpose.

3 Numerical Implementation Strategies

There are two (related) broad issues that need to be addressed in preparing to implement the mode-sum formula in practice (these issues arise, in some form, whether one uses the mode sum scheme or any of the other implementation methods based on the MiSaTaQuWa formalism). The first practical issue has often been referred to as the *gauge problem*: The mode-sum method (as the MiSaTaQuWa formalism underpinning it) is formulated in terms of the metric perturbation in the Lorenz gauge, while standard methods in black hole perturbation theory are formulated in other gauges. The second practical issue concerns the numerical treatment of the point-particle singularity; the problem takes a different form in the frequency domain and in the time domain, and we shall discuss these two frameworks separately below. We start, however, with a brief discussion of the gauge problem and the methods developed to address it.

3.1 Overcoming the Gauge Problem

The calculation of the gravitational SF requires direct information about the local metric perturbation near (and at) the particle. More specifically, one needs to be able to construct the metric perturbation, along with its derivatives, in the Lorenz gauge, at the particle’s location. The mode-sum approach allows us to do so without encountering infinities by considering individual multipoles of the perturbation, but the problem remains how to obtain these multipoles in the desired Lorenz gauge. Unfortunately, standard formulations of black hole perturbations invoke other gauges, favored for their algebraic simplicity. Such gauges are simple because they reflect faithfully the global symmetries of the underlying black hole spacetimes – unlike the Lorenz gauge, which is suitable for describing the locally isotropic particle singularity, but complies less well with the global symmetry of the background.

A common gauge choice for perturbation studies in Schwarzschild is Regge and Wheeler’s [88, 104, 121, 122], in which certain projections of the metric perturbation onto a tensor-harmonic basis are taken to vanish, resulting in a significant simplification of the perturbation equations. Another such useful “algebraic” gauge is the *radiation gauge* introduced by Chrzanowski [38], in which one sets to zero the projection of the perturbation along a principle null direction of the background black hole geometry. Perturbations of the *Kerr* geometry have been studied almost exclusively using the powerful formalism by Teukolsky [114], in which the perturbation is formulated in terms of the Newman–Penrose gauge-invariant scalars, rather than the metric. A reconstruction procedure for the metric perturbation out of the Teukolsky variables (in vacuum) was prescribed by Chrzanowski [38] (with later supplements by Wald [117] and Ori [92]), but only in the radiation gauge. It is not known how to similarly reconstruct the metric in the Lorenz gauge.

Several strategies have been proposed for dealing with this gauge-related difficulty. Some involve a deviation from the original MiSaTaQuWa notion of SF, while other seek to tackle the calculation of the Lorenz-gauge perturbation directly. Here is a survey of the main strategies.

3.1.1 Self-Force in an “Hybrid” Gauge

Equation 7 in Section 1 describes the gauge transformation of the SF. Let us refer here to a gauge as “regular” if the transformation from the Lorenz gauge yields a well-defined SF in that gauge [this requires that the expression in square brackets in Eq. 7 admits a definite and finite particle limit]. It has been shown [18] that the mode-sum formula maintains its form (20), with the same regularization parameters, for any such regular gauges. Namely, for the SF in a specific regular gauge “reg” we have the mode-sum formula

$$F_{\text{self}}^{\alpha}[\text{reg}] = \sum_{l=0}^{\infty} \left[F_{\pm}^{\alpha l}[\text{reg}] \mp LA^{\alpha} - B^{\alpha} \right], \quad (40)$$

where the full-force modes $F_{\pm}^{\alpha l}[\text{reg}]$ are these constructed from the full metric perturbation in the “reg” gauge (we have already set $C^{\alpha} = D^{\alpha} = 0$ here).

Now recall the radiative inspiral problem which motivates us here: Although the momentary SF is gauge dependent, the long-term radiative evolution of the orbit (as expressed, e.g., through the drift of the constants of motion) has gauge-invariant characteristics that should be accessible from the SF in whatever regular gauge. And so, as far as the physical inspiral problem is concerned, one might have hoped to circumvent the gauge problem by simply evaluating the SF in any gauge which is both regular and practical, using Eq. 40. Unfortunately, while the Lorenz gauge itself is “regular but not practical,” both the Regge–Wheeler gauge and the radiation gauge are generally “practical but not regular,” as demonstrated in Ref. [18].

However, a practical solution now suggests itself: Devise a gauge that is regular in the above sense, and yet practical in that it relates to one of the “practical” gauges – say, the radiation gauge – through a simple, explicit gauge transformation (unlike the Lorenz gauge itself). Heuristically, one may picture such a “hybrid” gauge (also referred to as “intermediate” gauge [22]) as one in which the metric perturbation retains its isotropic Lorenz-like form near the particle, while away from the particle it deforms so as to resemble the radiation-gauge perturbation. The SF in such a hybrid gauge would have the mode-sum formula

$$F_{\text{self}}^{\alpha}[\text{hyb}] = \sum_{l=0}^{\infty} \left[F_{\pm}^{\alpha l}[\text{rad}] \mp LA^{\alpha} - B^{\alpha} \right] - \delta D^{\alpha}, \quad (41)$$

where $F_{\pm}^{\alpha l}[\text{rad}]$ are the l modes of the full force in the radiation gauge, and the “counter term” δD^{α} is the difference $F_{\pm}^{\alpha l}[\text{rad}] - F_{\pm}^{\alpha l}[\text{hyb}]$, summed over l and evaluated at the particle. With a suitable choice of the hybrid gauge, the term δD^{α} can be calculated analytically, and Eq. 41 then prescribes a practical way of constructing the SF in a useful, regular gauge, out of the numerically calculated modes of the perturbation in the radiation gauge.

Different variants of this idea were studied by several authors [18,22,85,91,106], but it has not been implemented in full so far.

3.1.2 Generalized SF and Gauge Invariants

Another idea (set out in Ref. [18] and further developed in Ref. [57]) involves the generalization of the SF notion through the introduction of a suitable averaging over angular directions. In some gauges that are not strictly regular in the aforementioned sense, the SF could still be defined in a directional sense. Such is the case in which the expression in square brackets in Eq. 7 has a finite yet direction-dependent particle limit (upon transforming from the Lorenz gauge), and the resulting “directional” SF is bounded for any chosen direction. (This seems to be the situation in the Regge–Wheeler gauge, but not in the radiation gauge – in the latter, the metric perturbation from a point particle develops a stringlike one-dimensional singularity [18].) In such

cases, a suitable averaging over angular directions introduces a well-defined notion of an “average” SF, which generalized the original MiSaTaQuWa SF (it represents a generalization since, obviously, the average SF would coincide with the standard MiSaTaQuWa SF for all regular gauges, including Lorenz’s). The notion of an average SF could be a useful one if it can be used in a simple way to construct gauge-invariant quantities that describe the radiative motion. This is yet to be demonstrated.

A related method invokes the directional SF itself as an agent for constructing the desired gauge invariants. In this approach, one defines (for example) a “Regge–Wheeler” SF by taking a particular directional limit consistently throughout the calculation, and then using the value of this SF to construct the gauge invariants. This approach has been applied successfully, in combination with the mode-sum method, by Detweiler and others [41, 42] (see also Detweiler’s contribution in this volume).

3.1.3 Radiation-Gauge Regularization

Keidl et al. [69, 70] proposed the following construction: Starting with the Lorenz-gauge S field, construct the associated gauge-invariant Newman–Penrose scalar ψ_0^S (or ψ_4^S), and decompose it into spin-weighted spheroidal harmonics. Then obtain the harmonics of the retarded field ψ_0^{ret} by solving the Teukolsky equation with suitable boundary conditions, and (for each harmonic) define the R part through $\psi_0^R \equiv \psi_0^{\text{ret}} - \psi_0^S$. If ψ_0^S is known precisely, then ψ_0^R is a vacuum solution of the Teukolsky equation. To this solution, then, apply Chrzanowski’s reconstruction procedure to obtain a smooth radiation-gauge metric perturbation $h_{\alpha\beta}^R$ [rad], and use that to construct a “radiation gauge” SF. The relation between this definition of the radiation-gauge SF and the one obtained by applying the gauge transformation formula (7) to the standard Lorenz-gauge MiSaTaQuWa SF (the latter, recall, is not well defined) is yet to be investigated.

In reality, the S field is usually known only approximately, resulting in that ψ_0^R retains some nonsmoothness. How to apply Chrzanowski’s reconstruction to nonsmooth potentials is a matter of appreciable technical challenge.

3.1.4 Direct Lorenz-Gauge Implementation

In 2005, Barack and Lousto [15] succeeded in solving the full set of Lorenz-gauge perturbation equations in Schwarzschild, using numerical evolution in $1+1\text{D}$. This development opened the door for a direct implementation of the mode-sum formula in the Lorenz gauge. It later facilitated the first calculations of the gravitational SF for bound orbits in Schwarzschild [24–26].

In Section 4 we shall review this approach in detail. Here we just point to a few of its important advantages: (i) This direct approach obviously circumvents the gauge problem. The entire calculation is done within the Lorenz gauge, and

the mode-sum formula can be implemented directly, in its original form. (ii) The Lorenz-gauge perturbation equations take a fully hyperbolic form, making them particularly suitable for numerical implementation in the time domain. Conveniently, the supplementary gauge conditions (which take the form of elliptic “constraint” equations) can be made to hold automatically, as we discuss in Section 4. (iii) In this approach one solves directly for the metric perturbation components, without having to resort to complicated reconstruction procedures. This is an important advantage because metric reconstruction involves differentiation of the field variables, which inevitably results in loss of numerical accuracy. (iv) Working with the Lorenz-gauge perturbation components as field variables is also advantageous in that these behave more regularly near point particles than do Teukolsky’s or Moncrief’s variables. This has a simple manifestation, for example, within the 1+1D treatment in Schwarzschild: The individual multipole modes of the Lorenz-gauge perturbation are always *continuous* at the particle, just like in the simple example of Section 2.1; on the other hand, the multipole modes of Teukolsky’s or Moncrief’s variables are *discontinuous* at the particle, and so are, in general, the modes of the metric perturbation in the Regge-Wheeler gauge. Obviously, the regularity of the Lorenz-gauge modes makes them more suitable as numerical variables.

3.2 Numerical Representation of the Point Particle

Common to all numerical implementation methods is the basic preliminary task of solving the field equations (in whatever formulation) for the full (retarded) perturbation from a point particle in a specified orbit. This immediately brings about the practical issue of the numerical representation of the particle singularity. The particulars of the challenge depend on the methodological framework: In time-domain methods one faces the problem of dealing with the irregularity of the field variables near the worldline; in frequency-domain (spectral) treatments, such irregularity manifests itself in a problematic high-frequency behavior. We now survey some of the relevant methods.

3.2.1 Particle Representation in the Time Domain

Extended-Body Representations In the context of fully nonlinear Numerical Relativity, the problem of a binary black with a small mass ratio remains a difficult challenge, because of the large span of lengthscales intrinsic in this problem. (Current Numerical Relativistic technology can handle mass ratios as small as 1 : 10 [56] – still nothing near the $\sim 1 : 10^5$ ratio needed for LISA EMRI applications.) Bishop et al. [28] attempted a Numerical Relativistic treatment in which the particle is modeled by a quasi-rigid widely extended body whose “center” follows a geodesic. Comparison with perturbation results did not show sufficient accuracy, and the method requires further development.

An extended-body approach has also been implemented in perturbative studies. Khanna et al. [31, 71, 74] solved the Teukolsky equation in the time domain (i.e., in $2 + 1\text{D}$, for each azimuthal m mode) with a source “particle” represented by a narrow Gaussian distribution. This crude technique was much improved recently by Sundararajan et al. [112, 113] using a “finite impulse representation,” whereby the source is described by a series of spikes whose relative magnitude is carefully controlled so as to assure that the source has integral properties similar to that of a delta function. Such methods were demonstrated to reproduce wave-zone solutions with great accuracy (indeed, that is what they are designed to do), but they are likely to remain less useful for computing the accurate local perturbation near the particle as required for SF calculations.

Extended-body representations suffer from the inevitable trade-off between smoothness and localization: One can only smoothen the solution by making it less localized, and one can better localize it only by making it less smooth. In what follows we concentrate on methods in which the source particle is *precisely* localized on the orbit: The energy–momentum of the particle is represented by a delta-function source term (as in Eq. 10, or Eq. 44 below), and the delta distribution is treated analytically within the numerical scheme, in an exact manner.

Delta-Function Representation in 1+1D In full $3+1\text{D}$ spacetime, the full (retarded) metric perturbation obviously diverges toward the particle (on any given time slice). The divergence is asymptotically Coulomb-like in the Lorenz gauge (and can take a different form in other gauges). In spherically symmetric spacetimes, one can decompose the perturbation into tensor harmonics and solve a separated version of the field equations in $1+1\text{D}$ (time + radius) for each harmonic separately. In the particle problem this becomes beneficial not only thanks to the obvious dimensional reduction, but also because it mitigates the problematics introduced by the particle’s singularity: The angular integration involved in constructing the individual l modes effectively “smears” the Coulomb-like singularity across the surface of a 2-sphere, and the resulting l modes are finite even at the location of the particle. Furthermore, in the case of the metric perturbation in the Lorenz gauge, the individual l modes are also continuous at the particle (cf. Eq. 14 in Section 1). The corresponding l modes of the Teukolsky or Moncrief gauge-invariant variables are generally *not* continuous at the particle, and neither are the modes of the metric perturbation in non-Lorenz gauges.

The boundedness of the l modes is, of course, a crucial feature of the l -mode regularization scheme, as we have already discussed. That same feature also greatly simplifies the $1+1\text{D}$ numerical treatment of the particle. Lousto and Price [78] formulated a general method for incorporating a delta-function source in a finite-difference treatment of the field equations in $1+1\text{D}$. In this method, the finite-difference approximation at a numerical grid cell (in $t-r$ space) traversed by the particle’s world line is obtained, essentially, by integrating the field equation “by hand” over the grid cell at the required accuracy. The original (1D) delta function present in the source term thereby integrates out to contribute a finite term at each time step. The original Lousto–Price scheme (formulated with a first-order global convergence rate) was later improved by Martel and Poisson [80] (second-order

convergence) and Lousto [76] (fourth-order convergence). This simple but powerful idea is at the core of many of the 1+1D finite-difference implementations presented in the last few years [24, 59, 79], including the work discussed in Section 4.

Despite such advances, 1+1D particles remain numerically expensive to handle, because the nonsmoothness associated with them introduces a large scale variance in the solutions: The l -mode field gradients grow sharply near the particle, and, moreover, become increasingly more difficult to resolve with larger l (recall the l -mode gradient is $\propto l$ at large l). The mode-sum formula, recall, converges rather slowly (like $\sim 1/l$), and so requires one to compute a considerably large number of modes (typically ~ 20 with even a moderate accuracy goal). This proves to stretch the limit of what can be achieved today using finite differentiation on a fixed mesh.

Several methods have been proposed to address this problem in the current context. Sopena and collaborators [35, 36, 110, 111] explored the use of finite-element discretization. This technique is particularly powerful in dealing with multi-scale problems, and, being quasi-spectral, it benefits from an exponential convergence rate. So far it has been applied successfully for generic orbits in Schwarzschild, and higher-dimensional implementations (for Kerr studies) are currently being considered. A related quasi-spectral scheme was recently suggested by Field et al. [51]. Finally, Thornburg [115] very recently developed an adaptive mesh refinement algorithm for Lousto–Price’s finite differences scheme (with a global fourth-order convergence). This was successfully implemented for a scalar charge in a circular orbit in Schwarzschild, and generalizations are being considered.

Puncture Methods In Kerr spacetime, one no longer benefits from a 1+1D separability. The Lorenz-gauge perturbation equations are only separable into azimuthal m -modes, each a function of t, r, θ in a 2+1D space. The m modes are not finite on the worldline, but rather they diverge there logarithmically (see the discussion in Section II.C of Ref. [11]). Since the 2+1D numerical solutions are truly divergent, a direct finite-difference treatment becomes problematic. However, since the singular behavior of the perturbation can be approximated analytically, a simple remedy to this problem suggests itself.

The idea, which has recently been studied independently by several groups [11, 77, 116], is to utilize a new perturbation variable for the numerical time evolution (the “residual” field), constructed from the full (retarded) field by subtracting a suitable function (the “puncture” field), given analytically, which approximates the singular part of the perturbation well enough that the residual field is (at least) bounded at the particle. The perturbation equations are then recast with the residual field as their independent variable, and with a new source term (depending on the puncture field and its derivatives) which now extends off the worldline but contains no delta function. The equations are then solved for the residual field in the time domain, using, for example, standard finite differentiation.

Several variants of this method have been studied and tested with scalar-field codes in 1+1D [116] and 2+1D [11, 77], and also proposed for use in full 3+1D [69]. The various schemes differ primarily in the way they handle the puncture function far from the particle: Barack and Golbourn [11] introduce a puncture with a

strictly compact support around the particle, Detweiler and Vega [116] truncate it with a smooth attenuation function, and in Lousto and Nakano [77] the puncture is not truncated at all. We will discuss the puncture method in more detail in Section 5.

To obtain the necessary input for the SF mode-sum formula, the 2+1D (or 3+1D) numerical solutions need to be decomposed into l modes, in what then becomes a somewhat awkward procedure (we decompose the field into separate l modes just to add these modes all up again after regularization). Fortunately, there is a more direct alternative: Barack et al. [12] showed how the SF can be constructed directly, in a simple way, from the 2+1D m modes of the residual field (assuming only that these mode are differentiable at the particle, which is achieved by designing a suitable puncture). This direct “ m mode regularization” scheme, too, will be described in Section 5. It is hoped that this technique could provide a natural framework for calculations in Kerr. It is yet to be applied in practice.

3.2.2 Particle Representation in the Frequency Domain: the High-Frequency Problem and Its Resolution

The l modes required as input for the SF mode-sum formula can also be obtained using a spectral treatment of the field equations. This has the obvious advantage that one then only deals with ordinary differential equations, although constructing the l modes involves the additional step of summing over sufficiently many frequency modes. (See [27] for a recent analysis of the relative efficiencies of frequency vs. time domain treatments.) As with the time-domain methods discussed above, the representation of the particle in the frequency domain too brings about technical complications, but these now take a different form.

To illustrate the problem, consider the toy model of a scalar charge in Schwarzschild, allowing the particle to move on some bound (eccentric) geodesic of the background, with radial location given as a function of time by $r = r_p(t)$. Decompose the scalar field in spherical harmonics, and denote the multipolar mode functions by $\phi_{lm}(t, r)$. The time-domain modes ϕ_{lm} are continuous along $r = r_p(t)$ for each l, m . However, the derivatives $\phi_{lm,r}$ and $\phi_{lm,t}$ will generally suffer a finite jump discontinuity across $r = r_p(t)$ (recall our elementary example in Section 2.1), which reflects the presence of a source “shell” representing the l, m mode of the scalar charge. In particular, if the orbit is eccentric, the derivatives of ϕ_{lm} will generally be discontinuous functions of t at a fixed value of r along the orbit.

Now imagine trying to reconstruct $\phi_{lm}(t, r)$ (for some fixed r along the orbit) as a sum over its Fourier components:

$$\phi_{lm}(t, r) = \sum_{\omega} R_{lm\omega}(r) e^{-i\omega t}. \quad (42)$$

Since, for an eccentric orbit, $\phi_{lm}(t, r)$ is only a C^0 function of t at the particle’s worldline, it follows from standard Fourier theory [68] that the Fourier sum in Eq. 42 will only converge there like $\sim 1/\omega$. The actual situation is even worse, because for

SF calculations we need not only ϕ_{lm} but also its derivatives. Since, for example, $\phi_{lm,r}$ is a discontinuous function of t , we will inevitably face here the well known ‘‘Gibbs phenomenon’’: the Fourier sum will fail to converge to the correct value at $r \rightarrow r_p(t)$. Of course, the problematic behavior of the Fourier sum is simply a consequence of our attempt to reconstruct a discontinuous function as a sum over smooth harmonics.

From a practical point of view this would mean that (i) at the coincidence limit $r \rightarrow r_p(t)$ the sum over ω modes would fail to yield the correct one-sided values of $\phi_{lm,r}$, however many ω modes are included in the sum; and (ii) if we reconstruct $\phi_{lm,r}$ at a point $r = r_0$ off the worldline, then the Fourier series should indeed formally converge, however the number of ω modes required for achieving a prescribed precision would grow unboundedly as r_0 approaches $r_p(t)$, making it extremely difficult to evaluate $\phi_{lm,r}$ at the coincidence limit.

This technical difficulty is rather generic, and will show also in calculations of the local EM and gravitational fields. The situation is no different in the Kerr case, because there too the mode-sum formula requires as input the *spherical*-harmonic modes of the perturbation field, and for each such mode the source is represented by a δ -distribution on a thin shell, which renders the field derivatives discontinuous across that shell. The problem takes an even more severe form when considering EM or gravitational perturbations via the Teukolsky formalism: Here, the l, m modes of the perturbation fields (now the Newman–Penrose scalars) are not even continuous at the particle’s orbit – a consequence of the fact that the source term for Teukolsky’s equation involves derivatives of the electric four-current or the energy–momentum tensor (a single derivative in the EM case; a second derivative in the gravitational case). Again, a naive attempt to construct these multipoles as a sum over their ω modes will be hampered by the Gibbs phenomenon.

A simple and elegant way around the problem was proposed recently in Ref. [23]. It was shown how the desired values of the field and its derivatives at the particle can be constructed from a sum over properly weighted *homogeneous* (source-free) radial functions $R_{lm\omega}(r)$, instead of the actual inhomogeneous solutions of the frequency-domain equation. The Fourier sum of such homogeneous radial functions, which are smooth everywhere, converges exponentially fast. The Fourier sum of the derivatives, which are also smooth, is likewise exponentially convergent. The validity of the method (and the exponential convergence) was demonstrated in Ref. [23] with an explicit numerical calculation in the scalar-field monopole case ($l = 0$). It was later implemented in a frequency-domain calculation of the monopole and dipole modes of the Lorenz gauge metric perturbation for eccentric orbits in Schwarzschild [13, 25, 26]. The same method should be applicable for any of the other problems mentioned above, including the calculation of EM and gravitational perturbations using Teukolsky’s equation.

The method of Ref. [23] (dubbed *method of extended homogeneous solutions*) completely circumvents the problem of slow convergence (or the lack thereof) in frequency-domain calculations involving point sources. It makes the frequency-domain approach an attractive method of choice for SF calculations. The method is now being implemented in first calculations of the scalar-field SF for Kerr orbits [119].

4 An Example: Gravitational Self-Force in Schwarzschild Via 1+1D Evolution in Lorenz Gauge

As an example of a fully worked out calculation of the SF, we review here the work by Barack and Sago on eccentric geodesic in Schwarzschild [24, 26]. This work represents a direct implementation of the mode-sum formula in its original form (20). The decomposed Lorenz-gauge metric perturbation equations are integrated directly using numerical evolution in 1+1D. The numerical algorithm employs a straightforward fourth-order-convergent finite-difference scheme (using a version of the Lousto–Price method) on a staggered numerical mesh based on characteristic coordinates. Below we briefly describe the perturbation formalism, discuss the numerical implementation in some more detail, and present some results.

4.1 Lorenz-Gauge Formulation

The linearized Einstein equation in the Lorenz gauge takes the form

$$\square \bar{h}_{\alpha\beta} + 2R^\mu{}_\alpha{}^\nu{}_\beta \bar{h}_{\mu\nu} = -16\pi T_{\alpha\beta}, \quad (43)$$

where $\bar{h}_{\mu\nu}$ is the trace-reversed metric perturbation (Eq. 1), \square is the covariant D’Alembertian operator, $R_{\alpha\beta\gamma\delta}$ is the Riemann tensor of the background spacetime, and $T_{\alpha\beta}$ is the source’s energy–momentum tensor. For a point particle of mass μ moving on a timelike geodesic $x_p^\mu(\tau)$, the latter is modeled as

$$T_{\alpha\beta}(x^\mu) = \mu \int_{-\infty}^{\infty} (-g)^{-1/2} \delta^4[x^\mu - x_p^\mu(\tau)] u_\alpha u_\beta d\tau, \quad (44)$$

where g is the background metric’s determinant and u^α is the particle’s four-velocity. Eq. 43 is a linear, diagonal hyperbolic system, which admits a well-posed initial-value formulation on a spacelike Cauchy hypersurface (see, e.g., Theorem 10.1.2 of [118]). Furthermore, if the gauge conditions (Eq. 1) are satisfied on the initial Cauchy surface, then they are guaranteed to hold everywhere (assuming that Eq. 43 is satisfied everywhere and that $T_{\alpha\beta}{}^{;\beta} = 0$, as in our case).⁶

We now specialize to eccentric geodesics around a Schwarzschild black hole, and employ a Schwarzschild coordinate system (t, r, θ, φ) in which the orbit is equatorial ($\theta_p = \pi/2$). Such orbits constitute a two-parameter family; we may characterize each orbit by the radial “turning points” r_{\min} and r_{\max} , or alternatively by the “semi-latus rectum” $p \equiv 2(r_{\max} r_{\min}) / (r_{\max} + r_{\min})$ and “eccentricity” $e \equiv (r_{\max} - r_{\min}) / (r_{\max} + r_{\min})$.

⁶ The gauge conditions (1) do not fully specify the gauge: There is a residual gauge freedom within the family of Lorenz gauges, $h_{\alpha\beta} \rightarrow h_{\alpha\beta} + \xi_{\alpha;\beta} + \xi_{\beta;\alpha}$, with any ξ^μ satisfying $\square \xi^\mu = 0$. It is easy to verify that both Eqs. 1 and 43 remain invariant under such gauge transformations.

Barack and Lousto [15] decomposed the metric perturbation into tensor harmonics, in the form

$$\bar{h}_{\alpha\beta} = \frac{\mu}{r} \sum_{l,m} \sum_{i=1}^{10} \bar{h}^{(i)lm}(r,t) Y_{\alpha\beta}^{(i)lm}(\theta,\varphi), \quad (45)$$

and similarly for the source $T_{\alpha\beta}$. The harmonics $Y_{\alpha\beta}^{(i)lm}(\theta,\varphi)$ (whose components are constructed from ordinary spherical harmonics and their first and second derivatives) form a complete orthogonal basis for second-rank covariant tensors on a 2-sphere (see Appendix A of [15]). The time-radial functions $\bar{h}^{(i)lm}$ ($i = 1, \dots, 10$) form our basic set of perturbation fields, and serve as variables for the numerical evolution.⁷ The tensor-harmonic decomposition decouples Eq. 43 with respect to l, m , although not with respect to i : For each l, m , the variables $\bar{h}^{(i)lm}$ satisfy a coupled set of hyperbolic (in a 1+1D sense) scalar-like equations, which may be written in the form

$$\square^{(2)} \bar{h}^{(i)lm} + \mathcal{M}_{(j)}^{(i)l} \bar{h}^{(j)lm} = S^{(i)lm} \quad (i = 1, \dots, 10). \quad (46)$$

Here $\square^{(2)}$ is the 1+1D scalar field wave operator $\partial_{uv}^2 + V(r)$, where v and u are the standard Eddington–Finkelstein null coordinates, and $V(r) = \frac{1}{4}(1 - 2M/r) [2M/r^3 + l(l+1)/r^2]$ is an effective potential. The “coupling” terms $\mathcal{M}_{(j)}^{(i)l} \bar{h}^{(j)lm}$ involve first derivatives of the $\bar{h}^{(j)lm}$ ’s at most (no second derivatives), so that, conveniently, the set (46) decouples at its principal part. The decoupled source terms $S^{(i)lm}$ are each $\propto \delta[r - r_p(\tau)]$ (no derivatives of δ function) and, as a result, the physical solutions $\bar{h}^{(j)lm}$ are continuous even at the particle. Explicit expressions for the coupling terms and the source terms in Eq. 46 can be found in Ref. [15].

In addition to the evolution equations (46), the functions $\bar{h}^{(i)lm}$ also satisfy four first-order elliptic equations, which arise from the separation of the gauge conditions (1) into l, m modes. In the continuum initial-value problem, the solutions $\bar{h}^{(i)lm}$ satisfy these “constraints” automatically if only they satisfy them on the initial Cauchy surface. This is more difficult to guarantee in a finite-difference treatment, where (i) it is often impossible to prescribe exact initial data that satisfy the constraints, and (ii) discretization errors may amplify constraint violations over the evolution. Inspired by a remedy proposed for a similar problem in the context of nonlinear Numerical Relativity [58], Ref. [15] proposed the inclusion of “divergence dissipation” terms, $\propto h_{\alpha\beta}^{;\beta}$, in the original set (43), so designed to guarantee that any violations of the Lorenz-gauge conditions are efficiently damped during the evolution. These damping terms modify only the explicit form of the \mathcal{M} terms in Eq. 46 as shown in [15].

⁷ To simplify the appearance of Eq. 45 we have used here a slightly different normalization than that of [15] for the functions $\bar{h}^{(i)lm}$.

4.2 Numerical Implementation

The code developed by Barack and Sago [24, 26] solves the coupled set (46) (with constraint dissipation terms incorporated in the \mathcal{M} terms) via time evolution. The numerical domain, covering a portion of the external Schwarzschild geometry, is depicted in Fig. 3. The numerical grid is based on Eddington–Finkelstein null coordinates v, u , and initial data (the values of the 10 fields $\bar{h}^{(i)lm}$ for each l, m) are specified on two characteristic initial surfaces $v = \text{const}$ and $u = \text{const}$. Equations 46 are then discretized on this grid using a finite-difference algorithm which is globally fourth-order convergent. The numerical integrator solves for the various $\bar{h}^{(i)lm}$'s along consecutive $v = \text{const}$ rays. A particularly convenient feature of this setup is that no boundary conditions need be specified (the characteristic grid has no causal boundaries). Moreover, one need not be at all concerned with the determination of faithful initial conditions: It is sufficient to set $\bar{h}^{(i)lm} = 0$ on the initial surfaces and simply let the resulting spurious radiation (which emanate from the intersection of the particle's worldline with the initial surface) dissipate away over the evolution time. The early part of the evolution, which is typically dominated by such spurious radiation, is simply discarded.

The conservative modes $l = 0$ and $l = 1$ (the monopole and dipole, respectively) require a separate treatment, as they do not evolve stably using the above numerical scheme. (A naive attempt to evolve these modes leads to numerical instabilities which, so far, could not be cured.) Fortunately, the set (46) simplifies considerably for these modes, and solutions can be obtained in a semi-analytic manner based on physical considerations. Detweiler and Poisson [43] worked out the $l = 0, 1$ Lorenz-gauge solutions for circular orbits, and their work is generalized to eccentric orbits in Ref. [13], relying on the aforementioned method of extended homogeneous solutions. The calculation of [13] yields the values of the fields $\bar{h}^{(i)lm}$ and their derivatives for $l = 0, 1$.

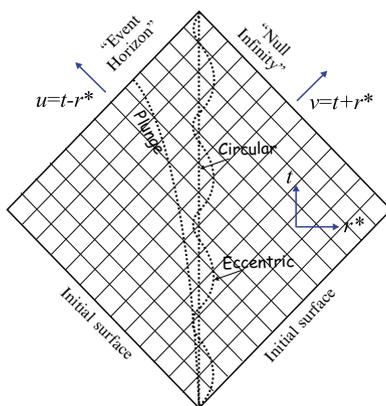


Fig. 3 The numerical 1+1D domain in the Barack–Sago code [24, 26]: A staggered mesh based on characteristic (Eddington–Finkelstein) coordinates v, u . t and r^* are, respectively, the Schwarzschild time and tortoise radial coordinates. The evolution proceeds from characteristic initial data on two null surfaces. Illustrated are a few sample geodesic orbits (radial plunge, circular, eccentric)

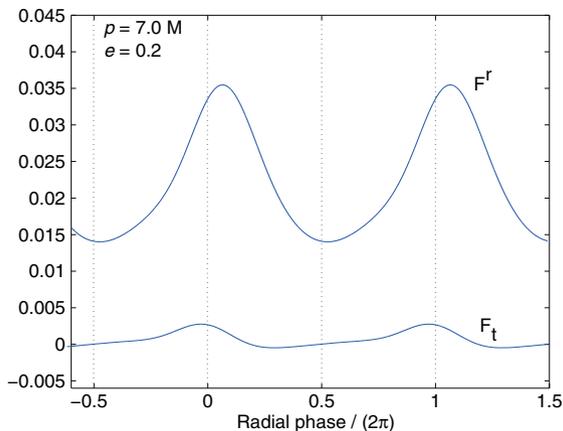


Fig. 4 The gravitational SF [in units of $(\mu/M)^2$] along a Schwarzschild geodesic with semi-latus rectum $p = 7M$ and eccentricity $e = 0.2$. The upper and lower graphs show F_{self}^r and F_t^{self} , respectively. Integer values on the horizontal axis correspond to periaapses ($r = r_{\text{min}}$)

To construct the l modes $F_{\pm}^{\alpha l}$ [the necessary input for the mode-sum formula (20)], one first substitutes the expansion (45) in the expression for the full force F^{α} (left-hand expression in Eq. 9), and then formally expands the result in spherical harmonics. The outcome is a formula for $F_{\pm}^{\alpha l}$ given in terms of the various fields $\bar{h}^{(i)l'm}$ and their derivatives (evaluated at the particle and summed over m), where, if the fixed- k extension is used, one finds in general $l - 3 \leq l' \leq l + 3$. One then uses the numerical values of the fields $\bar{h}^{(i)lm}$ and their derivatives, as calculated along the orbit using the above code, to construct $F_{\pm}^{\alpha l}$ for sufficiently many l modes. The mode-sum formula (20) then gives the SF. Figure 4 shows the final result for an eccentric geodesic with $p = 7M$ and $e = 0.2$.

The calculation we have just described represents a milestone in the SF program: We now finally have a code capable of tackling the generic EMRI-relevant SF problem in Schwarzschild. (Indeed, to address the fully generic physical problem would require a final generalization to the Kerr case!) Using this code, and similar codes developed independently by others [42], we can now begin to explore the physical consequences of the gravitational SF – particularly those effects associated with its conservative piece. Work done so far includes (i) study of gauge invariant SF effects on circular orbits [42], (ii) comparison of SF calculations and results from PN theory in the weak field regime [42, 109], (iii) comparison of SF results from different calculation schemes and using different gauges [109], and (iv) a calculation of the shift in the last stable circular orbit (and its frequency) due to the conservative piece of the SF [25]. Work is in progress to calculate SF-related precessional effects for eccentric orbits.

5 Toward Self-Force Calculations in Kerr: the Puncture Method and m -Mode Regularization

The time-domain Lorenz-gauge treatment of Section 4 relies crucially on the separability of the field equations into (tensorial) spherical harmonics, which is no longer possible in Kerr. In the Kerr problem one can at best separate the metric perturbation equations into azimuthal m modes, using the substitution⁸

$$\bar{h}_{\alpha\beta} = \sum_{m=-\infty}^{\infty} \bar{h}_{\alpha\beta}^m(t, r, \theta) e^{im\varphi}. \quad (47)$$

Then one faces solving the coupled set for the 2+1D variables $\bar{h}_{\alpha\beta}^m$. This, nowadays, is easily within the capability of even a standard desktop computer; indeed, over the past decade, 2+1D numerical evolution has been a method of choice in many studies of Kerr perturbations [11, 71–74, 77, 95, 112]. Alas, the introduction of a δ -function source in a 2+1D evolution is problematic, since the variables $\bar{h}_{\alpha\beta}^m$ then suffer a singularity near the particle. The *puncture* method, which we have mentioned briefly in Section 3.2, overcomes this technical difficulty. In this section we review this method (as implemented by Barack and Golbourn [11]) in some more detail. We also describe a new, ad hoc, regularization method – the *m-mode regularization scheme* – which enables a straightforward construction of the SF directly from the 2+1D numerical solutions, without resorting to a multipole decomposition.

5.1 Puncture Method in 2+1D

In Section 1 we have split the (trace-reversed) metric perturbation as $\bar{h}_{\alpha\beta} = \bar{h}_{\alpha\beta}^S + \bar{h}_{\alpha\beta}^R$, with the gravitational SF then obtained from the smooth field $\bar{h}_{\alpha\beta}^R$ as prescribed in Eq. 4. Here we introduce a new splitting,

$$\bar{h}_{\alpha\beta} = \bar{h}_{\alpha\beta}^S + \bar{h}_{\alpha\beta}^R = \bar{h}_{\alpha\beta}^{\text{punc}} + \bar{h}_{\alpha\beta}^{\text{res}}, \quad (48)$$

so that

$$\bar{h}_{\alpha\beta}^R = \bar{h}_{\alpha\beta}^{\text{res}} + (\bar{h}_{\alpha\beta}^{\text{punc}} - \bar{h}_{\alpha\beta}^S). \quad (49)$$

We denote the m modes of these various quantities, defined as in Eq. 47, by $\bar{h}_{\alpha\beta}^{\text{R},m}$, $\bar{h}_{\alpha\beta}^{\text{res},m}$, etc. The new splitting is defined (in a nonunique way) by devising a *puncture*

⁸ A full separation of variables in Kerr is possible within Teukolsky’s formalism, which, alas, brings about the metric reconstruction and gauge-related difficulties discussed in previous chapters. A full separation of the metric perturbation equations themselves, in Kerr, has not been formulated yet, to the best of our knowledge.

field $\bar{h}_{\alpha\beta}^{\text{punc}}$, given analytically, which approximates $\bar{h}_{\alpha\beta}^{\text{S}}$ near the particle well enough that the m modes of the resulting *residual* field $\bar{h}_{\alpha\beta}^{\text{res}}$ are continuous and differentiable along the particle’s worldline (and anywhere else). A particular such puncture is prescribed below (Eq. 56). The form of the puncture function away from the particle can be chosen as convenient; for example, in such a way that it can be decomposed into m modes explicitly, in analytic form (as in Ref. [11], for a scalar-field toy model).

The Lorenz-gauge perturbation equations (43) are now written in the form

$$\square \bar{h}_{\alpha\beta}^{\text{res}} + 2R^{\mu}{}_{\alpha}{}^{\nu}{}_{\beta} \bar{h}_{\mu\nu}^{\text{res}} = S_{\alpha\beta}^{\text{res}}, \quad (50)$$

with

$$S_{\alpha\beta}^{\text{res}} \equiv -16\pi T_{\alpha\beta} - \square \bar{h}_{\alpha\beta}^{\text{punc}} - 2R^{\mu}{}_{\alpha}{}^{\nu}{}_{\beta} \bar{h}_{\mu\nu}^{\text{punc}}. \quad (51)$$

The support of the source $S_{\alpha\beta}^{\text{res}}$ now extends beyond the particle’s worldline, but it contains no δ function on the worldline.⁹ The equations are separated into m modes, with the m -mode source given by

$$S_{\alpha\beta}^{\text{res},m}(t, r, \theta) \equiv \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{\alpha\beta}^{\text{res}}(t, r, \theta, \varphi) e^{-im\varphi} d\varphi. \quad (52)$$

If the puncture is sufficiently simple, the source $S_{\alpha\beta}^{\text{res},m}$ can be evaluated in closed form (as in [11]). The m -mode field equations for the variables $\bar{h}_{\alpha\beta}^{\text{res},m}(t, r, \theta)$, which are everywhere continuous and differentiable, can now be integrated numerically using straightforward finite differentiation on a $2+1\text{D}$ grid.

To ease the imposition of boundary conditions for $\bar{h}_{\alpha\beta}^{\text{res},m}$, it is convenient to suppress the support of the puncture $\bar{h}_{\alpha\beta}^{\text{punc}}$ away from the particle, so that the physical boundary conditions for $\bar{h}_{\alpha\beta}^{\text{res}}$ become practically identical to that of the full retarded field $\bar{h}_{\alpha\beta}$. In [11] this is achieved in a simple way by introducing an auxiliary “worldtube” around the particle’s worldline (in the $2+1\text{D}$ space): Inside this worldtube one solves the “punctured” m -mode equations for $\bar{h}_{\alpha\beta}^{\text{res},m}$, while outside the worldtube one uses the original, retarded-field m -modes $\bar{h}_{\alpha\beta}^m$ as evolution variables; the value of the evolution variables is adjusted on the boundary of the worldtube using $\bar{h}_{\alpha\beta}^m = \bar{h}_{\alpha\beta}^{\text{res},m} + \bar{h}_{\alpha\beta}^{\text{punc},m}$.

Two very similar variants of the puncture scheme have been developed and implemented independently by two groups [11, 77] – both for the toy model of a scalar charge on a circular orbit around a Schwarzschild black hole (refraining from a spherical harmonics decomposition and instead working in $2+1\text{D}$). The method is yet to be applied to Kerr and to gravitational perturbations.

⁹ This can be shown by integrating $S_{\alpha\beta}^{\text{res}}$ over a small 3-ball containing the particle (at a given time), and then inspecting the limit as the radius of the ball tends to zero [11].

5.2 *m-Mode Regularization*

The 2+1D numerical solutions obtained using the puncture method can be used to construct the input modes $F_{\pm}^{\alpha l}$ for the mode-sum formula (20), but this would require the additional step of a decomposition in spherical harmonics. It appears, however, that there is a simple way to construct the SF directly from the residual modes $\bar{h}_{\alpha\beta}^{\text{res},m}$, without resorting to a multipole decomposition. Such “*m*-mode regularization” procedure was prescribed recently in Ref. [12] for the scalar, EM, and gravitational SFs. We describe it here as applied to the gravitational SF.

In analogy with Eq. 9, let us define the “force” fields

$$F_{\text{res}}^{\alpha}(x) = \mu k^{\alpha\beta\gamma\delta} \bar{h}_{\beta\gamma;\delta}^{\text{res}}, \quad F_{\text{punc}}^{\alpha}(x) = \mu k^{\alpha\beta\gamma\delta} \bar{h}_{\beta\gamma;\delta}^{\text{punc}}. \quad (53)$$

Then, recalling Eqs. 4 and 49, we may write the SF at a point z along the orbit as

$$F_{\text{self}}^{\alpha}(z) = \lim_{x \rightarrow z} \left[F_{\text{res}}^{\alpha}(x) + \left(F_{\text{punc}}^{\alpha}(x) - F_{\text{S}}^{\alpha}(x) \right) \right]. \quad (54)$$

Recall that the expression in square brackets ($= \mu k^{\alpha\beta\gamma\delta} \bar{h}_{\beta\gamma;\delta}^{\text{R}}$) is a smooth (analytic) function of x , and so the limit $x \rightarrow z$ is well defined.

Now consider a particular puncture function, prescribed as follows. For an arbitrary spacetime point x outside the black hole, let Σ_x be the spatial hypersurface $t = \text{const}$ containing x , let $\bar{\tau}(t)$ be the value of τ at which Σ_x intersects the particle’s worldline, and denote $\bar{z}(t) \equiv z[\bar{\tau}(t)]$ and $\bar{u}^{\alpha} \equiv u^{\alpha}(\bar{z})$. For given x define the coordinate distance $\delta x^{\alpha} \equiv x^{\alpha} - \bar{z}^{\alpha}(t)$ [with Boyer–Lindquist components $(0, \delta r, \delta\theta, \delta\varphi)$], and construct the three quantities \bar{S}_0 , \bar{S}_1 , and $\delta\bar{u}_{\alpha}$ defined just as S_0 , S_1 , and δu_{α} in Eqs. 23–25, with z replaced by \bar{z} . For $|\delta x|$ *very small*, the sum $\bar{S}_0 + \bar{S}_1$ approximates the squared geodesic distance from x to the worldline, and $\bar{u}_{\alpha}(z) + \delta\bar{u}_{\alpha}$ approximates the four-velocity \hat{u}_{α} parallelly propagated from \bar{z} to x :

$$\bar{S}_0 + \bar{S}_1 = \epsilon^2 + O(\delta x^4), \quad \bar{u}_{\alpha} + \delta\bar{u}_{\alpha} = \hat{u}_{\alpha} + O(\delta x^2) \quad (55)$$

(recall Eq. 22). We emphasize, however, that here the quantities \bar{S}_0 , \bar{S}_1 , and $\delta\bar{u}_{\alpha}$ are given globally, in closed form, for any x outside the black hole. The puncture function we wish to consider is given, for any x , by

$$\bar{h}_{\alpha\beta}^{\text{punc}}(x) = \frac{4\mu (\bar{u}_{\alpha}\bar{u}_{\beta} + \delta\bar{u}_{\alpha}\bar{u}_{\beta} + \bar{u}_{\alpha}\delta\bar{u}_{\beta})}{(\bar{S}_0 + \bar{S}_1)^{1/2}}. \quad (56)$$

This function can be attenuated far from the particle in order to control its global properties, but such modifications will not affect our discussion here.

We now recall the form of the “physical” singular field $\bar{h}_{\alpha\beta}^{\text{S}}$ expressed in Eq. 5. This expression applies to any nearby points z and x , but here we wish to apply it with x being a given point near the worldline and $z = \bar{z}(t)$ being the associated

worldline point on Σ_x . It is not difficult to show [12] that, near the particle, the difference between the puncture (56) and the S field has the form

$$\bar{h}_{\alpha\beta}^{\text{punc}} - \bar{h}_{\alpha\beta}^{\text{S}} = \bar{\epsilon}_0^{-3} P_{\alpha\beta}^{(4)}(\delta x) + \text{const} + O(\delta x^2), \quad (57)$$

where $\bar{\epsilon}_0 \equiv \bar{S}_0^{1/2}$, and $P_{\alpha\beta}^{(4)}$ is a smooth function of the coordinate differences δx^α , of homogeneous order $(\delta x)^4$. It follows that the difference between the corresponding force fields has the local form

$$F_{\text{punc}}^\alpha(x) - F_S^\alpha(x) = \bar{\epsilon}_0^{-5} P_{(5)}^\alpha(\delta x) + O(\delta x) \quad (58)$$

(for any smooth k -extension), where $P_{(5)}^\alpha$ is yet another smooth function, now of homogeneous order $(\delta x)^5$. Notice that $F_{\text{punc}}^\alpha - F_S^\alpha$ is bounded but generally discontinuous (direction dependent) at $x \rightarrow \bar{z}$. From Eq. 54 it then follows that F_{res}^α too is bounded but discontinuous at $x \rightarrow \bar{z}$ [since the limit of the entire expression in square brackets in (54) is known to be definite].

We now arrive at the crucial step of our discussion. In Eq. 54, for a given point z along the particle's worldline, we express the (analytic) function in square brackets as a formal sum over m modes, in the form

$$F_{\text{self}}^\alpha(z) = \lim_{x \rightarrow z} \sum_{m=-\infty}^{\infty} \left[F_{\text{res}}^{\alpha,m}(x) + \left(F_{\text{punc}}^{\alpha,m}(x) - F_S^{\alpha,m}(x) \right) \right], \quad (59)$$

where

$$F_{\text{res}}^{\alpha,m}(x) \equiv \frac{1}{2\pi} \int_{-\pi}^{\pi} F_{\text{res}}^\alpha(t, r, \theta, \varphi'; \bar{z}) e^{im(\varphi - \varphi')} d\varphi', \quad (60)$$

and similarly for $F_{\text{punc}}^{\alpha,m}$ and $F_S^{\alpha,m}$. Since the m -mode sum is formally a Fourier expansion, and since the Fourier expansion of an analytic function converges uniformly, we may replace the order of limit and summation in Eq. 59:

$$F_{\text{self}}^\alpha(z) = \sum_{m=-\infty}^{\infty} \lim_{x \rightarrow z} \left[F_{\text{res}}^{\alpha,m}(x) + \left(F_{\text{punc}}^{\alpha,m}(x) - F_S^{\alpha,m}(x) \right) \right]. \quad (61)$$

From Eq. 58, omitting the terms $O(\delta x)$ as they cannot possibly affect the final value of the SF in Eq. 54, we have

$$\lim_{x \rightarrow z} \left(F_{\text{punc}}^{\alpha,m} - F_S^{\alpha,m} \right) = \lim_{x \rightarrow \bar{z}} \frac{1}{2\pi} e^{im\varphi} \int_{-\pi}^{\pi} \bar{\epsilon}_0^{-5} P_{(5)}^\alpha e^{-im\varphi'} d\varphi', \quad (62)$$

where the integrand is evaluated at $\varphi = \varphi'$ and where we have used the fact that $x \rightarrow z$ implies also $x \rightarrow \bar{z}$. Crucially, one finds [12] that the last integral *vanishes* at

the limit $x \rightarrow \bar{z}$, for any m , regardless of the explicit form of $P_{(5)}^\alpha$.¹⁰ Hence, Eq. 61 reduces to

$$F_{\text{self}}^\alpha(z) = \sum_{m=-\infty}^{\infty} F_{\text{res}}^{\alpha,m}(z). \quad (63)$$

Here the substitution $x = z$ is allowed since the limit $x \rightarrow z$ is known to be well defined (and, in particular, direction independent). Phrased differently, the modes $F_{\text{res}}^{\alpha,m}$ corresponding to our puncture (56) are continuous at the particle, as desired.

Our result (63) can be written explicitly in terms of the m modes of the residual field $\bar{h}_{\beta\gamma}^{\text{res}}$, and easily so if we use the fixed- k extension introduced in Section 2.3 (the choice of k -extension may affect the value of the individual m modes, but not the eventual value of the SF). We then have

$$F_{\text{self}}^\alpha = \sum_{m=-\infty}^{\infty} \mu k^{\alpha\beta\gamma\delta} \left(\bar{h}_{\beta\gamma}^{\text{res},m} e^{im\varphi} \right)_{;\delta}, \quad (64)$$

where, of course, the derivatives are evaluated at the particle.

Equation 64 prescribes an extremely simple algorithm for constructing the SF in a 2+1D framework. In the puncture scheme we effectively “regularize” the field equations themselves (not their solutions, as in the standard l -mode regularization method), by writing them in the form (50) with a sufficiently accurate puncture function (like the one prescribed in 56). Once the m modes of the residual perturbation have been solved for, the SF is constructed directly from these mode with no further regularization needed.

The m -mode method is yet to be applied in actual calculations of the SF. It has the potential of providing a robust, simple, and efficient framework for calculations of the gravitational SF in Kerr spacetime.

6 Reflections and Prospects

We have attempted here to give a snapshot of the activity surrounding the development of reliable, efficient, and accurate computational methods for the SF in black hole spacetimes. The problem still attracts considerable attention (more than half of the papers on our reference list date ≥ 2005), with a multitude of different approaches being studied by different groups. This multitude is, of course, a great blessing, as it offers the opportunity for cross-validation of techniques and

¹⁰ It should not come as a surprise that at $x = z$ the sum over m -modes $F_{\text{punc}}^{\alpha,m} - F_{\text{S}}^{\alpha,m}$ (which are all zero) fails to recover the original function $F_{\text{punc}}^\alpha - F_{\text{S}}^\alpha$ (which is discontinuous and hence indefinite). Recall that the formal Fourier sum at a step discontinuity (where the function itself is indefinite) is in fact convergent: it yields the two-side average value of the function at the discontinuity. Technically, this peculiarity of the formal Fourier expansion is due simply to the noninterchangeability of the sum and limit at the point of discontinuity.

results – a particularly important prospect given the intricate nature of the numerics involved and the fact that SF calculations explore a new territory in black hole physics, yet uncharted by neither PN theory nor Numerical Relativity.

Indeed, the field has by now matured enough that such cross-validation exercises are becoming possible: A first such comparison between results from different calculations (using different gauges and different numerical methods) was presented very recently in Ref. [109]. We are now able to use SF codes to explore, for the first time, the conservative post-geodesic dynamics of strong-field orbits around a Schwarzschild black hole [25,41,42,109]. We can compute physical gauge-invariant post-geodesic effects and test them directly against results from PN theory in the weak-field regime [41, 42, 109]. Indeed, we can now start to use strong-field SF results in order to calibrate PN methods and assess their performance [25]. The exciting prospects for synergistic interaction between SF and PN theories are beginning to materialize, with much scope for important developments in the coming years (see Detweiler’s contribution in this volume).

At present, the state of the art in SF calculations is a code that can calculate the gravitational SF along any bound geodesic of the Schwarzschild geometry (currently at great computational cost, which future developments may help reduce). This code, as many others mentioned in our review, is an implementation of the mode-sum regularization method, which has proven a useful framework for calculations in Schwarzschild.

The Kerr problem, on the other hand, has hardly been tackled so far, and it represents the next significant challenge. Although the standard mode-sum method is in principle applicable to the gravitational SF in Kerr spacetime, the details of its numerical implementation in this case are yet to be worked out. It is possible that higher-dimension techniques (like the m -mode scheme discussed in Section 5) could provide an attractive alternative to standard mode-sum in the Kerr case.

In the short term, activity is likely to focus on the following tasks: (i) continue to improve the computational efficiency of SF calculations using advanced numerical techniques (mesh refinement, spectral methods, etc.); (ii) tackle the Kerr problem; (iii) use SF codes to study post-geodesic physical effects (such as the finite-mass correction to the orbital precession rate), and in particular assess the relative importance of conservative SF effects in the EMRI problem; (iv) explore what can be learned from a comparison between SF and PN results.

Within the context of the LISA EMRI problem, the computation of the SF on momentarily geodesic particles is only one essential ingredient. There is still much more to understand before a faithful model of an astrophysical inspiral can be announced. Most crucially, a reliable method must be devised for calculating the long-term evolution of the inspiral orbit. Work to address this problem has begun recently [57, 65, 99, 100] but much further study is needed.

Acknowledgements We acknowledge support from PPARC/STFC through grant number PP/D001110/1.

Appendix: The Regularization Parameter B^α in Kerr

We state here the explicit value of the regularization parameter B^α for the gravitational SF at a point $(t_0, r_0, \theta_0, \varphi_0)$ along a generic (geodesic) orbit in Kerr. In what follows u^α represents the four-velocity at that point, and $g_{\alpha\beta}$ and $\Gamma_{\alpha\beta}^\lambda$ are the Kerr metric and connection coefficients there. The derivation of all regularization parameters in Kerr was reported in Ref. [22], with the explicit value of B^α (reproduced here) given in the online arXiv version of that paper [20].

Using the method sketched in Section 2.3, one obtains

$$B^\alpha = \mu^2 (2\pi)^{-1} P_{abcd}^\alpha I^{abcd}, \quad (65)$$

where hereafter roman indices (a, b, c, \dots) run over the two Boyer–Lindquist angular coordinates θ, φ . The coefficients P_{abcd}^α are given by

$$P_{abcd}^\alpha = (3P_a^\alpha P_{be} - P_e^\alpha P_{ab}) C_{cd}^e + \frac{1}{2} [3P_d^\alpha P_{abc} - (2P_{ab}^\alpha + P_{ab}^\alpha) P_{cd}], \quad (66)$$

where

$$P_{\alpha\beta} \equiv g_{\alpha\beta} + u_\alpha u_\beta, \quad P_{\alpha\beta\gamma} \equiv (u_\lambda u_\gamma \Gamma_{\alpha\beta}^\lambda + g_{\alpha\beta, \gamma} / 2), \quad (67)$$

and

$$C_{\varphi\varphi}^\theta = \frac{1}{2} \sin \theta_0 \cos \theta_0, \quad C_{\theta\theta}^\varphi = C_{\varphi\theta}^\varphi = -\frac{1}{2} \cot \theta_0, \quad (68)$$

with all other coefficients C_{cd}^e vanishing. The quantities I^{abcd} are

$$I^{abcd} = (\sin \theta_0)^{-N} \int_0^{2\pi} G(\gamma)^{-5/2} (\sin \gamma)^N (\cos \gamma)^{4-N} d\gamma,$$

where $N \equiv N(abcd)$ is the number of times the index φ occurs in the combination (a, b, c, d) , namely

$$N = \delta_\varphi^a + \delta_\varphi^b + \delta_\varphi^c + \delta_\varphi^d, \quad (69)$$

and

$$G(\gamma) = P_{\tilde{\varphi}\tilde{\varphi}} \sin^2 \gamma + 2P_{\theta\tilde{\varphi}} \sin \gamma \cos \gamma + P_{\theta\theta} \cos^2 \gamma, \quad (70)$$

where

$$P_{\tilde{\varphi}\tilde{\varphi}} \equiv P_{\varphi\varphi} / \sin^2 \theta_0, \quad P_{\theta\tilde{\varphi}} \equiv P_{\theta\varphi} / \sin \theta_0. \quad (71)$$

To write down I^{abcd} more explicitly, denote

$$\alpha \equiv P_{\theta\theta} / P_{\tilde{\varphi}\tilde{\varphi}} - 1, \quad \beta \equiv 2P_{\theta\tilde{\varphi}} / P_{\tilde{\varphi}\tilde{\varphi}}. \quad (72)$$

Then I^{abcd} takes the form

$$I^{abcd} = \frac{(\sin \theta_0)^{-N}}{d} \left[Q I_K^{(N)} \hat{K}(w) + I_E^{(N)} \hat{E}(w) \right], \quad (73)$$

where

$$Q = \alpha + 2 - (\alpha^2 + \beta^2)^{1/2}, \quad (74)$$

$$d = 3P_{\tilde{\varphi}\tilde{\varphi}}^{5/2} (\alpha^2 + \beta^2)^2 (4\alpha + 4 - \beta^2)^{3/2} (Q/2)^{1/2}, \quad (75)$$

$\hat{K}(w)$ and $\hat{E}(w)$ are complete elliptic integrals of the first and second kinds, respectively, and the argument is

$$w = \frac{2(\alpha^2 + \beta^2)^{1/2}}{\alpha + 2 + (\alpha^2 + \beta^2)^{1/2}}. \quad (76)$$

The ten coefficients $I_K^{(N)}, I_E^{(N)}$ are given by

$$\begin{aligned} I_K^{(0)} &= 4 [12\alpha^3 + \alpha^2(8 - 3\beta^2) - 4\alpha\beta^2 + \beta^2(\beta^2 - 8)], \\ I_E^{(0)} &= -16[8\alpha^3 + \alpha^2(4 - 7\beta^2) + \alpha\beta^2(\beta^2 - 4) - \beta^2(\beta^2 + 4)], \end{aligned} \quad (77)$$

$$\begin{aligned} I_K^{(1)} &= 8\beta [9\alpha^2 - 2\alpha(\beta^2 - 4) + \beta^2], \\ I_E^{(1)} &= -4\beta[12\alpha^3 - \alpha^2(\beta^2 - 52) + \alpha(32 - 12\beta^2) + \beta^2(3\beta^2 + 4)], \end{aligned} \quad (78)$$

$$\begin{aligned} I_K^{(2)} &= -4 [8\alpha^3 - \alpha^2(\beta^2 - 8) - 8\alpha\beta^2 + \beta^2(3\beta^2 - 8)], \\ I_E^{(2)} &= 8[4\alpha^4 + \alpha^3(\beta^2 + 12) - 2\alpha^2(\beta^2 - 4) \\ &\quad + 3\alpha\beta^2(\beta^2 - 4) + 2\beta^2(3\beta^2 - 4)], \end{aligned} \quad (79)$$

$$\begin{aligned} I_K^{(3)} &= 8\beta [\alpha^3 - 7\alpha^2 + \alpha(3\beta^2 - 8) + \beta^2], \\ I_E^{(3)} &= -4\beta[8\alpha^4 - 4\alpha^3 + \alpha^2(15\beta^2 - 44) + 4\alpha(5\beta^2 - 8) \\ &\quad + \beta^2(3\beta^2 + 4)], \end{aligned} \quad (80)$$

$$\begin{aligned} I_K^{(4)} &= -4[4\alpha^4 - 4\alpha^3 + \alpha^2(7\beta^2 - 8) + 12\alpha\beta^2 - \beta^2(\beta^2 - 8)], \\ I_E^{(4)} &= 16[4\alpha^5 + 4\alpha^4 + \alpha^3(7\beta^2 - 4) + \alpha^2(11\beta^2 - 4) \\ &\quad + (2\alpha + 1)\beta^2(\beta^2 + 4)]. \end{aligned} \quad (81)$$

References

1. P. Amaro-Seoane, J.R. Gair, M. Freitag, M. Coleman Miller, I. Mandel, C.J. Cutler, S. Babak, *Class. Q. Grav.* **24**, R113 (2007)
2. P.R. Anderson, B.L. Hu, *Phys. Rev. D* **69**, 064039 (2004); Erratum, *ibidem* **75**, 129901(E) (2007)
3. P.R. Anderson, A. Eftekharzadeh, B.L. Hu, *Phys. Rev. D* **73**, 064023 (2006)
4. W.G. Anderson, A.G. Wiseman, *Class. Q. Grav.* **22**, S783 (2005)
5. W.G. Anderson, E.E. Flanagan, A.C. Ottewill, *Phys. Rev. D* **71**, 024036 (2005)
6. L. Barack, *Phys. Rev. D* **62**, 084027 (2000)
7. L. Barack, *Phys. Rev. D* **64**, 084021 (2001)
8. L. Barack, L.M. Burko, *Phys. Rev. D* **62**, 084040 (2000)
9. L. Barack, C. Cutler, *Phys. Rev. D* **69**, 082005 (2004)
10. L. Barack, C. Cutler, *Phys. Rev. D* **75**, 042003 (2007)
11. L. Barack, D.A. Golbourn, *Phys. Rev. D* **76**, 044020 (2007)
12. L. Barack, D.A. Golbourn, N. Sago, *Phys. Rev. D* **76**, 124036 (2007)
13. L. Barack, D.A. Golbourn, N. Sago, in preparation
14. L. Barack, C.O. Lousto, *Phys. Rev. D* **66**, 061502 (2002)
15. L. Barack, C.O. Lousto, *Phys. Rev. D* **72**, 104026 (2005)
16. L. Barack, Y. Mino, H. Nakano, A. Ori, M. Sasaki, *Phys. Rev. Lett.* **88**, 091101 (2002)
17. L. Barack, A. Ori, *Phys. Rev. D* **61**, 061502 (2000)
18. L. Barack, A. Ori, *Phys. Rev. D* **64**, 124003 (2001)
19. L. Barack, A. Ori, *Phys. Rev. D* **66**, 084022 (2002)
20. L. Barack, A. Ori, arXiv:gr-qc/0212103v2 (2003)
21. L. Barack, A. Ori, *Phys. Rev. D* **67**, 024029 (2003)
22. L. Barack, A. Ori, *Phys. Rev. Lett.* **90**, 111101 (2003)
23. L. Barack, A. Ori, N. Sago, *Phys. Rev. D* **78**, 084021 (2008)
24. L. Barack, N. Sago, *Phys. Rev. D* **75**, 064021 (2007)
25. L. Barack, N. Sago, *Phys. Rev. Lett.* **102**, 191101 (2009)
26. L. Barack, N. Sago, *Phys. Rev. D* **81**, 084021 (2010)
27. J.L. Barton, D.J. Lazar, D.J. Kennefick, G. Khanna, L.M. Burko, *Phys. Rev. D* **78**, 064042 (2008)
28. N.T. Bishop, R. Gomez, S. Husa, L. Lehner, J. Winicour, *Phys. Rev. D* **68**, 084015 (2003)
29. L.M. Burko, *Class. Q. Grav.* **17**, 227 (2000)
30. L.M. Burko, *Phys. Rev. Lett.* **84**, 4529 (2000)
31. L.M. Burko, G. Khanna, *Europhys. Lett.* **78**, 60005 (2007)
32. L.M. Burko, Y.T. Liu, *Phys. Rev. D* **64**, 024006 (2001)
33. L.M. Burko, Y.T. Liu, Y. Soen, *Phys. Rev. D* **63**, 024015 (2001)
34. L.M. Burko, A.I. Harte, E. Poisson, *Phys. Rev. D* **65**, 124006 (2002)
35. P. Canizares, C.F. Sopuerta, *J. Phys. Conf. S.* **154**, 012053 (2009)
36. P. Canizares, C.F. Sopuerta, *Phys. Rev. D* **79**, 084020 (2009)
37. M. Casals, S.R. Dolan, A.C. Ottewill, B. Wardell, *Phys. Rev. D* **79**, 124043 (2009)
38. P. L. Chrzanowski, *Phys. Rev. D* **11**, 2042 (1975)
39. N.A. Collins, S.A. Hughes, *Phys. Rev. D* **69**, 124022 (2004)
40. S. Detweiler, *Phys. Rev. Lett.* **86**, 1931 (2001)
41. S. Detweiler, *Class. Q. Grav.* **22**, S681 (2005)
42. S. Detweiler, *Phys. Rev. D* **77**, 124026 (2008)
43. S. Detweiler, E. Poisson, *Phys. Rev. D* **69**, 084019 (2004)
44. S. Detweiler, B.F. Whiting, *Phys. Rev. D* **67**, 024025 (2003)
45. S. Detweiler, E. Messaritaki, B.F. Whiting, *Phys. Rev. D* **67**, 104016 (2003)
46. B.S. DeWitt, R.W. Brehme, *Ann. Phys. (N.Y.)* **9**, 220 (1960)
47. B.S. DeWitt, C.M. DeWitt, *Physics (Long Island City, NY)* **1**, 3 (1964)
48. L.M. Diaz-Rivera, E. Messaritaki, B.F. Whiting, S. Detweiler, *Phys. Rev. D* **70**, 124018 (2004)
49. S. Drasco, E.E. Flanagan, S.A. Hughes, *Class. Q. Grav.* **22**, S801 (2005)

50. S. Drasco, S.A. Hughes, *Phys. Rev. D* **73**, 024027 (2006)
51. S.E. Field, J.S. Hesthaven, S.R. Lau, *Class. Quant. Grav.* **26**, 165010 (2009)
52. J.R. Gair, *Class. Quant. Grav.* **26**, 094034 (2009)
53. J.R. Gair, L. Barack, T. Creighton, C. Cutler, S.L. Larson, E.S. Phinney, M. Vallisneri, *Class. Q. Grav.* **21**, S1595 (2004)
54. J.R. Gair, C. Li, I. Mandel, *Phys. Rev. D* **77**, 024035 (2008)
55. K. Ganz, W. Hikida, H. Nakano, N. Sago, T. Tanaka, *Prog. Theor. Phys.* **117**, 1041 (2007)
56. J.A. Gonzalez, U. Sperhake, B. Bruegmann, *Phys. Rev. D* **79**, 124006 (2009)
57. S.E. Gralla, R.M. Wald, *Class. Q. Grav.* **25**, 205009 (2008)
58. C. Gundlach, G. Calabrese, I. Hinder, J.M. Martín-García, *Class. Q. Grav.* **22**, 3767 (2005)
59. R. Haas, *Phys. Rev. D* **75**, 124011 (2007)
60. R. Haas, Presentation at the 11th Capra meeting, Orléans, June 2008
61. R. Haas, E. Poisson, *Class. Q. Grav.* **22**, S739 (2005)
62. R. Haas, E. Poisson, *Phys. Rev. D* **74**, 044009 (2006)
63. W. Hikida, S. Jhingan, H. Nakano, N. Sago, M. Sasaki, T. Tanaka, *Prog. Theor. Phys.* **111**, 821 (2004)
64. W. Hikida, S. Jhingan, H. Nakano, N. Sago, M. Sasaki, T. Tanaka, *Prog. Theor. Phys.* **113**, 283 (2005)
65. T. Hinderer, E.E. Flanagan, *Phys. Rev. D* **78**, 064028 (2008)
66. J.M. Hobbs, *Ann. Phys. (N.Y.)* **47**, 141 (1968)
67. S.A. Hughes, S. Drasco, E.E. Flanagan, J. Franklin, *Phys. Rev. Lett.* **94**, 221101 (2005)
68. G. James, *Advanced Modern Engineering Mathematics*, 3rd edn., Section 4.2.8. (Pearson, Harlow, 2004)
69. P. Jaranowski et al., *Class. Q. Grav.* **25**, 114020 (2008)
70. T.S. Keidl, J.L. Friedman, A.G. Wiseman, *Phys. Rev. D* **75**, 124009 (2007)
71. G. Khanna, *Phys. Rev. D* **69**, 024016 (2004)
72. W. Krivan, P. Laguna, P. Papadopoulos, *Phys. Rev. D* **54**, 4728 (1996)
73. W. Krivan, P. Laguna, P. Papadopoulos, N. Andersson, *Phys. Rev. D* **56**, 3395 (1997)
74. R. Lopez-Aleman, G. Khanna, J. Pullin, *Class. Q. Grav.* **20**, 3259 (2003)
75. C.O. Lousto, *Phys. Rev. Lett.* **84**, 5251 (2000)
76. C.O. Lousto, *Class. Q. Grav.* **22**, S543 (2005)
77. C.O. Lousto, H. Nakano, *Class. Q. Grav.* **25**, 145018 (2008)
78. C.O. Lousto, R.H. Price, *Phys. Rev. D* **56**, 6439 (1997)
79. K. Martel, *Phys. Rev. D* **69**, 044025 (2004)
80. K. Martel, E. Poisson, *Phys. Rev. D* **66**, 084001 (2002)
81. Y. Mino, *Phys. Rev. D* **67**, 084027 (2003)
82. Y. Mino, *Class. Q. Grav.* **22**, S717 (2005)
83. Y. Mino, *Prog. Theor. Phys.* **115**, 43 (2006)
84. Y. Mino, *Phys. Rev. D* **77**, 044008 (2008)
85. Y. Mino, H. Nakano, *Prog. Theor. Phys.* **100**, 507 (1998)
86. Y. Mino, M. Sasaki, T. Tanaka, *Phys. Rev. D* **55**, 3457 (1997)
87. Y. Mino, H. Nakano, M. Sasaki, *Prog. Theor. Phys.* **108**, 1039 (2003)
88. V. Moncrief, *Ann. Phys. (N.Y.)* **88**, 323 (1974)
89. H. Nakano, M. Sasaki, *Prog. Theor. Phys.* **105**, 197 (2001)
90. H. Nakano, Y. Mino, M. Sasaki, *Prog. Theor. Phys.* **106**, 339 (2001)
91. H. Nakano, N. Sago, M. Sasaki, *Phys. Rev. D* **68**, 124003 (2003)
92. A. Ori, *Phys. Rev. D* **67**, 124010 (2003)
93. A.C. Ottewill, B. Wardell, *Phys. Rev. D* **77**, 104002 (2008)
94. A.C. Ottewill, B. Wardell, *Phys. Rev. D* **79**, 024031 (2009)
95. E. Pazos-Avalos, C.O. Lousto, *Phys. Rev. D* **72**, 084022 (2005)
96. M.J. Pfenning, E. Poisson, *Phys. Rev. D* **65**, 084001 (2002)
97. E. Poisson, *Living Rev. Rel.* **7**, URL: <http://www.livingreviews.org/lrr-2004-6>
98. A. Pound, E. Poisson, B.G. Nickel, *Phys. Rev. D* **72**, 124001 (2005)
99. A. Pound, E. Poisson, *Phys. Rev. D* **77**, 044012 (2008)
100. A. Pound, E. Poisson, *Phys. Rev. D* **77**, 044013 (2008)

101. L. Price, Private communication
102. T.C. Quinn, Phys. Rev. D **62**, 064029 (2000)
103. T.C. Quinn, R.M. Wald, Phys. Rev. D **56**, 3381 (1997)
104. T. Regge, J.A. Wheeler, Phys. Rev. **108**, 1063 (1957)
105. A.G. Smith, C.M. Will, Phys. Rev. D **22**, 1276 (1980)
106. N. Sago, H. Nakano, M. Sasaki, Phys. Rev. D **67**, 104017 (2003)
107. N. Sago, T. Tanaka, W. Hikida, H. Nakano, Prog. Theor. Phys. **114**, 509 (2005)
108. N. Sago, T. Tanaka, W. Hikida, K. Ganz, H. Nakano, Prog. Theor. Phys. **115**, 873 (2006)
109. N. Sago, L. Barack, S. Detweiler, Phys. Rev. D **78**, 124024 (2008)
110. C.F. Sopuerta, P. Laguna, Phys. Rev. D **73**, 044028 (2006)
111. C.F. Sopuerta, P. Sun, P. Laguna, J. Xu, Class. Q. Grav. **23**, 251 (2006)
112. P.A. Sundararajan, G. Khanna, S.A. Hughes, Phys. Rev. D **76**, 104005 (2007)
113. P.A. Sundararajan, G. Khanna, S.A. Hughes, S. Drasco, Phys. Rev. D **78**, 024022 (2008)
114. S.A. Teukolsky, Phys. Rev. Lett. **29**, 1114 (1972)
115. J. Thornburg, arXiv:0909.0036v3 [gr-qc] (2010); arXiv:1006.3788v1 [gr-qc] (2010)
116. I. Vega, S. Detweiler, Phys. Rev. D **77**, 084008 (2008)
117. R.M. Wald, Phys. Rev. Lett. **41**, 203 (1978)
118. R.M. Wald, *General Relativity* (Chicago Press University, Chicago, 1984)
119. N. Warburton, L. Barack, Phys. Rev. D **81**, 084039 (2010)
120. A.G. Wiseman, Phys. Rev. D **61**, 084014 (2000)
121. F.J. Zerilli, J. Math. Phys. **11**, 2203 (1970)
122. F.J. Zerilli, Phys. Rev. D **2**, 2141 (1970)

Radiation Reaction and Energy–Momentum Conservation

Dmitri Gal'tsov

Abstract We discuss subtle points of the momentum balance for radiating particles in flat and curved space-times. An instantaneous balance is obscured by the presence of the Schott term which is a finite part of the bound field momentum. To establish the balance, one has to take into account the initial and final conditions for acceleration, or to apply averaging. In curved space-time, an additional contribution arises from the tidal deformation of the bound field. This force is shown to be the finite remnant from the mass renormalization and it is different both from the radiation recoil force and the Schott force. For radiation of nongravitational nature from point particles in curved space-time the reaction force can be computed by substituting the retarded field directly to the equations of motion. A similar procedure is applicable to gravitational radiation in a vacuum space-time, but fails in the non-vacuum case. The existence of the gravitational quasilocal reaction force in this general case seems implausible, though it still exists in the nonrelativistic approximation. We also explain the putative antidamping effect for gravitational radiation under non-geodesic motion and derive the nonrelativistic gravitational quadrupole Schott term. Radiation reaction in curved space of dimension other than four is also discussed.

1 Introduction

One of the major tasks of gravitational wave astronomy is the precise theoretical prediction and observational measurement of gravitational waveforms from an inspiral fall of compact bodies into a supermassive black hole. This requires knowledge of the orbits with account for gravitational radiation reaction. Radiation gives rise to the reaction force, which can be incorporated into the equations of motion. The standard strategy to get the reaction force consists in substitution of the retarded field produced by the body into its equation of motion. The resulting equation is believed

D. Gal'tsov (✉)

Department of Physics, Moscow State University, Vorobjevy Gory, Moscow 119899, Russia

e-mail: galtsov@phys.msu.ru

to give a correct description of the instantaneous effect of radiation on the motion of the body. If the effect of reaction is small with respect to the main force, one can treat it adiabatically. But when the reaction force is not small, the situation is more subtle: the balance equations involve not only the kinetic energy–momentum of the body and that of radiation, but also a variable contribution of the bound field. Generically, an instantaneous loss of the energy–momentum by the body is not equal to the energy–momentum carried away by radiation.

This situation looks particularly simple in flat space Maxwell–Lorentz electrodynamics, where due to linearity of the equations one can use the notion of the point-like particle described by the delta-function. The Lorentz–Dirac equation obtained by substituting the retarded field into the equation of motion and performing the mass renormalization is expected to describe the motion of the particle with account for radiation loss. However, it turns out that the momentum loss due to radiation gives only a part of the reaction force. Although the initial system of the charges and the Maxwell field obey the overall momentum conservation equations, the Lorentz–Dirac equation, treated as the particle equation of motion, violates the naively expected balance between the particle momentum and the momentum of radiation. The difference is given by the so-called Schott term, which is the third-order total derivative of the coordinate. A careful comparison of the Dirac derivation [7, 17] and the Rohrlich analysis [8, 23, 24] has led Teitelboim [28] (see also [29]) to interpret the Schott term as the finite contribution of the bound (non-radiated) momentum of the charge remaining after the mass renormalization (an explicit proof of this claim was given in [11]). This may look strange, since we are used to thinking about the bound field as a stable Coulomb “coat,” which is spherically symmetric in the instantaneous rest frame, or pancake shaped for a relativistic charge. But this simple picture is valid only for constant velocity. Once acceleration is nonzero, the energy–momentum carried by the Coulomb coat becomes variable. The most surprising fact is that this momentum is simply proportional to the acceleration, and thus its derivative (the force) is the third derivative of the particle coordinate. Of course, the split of the total field into radiation and the bound field has to be done at all distances from the charge, not just in the wave zone. This split was given by Rohrlich [24]; for more recent discussion see the book by B. Kosyakov [19]. It is worth noting that in this case the finite part of the reaction force is entirely given by the time-antisymmetric part of the particle field (half-difference of the retarded and advanced potentials).

Thus, the energy–momentum balance of the system consisting of the accelerated charge and its Maxwell field includes three, not just two, ingredients: the particle momentum, the momentum carried by radiation, and the bound electromagnetic momentum. The radiation momentum can be extracted both from the particle momentum and, indirectly, from the bound momentum. This explains the origin of radiation of the uniformly accelerated charge, in which case the total reaction force is zero and thus the kinetic particle momentum is constant. While the charge is undergoing a constant acceleration, its bound electromagnetic momentum decreases and is transferred to radiation. Physically, however, the acceleration has to start at some moment and to finish at some moment, and during the stages of acquiring

and losing the acceleration the bound momentum is exchanged with the kinetic momentum. Therefore, the total energy–momentum loss of the charge will be equal to the momentum carried away by radiation. But an instantaneous balance is obscured by the presence of the Schott term. Another simple situation is periodic motion. Since the ambiguous Schott term is a total derivative, its contribution vanishes if one integrates over the period, or, equivalently, averages over the period.

For radiation of linear fields of nongravitational nature or of the linearized gravitational field in curved space-time the situation is more complicated because there are no local conservation laws unless Killing symmetries are present, and because the tail term cannot be found in the closed form [5, 16] (for a review see [22]). Another new feature in this case is that the reaction force contains a finite time-symmetric contribution coming from the half-sum of the retarded and advanced potentials – the force found by DeWitt–DeWitt [6] in linearized gravity and later rediscovered in the full General Relativity by Smith and Will [26] (the DDSW force). This force, however, has nothing to do with radiation and so the work done by this force is not expected to contribute to the energy balance for radiation. In the stationary space-time the total energy of the particle and the field is conserved, so one can expect that the energy balance between radiation and the particle energy loss will hold integrally or in average. In the case of axial symmetry, similar considerations apply to an associated angular momentum. An explicit proof of this balance for scalar, electromagnetic, and linearized gravitational radiation in the Kerr space-time was given in [9]. Apparently this work was not properly understood and was criticized in a number of papers until 2005, when analogous calculations were performed by other people (for a review and further references see [27]).

The case of linearized gravity is not entirely similar to other linear fields in the curved background, however. In fact, linearized gravity on the *non-vacuum* background is not a consistent linear theory since the full Bianchi identities do not allow for the harmonic gauge necessary to locally disentangle the linearized Einstein equations [14]. Physically, this means that the proper source of the gravitational radiation from the point particle is not just its energy–momentum tensor, but it also includes a contribution from the perturbed source of the background. Thus, the total source is nonlocal, which raises doubts that there might exist a local equation describing radiation reaction (here we mean the nonlocality stronger than that of the tail term, which is still localized on the world line of the point particle).

2 Energy–Momentum Balance Equation

We start by considering an interacting system of N point charges and the Maxwell field in Minkowski space-time, which is described by the coupled system of equations

$$\partial_\nu F^{\mu\nu} = 4\pi \int \sum_{a=1}^N \dot{z}^\mu \delta(x - z_a(\tau)) d\tau, \quad m_a^0 z_a^\mu = e_a F^\mu{}_\nu \dot{z}_a^\nu. \quad (1)$$

It has $3N + \infty$ degrees of freedom, where ∞ stands for the Maxwell field. The corresponding energy–momentum conservation equation is

$$\partial_\nu \left(\frac{m}{T}{}^{\mu\nu} + \frac{F}{T}{}^{\mu\nu} \right) = 0, \quad (2)$$

where

$$\frac{m}{T}{}^{\mu\nu} = \int \sum_{a=1}^N m_a^0 z_a^\mu \dot{z}_a^\nu \delta(x - z_a(\tau)) d\tau, \quad \frac{F}{T}{}^{\mu\nu} = \frac{1}{4\pi} \left(F^{\mu\lambda} F_\lambda^\nu + \frac{\eta^{\mu\nu}}{4} F^{\alpha\beta} F_{\alpha\beta} \right). \quad (3)$$

We have introduced the bare masses m_a^0 in anticipation of mass renormalization. Since the Maxwell equation is linear, one can decompose the total field in the vicinity of any given charge e_a into the sum of the field generated by the other $N - 1$ charges, $F_{\text{ext}}^{\mu\nu}$ (regular at its location) and the retarded field of e_a , $F_{a\text{ret}}^{\mu\nu}$. In spite of the fact that the total field acting on e_a , $F_a^{\mu\nu} = F_{\text{ext}}^{\mu\nu} + F_{a\text{ret}}^{\mu\nu}$ diverges at $x^\mu = z_a^\mu$, energy–momentum conservation is ensured by the equations of motion. The required mass renormalization does not change this statement.

The field $F_{a\text{ret}}^{\mu\nu}$ describes radiation and the bound field, which both contribute to the field energy–momentum. The overall conservation equation does not distinguish between these two parts, so an additional analysis is needed. The retarded potential at a given point x of space-time depends on the world-line variables taken at the moment of proper time $s_{\text{ret}}(x)$ defined as the solution to the equation

$$R^\mu R_\mu = 0, \quad R^\mu = x^\mu - z^\mu(s_{\text{ret}}), \quad (4)$$

satisfying $x^0 > z^0$. The advanced solution to the same equation with $z^0 > x^0$ refers to the advanced proper time $s_{\text{adv}}(x)$. Introducing the invariant distance

$$\rho = v_\mu(s_{\text{ret}}) R^\mu, \quad v^\mu = \frac{dz^\mu}{ds}, \quad (5)$$

which is equal to the spatial distance $|\mathbf{R}| = |\mathbf{x} - \mathbf{z}(s_{\text{ret}})|$ between the points of emission and observation in the momentarily co-moving Lorentz frame at the time moment $x^0 = z^0(s_{\text{ret}})$, one can present the retarded potential as (we omit the index a):

$$A_{\text{ret}}^\mu(x) = \frac{ev^\mu}{\rho} \Big|_{s_{\text{ret}}(x)}. \quad (6)$$

Introduce the normalized null vector $c^\mu = R^\mu/\rho$, such that $vc = 1$, and the unit space-like vector $u^\mu = c^\mu - v^\mu$, $u^2 = -1$. The following differentiation rules then hold:

$$c_\mu = \partial_\mu s_{\text{ret}}(x), \quad \partial_\mu \rho = v_\mu + \lambda c_\mu, \quad \partial_\mu c^\nu = \frac{1}{\rho} (\delta_\mu^\nu - v_\mu c^\nu - c_\mu v^\nu - \lambda c_\mu c^\nu), \quad (7)$$

where $\lambda = \dot{\rho} = \rho(ac) - 1$. The retarded field strength will read:

$$F_{\text{ret}}^{\mu\nu} = \frac{e(\rho(ac) - 1)}{\rho^2} v^{[\mu} c^{\nu]} - \frac{e}{\rho} a^{[\mu} c^{\nu]}, \quad (8)$$

where antisymmetrization is defined by e.g. $v^{[\mu} c^{\nu]} = v^{\mu} c^{\nu} - v^{\nu} c^{\mu}$. The retarded potential in Minkowski space admits a natural decomposition with respect to T-parity:

$$A_{\text{ret}}^{\mu} = A_{\text{self}}^{\mu} + A_{\text{rad}}^{\mu}, \quad (9)$$

where the radiative part $A_{\text{rad}}^{\mu} = \frac{1}{2}(A_{\text{ret}}^{\mu} - A_{\text{adv}}^{\mu})$ obeys an homogeneous wave equation, while the self part $A_{\text{self}}^{\mu} = \frac{1}{2}(A_{\text{ret}}^{\mu} + A_{\text{adv}}^{\mu})$ has a source at $x = z(s)$. One could expect that only T-symmetric A_{self}^{μ} corresponds to the *bound* field, but it is not so. For an accelerated charge the situation is more subtle.

2.1 Decomposition of the Stress Tensor

Constructing the energy–momentum tensor $T^{\mu\nu}$ with the retarded field $F_{\text{ret}}^{\mu\nu}$, one finds that it admits a natural decomposition:

$$T^{\mu\nu} = T_{\text{emit}}^{\mu\nu} + T_{\text{bound}}^{\mu\nu}, \quad (10)$$

where the first term is selected by its dependence on ρ as ρ^{-2} :

$$T_{\text{emit}}^{\mu\nu} = -\frac{((ac)^2 + a^2)c_{\mu}c_{\nu}}{\rho^2}, \quad (11)$$

while the second contains higher powers of ρ^{-1} :

$$T_{\text{bound}}^{\mu\nu} = \frac{a^{(\mu} c^{\nu)} + 2(ac)c^{\mu}c^{\nu} - (ac)v^{(\mu}c^{\nu)}}{\rho^3} + \frac{v^{(\mu}c^{\nu)} - c^{\mu}c^{\nu} - \eta^{\mu\nu}}{2\rho^4}, \quad (12)$$

where symmetrization without 1/2 is understood, say $v^{(\mu}c^{\nu)} = v^{\mu}c^{\nu} + v^{\nu}c^{\mu}$. The “emit” part (11) has the following properties:

- It is the tensor product of two null vectors c^{μ} .
- It is traceless.
- It falls off as $|\mathbf{x}|^{-2}$ when $|\mathbf{x}| \rightarrow \infty$.
- As follows from the differentiation rules (7), it is divergence free without assuming the validity of the equations of motion:

$$\partial_{\mu} T_{\text{emit}}^{\mu\nu} = 0. \quad (13)$$

All these features indicate that $T_{\text{emit}}^{\mu\nu}$ describes the outgoing radiation.

Since the total energy–momentum tensor including the contribution of charges is (on shell) divergence free, with account for (13) we find that

$$\partial_\mu \overset{F}{T}_{\text{bound}}^{\mu\nu} + \partial_\mu \overset{m}{T}^{\mu\nu} = 0, \quad (14)$$

so the bound field momentum can be exchanged with the particle momentum. Note, that outside the world line the bound stress tensor is divergence free. It is also worth noting, that Eq. 13 does not mean that there is no reaction force acting on a particle that counterbalances the emitted momentum.

Consider now the total balance of forces. The conservation of the total four-momentum (2) implies that the sum of the mechanical momentum and the momentum carried by the electromagnetic field is constant (for simplicity we do not include the external field):

$$\frac{dp_{\text{mech}}^\mu}{ds} + \frac{dp_{\text{em}}^\mu}{ds} = 0. \quad (15)$$

Here the mechanical part is proportional to the bare mass of the charge

$$p_{\text{mech}}^\mu = \int \overset{m}{T}^{\mu\nu} d\Sigma_\nu = m^0 v^\mu, \quad (16)$$

while the field part is given by

$$p_{\text{em}}^\mu = \int \overset{F}{T}^{\mu\nu} d\Sigma_\nu, \quad (17)$$

where integration of the electromagnetic stress tensor is performed over a space-like hypersurface whose choice will be specified later on. It has to be emphasized that the stress tensor of the electromagnetic field is constructed in terms of the physical retarded field. According to the above splitting, we can write

$$\frac{dp_{\text{mech}}^\mu}{ds} = f_{\text{emit}}^\mu + f_{\text{bound}}^\mu, \quad (18)$$

$$f_{\text{emit}}^\mu = - \int \overset{F}{T}_{\text{emit}}^{\mu\nu} d\Sigma_\nu, \quad f_{\text{bound}}^\mu = - \int \overset{F}{T}_{\text{bound}}^{\mu\nu} d\Sigma_\nu. \quad (19)$$

On the other hand, the derivative of the bare mechanical momentum of the charge can be found by substituting the retarded field into the equation of motion. In this case it is useful to decompose the retarded field according to (9), obtaining another split of the mechanical momentum:

$$\frac{dp_{\text{mech}}^\mu}{ds} = e F_{\text{ret}}^{\mu\nu} v_\nu = e (F_{\text{self}}^{\mu\nu} + F_{\text{rad}}^{\mu\nu}) v_\nu = f_{\text{self}}^\mu + f_{\text{rad}}^\mu. \quad (20)$$

Now, somewhat unexpectedly, $f_{\text{rad}}^\mu \neq f_{\text{emit}}^\mu$ and $f_{\text{self}}^\mu \neq f_{\text{bound}}^\mu$, the difference being called the Schott term [11]:

$$f_{\text{rad}}^\mu = f_{\text{emit}}^\mu + f_{\text{Schott}}^\mu, \quad f_{\text{self}}^\mu = f_{\text{bound}}^\mu - f_{\text{Schott}}^\mu. \quad (21)$$

Clearly,

$$f_{\text{self}}^\mu + f_{\text{rad}}^\mu = f_{\text{bound}}^\mu + f_{\text{emit}}^\mu, \quad (22)$$

as expected. Note that both f_{bound}^μ and f_{self}^μ contain divergences that mutually cancel in Eq. 22.

The forces f_{self}^μ and f_{rad}^μ can be found using the Green functions [17]

$$G_{\text{self}}(Z) = \delta(Z^2), \quad G_{\text{rad}}(Z) = \frac{Z^0}{|Z^0|} \delta(Z^2), \quad (23)$$

where $Z^\mu = Z^\mu(s, s') = z^\mu(s) - z^\mu(s')$. Substituting the value of the electromagnetic field generated by the charge on its world line one obtains

$$f^\mu(s) = 2e^2 \int Z^{[\mu}(s, s') v^{\nu]}(s') v_\nu(s) \frac{d}{dZ^2} G(Z) ds', \quad (24)$$

for both f_{self}^μ and f_{rad}^μ . Due to delta-functions, only a finite number of Taylor expansion terms in $\sigma = s - s'$ contribute to the integral. In the four-dimensional case, it is sufficient to retain the terms up to σ^3 :

$$2Z^{[\mu}(s, s') v^{\nu]}(s') v_\nu(s) = \dot{v}^\mu \sigma^2 - \frac{2}{3} (\ddot{v}^\mu + v^\mu \dot{v}^2) \sigma^3 + O(\sigma^4). \quad (25)$$

Taking into account that $Z^2 = \sigma^2 + O(\sigma^4)$, the leading terms in the expansions of derivatives of the Green functions will be

$$\frac{d}{dZ^2} G_{\text{self}}(Z) = \frac{d}{d\sigma^2} \delta(\sigma^2), \quad \frac{d}{dZ^2} G_{\text{rad}}(Z) = \frac{d}{d\sigma^2} \left(\frac{\sigma}{|\sigma|} \delta(\sigma^2) \right). \quad (26)$$

Regularizing the delta-functions of σ^2 by point-splitting

$$\delta(\sigma^2) = \lim_{\varepsilon \rightarrow +0} \delta(\sigma^2 - \varepsilon^2) = \lim_{\varepsilon \rightarrow +0} \frac{\delta(\sigma - \varepsilon) + \delta(\sigma + \varepsilon)}{2\varepsilon}, \quad (27)$$

with a prescription that the limit should be taken after evaluating the integrals, one finds

$$f_{\text{self}}^\mu = -\frac{e^2}{2\varepsilon} a^\mu, \quad f_{\text{rad}}^\mu = \frac{2e^2}{3} (v^\mu a^2 + \dot{a}^\mu). \quad (28)$$

After the mass renormalization, $m_0 + \frac{1}{2\varepsilon} = m$, we get the Lorentz–Dirac equation

$$ma^\mu = \frac{2e^2}{3} (v^\mu a^2 + \dot{a}^\mu). \quad (29)$$

The first term at the right-hand side is equal to the derivative of the momentum carried away by radiation,

$$f_{\text{emit}}^{\mu} = -\frac{dp_{\text{emit}}^{\mu}}{ds} = \frac{2e^2}{3}a^2v^{\mu}. \quad (30)$$

Its independent evaluation by integration of $T_{\text{emit}}^{\mu\nu}$ can be found in [11]. The second total derivative term is the Schott term. It is worth noting, that within the local calculation, the Schott term originates from the T-odd part of the retarded field.

2.2 Bound Momentum

An explicit evaluation of the bound momentum associated with a given moment of proper time s on the particle world-line,

$$p_{\text{bound}}^{\mu}(s) = \int_{\Sigma(s)} T_{\text{bound}}^{\mu\nu} d\Sigma_{\nu}, \quad (31)$$

was given in [11], which we follow here. First of all one has to choose the space-like hypersurface $\Sigma(s)$ intersecting the world line at $x^{\mu} = z^{\mu}(s)$. A convenient choice will be the hyperplane orthogonal to the world line

$$v_{\mu}(s) (x^{\mu} - z^{\mu}(s)) = 0. \quad (32)$$

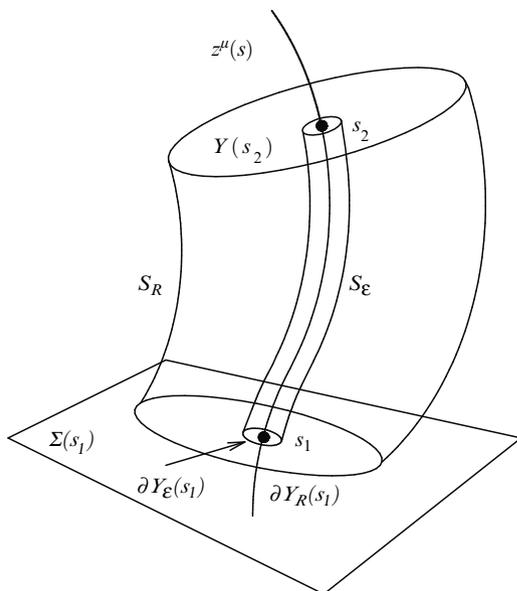
The integral (31) is divergent on the world line. We introduce the small length parameter ε , the radius of the 2-sphere $\partial Y_{\varepsilon}(s)$ (Fig. 1), defined as the intersection of the hyperplane (32) with the hyperboloid $(x - z(s))^2 = -\varepsilon^2$. We also introduce the sphere $\partial Y_R(s)$ of the large radius R defined as the intersection of $\Sigma(s)$ with the hyperboloid $(x - z(s))^2 = -R^2$. The total field momentum can be obtained as the limit $\varepsilon \rightarrow 0$, $R \rightarrow \infty$ of the integral over the domain $Y(s) \subset \Sigma(s)$ between the boundaries $\partial Y_{\varepsilon}(s)$ and $\partial Y_R(s)$.

Let us evaluate the variation of this quantity between the instants s_1 and s_2 of the proper time on the world line of the charge

$$\Delta p_{\text{em}}^{\mu} = \int_{Y(s_2)} T^{\mu\nu} d\Sigma_{\nu} - \int_{Y(s_1)} T^{\mu\nu} d\Sigma_{\nu}. \quad (33)$$

For the bound momentum it is convenient to consider the tubes S_{ε} and S_R formed as sequences of the spheres $\partial Y_{\varepsilon}(s)$ and $\partial Y_R(s)$ on the interval $s \in [s_1, s_2]$ and to transform this quantity to

Fig. 1 Integration of the bound electromagnetic momentum. Here $\Sigma(s_1)$ is the space-like hyperplane transverse to the world line $z^\mu(s)$ intersecting it at the proper time s_1 (similarly $\Sigma(s_2)$). The hypersurfaces S_ε and S_R are small and large tubes around the world line formed by sequences of the 2-spheres $\partial Y_\varepsilon(s)$ and $\partial Y_R(s)$ for $s \in [s_1, s_2]$. The domain $Y(s_2) \subset \Sigma(s_2)$ (similarly $Y(s_1)$) is the 3-annulus between $\partial Y_R(s_2)$ and $\partial Y_\varepsilon(s_2)$



$$\Delta p_{\text{bound}}^\mu = \int_{S_R} T_{\text{bound}}^{\mu\nu} dS_\nu - \int_{S_\varepsilon} T_{\text{bound}}^{\mu\nu} dS_\nu, \tag{34}$$

in view of the conservation equation for $T_{\text{bound}}^{\mu\nu}$ outside the world line, see the remark after Eq. 14. Here the integration elements dS_ν are directed outward to the world line. The contribution from the distant surface S_R vanishes if one assumes that the acceleration is zero in the limit $s \rightarrow -\infty$ [28]. This is nontrivial: though the stress tensor (12) decays as R^{-3} at spatial infinity, the corresponding flux does not vanish a priori, because the surface element contains a term (proportional to the acceleration) which asymptotically grows as R^3 . As a consequence, the surviving term will be proportional to the acceleration taken at the instant s_{ret} of the proper time, where $s_{\text{ret}} \rightarrow -\infty$ in the limit $R \rightarrow \infty$. Finally, we are left with the integral over the inner boundary S_ε only. To find an integration measure on S_ε we foliate the space-time domain in Fig. 1 by the hypersurfaces $\Sigma(s)$ parameterized by the spherical coordinates $r, \theta_1 = \theta, \theta_2 = \varphi$. Introducing the unit space-like vector $n^\mu(s, \theta_i)$, $n_\mu n^\mu = -1$ transverse to v^μ , we use the coordinate transformation $x^\mu = z^\mu(s) + r n^\mu(s, \theta_i)$. The induced metric on S_ε reads $dS_\mu = \varepsilon^2 [1 - \varepsilon(an)] n_\mu ds d\Omega$, and hence

$$\frac{dp_{\text{bound}}^\mu}{ds} = - \int_{S_\varepsilon} \varepsilon^2 [1 - \varepsilon(an)] T_{\text{bound}}^{\mu\nu} n_\nu d\Omega, \tag{35}$$

where the limit $\varepsilon \rightarrow 0$ has to be taken. One has to expand $T_{\text{bound}}^{\mu\nu}$ in terms of ε . In fact, the energy–momentum tensor depends on the space-time point x^μ through the quantity ρ , depending directly on x^μ , and also through the retarded proper time s_{ret} . We have to express the resulting quantity as a function of the proper time s corresponding to the intersection of the world line with the space-like hypersurface. We write $T_{\mu\nu}^{\text{bound}}$ in terms of the null vector $R^\mu = c^\mu \rho$:

$$\frac{4\pi}{e^2} T_{\text{bound}}^{\mu\nu} = \frac{a^{(\mu} R^{\nu)}}{\rho^4} + \frac{(2(aR) - 1) R^\mu R^\nu}{\rho^6} + \frac{(1 - (aR)) v^{(\mu} R^{\nu)}}{\rho^5} - \frac{\eta^{\mu\nu}}{2\rho^4}, \quad (36)$$

and expand R^μ as

$$R^\mu = x^\mu - z^\mu(s_{\text{ret}}) = \varepsilon n^\mu + v^\mu \sigma - \frac{1}{2} a^\mu \sigma^2 + \frac{1}{6} \dot{a}^\mu \sigma^3 + \mathcal{O}(\sigma^4), \quad (37)$$

where $\sigma = s - s_{\text{ret}} > 0$ and all the vectors are taken at s . This expansion in powers of σ has to be rewritten in terms of ε . The relation between the two can be found from the condition $R^2 = 0$:

$$\sigma = \varepsilon + \frac{an}{2} \varepsilon^2 + (9(an)^2 + a^2 - 4\dot{a}n) \frac{\varepsilon^3}{24} + \mathcal{O}(\varepsilon^4). \quad (38)$$

Substituting this into Eq. 37 and further to (36) we find:

$$\begin{aligned} \Delta p_{\text{bound}}^\mu = \frac{e^2}{4\pi} \int_{s_1}^{s_2} ds \left\{ \frac{-n^\mu}{2\varepsilon^2} + \frac{a^\mu}{2\varepsilon} + [(an)^2 + a^2/3] v^\mu \right. \\ \left. + ((an)^2 + a^2/2) n^\mu - 2\dot{a}^\mu/3 + 3(an)a^\mu/4 \right\} d\Omega. \quad (39) \end{aligned}$$

The leading divergent term $1/\varepsilon^2$ disappears after angular integration. Thus we obtain [11]:

$$f_{\text{bound}}^\mu = -\frac{dp_{\text{bound}}^\mu}{ds} = -\frac{e^2 a^\mu}{2\varepsilon} + \frac{2e^2}{3} \dot{a}^\mu. \quad (40)$$

Here the first divergent term has to be absorbed by the renormalization of mass, while the second is the finite Schott term. Comparing this with (28) we confirm the identity (22). Note that a priori the regularization parameter ε here (the radius of the small tube) is not related to the splitting parameter of the delta-function in the previous local force calculation. But actually they give the same form for divergent terms, for which reason we use the same symbol ε for both of them. With this convention, the divergent terms in the momentum conservation identity (22) mutually cancel.

The significance of the Schott term in the balance of momentum between the radiating charge and the emitted radiation is not always recognized in the textbooks on

classical electrodynamics. Instead, its presence is often interpreted as a drawback of classical theory, since formally it may lead to self-accelerating solutions. Meanwhile such solutions must be discriminated as unphysical since they do not satisfy proper initial/final conditions that should be imposed on the third-order equation of motion [17]. From the above analysis it is clear that the Lorentz–Dirac equation, although formulated in terms of particle variables, actually describes the composite system consisting of a charge and its bound electromagnetic momentum. The redefinition of the particle momentum joining to it the bound electromagnetic momentum obscures the problem of interpretation. It is better to think of the Schott term as the field degree of freedom and interpret the Lorentz–Dirac equation as the momentum balance equation for the total system including the electromagnetic field.

An instantaneous momentum balance is not just the balance between the particle and radiation, the energy–momentum can be also transferred between the particle and the field coat bound to it. Or, the radiated momentum is not always taken from the mechanical momentum of the charge, but by virtue of the Schott term, it can be extracted from the field coat too. This is what happens in the case of the uniformly accelerated charge, when the total reaction force instantaneously is zero, while radiation carries the momentum away at a constant rate. The balance is ensured by the Schott term. However, the constant acceleration cannot last infinitely long. One has to consider the switching on/off processes in order to understand that finally the energy–momentum of radiation is taken from the particle. This consideration clarifies the necessity of the time averaging or integration over time needed to establish the momentum balance between radiation and the source particle. The equation of motion including the reaction force instantaneously does not imply the equality of the radiative momentum loss and the particle momentum. This feature is general enough, it is also applicable to radiation of nongravitational nature from particles moving along the geodesics in curved space-time, as well as to the gravitational radiation.

2.3 The Rest Frame (Nonrelativistic Limit)

In the rest frame of a charge the recoil force has no spatial component. This is due to the fact that radiation in two opposite directions is the same so that the spatial momentum is not lost by radiation, though the energy is lost. Hence the total spatial component of the reaction force is presented by the Schott term, namely

$$\mathbf{f}_{\text{Schott}} = \frac{2}{3}e^2\dot{\mathbf{a}}. \quad (41)$$

The work done by this force,

$$\int \mathbf{f}_{\text{Schott}} \cdot \mathbf{v} dt = \int \frac{2}{3}e^2\dot{\mathbf{a}} \cdot \mathbf{v} dt = - \int \frac{2}{3}e^2\mathbf{a}^2 dt + \text{boundary terms}, \quad (42)$$

correctly reproduces the radiative loss in the rest frame. (Boundary terms should vanish by appropriate asymptotic switching on/off or periodicity conditions.)

3 Flat Dimensions Other than Four

Recent interest to models with large extra dimensions motivates the study of radiation and radiation reaction in dimensions other than four. It turns out that the radiation picture is substantially different in even and odd dimensions because of the different structure of the retarded Green's functions for massless fields in the coordinate representation [10] (they still look similarly in all dimensions in the momentum representation). In even dimensions the retarded potential is localized on the past light cone (Huygens principle) so the situation is qualitatively similar to that in the $4D$ case. In odd dimensions it is nonzero also inside the past null cone though radiation still propagates along the null rays. In $3D$, for instance, the scalar Green's function reads

$$G_{\text{ret}}^{3D}(X) = \vartheta(X^0)\vartheta(X^2)(X^2)^{-1/2}, \quad X^\mu = x^\mu - x'^\mu. \quad (43)$$

It does not contain the "direct" part singular on the light cone. Green's functions in higher odd dimensions $D = 2n + 1$ can be obtained by the recurrent relation [10, 12]

$$G_{\text{ret}}^{2n+1}(X) \sim \frac{dG_{\text{ret}}^{2n-1}}{dX^2}, \quad (44)$$

In particular, in $5D$

$$G_{\text{ret}}^{5D}(X) \sim \vartheta(X^0) \left(\frac{\delta(X^2)}{(X^2)^{1/2}} - \frac{1}{2} \frac{\vartheta(X^2)}{(X^2)^{3/2}} \right), \quad (45)$$

both the direct and the tail parts are present. It turns out that the direct part regularizes the tail contribution to the field stress proportional to the derivative of G that otherwise would be singular outside the world line.

In even dimensional space-times the split of the retarded potentials into the time symmetric and the radiative parts leads to purely divergent self-force and a finite radiative part:

$$f_{\text{self}}^\mu = f_{\text{div}}^\mu, \quad f_{\text{rad}}^\mu = \text{finite}. \quad (46)$$

Divergent terms are Lagrangian type and can be absorbed by introducing suitable counter-terms. Since the Coulomb dependence is more singular at the location of the source in higher dimensions, the self-action gives rise to larger number of divergent terms $1/\varepsilon^n$, where n is changing from unity to the integer part of $D/2 - 1$. The highest divergence is absorbed by the renormalization of mass, while to absorb other divergences additional counter-terms are needed depending on higher derivatives of the velocity. These are not present in the initial action, so higher-dimensional classical theories are not renormalizable. In $6D$ one has two divergent terms (which in terms of the field split correspond to f_{self}^μ [18]):

$$f_{\text{div}}^\mu = -\frac{1}{6\varepsilon^3}a_\mu + \frac{1}{2\varepsilon} \left(\frac{3}{4}v_\mu(\dot{a}a) + \frac{3}{8}a^2a_\mu + \frac{1}{4}\ddot{a}_\mu \right), \quad (47)$$

the leading term being eliminated by the mass renormalization and the subleading requiring the counter-term [18]

$$S_1 = -\kappa_0^{(1)} \int (\ddot{z})^2 ds, \tag{48}$$

which leads to the Frenet–Serret dynamics [1] unless the renormalized value $\kappa_0^{(1)} = 0$. For each two space-time dimensions one new higher-derivative counter-term is needed to absorb divergences.

The split of the field stress tensor built with the retarded field into the sum of the emitted and bound terms is also possible in all even dimensions, and one always has the relation (22). In $6D$, for example, the radiation recoil force in $6D$ is

$$f_{\text{emit}}^\mu = \frac{4}{45} e^2 \left(\dot{a}^2 v^\mu + \frac{2}{21} (a\dot{a}) a^\mu - \frac{2}{9} a^4 v^\mu - \frac{2}{105} a^2 \dot{a}^\mu \right), \tag{49}$$

and the Schott terms is

$$f_{\text{Schott}}^\mu = -\frac{4e^2}{45} \left(\ddot{a}^\mu + \frac{16}{7} a^2 \dot{a}^\mu + \frac{60}{7} (a\dot{a}) a_\mu + 4\dot{a}^2 v^\mu + 4(a\ddot{a}) v^\mu \right), \tag{50}$$

the sum of two being orthogonal to the 6-velocity. The Schott terms is again given by the finite part of the integrated bound momentum.

In odd dimensions one always has tail terms. A split of the retarded field into the self and the radiative parts is always possible, and the substitution into the equations of motion leads to divergent and finite terms. Divergent ones can again be absorbed introducing the counter-terms. The split of the stress–tensor into the emitted and the bound parts is more subtle. The situation is obscured by the fact that though the free field is massless and thus propagates along the null cone, the full retarded potential fills the interior of the past light cone. Still one is able to obtain the general formula for radiation momentum that is no more associated with the retarded proper time on the world line [13].

4 Local Method for Curved Space-Time

An approach initiated by DeWitt and Brehme and further applied to linearized gravity in [20] appeals to computation of the integral of the stress tensor in the world-tube surrounding the world line. This is similar to our calculation of the bound momentum. However, in curved space-time the split of the stress tensor into the emitted and the bound parts becomes problematic, so the complete analysis of the balance between radiation, the kinetic momentum, and the bound momentum is not available. Meanwhile, to compute the total reaction force one can use a much simpler calculation substituting the retarded field directly into the equations of motion [14]. This approach was also formulated in higher even-dimensional space-time in the paper [12] which we follow here.

4.1 Hadamard Expansion in Any Dimensions

As in $D = 4$ [5], the curved space Green's functions for massless fields in other dimensions can be constructed starting with the Hadamard solution. For simplicity we consider here the scalar case. The scalar Hadamard Green's function $G_H(x, x')$ is a solution of the homogeneous wave equation $\square_x G_H(x, x') = 0$, where $\square = g_{\mu\nu} \nabla^\mu \nabla^\nu$. The procedure consists in expanding $G_H(x, x')$ in terms of the Synge world function $\sigma(x, x')$. For $D = 4$, the Hadamard expansion contains two terms singular in σ , namely, σ^{-1} and $\ln \sigma$. In higher dimensions one has to add other singular terms, and by dimensionality it is easy to guess that each dimension introduces an additional factor $\sigma^{-1/2}$. Thus, the Hadamard expansion in $D = 2d$ dimensions ($d \geq 3/2$ is integer or half-integer) generically must read

$$G_H(x, x') = \frac{1}{(2\pi)^d} \left[\sum_{n=1}^D g_n \sigma^{1-n/2} + \nu \ln \sigma \right], \quad (51)$$

where $g_n = g_n(x, x')$, $\nu = \nu(x, x')$ are two-point functions. It can be shown, that in odd dimensions we actually have only odd powers of $\sigma^{-1/2}$, and in even dimensions – only even powers, that is, an expansion in terms of inverse integer powers of σ . The logarithmic term is present only in even dimensions.

Substituting (51) into the wave equation, in the leading singular order we will have $g_D = \Delta^{1/2}$. In the next to leading order we obtain the equation:

$$2 \partial_\mu g_{D-1} \sigma^\mu + g_{D-1} \square \sigma - (D-1) g_{D-1} = 0. \quad (52)$$

which does not have analytic solutions, so $g_{D-1} = 0$. For $D = 3$ this means the absence of the logarithmic term. Similarly, considering the equation for $g_{(D-1-2k)}$, $k \in \mathbb{N}$, we find $g_{D-1-2k} = 0$. This means that for an even-dimensional space-time the Hadamard Green's function contains only integer negative powers of σ plus logarithm and a regular part, while in the odd-dimensional case – only half-integer powers of σ plus a regular part.

For the sequence of Green's functions in the flat space-time, the one in $D + 2$ dimension is proportional to the derivative of the Green's function in the twice preceding dimension D . In fact, in even dimensions the symmetric Green's function is the derivative of the order $d - 2$ of the delta-function: $G^D \sim \delta^{d-2}(\sigma)$, $\sigma = (x - x')^2/2$ and thus, $G^{D+2} \sim dG^D/d\sigma$. Applying regularization, $\delta((x - x')^2) = \lim_{\varepsilon \rightarrow +0} \delta(|(x - x')^2| - \varepsilon^2)$, we obtain

$$G^{D+2} \sim dG^D/d\varepsilon^2. \quad (53)$$

This relation has a consequence that the Laurent expansion of the Lorentz–Dirac force in terms of ε in the even-dimensional Minkowski space has only odd negative powers, and no even terms. So the number of divergent terms in the self-action increases by one for each next even dimension. In curved space passing to the standard notation for g_n we have

$$G_H = \frac{1}{(2\pi)^d} \left(\sum_{k=0}^{d-2} \frac{u_k}{\sigma^{d-1-k}} + v \ln \sigma + w \right), \tag{54}$$

where $u_0 = \Delta^{1/2}$ and we denoted $v = u_{d-1}$, $w = u_d$. Applying \square with respect to x^μ , we obtain the system of recurrent differential equations for $u_i(x, x')$. Integrating them along the geodesic connecting the points x, x' , one can uniquely express $u_1(x, x')$ through $u_0(x, x')$. Furthermore, u_2 is expressed through u_1 , etc.

To relate the coefficient functions $u_i^D(x, x')$ in different dimensions we first observe that $u_0 = \Delta^{1/2}$ for any D . It is worth noting, however, that in the expansion in terms of σ

$$u_0^D = \Delta^{1/2} = 1 + 1/12 R_{\alpha\beta} \sigma^\alpha \sigma^\beta + \dots, \tag{55}$$

where the tensor indices α, β run through all values corresponding to D . With this in mind, we can write $u_0^{D'} = u_0^D$ for any D, D' . The next equation in the recurrence gives for $D, D' \geq 5$,

$$u_1^{D'} = u_1^D \frac{d-2}{d'-2}. \tag{56}$$

Similarly for $D \geq 7, D' = D + 2$ one obtains:

$$u_2^{D+2} = u_2^D \frac{d-3}{d-1}. \tag{57}$$

Continuing this process further one finds

$$-\frac{1}{d-1} \frac{\partial G_H^D}{\partial \sigma} = (2\pi) G_H^{\text{dir} D+2}, \tag{58}$$

where the “direct” part of the Hadamard function is

$$G_{H \text{ dir}} = \frac{1}{(2\pi)^d} \sum_{k=0}^{d-2} \frac{u_k}{\sigma^{d-1-k}}. \tag{59}$$

4.2 Divergences

For the retarded Green’s function one finds:

$$G_{\text{ret}} = \frac{1}{2(2\pi)^{d-1}} \Theta(x', \Sigma(x)) \left(\sum_{m=0}^{d-2} \frac{(-1)^m u_{d-2-m} \delta^{(m)}(\sigma)}{m!} - v \theta(-\sigma) \right). \tag{60}$$

The first term constitutes the direct part of the retarded function with the support on the light cone:

$$G_{\text{dir}} = \frac{1}{2(2\pi)^{d-1}} \Theta[\Sigma] \sum_{m=0}^{d-2} \frac{(-1)^m u_{d-2-m} \delta^{(m)}(\sigma)}{m!}. \tag{61}$$

Similarly to the recurrent relations of the previous section, we obtain for the retarded Green's functions:

$$\frac{\partial G_{\text{ret}}^D}{\partial \sigma} = -2\pi (d-1) G_{\text{dir}}^{D+2}. \quad (62)$$

Using the regularization $\delta(\sigma) = \lim_{\varepsilon \rightarrow +0} \delta(\sigma - \mathcal{E})$, where $\mathcal{E} = \varepsilon^2/2$, we obtain for the direct part of the reaction force

$$f_{D+2}^{\mu \text{ dir}} = -\frac{1}{D-1} \frac{\partial f_D^{\mu \text{ dir}}(s, \mathcal{E})}{\partial \mathcal{E}}. \quad (63)$$

The limit $\mathcal{E} \rightarrow +0$ has to be taken after the differentiation. The direct force is due to the light cone part of the retarded Green's function. This is not the full local contribution to the Lorentz–Dirac force. An additional contribution comes from the differentiation of the theta-function in the tail term $v\theta(\sigma)$. In the scalar case this contribution vanishes, but in the electromagnetic case an extra local term arises:

$$f_{\text{loc}}^{\mu} = e^2 ([v_{\mu\alpha}] \dot{z}_\nu - [v_{\nu\alpha}] \dot{z}_\mu) \dot{z}^\nu \dot{z}^\alpha, \quad (64)$$

where the coincidence limit $[v_{\nu\alpha}]$ depends on the dimension. The remaining contribution from the tail term will have the form of an integral along the past half of the particle world line.

The direct part of the Lorentz–Dirac force contains divergences. To separate the divergent terms one can use the decomposition of the retarded potential suggested in the case of four dimensions by Detweiler and Whiting [3, 4]. In higher even dimensions we can follow essentially the same procedure. We define the “singular” part G_S of the retarded Green's function as the sum of the symmetric part (self) and the tail function v as follows

$$G_S(x, x') = G_{\text{self}}(x, x') + \frac{v(x, x')}{4(2\pi)^{d-1}} = G_{\text{self dir}}(x, x') + \frac{v(x, x') \theta(-\sigma)}{4(2\pi)^{d-1}}. \quad (65)$$

Here the direct part of the self function means its part without the tail v -term. The remaining part of the Green's function $G_R(x, x') = G_{\text{ret}}(x, x') - G_S(x, x')$ satisfies a free wave equation and is regular. Taking into account $\square v = 0$, it is clear that G_S satisfies the same inhomogeneous equation as G_{self} . The v -term in the second line of Eq. 65 is localized outside the light cone. Therefore the corresponding field (for instance, scalar), at an arbitrary point x will be given by

$$\phi_S(x) = \phi_{\text{self dir}} + \frac{m_0 q}{4(2\pi)^{d-1}} \int_{\tau_{\text{ret}}}^{\tau_{\text{adv}}} v(x, z(\tau)) d\tau, \quad (66)$$

where the retarded and advanced proper time values $\tau_{\text{ret}}(x)$, $\tau_{\text{adv}}(x)$ are the intersection points of the past and future light cones centered at x with the world line.

Inserting (66) into the Lorentz–Dirac force defined on the world line $x = z(\tau)$, we observe that the integral contribution from the tail term vanishes and so the divergent part is entirely given by $\phi_{\text{self dir}}$. Using for delta-functions the point-splitting regularization we find that all divergent terms arise as negative powers of ε . To prove the existence of the counter-terms we consider the interaction term in the action substituting the field as the S-part of the retarded solution to the wave equation. In the scalar case we will have:

$$S_S = \frac{1}{2} \int G_S(x, x') \rho(x) \rho(x') \sqrt{g(z)g(z')} dx dx', \tag{67}$$

where a factor one-half is introduced to avoid double counting when self-interaction is considered. Substituting the currents we get the integral over the world line. Since the Green’s function is localized on the light cone we expand the integrand in terms of the difference $t = \tau - \tau'$ around the point $z(\tau)$:

$$S_S \sim \int d\tau \int \sum_{k,l} B_{kl}(\tau) \delta^{(k)}(t^2 - \varepsilon^2) t^l dt. \tag{68}$$

Here the coefficients $B_{kl}(z)$ depend on the curvature, while the delta-functions are flat: $\delta^{(k)}(t^2 - \varepsilon^2)$. By virtue of parity, the integrals with odd l vanish, so only the odd inverse powers of ε will be present in the expansion. Moreover, once we know the divergent terms in some dimension D , we can obtain by differentiation all divergent terms in $D + 2$, except for $1/\varepsilon$ term. The linearly divergent term corresponds to $l = 2k$. The integral is equal to

$$\int_0^\infty \delta^{(k)}(t^2 - \varepsilon^2) t^{2k} dt = \frac{(-1)^k (2k - 1)!!}{2^{k+1}} \frac{1}{\varepsilon}.$$

In four dimensions this term is unique. Applying our recurrence chain we obtain the inverse cubic divergence in six dimensions and calculate again the linearly divergent term. Thus, in $D = 2d$ dimensions we will get $d - 1$ divergent terms from which $d - 2$ can be obtained by the differentiation of the previous-dimensional divergence, and the linearly divergent will be new. This linearly divergent self-action term in the action will have generically the form

$$S_S^{(-1)} \sim \frac{1}{\varepsilon} \int \sum_{k=0}^{d-2} \frac{(-1)^k (2k - 1)!!}{2^{k+1}} B_{k,2k}(\tau) d\tau.$$

Here the coefficient functions are obtained taking the coincidence limits of the two-point tensors involved in the expansion of the Hadamard solution. They actually depend on the derivatives of the world-line embedding function $z(\tau)$ as well as the curvature terms taken on the world line:

$$B_{k,2k}(\tau) = B_{k,2k}(\dot{z}, \ddot{z}, \dots, R(z(\tau)), R_{\mu\nu}(z(\tau)), \dots).$$

The vector case is technically the same, now one has to expand in powers of t in the integrand of

$$S^S = \frac{1}{2} \int G_{\mu\alpha}^S(x, x') j^\mu(x) j^\alpha(x') \sqrt{g(z)g(z')} dx dx'.$$

From this analysis it follows that in any dimension the highest divergent term can be absorbed by the renormalization of the mass as in the generating four-dimensional case. To absorb the remaining $d - 2$ divergences one has to add to the initial action the sum of $d - 2$ counter-terms depending on higher derivatives of the particle velocity. Typically the counter-terms depend on the Riemann tensor of the background.

4.3 Four Dimensions

In four dimensions the scalar retarded Green's function contains a single direct term localized on the light cone and a tail term:

$$G_{\text{ret}} = \frac{1}{4\pi} \theta[\Sigma(x), x'] \left[\Delta^{1/2} \delta(\sigma) - v\theta(\sigma) \right]. \tag{69}$$

The retarded solution for the scalar field reads

$$\phi_{\text{ret}}(x) = m_0 q \int_{-\infty}^{\tau_{\text{ret}}(x)} \left[-\Delta^{1/2} \delta(\sigma) + v\theta(\sigma) \right] d\tau'. \tag{70}$$

Differentiating this expression we obtain $\phi_v = \partial_v \phi$ on the world line:

$$\phi_v(z(\tau)) = m_0 q \int_{-\infty}^{\tau} \left[-\Delta^{1/2} \delta'(\sigma) \sigma_v - \Delta_{;v}^{1/2} \delta(\sigma) - v\delta(\sigma) \sigma_v + v_v \right] d\tau', \tag{71}$$

where integration is performed along the past history of the particle. All the two-point functions are taken on the world line at the points $x = z(\tau)$ (observation point) and $z' = z(\tau')$ (emission point).

To compute local contributions from the terms proportional to a delta-function or its derivative, it is enough to expand the integrand in terms of the difference $s = \Delta\tau = \tau - \tau'$ around the point $z(\tau)$. The Taylor (covariant) expansion of the fundamental biscalar $\sigma(z(\tau), z(\tau'))$ is given by [2]:

$$\sigma(z(\tau), z(\tau')) = \sum_{k=0}^{\infty} \frac{1}{k!} D^k \sigma(\tau, \tau) (\tau - \tau')^k, \tag{72}$$

where we denote as $Q(\tau, \tau')$ the quantity $Q(z(\tau), z(\tau'))$ for any Q taken on the world line, and D is a covariant derivative along the world line (also denoted by a dot):

$$D\sigma \equiv \dot{\sigma} = \sigma_\alpha \dot{z}^\alpha, \quad D^2\sigma \equiv \ddot{\sigma} = \sigma_{\alpha\beta} \dot{z}^\alpha \dot{z}^\beta + \sigma_\alpha \ddot{z}^\alpha, \text{ etc.}$$

Such an expansion exists since the difference $s = \tau - \tau'$ is a two-point scalar itself: this is the integral from the scalar function $\int (-\dot{z}^2)^{1/2} ds$, along the world line from $z(\tau)$ to $z(\tau')$. Taking the limits and using $\dot{z}^2(\tau) = -1$, we find:

$$\sigma(s) = -\frac{s^2}{2} - \ddot{z}^2(\tau) \frac{s^4}{24} + \mathcal{O}(s^5). \tag{73}$$

To obtain an expansion of the derivative of σ over $z^\mu(\tau)$ one can expand $\sigma^\mu(\tau, \tau - s)$ in powers of s . This quantity transforms as a vector at $z(\tau)$ and a scalar at $z(\tau')$,

$$\sigma^\mu(s) = s \left(\dot{z}^\mu - \ddot{z}^\mu \frac{s}{2} + \dddot{z}^\mu \frac{s^2}{6} \right) + \mathcal{O}(s^4), \tag{74}$$

where the index μ corresponds to the point $z(\tau)$: $\sigma^\mu = \partial\sigma(z, z')/\partial z_\mu$. Recall that the initial Greek indices correspond to $z(\tau')$. The expansion of $\delta(-\sigma)$ will read:

$$\delta(-\sigma) = \delta(s^2/2) + s^4 \frac{\ddot{z}^2(\tau)}{24} \delta'(s^2/2) + \dots, \tag{75}$$

where the derivative of the delta-function is taken with respect to the full argument. Since the most singular term is $\Delta^{1/2} \delta'(\sigma) \sigma_\nu$, the maximal order giving the nonzero result after the integration is s^3 . (Note, that in order to use our dimensional recurrent relations to obtain a reaction force in higher dimensions we should perform an expansion up to higher orders in s .) Thus, with the required accuracy, $\delta(-\sigma) = 2\delta(s^2)$, and all the integrals for the delta-derivatives are the same as in the flat space-time. This allows us to use the same regularization of the delta-functions with double roots $\delta(s^2)$ by the point-splitting. Expanding the biscalar $\Delta^{1/2}$ and its gradient at z we have:

$$\Delta^{1/2} = 1 + \frac{s^2}{12} R_{\sigma\tau} \dot{z}^\tau \dot{z}^\sigma, \quad \partial_\nu \Delta^{1/2} = \frac{s}{6} R_{\nu\tau} \dot{z}^\tau. \tag{76}$$

Combining all the contributions we obtain finally for the field strength on the world line:

$$\phi_\nu \Big|_{z(\tau)} = m_0 q \left(\frac{1}{2\varepsilon} \ddot{z}_\nu - \frac{1}{3} \dddot{z}_\nu - \frac{1}{6} R_{\nu\tau} \dot{z}^\tau - \frac{1}{6} R_{\gamma\delta} \dot{z}^\gamma \dot{z}^\delta \dot{z}_\nu + \frac{1}{12} R \dot{z}_\nu + \int_{-\infty}^{\tau} v_\nu d\tau' \right). \tag{77}$$

After the renormalization of mass one recovers the familiar equation.

Similarly, in the electromagnetic case one finds the retarded Maxwell tensor on the world line $x = z(\tau)$:

$$F_{\mu\nu}^{\text{ret}} \Big|_{z(\tau)} = e \int_{-\infty}^{\tau} [u_{\mu\alpha}; \nu \delta(\sigma) + u_{\mu\alpha} \sigma_\nu \delta'(\sigma) + v_{\nu\alpha}; \mu + v_{\mu\alpha} \sigma_\nu \delta(\sigma) - \{\mu \leftrightarrow \nu\}] \dot{z}^\alpha d\tau'. \quad (78)$$

Performing expansions on the world line, one easily recover the DeWitt–Brehme–Hobbs result

$$f_{\text{em}}^\mu = e^2 \left[-\frac{\ddot{z}^\mu}{2\varepsilon} + \Pi^{\mu\nu} \left(\frac{2}{3} \ddot{z}_\nu + \frac{1}{3} R_{\nu\alpha} \dot{z}^\alpha \right) + \dot{z}^\nu(\tau) \int_{-\infty}^{\tau} (v_{\alpha;\nu}^\mu - v_{\nu\alpha}{}^{;\mu}) \dot{z}^\alpha(\tau') d\tau' \right], \quad (79)$$

where $\Pi^{\mu\nu} = g^{\mu\nu} - \dot{z}^\mu \dot{z}^\nu$. For geodesic motion the entire reaction force is given by the tail term.

4.4 Self and Radiative Forces in Curved Space-Time

Apart from the tail term, another new feature of the instantaneous momentum balance between the radiating charge and radiation is the presence of the finite contribution in the self part of the reaction force originating from the half-sum of the retarded and advanced potentials. As we have seen in the previous section, in the case of Minkowski space the self part in four and other even dimensions is a pure divergence which has to be absorbed by renormalization of the mass and (in higher dimensions) the bare coupling constants in the higher derivative counter-terms. In curved space, as was first shown in [5], the tail term in the equation for radiating charge moving along the geodesic with nonrelativistic velocity (in weak gravitational field) contains apart from the dissipative term also the conservative force. This conservative force was later found for a static charge in the Schwarzschild metric [26]. Here we would like to explore the significance of this result in view of the above analysis of self/radiative decomposition. In the paper [6] the splitting of the retarded field into the self and radiative parts was not used. Considering the geodesic motion we have the equation:

$$m \ddot{z}^\alpha = e^2 \dot{z}^\beta \int_{-\infty}^{\tau} v_{\text{ret}\beta\gamma}^\alpha \dot{z}^\gamma(s') ds'. \quad (80)$$

Splitting the tail function with respect to T-parity

$$v_{\text{ret}\beta\gamma}^\alpha \dot{z}^\gamma = v_{\text{self}\beta\gamma}^\alpha \dot{z}^\gamma + v_{\text{rad}\beta\gamma}^\alpha \dot{z}^\gamma, \quad (81)$$

and repeating the calculations along the lines of [6] one finds in the weak-field nonrelativistic case the corresponding (spatial) parts of the reaction force in terms of the flat space theory:

$$\mathbf{f}_{\text{self}} = \mathbf{f}_{\text{div}} + \mathbf{f}_{\text{DDSW}}, \quad \mathbf{f}_{\text{rad}} = \mathbf{f}_{\text{Schott}}, \quad (82)$$

where the divergent term is the same as in the flat space. The two finite terms

$$\mathbf{f}_{\text{DDSW}} = \frac{GM e^2}{r^4} \mathbf{r}, \quad \mathbf{f}_{\text{Schott}} = \frac{2}{3} e^2 \dot{\mathbf{a}} \quad (83)$$

are the DDSW force and the Schott force. Therefore, the DDSW force is a finite part of the T-even (self) contribution to the tail, while the T-odd (rad) part reproduces the Schott term which is precisely the same as in the flat space. Similar result holds for the scalar radiation.

5 Gravitational Radiation

Gravitational radiation in the framework of linearized gravity on the curved background may seem similar to scalar or electromagnetic radiation, but the similarity is incomplete. The difference is that radiation of nongravitational nature emitted by bodies moving along geodesics in the fixed background is not influenced by the nonlinear nature of the Einstein equations. For gravitational radiation this is not so. The scalar or vector linear field equations in curved space imply vanishing of the covariant divergence of the field stress tensor of matter field with respect to the background metric. In the gravitational case one has the Bianchi identity that generally does not imply the covariant conservation of the matter perturbation stress tensor alone unless the background is vacuum. In most of the literature on gravitational radiation reaction the background is assumed to be vacuum. This seems to be satisfactory for the Schwarzschild or Kerr metrics. But physically we have to deal not with an eternal black hole, but with the collapsing body that is not globally vacuous. Meanwhile the conservation laws that may hold in the asymptotically flat case has to be considered globally. It turns out that we must take into account the contribution from the (perturbed) source of the background field as well. This is directly implied by the Bianchi identities.

5.1 Bianchi Identity

Let the background be generated by the stress tensor $T^{\mu\nu}$. We are interested by gravitational radiation emitted by a point particle of mass m moving along the geodesic of the background. Perturbations caused by the particle are assumed to be small so the particle stress tensor,

$$\overset{m}{T}{}^{\mu\nu} = m \int \dot{z}^\mu(\tau) \dot{z}^\nu(\tau) \frac{\delta(x - z(\tau))}{\sqrt{-g}} d\tau, \quad (84)$$

is the first-order quantity with respect to $\overset{B}{T}{}^{\mu\nu}$. By construction, the tensor (84) is divergence free provided the particle follows the geodesic in the space-time. Since it is the first-order quantity, it must be divergence free with respect to the background covariant derivative up to the terms of the second order. Expanding the full metric $\hat{g}_{\mu\nu} = g_{\mu\nu} + \kappa h_{\mu\nu}$, where $\kappa^2 = 8\pi G$, and the Einstein tensor

$$\hat{G}{}^{\mu\nu} = \overset{B}{G}{}^{\mu\nu} + \overset{1}{G}{}^{\mu\nu}, \quad (85)$$

we could naively expect the full Einstein equations to be of the form

$$\hat{G}{}^{\mu\nu} = \kappa^2 \left(\overset{B}{T}{}^{\mu\nu} + \overset{1}{T}{}^{\mu\nu} \right). \quad (86)$$

Since $\overset{B}{G}{}^{\mu\nu} = \kappa^2 \overset{B}{T}{}^{\mu\nu}$, we then should have

$$\overset{1}{G}{}^{\mu\nu} = \kappa^2 \overset{m}{T}{}^{\mu\nu}. \quad (87)$$

But the left-hand side of this equation is not divergence free with respect to the background covariant derivative. Expanding the full covariant derivative as

$$\hat{\nabla}_\mu = \overset{B}{\nabla}_\mu + \overset{1}{\nabla}_\mu, \quad (88)$$

and taking into account the Bianchi identity for the background $\overset{B}{\nabla}_\mu \overset{B}{G}{}^{\mu\nu} = 0$, we obtain in the first order

$$\overset{B}{\nabla}_\mu \overset{1}{G}{}^{\mu\nu} = - \overset{1}{\nabla}_\mu \overset{B}{G}{}^{\mu\nu}. \quad (89)$$

Therefore

$$\overset{B}{\nabla}_\mu \overset{m}{T}{}^{\mu\nu} = - \overset{1}{\nabla}_\mu \overset{B}{G}{}^{\mu\nu}, \quad (90)$$

where the right-hand side is the first-order quantity. Thus Eq. 87 is contradictory. Physically the reason is that we have to take into account the perturbation of the background $\delta T^{\mu\nu}$ caused by the particle, so that the correct equation should be

$$\overset{1}{G}{}^{\mu\nu} = \kappa^2 \left(\overset{m}{T}{}^{\mu\nu} + \delta T^{\mu\nu} \right). \quad (91)$$

But this is not an equation for $h_{\mu\nu}$, since in order to find $\delta T^{\mu\nu}$ one has to consider the matter field equations for the background metric. The problem thus becomes essentially nonlocal. This nonlocality does not reduce to that of the tail term in the DeWitt–Brehme equation.

5.2 Vacuum Background

If $T^{\mu\nu} = 0$ the above obstacle is removed and one can proceed further with the linearized equations for $h_{\mu\nu}$. The derivation of the reaction force initiated in [21] and completed in [20] was based on the DeWitt–Brehme type calculation involving the integration of the field momentum over a small tube surrounding the particle world line. As was noted later [25], this derivation had some drawbacks (for more recent discussion and further references see [15]). One problem consisted in computing the contributions of “caps” at the ends of the chosen tube segment which were not rigorously calculated. Another problem was the singular integral over the internal boundary of the tube which was simply discarded. In addition, the usual mass renormalization is not directly applicable in the gravitational case: due to the equivalence principle the mass does not enter into the geodesic equations.

In [14] the local derivation of the gravitational reaction force was given which is free from the above problems and, in addition, is much simpler technically. It deals with the quantities defined only on the world line and does not involve the ambiguous volume integrals over the world-tube at all. The elimination of divergences amounts to the redefinition of the affine parameter on the world-line.

We start with reparametrization invariant form of the particle action introducing the einbein $e(\tau)$ on the world line acting as a Lagrange multiplier:

$$S[z^\mu, e] = -\frac{1}{2} \int \left[e(\tau) g_{\mu\nu} \dot{z}^\mu \dot{z}^\nu + \frac{m^2}{e(\tau)} \right] d\tau. \tag{92}$$

Variation with respect to $z^\mu(\lambda)$ and $e(\tau)$ gives the equations

$$\frac{D}{d\tau}(e\dot{z}^\mu) = 0, \quad e = \frac{m}{\sqrt{-g_{\mu\nu} \dot{z}^\mu \dot{z}^\nu}}, \tag{93}$$

and we obtain the geodesic equation in a manifestly reparametrization invariant form

$$\frac{D}{d\tau} \left(\frac{\dot{z}^\lambda}{\sqrt{-g_{\mu\nu} \dot{z}^\mu \dot{z}^\nu}} \right) = 0. \tag{94}$$

Assuming now that the particle motion with no account for radiation reaction is geodesic on the background metric, the perturbed equation in the leading order in κ will read

$$\ddot{z}^\mu = \frac{\kappa}{2} \left(g^{\mu\nu} - \frac{\dot{z}^\mu \dot{z}^\nu}{\dot{z}^2} \right) (h_{\lambda\rho; \nu} - 2h_{\nu\lambda; \rho}) \dot{z}^\lambda \dot{z}^\rho, \tag{95}$$

where contractions are with the background metric.

The particle energy–momentum tensor in our formulation will read

$$T^{\mu\nu} = \int e(\tau) \dot{z}^\mu(\tau) \dot{z}^\nu(\tau) \frac{\delta(x - z(\tau))}{\sqrt{-g}} d\tau, \tag{96}$$

and we choose the non-perturbed ein-bein $e_0 = \text{const}$ as a bare parameter. After the calculation similar to that of the preceding section we obtain

$$\ddot{z}^\mu = \kappa^2 e_0 \left\{ \frac{7}{2\varepsilon} \ddot{z}^\mu + \frac{1}{4} \Pi^{\mu\nu} \int_{-\infty}^{\tau} \left[4v_{\nu\lambda\alpha\beta;\rho} - 2 \left(g_{\nu\lambda} v_{\sigma\alpha\beta;\rho}^\sigma + v_{\lambda\rho\alpha\beta;\nu} \right) - g_{\lambda\rho} v_{\sigma\alpha\beta;\nu}^\sigma \right] \dot{z}'^\alpha \dot{z}'^\beta \dot{z}'^\lambda \dot{z}'^\rho d\tau' \right\}. \quad (97)$$

Renormalization of the einbein is

$$\left(\frac{1}{e_0} - \frac{7\kappa^2}{2\varepsilon} \right) \ddot{z}^\mu = \frac{1}{e} \ddot{z}^\mu. \quad (98)$$

Finally, we choose the renormalized affine parameter so that $\dot{z}^2 = -1$, which is equivalent to setting $e = m$ and obtain the MiSaTaQuWa equation. As was shown in [14] this equation remains valid in a class of non-vacuum metrics, in particular, for Einstein spaces.

5.3 Gravitational Radiation for Non-Geodesic Motion

If the particle world line is non-geodesic, the radiation reaction force contains a putative antidamping term which is a local part of the rad contribution to the self-force:

$$f_{\text{rad}}^\mu = -\frac{11\kappa^2}{3} (g^{\mu\nu} - \dot{z}^\mu \dot{z}^\nu) \ddot{z}_\nu + \text{tail}. \quad (99)$$

The reason is simply that the source of gravitational radiation is incomplete and the stress tensor is not divergence free as required. Indeed, if the force is nongravitational, one has to take into account the contribution of stresses of the field causing the body to accelerate. For instance, to describe gravitational radiation of an electron in the atom, one has to add the contribution from the Maxwell field stresses (spatial components, nonrelativistic motion):

$$\square \psi^{ij} = -\kappa^2 G T^{ij}, \quad T^{ij} = \overset{m}{T}{}^{ij} + \overset{st}{T}{}^{ij}, \quad (100)$$

where

$$\overset{m}{T}{}^{ij} = \sum_{a=1,2} m_a \dot{z}_a^i \dot{z}_a^j \delta^3(X_a), \quad \overset{st}{T}{}^{ij} = -\frac{e_1 e_2}{4\pi} \frac{X_1^i X_2^j}{(X_1^2 X_2^2)^{3/2}} + (i \leftrightarrow j), \quad (101)$$

and $X_a^i = x^i - z_a^i(t)$. Using this source one can calculate the gravitational force and find the gravitational Schott term

$$f_{\text{Gshott}}^i = -\frac{G\mu}{15} \frac{d^5 D^{ij}}{dt^5} x_j, \quad \mu = \frac{m_1 m_2}{m_1 + m_2}, \quad (102)$$

where D^{ij} is the quadrupole moment.

Note that this derivation of the Schott term is not based on the local calculation, the two-body treatment was necessary. These features seem to be general: gravitational radiation reaction from a non-geodesically moving particle cannot be described by some DeWitt–Brehme-type equation.

6 Conclusions

The purpose of this lecture was to discuss some subtle points associated with interpretation of the radiation reaction force. We have shown that the Lorentz–Dirac equation in classical electrodynamics describes the balance of three and not just two momenta: the mechanical momentum of the particle, the momentum of emitted radiation, and the momentum carried by the electromagnetic field bound to the charge. The total momentum is conserved, but this does not imply an instantaneous balance of the emitted momentum and that of the particle. The bound field momentum described by the Schott terms destroys the local balance. The total balance, however, is restored if one consider the situation when the charge has zero acceleration at the initial and final moments, or for a periodic motion subject to averaging. These considerations are equally applicable to radiation reaction of a charge in curved space-time and for gravitational radiation reaction. This explains, in particular, the necessity of averaging in calculating the evolution of the Carter constant in the Kerr field [27].

A novel feature related to curved space is the existence of the finite DDSW force arising due to the tidal deformation of the bound electromagnetic field of the charge. This force is often interpreted as part of radiation reaction force, but one has clearly understood, however, that it has nothing to do either with radiation or with the Schott force. As we have shown, it is given by the T-even part of the retarded field, and thus represents a finite remnant from the mass renormalization.

Derivation of the reaction force of nongravitational nature acting on a charge moving (both geodesically and non-geodesically) in curved space-time can be computed by directly substituting the retarded field into the equations of motion, as in the Minkowski space. The regularization is easily achieved by the point-splitting, and divergences are eliminated by renormalization of mass. In higher dimensions one needs counter-terms depending on higher derivatives of the velocity. Divergences may contain the Riemann tensor of the background.

Gravitational radiation reaction force can be obtained in a way similar to a nongravitational one only in vacuum space-time. In non-vacuum background the

source of radiation apart from the local contribution from the particle must contain the contribution from the perturbed background. This can be seen from the analysis of the Bianchi identity. This second contribution is nonlocal, so the possibility to obtain the equation of the DeWitt–Brehme type seems implausible.

For a non-geodesically moving mass the formal derivation of the reaction force leads to putative antidamping effect. To cure this problem one has to take into account the contribution of stresses forcing the mass to accelerate. Then in the nonrelativistic case one derives the gravitational quadrupole Schott term, but the derivation is nonlocal. This is another example when the (quasi)local equation of motion with the reaction force does not exist. Here by quasilocality we mean the possibility of the tail term.

Acknowledgements The author is grateful to the Organizing Committee for invitation and support in Orléans. Useful discussions with the participants of the School on Mass and Capra conference are acknowledged. Most of the results presented in this lecture were obtained in collaboration with Pavel Spirin, to whom the author is indebted. The work was supported by the RFBR project 08-02-01398-a.

References

1. G. Arreaga, R. Capovilla, J. Guven, *Class. Quant. Grav.* **18**, 5065 (2001)
2. S.M. Christensen, *Phys. Rev. D* **14**, 2490 (1976)
3. S. Detweiler, *Class. Q. Grav.* **22**, S681 (2005)
4. S. Detweiler, B. F. Whiting, *Phys. Rev. D* **67**, 024025 (2003)
5. B.S. DeWitt, R.W. Brehme, *Ann. Phys. (N.Y.)* **9**, 220 (1960)
6. C.M. DeWitt, B.S. DeWitt, *Physics* **1**, 3 (1964)
7. P. Dirac, *Proc. R. Soc. Lond. A* **167**, 148 (1938)
8. T. Fulton, F. Rohrlich, *Ann. Phys. (N.Y.)* **9**, 499 (1960)
9. D.V. Gal'tsov, *J. Phys. A* **15**, 3737 (1982)
10. D.V. Galtsov, *Phys. Rev. D* **66**, 025016 (2002)
11. D.V. Gal'tsov, P. Spirin, *Grav. Cosmol.* **12**, 1 (2006)
12. D.V. Galtsov, P.A. Spirin, *Grav. Cosmol.* **13**, 241 (2007)
13. D.V. Galtsov, P.A. Spirin, Radiation in odd dimensions, in preparation
14. D. Gal'tsov, P. Spirin, S. Staub, in *Gravitation and Astrophysics*, ed. by J.M. Nester, C.-M. Chen, J.-P. Hsu (World Scientific, Singapore, 2006), p. 345, arXiv:gr-qc/0701004v1 (2006)
15. S.E. Gralla, R.M. Wald, *Class. Q. Grav.* **25**, 205009 (2008)
16. J.M. Hobbs, *Ann. Phys. (N.Y.)* **47**, 141 (1968)
17. D. Ivanenko, A. Sokolov, *Sov. Phys. Doklady* **36**, 37 (1940); in Russian, *Classical Field Theory* (Gostehizdat, Moskva, 1948); in German, *Klassische feldtheorie* (Akademie Verlag, Berlin, 1953)
18. B.P. Kosyakov, *Theor. Math. Phys.* **199**, 493 (1999)
19. B.P. Kosyakov, *Introduction to the Classical Theory of Particles and Fields* (Springer, Berlin, 2007)
20. Y. Mino, M. Sasaki, T. Tanaka, *Phys. Rev. D* **55**, 3457 (1997)
21. C. Morette-DeWitt, J. L. Ging, *C. R. Acad. Sci. Paris* **251**, 1868 (1960)
22. E. Poisson, *Living Rev. Rel.* **7**, URL: <http://www.livingreviews.org/lrr-2004-6>
23. F. Rohrlich, *Nuovo Cimento* **21**, 811 (1961)
24. F. Rohrlich, *Classical Charged Particles*, 2nd edn. (Addison-Wesley, Reading, MA, 1965; Redwood City, CA, 1990)

25. J.M. Sanchez, E. Poisson, arXiv:gr-qc/0512111v2 (2007)
26. A.G. Smith, C.M. Will, Phys. Rev. D **22**, 1276 (1980)
27. T. Tanaka, Prog. Theor. Phys. Suppl. **163**, 120 (2006)
28. C. Teitelboim, Phys. Rev. D **1**, 1572 (1970)
29. C. Teitelboim, D. Villarroel, Ch. G. van Weert, Riv. Nuovo Cimento **3**, 1 (1980)

The State of Current Self-Force Research

Lior M. Burko

Abstract We briefly review some of the issues relevant to self-force research. We consider frequency-domain and time-domain approaches, adiabatic and post-adiabatic waveforms, the need for a revised regularization method for 2+1D numerical-simulations, the so-called m -mode regularization, and the need for second-order self-forces and their effect on the gravitational waveforms.

1 Introduction

The two-body problem in General Relativity has been practically solved only recently, 90 years after the formulation of the theory [3, 16, 42]. The solution for two bound gravitating bodies, including their combined gravitational field, is numerical, and exhibits the rich strong nonlinear dynamics that characterizes the problem. Indeed, numerical relativity has made impressive progress over the last couple of years, and is now routinely used to study phenomena such as spin flips and kick velocities in binary black hole mergers (see, e.g., [2, 17] and references cited therein).

The typical numerical relativity problem involves two objects of comparable masses. Therefore, the dynamical timescale for the problem in the strong-field regime is comparable to the orbital period. Specifically, the binary merges within a few orbits, starting with initial separation comparable to the system's mass. The situation is quite different when the masses of the two members of the binary are very different. In the extreme mass ratio case the system exhibits two distinct timescales, namely the orbital period timescale and the binary evolution timescale, the latter being much greater than the former. Full numerical relativity simulations would therefore be a wasteful approach to undertake: the binary evolves over very many orbits. Indeed, for typical extreme mass-ratio inspirals (EMRIs) of astrophysical

L.M. Burko (✉)

Department of Physics and the Center for Space Plasma and Aeronomic Research,
University of Alabama in Huntsville, Huntsville, AL 35899, USA
e-mail: burko@uah.edu

interest the expected number of orbits is in the range $10^5 - 5 \times 10^5$. Specifically, for a binary with masses M and $\mu \ll M$, the orbital timescale $\sim M$, and the inspiral timescale $\sim M^2/\mu$, which is longer than the orbital timescale by a factor of $M/\mu \gg 1$. The number of orbits therefore scales with M/μ . Following a full numerical relativity simulation over this many orbits is as yet impossible, and certainly is impractical. Instead, the two very different timescales suggest a different approach, based on adiabatic evolution of the system. The extreme mass ratio of the two members of the binary suggests the use of perturbation theory, with the smallness parameter being the binary's mass ratio μ/M .

Perturbation theory simplifies the problem in some sense, although practically the problem is as yet unsolved, as can be attested by the papers in these Proceedings. The introduced simplicity of a point particle brings with it an entire Pandora's box of problems: Is the model of a point particle self-consistent within General Relativity?

In this chapter we focus our overview on the frequency-domain (FD) and time-domain (TD) approaches to calculate EMRIs, with the objective of finding accurate gravitational wave templates. Regularization methods are discussed elsewhere. An important requirement for such templates is that the source parameters span a very wide range in eccentricities. EMRIs in the good-sensitivity frequency band for LISA are expected to span a wide range in eccentricities. Specifically, such sources are believed to be created as a result of the compact object's scattering by multi-body interactions onto a highly eccentric orbit in the spacetime of the central black hole. For central black holes with mass of $M = 3 \times 10^6 M_\odot$ it was shown by Hopman and Alexander [31] that the probability distribution function of compact objects entering the LISA band with eccentricity ε in the range $0 \leq \varepsilon \leq 0.81$ peaks for $\varepsilon \sim 0.6-0.7$. Intermediate mass black holes would have even higher eccentricities. Specifically, it was shown in [31] that for a central black hole of $M = 10^3 M_\odot$ the maximal eccentricity is $\varepsilon_{\max} = 0.998$ and all inspiraling compact objects (except for white dwarfs that are likely to be tidally torn) are likely to have eccentricities close to the maximal value. The wide range of eccentricities of possible LISA EMRI sources raises the question of how the construction of theoretical templates depends on the parameters of the EMRI. Specifically, we are interested in how different approaches to compute theoretical gravitational-wave templates depend on the eccentricity.

Construction of theoretical templates is important both for detection of EMRIs gravitational waves and for accurate parameter estimation. Numerical waveforms can be constructed using the FD or the TD approaches. The former approach has been developed to very high accuracy, and is considered robust and accurate [1, 18–20, 26, 28, 32, 33, 38, 39, 46, 49–51]. On the other hand, advances to the TD approach have been hindered first by the success of the FD approach [29, 32], and by the crudity of the initial attempts to evolve numerically the fields coupled to a point-like source with the Teukolsky equation [34, 37]. Significant improvement in the accuracy of TD solutions of the inhomogeneous Teukolsky equation, that is, the 2+1D solution of the Teukolsky equation coupled to a point mass, was recently achieved in [15, 47]. For the first time, it was shown that TD calculations can be as accurate as FD calculations. The TD method of [15] was improved with the introduction of the “discrete delta” model of the source [47] and an appropriate low pass fil-

ter that makes the discrete delta useful also for eccentric or inclined orbits [48]. Specifically, correlation integrals of gravitational waveforms sourced by generic orbits (i.e., inclined and eccentric orbits) done for the same system in the FD and TD approaches show that the two agree to a high level [48]. One may therefore argue that the two methods are comparable in the results they are capable of producing. We therefore contend that the viewpoint that the TD solution of the inhomogeneous Teukolsky equations is far from being competitive from the FD solution can no longer be supported. However, we believe that one should not seek *competition* of the two approaches, but rather how they *complement* each other, as either method has nonoverlapping strengths.

The organization of this paper is as follows. In Section 2 we discuss the Teukolsky equation, and its use for finding adiabatic waveforms, followed by a procedure based on the linearized Einstein equations in the Lorenz gauge. In Section 3 we discuss the FD approach for calculating the self-force, and in Section 4 the TD counterpart. Finally, in Section 5 we discuss post-adiabatic waveforms and the importance of second-order self-forces.

2 The Teukolsky Equation

2.1 The Inhomogeneous Teukolsky Equation with a Distributional Source

The Teukolsky equation, that governs the perturbations of Kerr geometry in Boyer–Lindquist coordinates with a matter-source term, is given by [52]

$$\begin{aligned} & \left[\frac{(r^2 + a^2)^2}{\Delta} - a^2 \sin^2 \theta \right] \frac{\partial^2 \psi}{\partial t^2} + \frac{4Mar}{\Delta} \frac{\partial^2 \psi}{\partial t \partial \phi} + \left[\frac{a^2}{\Delta} - \frac{1}{\sin^2 \theta} \right] \frac{\partial^2 \psi}{\partial \phi^2} \\ & - \Delta^{-s} \frac{\partial}{\partial r} \left(\Delta^{s+1} \frac{\partial \psi}{\partial r} \right) - \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \psi}{\partial \theta} \right) - 2s \left[\frac{a(r-M)}{\Delta} + \frac{i \cos \theta}{\sin^2 \theta} \right] \frac{\partial \psi}{\partial \phi} \\ & - 2s \left[\frac{M(a^2 - r^2)}{\Delta} - r - ia \cos \theta \right] \frac{\partial \psi}{\partial t} + [s^2 \cot^2 \theta - s] \psi = 4\pi(r^2 + a^2 \cos^2 \theta)T, \quad (1) \end{aligned}$$

where $\Delta = r^2 - 2Mr + a^2$ and $T = 2\rho^{-4} T_4$, where

$$\begin{aligned} T_4 = & (\Delta + 3\gamma - \gamma^* + 4\mu + \mu^*) [(\delta^* - 2\tau^* + 2\alpha) T_{nm^*} \\ & - (\Delta + 2\gamma - 2\gamma^* + \mu^*) T_{m^*m^*}] + (\delta^* - \tau^* + \beta^* + 3\alpha \\ & + 4\pi) [(\Delta + 2\gamma + 2\mu^*) T_{nm^*} - (\delta^* - \tau^* + 2\beta^* + 2\alpha) T_{nm}]. \quad (2) \end{aligned}$$

Here, $\rho := -1/(r - ia \cos \theta)$. All the symbols used above are defined in [52]. By choosing their values in Boyer–Lindquist coordinates and expressing Eq. 1

explicitly, we obtain an explicit but quite complicated form for the source term. The Teukolsky function ψ is a *gauge-invariant* projection of the Weyl tensor on the null legs of the Kinnersley tetrad. Specifically, it is $\psi_4 := -C_{\alpha\beta\gamma\delta}n^\alpha\bar{m}^\beta n^\gamma\bar{m}^\delta$. There is also a comparable equation for $\psi_0 := -C_{\alpha\beta\gamma\delta}\ell^\alpha m^\beta \ell^\gamma m^\delta$. Note, that although ψ_0 and ψ_4 are gauge invariant, they are *tetrad dependent*. It is therefore important that they are projected off the Weyl tensor in the Kinnersley tetrad. The major problem with evaluating self-forces using the Teukolsky equation approach is the reconstruction of the metric perturbations needed for the determination of the self-force, including the problem of finding the contributions of the non-radiative modes $\ell = 0, 1$, where ℓ is the colatitude number.

One technical difficulty associated with the source term on Eq. 1 is that it includes not just delta functions, but also their gradients. Several approaches to the numerical modeling of these distributions have been attempted [15, 47].

2.2 Adiabatic Waveforms

The source term for the Teukolsky equation (1) depends on an integral over the world line of the orbiting particle, specifically the various projections of the stress-energy tensor on the null tetrad vectors. The most immediate problem is that the complete history of the source term includes the radiation-reaction affected world line, which is the very quantity we are seeking.

Several approaches to address this difficulty have been proposed. The simplest approach, undertaken in [32], is to approximate the world line as a Kerr geodesic orbit. This approximation is valid in the adiabatic limit, and is useful to find the first-order radiative correction. The approach suggested in [32] is to model the short timescale by geodesic Kerr motion, with constant constants of motion. The latter vary appreciably only on longer timescales, such that they are not constants of the motion in the strict sense, but would be in the absence of radiation-reaction effects. The short timescale pieces are then connected by making small adjustments to the constants of the motion, in a manner that preserves global conservation laws by means of the fluxes of energy and angular momentum to infinity and down the event horizon of the Kerr black hole. This approach is complicated by the third constant of the motion, namely the Carter constant. The latter, being essentially the square of the angular momentum, is a nonadditive constant of motion in the absence of radiation reaction. One may therefore not associate “this much” Carter constant to this part of space, and “that much” to another because of essentially self-interference effects. The evaluation of the rate of change of Carter’s constant normally requires the local self-force, and therefore the so-called radiation reaction without radiation-reaction approach has limited applicability, and cannot address generic cases. The nonadditivity of Carter’s constant implies that one may not use balance arguments as with the fluxes of energy and angular momentum, and one must then seek alternative methods for evaluating the rate of change of Carter’s constant.

2.3 Numerical Solution of the Teukolsky Equation

Direct numerical solutions of the Teukolsky equation (1) is problematic because of its long-range effective potential. At great distances, $r_* \rightarrow \infty$, where r_* is the “tortoise coordinate,” the outgoing and incoming solutions behave like $e^{i\omega r_*}$ and $r^{-4} e^{-i\omega r_*}$, respectively. The latter then are swamped by the former, and the solution becomes inaccurate. The solution for this problem is to solve in practice the Sasaki–Nakamura equation, that enjoys a shorter effective potential, and therefore does not share the technical problem of its Teukolsky counterpart. The solutions to the Sasaki–Nakamura equation and the Teukolsky equation are related by a transformation that can be calculated accurately anywhere, including the radiation zone. Therefore, one may in practice solve the Sasaki–Nakamura equation, and then transform the solutions to the solutions of the Teukolsky equation, without ever needing to solve directly the latter.

For generic orbits, when working in the FD one needs to sum over harmonics of the three fundamental frequencies, namely, Ω_ϕ , Ω_θ , and Ω_r . When the orbits are relatively simple, the number of modes to be summed over is easily manageable, and the FD approach leads to a fast and accurate solution, exhibiting the so-called voices of the gravitational waveforms. It is important to emphasize that for such orbits the FD calculation is much more computationally efficient than its TD counterpart. However, for generic orbits, the situation is quite different. Already for equatorial orbits, the number of modes to be calculated for waveforms calculated at a given accuracy level grows fast with the eccentricity of the orbit, as was shown recently by Barton et al. [11]. Indeed, Drasco has shown how the different overtones are excited and evolve during the motion, and in particular how their number and distribution evolve [25].

In contrast, the TD calculation time does not show such a strong dependence on the eccentricity of the orbit. When a fixed spatial and temporal grid density is used, the calculation time is even independent of the orbital parameters [11]. Weak dependence of the computation time is found, however, when the grid parameters are optimized for the orbital parameters, for example, when the grid parameters are taken to be the coarsest that give rise to the desired accuracy level. The TD calculation may be faster than a FD calculation for the same orbit and the same accuracy level, specifically when the orbital eccentricity is very high. At lower eccentricities, the FD calculation would be faster. It therefore makes sense that in practice certain parts of the parameter space would be calculated with one method, and other parts of the parameter space with another. Figure 1 shows the comparison of the computational time for individual azimuthal modes m for a particular orbit for FD and TD calculations. Typically, the TD computation time is insensitive to the value of m , but the FD computation time increases fast with m . One should bear in mind, however, that with a fixed grid resolution, one is in practice limited in the resolution of high m (and therefore also ℓ) modes, so that the θ resolution is limited. Increasing the grid density to enhance the θ resolution would shift the entire TD curve upward. One may of course adapt the grid angular resolution to the calculated mode, which is expected to result in an increasing computational time as a function of the mode number m . Quantitative studies of this question are yet to be done.

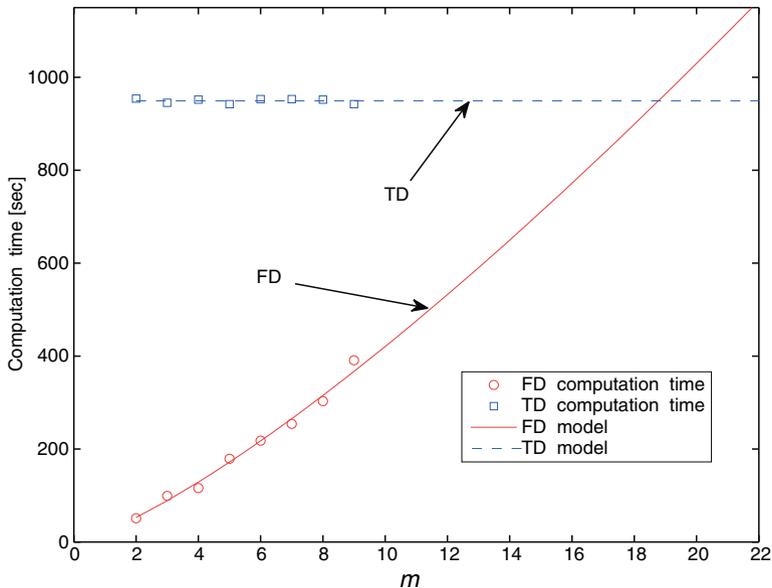


Fig. 1 The computation time of individual azimuthal m modes in a frequency-domain (FD) and a time-domain (TD) computation. The orbit has semilatus rectum $p = 4.64M$, and eccentricity $\varepsilon = 0.5$, and the black hole spin is $a = 0.5M$. The FD results (in 2+1D simulations) are shown in circles, and the TD results in squares. For more detail, see [11]

One difficulty of the inhomogeneous Teukolsky equation is that the source term includes derivatives of delta functions. These derivatives are a source for high-frequency numerical noise in the TD, which needs to be removed either by the introduction of a low pass filter (as was done in [47]), or by introducing artificial viscosity.

2.4 The Linearized Einstein Equations

Another approach is to write directly the linearized Einstein equations in the Lorenz gauge. The source term for the linearized Einstein equations includes only delta functions, but not their derivatives. Therefore, there is reason to expect numerical noise production to be lower than in the case of the Teukolsky equation. Of course, the solution of the linearized Einstein equations involves the solution of 10 coupled wave equations, in addition to 4 gauge conditions that need to be monitored or even enforced. The field equations take the form

$$D^2 \bar{h}_{\alpha\beta} + 4g^{\gamma\delta} \Gamma_{\gamma(\alpha}^{\lambda} \bar{h}_{\beta)\lambda,\delta} + 2\bar{h}_{\lambda(\alpha} \mathcal{G}_{\beta)}^{\lambda} + \bar{h}_{\lambda\sigma} \mathcal{F}_{(\alpha\beta)}^{\lambda\sigma} + 2R^{\mu}{}_{\alpha}{}^{\nu}{}_{\beta} \bar{h}_{\mu\nu} = -16\pi T_{\alpha\beta}, \quad (3)$$

where

$$\mathcal{G}_\alpha^\lambda := g^{\gamma\delta} \left[2\Gamma_{\sigma(\gamma}\Gamma_{\alpha)\delta}^\sigma - \Gamma_{\alpha\gamma,\delta}^\lambda \right] \tag{4}$$

$$\mathcal{F}_{\alpha\beta}^{\lambda\sigma} := 2g^{\gamma\delta}\Gamma_{\gamma(\alpha}\Gamma_{\beta)\delta}^\sigma. \tag{5}$$

Here, D^2 is the scalar field wave operator, that is,

$$D^2 := g^{\gamma\delta} \partial_\gamma \partial_\delta - g^{\gamma\delta} \Gamma_{\delta\gamma}^\sigma \partial_\sigma. \tag{6}$$

The source is given by

$$T_{\alpha\beta} = \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{-g}} \delta^4 \left[x^\mu - x_p^\mu(\tau) \right] u_\alpha u_\beta d\tau, \tag{7}$$

and includes only delta functions but not their derivatives. This approach is now being undertaken [36].

3 Frequency-Domain Calculations of the Self-Force

The most popular approach for calculating the self-force in the FD is to use mode-sum regularization or the Detweiler–Whiting decomposition. Specifically, one calculates directly the colatitude ℓ modes of the divergent, bare self-force, and then subtracts the divergent piece. If this is done carefully enough, one obtains the unique finite, physical self-force.

The simplest case is that of scalar field self-force sourced by a Schwarzschild circular orbit. Indeed, this case was the first to be studied [12, 23]. In principle, one may find numerically the large ℓ behavior of the individual field modes, and perform mode-sum regularization using the numerically obtained values. While this approach is not impossible, one can obtain better and more accurate results using the analytically obtained regularization parameters. The latter were calculated for generic Kerr geodesics, so that they are readily available [8]. A similar approach was successfully applied also for the gravitational self-force (radial Schwarzschild trajectories) [7].

3.1 Mode-Sum Regularization

In its simplest form, the mode-sum approach requires finding the contributions to the self-force mode by mode. The contribution to the physical self-force from the tail part of the Green’s function can be decomposed into stationary Teukolsky modes, and then summed over the frequencies ω and the azimuthal numbers m .

The self-force equals then the limit $\varepsilon \rightarrow 0^-$ of the sum over all ℓ modes, of the difference between the force sourced by the entire world line (the bare force ${}^{\text{bare}}f_\mu^\ell$) and the force sourced by the half-infinite world line to the future of ε , where the particle has proper time $\tau = 0$, and ε is an event along the past ($\tau < 0$) world line. Next, we seek a regularization function h_μ^ℓ which is independent of ε , such that the series

$$\sum_\ell \left({}^{\text{bare}}f_\mu^\ell - h_\mu^\ell \right) \quad (8)$$

converges. Once such a regularization function is found, the regularized self-force is then given by

$${}^{\text{reg}}f_\mu = \sum_\ell \left({}^{\text{bare}}f_\mu^\ell - h_\mu^\ell \right) - D_\mu, \quad (9)$$

where D_μ is a finite valued function of the metric of the space-time of the background, and of the orbital parameters. It has been shown that the regularization function h_μ^ℓ has the general form

$$h_\mu^\ell = \left(\ell + \frac{1}{2} \right) A_\mu + B_\mu + \left(\ell + \frac{1}{2} \right)^{-1} C_\mu. \quad (10)$$

The functions A_μ , B_μ , C_μ , and D_μ are the regularization parameters, and they are functions of the background metric and of the orbital parameters. For generic Kerr geodesics, it was shown that $C_\mu = 0$ (which guarantees that the self-force has no logarithmic divergences), and that $D_\mu = 0$. Note, that unless it is known that $D_\mu = 0$ (or any other function for that matter), one may not calculate the self-force from just the numerically found asymptotic behavior of the ℓ modes of the *bare* force. It is therefore fortuitous that $D_\mu = 0$ for all Kerr geodesics, which may enable practical numerical evaluation of the regularization parameters. However, such calculations are slow, and require many ℓ modes, which is computationally expensive.

A much better way to proceed is to evaluate the regularization parameters A_μ , B_μ analytically, which can be done because they depend only on the local behavior of the Green's function. One may even further speed up the convergence of the sum over all ℓ modes by including the contributions of terms that are higher orders in $1/(\ell + 1/2)$. The first use in this technique was made in [12], and it was further developed in [23].

3.2 The Detweiler–Whiting Regular Part of the Self-Force

In [23], the approach to mode-sum regularization was different: The retarded scalar field ψ^{ret} was divided to a singular part ψ^{S} and a regular part ψ^{R} , such that

$$\psi^{\text{ret}} = \psi^{\text{S}} + \psi^{\text{R}}. \quad (11)$$

Since both the retarded and the singular fields are solutions to the inhomogeneous wave equation that is sourced by the point particle, it immediately follows that the regular field is a solution of the *homogeneous* wave equation. It was shown by Detweiler and Whiting [24], that the correct physical self-force is found when the regular piece ψ^R is used instead of the tail part as in the mode-sum approach discussed above. The Detweiler–Whiting regular piece of the field, that satisfies the homogeneous wave equation, provides us with an alternative, and for many purposes clearer, conceptual viewpoint of the origin of the self-force. The latter originates from the free waves that the particle emitted at retarded times. The particle moves on a geodesic orbit in a perturbed space-time that is regular everywhere in a neighborhood of the world line. This viewpoint complements the older self-force viewpoint, that maintains that the particle moves along an accelerated world line in the background space-time. Either viewpoint has advantages, and we contend that using them complementarily may add insight to practical applications.

The gravitational self-force for circular Schwarzschild orbits was calculated by Detweiler in [22]. The calculation was done in the FD and in the Regge–Wheeler gauge, and was done from the viewpoint of geodesic motion in a (smooth) perturbed spacetime. That is, the particle is following geodesic motion not in Schwarzschild, but rather in Schwarzschild endowed with a linear perturbation field that is sourced by the particle itself at retarded times. Detweiler also considered observables, that is, gauge independent quantities, that are the only meaningful quantities to seek in the gravitational waveforms. Specifically, Detweiler found in [22] the effects on the orbital frequency and on the rate of passage of proper time along the world line. Detweiler was also able to show agreement of his results with the post-Newtonian results in the weak-field limit. Very importantly, Detweiler’s results were also shown to be in agreement, to within numerical accuracy, with the TD calculations done from the complementary viewpoint, that is, that of accelerated motion on a fixed background, that we discuss below in the following section [45].

4 Time-Domain Calculations of the Self-Force

TD calculations have a number of advantages over FD calculations. First, one is freed from the dependence of the computation time on the orbital parameters. Second, it is quite straightforward to specify any world line, and the computation would continue much in the same way. Last, back reaction of the self-force on the world line appears to be done in a more natural way in the TD.

4.1 1+1D Numerical Simulations

In the case of the Teukolsky equation progress on the TD front was lagging behind the FD approach, as discussed above. For self-force calculations, however, the first TD computation of the self-force appeared shortly after the first FD calculation,

albeit for the very simplified problem of scalar field self-force, and radial free fall in Schwarzschild [4]. In [4], the summation over all ℓ modes of the self-force was expedited by approximating the remainder of the partial sum over ℓ modes with the use of the trigamma function.

The TD calculation of scalar field self-force in the case of radial fall in Schwarzschild was then followed in [7] for its gravitational self-force counterpart. The most significant added difficulty in the gravitational self-force case is the gauge problem. Specifically, the mode-sum regularization approach is based on the regularization parameters being calculated in the Lorenz gauge. However, the full, or bare force is most easily calculated in a gauge that simplifies the field equations, namely the Regge–Wheeler gauge for the Schwarzschild space-time. To do the subtraction meaningfully, one needs to have both pieces of the self-force be given in the same gauge. In [7] this problem was fortuitously avoided, because of an accidental coincidence of the results for the regularization parameters in the two gauges for the particular world line of interest. This fortunate coincidence occurs because for radial Schwarzschild geodesics the transformation between the Lorenz gauge and the Regge–Wheeler gauge is regular. For more general orbits this transformation is singular. Other advantages to working in the Lorenz gauge are that it makes hyperbolicity manifest, and that the local singularity of the particle’s field is isotropic and isolated.

Circular Schwarzschild orbits were first considered for the gravitational self-force in [10]. The entire calculation was done in [10] in the Lorenz gauge in the TD. The calculation in [10] was done in 1+1D, which simplified the calculations considerably, and facilitated the successful calculation of the self-force. However, this approach is not generalizable to more complicated orbits, because of the direct use made of the spherical symmetry of the background to reduce the number of dimensions.

4.2 2+1D Numerical Simulations

One therefore needs 2+1D simulations. The problem with 2+1D calculations of the self-force is twofold. First, obviously the computational problem is harder in 2+1D than in 1+1D. The bigger problem, however, is that the mode-sum regularization method was developed for the decomposition of the full (retarded) perturbation field into multipole modes, specifically into ℓ modes, followed by the application of a certain mode-by-mode regularization procedure that depends crucially on the large ℓ behavior of the individual modes. Decomposition of the field is most naturally done in 2+1D in azimuthal m modes, not multipoles ℓ . One therefore needs to first have a robust regularization scheme that is based on the field being decomposed into azimuthal m modes. Such a regularization scheme was indeed presented in [6].

4.2.1 *m*-Mode Regularization

Letting the space-time of a rotating black hole be the background metric $g_{\mu\nu}$ and the (assumed small) perturbation due to the particle be the metric $h_{\mu\nu}$, one may write the vacuum Einstein equations as $G_{\alpha\beta}(g_{\mu\nu} + h_{\mu\nu}) = 8\pi T_{\alpha\beta}$, which can be written as a linearized perturbation over a fixed background in the Lorenz gauge (for the trace-reversed metric perturbations) as follows (see [10] for details and definitions)

$$\nabla_\beta \nabla^\beta \bar{h}_{\mu\nu}(x^\alpha) + 2R^\rho{}_\sigma \bar{h}_{\rho\sigma} = -16\pi T_{\mu\nu}, \tag{12}$$

where the stress-energy tensor of a point particle of mass m moving along an arbitrary world-line $x_p^\alpha(\tau)$ is given by Eq. 7.

As the Kerr space-time is axisymmetric, TD solutions are most convenient when the computational domain is at least 2+1 dimensional (two space dimensions and time). While 3 space dimensions are possible, the computational costs involved would be high, in addition to the lack of a viable regularization method without any mode decomposition. In 2+1 dimensions one solves for the azimuthal modes of the full field. The linearized Einstein equations then take a form similar to Eq. 3. In the context of the Schwarzschild space-time, the numerical solution was done in double-null coordinates [5], which allows for an explicit expression of the solution at a grid point in which it is unknown, when all expressions are evaluated at the center of each computational cell. As the total field is decomposed into azimuthal modes, the divergent behavior of each m mode can be found analytically. Indeed, in the much simpler context of a Schwarzschild black hole background, this analytical behavior was found explicitly in terms of a perturbation expansion [5]. Specifically, the typical metric function h diverges approaching the particle μ as

$$h(x) \sim \frac{\mu}{d}, \tag{13}$$

where d is a measure of the distance from the evaluation point x to the point-like particle (along a space-like geodesic). It can then be shown, that the m mode of the field diverges logarithmically [5]. This situation is not as convenient as in the case that the bare field is decomposed into spherical harmonic modes. In the latter case, each ℓ, m mode of the field is regular, even though the total field diverges. In our case, although each m mode in itself is singular, the singular piece can be found analytically. In the context of a Schwarzschild black hole, this logarithmic divergence was handled by applying a so-called puncture scheme [5]. Specifically, outside a world tube the numerical solution of the bare field is found. The world tube thickness allows for a natural cut off to be introduced, so that no numerical divergences occur. Inside the world tube, however, where the bare field grows unboundedly, one can introduce a “regular field mode” so that

$$h_R^m := h^m - h_p^m, \tag{14}$$

which can be solved numerically inside the world tube as it is regular. The key point is that h_p^m is known analytically (i.e., it has the same logarithmic divergence as the full field). One can then use the known analytic relation of the bare and regular fields, and match smoothly across the surface of the world tube.

4.2.2 The Square of the Geodesic Distance

The square of the geodesic distance is given by

$$S = S_0 + S_1 + S_2 + \dots, \quad (15)$$

where

$$S_0 = (g_{\alpha\beta} + u_\alpha u_\beta) \delta x^\alpha \delta x^\beta, \quad (16)$$

$$S_1 = (g_{\alpha\beta} + u_\alpha u_\beta) \Gamma_{\rho\sigma}^\beta \delta x^\alpha \delta x^\rho \delta x^\sigma. \quad (17)$$

For a circular Schwarzschild orbit,

$$\begin{aligned} S_0 = & \frac{r-2M}{r-3M} \left(\frac{M}{r} \delta x^0 \delta x^0 - 2\sqrt{Mr} \delta x^0 \delta x^3 \right) + \frac{r}{r-2M} \delta x^1 \delta x^1 \\ & + r^2 \delta x^2 \delta x^2 + r^2 \left(\frac{r-2M}{r-3M} \right) \delta x^3 \delta x^3, \end{aligned} \quad (18)$$

$$\begin{aligned} S_1 = & -\frac{M}{(r-2M)^2} \delta x^1 \delta x^1 \delta x^1 \\ & + \frac{\delta x^1}{r^2(r-3M)} \left[M(r-M) \delta x^0 \delta x^0 + r^3(r-3M) \delta x^2 \delta x^2 \right. \\ & \left. + r^3(r-M) \delta x^3 \delta x^3 - 2r(r-M)\sqrt{Mr} \delta x^0 \delta x^3 \right]. \end{aligned} \quad (19)$$

We would normally take a $t = \text{const}$ hypersurface, and evaluate the square of the geodesic distance on that hypersurface, that is, $\delta x^0 = 0$. Our expressions then simplify to

$$S_0 = \frac{r}{r-2M} \delta x^1 \delta x^1 + r^2 \delta x^2 \delta x^2 + r^2 \left(\frac{r-2M}{r-3M} \right) \delta x^3 \delta x^3, \quad (20)$$

$$\begin{aligned} S_1 = & -\frac{M}{(r-2M)^2} \delta x^1 \delta x^1 \delta x^1 + \frac{\delta x^1}{r^2(r-3M)} \left[r^3(r-3M) \delta x^2 \delta x^2 \right. \\ & \left. + r^3(r-M) \delta x^3 \delta x^3 \right]. \end{aligned} \quad (21)$$

Finally $\varepsilon_P = \sqrt{S_0 + S_1}$, neglecting terms of higher order.

4.2.3 The Puncture Function

One introduces next the puncture function

$$\Psi_{\alpha\beta}^{\text{P}} = \frac{4\mu}{\varepsilon_{\text{P}}(x)} \left[\bar{u}_{\alpha}\bar{u}_{\beta} + \left(\bar{\Gamma}_{\alpha\gamma}^{\lambda}\bar{u}_{\beta} + \bar{\Gamma}_{\beta\gamma}^{\lambda}\bar{u}_{\alpha} \right) \bar{u}_{\lambda} \delta x^{\gamma} \right]. \quad (22)$$

Here, the barred quantities are defined as follows. First, the point $\bar{z}^{\alpha}(t)$ is a point along the particle's world line, at the time that the field point x^{α} is evaluated. That is, it is the intersection of the spatial slice containing x^{α} with the particle's world line. Then, \bar{u}_{α} and $\bar{\Gamma}_{\alpha\beta}^{\lambda}$ are evaluated on the particle's world line at the same point. Finally, here $\delta x^{\alpha} := x^{\alpha} - \bar{z}^{\alpha}$.

The puncture function is applied for the trace-reverse metric perturbations, that is, $\Psi_{\alpha\beta}^{\text{P}} \equiv \bar{h}_{\alpha\beta}^{\text{P}}$. We define the *residual field*

$$\bar{h}_{\alpha\beta}^{\text{res}} := \bar{h}_{\alpha\beta} - \bar{h}_{\alpha\beta}^{\text{P}}. \quad (23)$$

Therefore,

$$\begin{aligned} \nabla_{\gamma} \nabla^{\gamma} \bar{h}_{\alpha\beta} + 2R^{\mu}{}_{\alpha}{}^{\nu}{}_{\beta} \bar{h}_{\mu\nu} \\ &= \nabla_{\gamma} \nabla^{\gamma} \left(\bar{h}_{\alpha\beta}^{\text{res}} + \bar{h}_{\alpha\beta}^{\text{P}} \right) + 2R^{\mu}{}_{\alpha}{}^{\nu}{}_{\beta} \left(\bar{h}_{\mu\nu}^{\text{res}} + \bar{h}_{\mu\nu}^{\text{P}} \right) \\ &= \left(\nabla_{\gamma} \nabla^{\gamma} \bar{h}_{\alpha\beta}^{\text{res}} + 2R^{\mu}{}_{\alpha}{}^{\nu}{}_{\beta} \bar{h}_{\mu\nu}^{\text{res}} \right) + \left(\nabla_{\gamma} \nabla^{\gamma} \bar{h}_{\alpha\beta}^{\text{P}} + 2R^{\mu}{}_{\alpha}{}^{\nu}{}_{\beta} \bar{h}_{\mu\nu}^{\text{P}} \right) \\ &= -16\pi T_{\alpha\beta}, \end{aligned} \quad (24)$$

so that the residual field satisfies the equation

$$\nabla_{\gamma} \nabla^{\gamma} \bar{h}_{\alpha\beta}^{\text{res}} + 2R^{\mu}{}_{\alpha}{}^{\nu}{}_{\beta} \bar{h}_{\mu\nu}^{\text{res}} = -16\pi T_{\alpha\beta} - \left(\nabla_{\gamma} \nabla^{\gamma} \bar{h}_{\alpha\beta}^{\text{P}} + 2R^{\mu}{}_{\alpha}{}^{\nu}{}_{\beta} \bar{h}_{\mu\nu}^{\text{P}} \right) := Z_{\alpha\beta}^{\text{res}} \quad (25)$$

Notice, that the puncture field is known analytically. Numerically, Eq. 25 is the one that needs to be evaluated.

Having found the regularized metric perturbations $\bar{h}_{\alpha\beta}^{\text{res}}$, one can now find the self-force. Specifically,

$$f^{\alpha} = -\frac{1}{2}\mu \left(g^{\alpha\lambda} + u^{\alpha}u^{\lambda} \right) \left(h_{\lambda\mu;\nu}^{\text{res}} + h_{\lambda\nu;\mu}^{\text{res}} - h_{\mu\nu;\lambda}^{\text{res}} \right) u^{\mu}u^{\nu}, \quad (26)$$

or, alternatively,

$$f^{\alpha} = \mu k^{\alpha\beta\gamma\delta} \bar{h}_{\beta\gamma;\delta}^{\text{res}}, \quad (27)$$

where

$$k^{\alpha\beta\gamma\delta} = \frac{1}{2}g^{\alpha\delta}u^{\beta}u^{\gamma} - g^{\alpha\beta}u^{\gamma}u^{\delta} - \frac{1}{2}u^{\alpha}u^{\beta}u^{\gamma}u^{\delta} + \frac{1}{4}g^{\beta\gamma}u^{\alpha}u^{\delta} + \frac{1}{4}g^{\alpha\delta}g^{\beta\gamma}. \quad (28)$$

Notice, that Eq. 26 is written for the residual piece of the metric perturbations, while Eq. 27 is written for the trace-reversed perturbations.

This approach was applied successfully by Barack and Golbourn [5] for the scalar field self-force for a circular Schwarzschild orbit. Most importantly, even though the Schwarzschild background is spherically symmetric, no use of this symmetry was done explicitly in [5]. Therefore, the generalization to the Kerr case, and to more generic orbits, should be possible. The gravitational self-force counterpart is now in progress [35].

5 Post-adiabatic Self-Force-Driven Orbital Evolution

5.1 The Importance of Second-Order Self-Forces

Adiabatic templates can describe EMRI waveforms to $O(\mu/M)^{-1}$. To obtain the self-force $O(\mu/M)^0$ corrections, it is not sufficient to include the self-force to leading order in the small body's mass μ : one needs also to include the self force to the next order, that is, to $O(\mu^2)$. In support of this claim, we present here a simple argument, based on [14], for the simplest case of quasi-circular Schwarzschild orbits. One should bear in mind, however, that the regularization problem of second-order self forces is a rather difficult one, because it includes highly singular source terms that make the standard retarded solutions diverge [44].

Let a nonspinning test body of mass μ be accelerated because of its self-force f_α^{SF} in a quasi-circular orbit around a Schwarzschild black hole of mass M (satisfying $\mu/M \ll 1$), such that its four-velocity is u^α and its proper time is τ . The equations of motion (hereafter EOM) are given by

$$\mu \frac{Du^\alpha}{d\tau} = f_\beta^{\text{SF}} g^{\alpha\beta}, \quad (29)$$

which we solve perturbatively order by order in μ/M . The metric in the usual Schwarzschild coordinates is given by

$$ds^2 = -F(r) dt^2 + F^{-1}(r) dr^2 + r^2 d\Omega^2, \quad (30)$$

where $F(r) = 1 - 2M/r$. Here, $d\Omega^2 = d\vartheta^2 + \sin^2\vartheta d\phi^2$.

As the orbit is planar, the ϑ component of the EOM is trivial. We use the normalization condition for u^α , namely $u^\alpha u_\alpha = -1$, to eliminate u^t from the EOM $Du^i/d\tau = \mu^{-1} f_k^{\text{SF}} g^{ik}$, $i, k = t, r, \phi$, where D denotes covariant differentiation compatible with the metric (30). We next use the t component of the EOM to eliminate \dot{u}^t . We can simplify the EOM to first-order (nonlinear) ordinary differential equations by taking $\dot{r} = V(r)$, $\dot{x} = Vx'(r)$, where x denotes any quantity. We find the EOM to be

$$\begin{aligned}
 VV' - \frac{3MV^2}{r(r-2M)} - (r-2M)\sigma \\
 - \frac{1}{\mu(u^t)^2} \left[\left(1 - \frac{2M}{r}\right) f_r^{\text{SF}} + \frac{V}{1-2M/r} f_t^{\text{SF}} \right] = 0,
 \end{aligned} \tag{31}$$

and

$$\begin{aligned}
 V\sigma' - \frac{3MV}{r^4} + 2\frac{M/r^3 + \sigma}{1-2M/r} \left[\frac{2}{r}V \left(1 - \frac{3M}{r}\right) - \frac{f_t^{\text{SF}}}{\mu(u^t)^2} \right] \\
 - \frac{2(M/r^3 + \sigma)^{1/2}}{\mu(u^t)^2 r^2} f_\phi^{\text{SF}} = 0,
 \end{aligned} \tag{32}$$

where $(u^t)^2 = 1/[1 - 3M/r - r^2\sigma - V^2/(1 - 2M/r)]$. We denote by Ω the orbital frequency as measured at infinity, and by an overdot and a prime (partial) differentiation with respect to coordinate time t and r , respectively. Here, σ measures the deviation from Kepler’s law, that is,

$$\Omega^2 = M/r^3 + \sigma(r). \tag{33}$$

This last relation is gauge independent, although each of the terms on the right-hand side (RHS) are separately gauge dependent. Specifically, we fix the gauge by choosing the first term on the RHS to include the radial Schwarzschild coordinate r : under an infinitesimal coordinate transformation $x^\alpha \rightarrow x'^\alpha = x^\alpha + \xi^\alpha$ where the gauge vector $\xi^\alpha = O(\mu)$, the metric perturbations $h_{\alpha\beta} \rightarrow h'_{\alpha\beta} = h_{\alpha\beta} - 2\nabla_{(\alpha}\xi_{\beta)}$. In the last expression, the covariant derivative operator ∇_α is compatible with the background metric (30). A gauge choice of the metric perturbations $h_{\alpha\beta}$ (e.g., the Regge–Wheeler or the Lorenz gauges) is therefore equivalent to a condition on the gauge vector ξ^α . The latter may have a radial component, that changes the radial coordinate description of the orbit. As noted by Detweiler [21], while the angular velocity Ω is gauge invariant, the radius of the orbit r is not. This means that under a gauge transformation the different terms on the RHS of Eq. 33 change, but in such a manner that their combination, or the left-hand side (LHS) of (33) is invariant. One may, therefore, fix the gauge (possibly up to a residual gauge freedom) by determining the ratio of the terms on the RHS of (33), or, alternatively, by fixing the geometrical meaning of the symbol r appearing on the RHS of (33). In particular, the latter may be fixed to equal the Schwarzschild radial coordinate of the unperturbed Schwarzschild background (30). By clarifying the geometrical meaning of r we therefore effectively fix the gauge. Notice that this fixing is *not* equivalent to merely choosing which coordinates are used to describe the background geometry. Our analysis below does not depend, however, on the choice of gauge. One may indeed, for the purpose of the present argument, just argue that the LHS of (33) is determined in some gauge.

We next expand in powers of $\varepsilon := \mu/M$ (do not confuse with the symbol ε used above, denoting the orbital eccentricity) as $\sigma(r) = \sigma^{(1)} + \sigma^{(2)}$, $V = V^{(1)} + V^{(2)}$,

and $a_i = a_i^{(1)} + a_i^{(2)}$, $x^{(j)}$ denoting the term in x which is at $O(\varepsilon^j)$, and a_i being the (self) acceleration. We then expand the self-force as $f_i^{\text{SF}} = f_i^{(1)} + f_i^{(2)}$, where $f_i^{(j)} = \mu a_i^{(j)}$. Solving perturbatively, we find the first-order terms to be

$$\sigma^{(1)} = -\frac{r-3M}{\mu r^2} f_r^{(1)}, \quad (34)$$

$$V^{(1)} = \frac{2r}{\mu M} \frac{r-3M}{r-6M} \left[\left(\frac{M}{r} \right)^{\frac{1}{2}} \left(1 - \frac{2M}{r} \right) f_\phi^{(1)} + M f_t^{(1)} \right], \quad (35)$$

and the second-order terms to be

$$\begin{aligned} \sigma^{(2)} = & -\frac{r-3M}{\mu r^2} \left[f_r^{(2)} + r^2 \frac{V^{(1)} f_t^{(1)}}{(r-2M)^2} \right] \\ & + \frac{r}{\mu} \sigma^{(1)} f_r^{(1)} - \frac{3M V^{(1)2}}{r(r-2M)^2} + \frac{V^{(1)} V^{(1)'}}{r-2M}, \end{aligned} \quad (36)$$

$$\begin{aligned} V^{(2)} = & \frac{r(r-3M)}{\mu^2 M^2 (r-6M)^2} \left[2 \left(\frac{M}{r} \right)^{\frac{1}{2}} f_\phi^{(1)} f_r^{(1)'} r(r-2M)^2 \right. \\ & \times (r-3M) + \left(\frac{M}{r} \right)^{\frac{1}{2}} f_\phi^{(1)} f_r^{(1)} (5r-6M)(r-2M) \\ & \times (r-3M) + 2M f_t^{(1)} f_r^{(1)'} r^2 (r-2M)(r-3M) \\ & + 4M f_t^{(1)} f_r^{(1)} r^2 (r-3M) + 2\mu M^2 f_t^{(2)} (r-6M) \\ & \left. + 2\mu \left(\frac{M}{r} \right)^{\frac{3}{2}} f_\phi^{(2)} (r-2M)(r-6M) \right]. \end{aligned} \quad (37)$$

The first-order corrections for the orbital parameters, given in Eqs. 34 and 35, involve dissipative and conservative effects. Dissipation is included in Eq. 35, whereas Eq. 34 involves conservative changes to the particle's world line. Indeed, the latter depends only on $f_r^{(1)}$, that for circular Schwarzschild orbits is purely conservative.

The second-order corrections to the orbital parameters, given by Eqs. 36 and 37 mix conservative and dissipative effects, and includes terms that are quadratic in $f_i^{(1)}$, in addition to terms that are linear in the second-order self-force, namely $f_i^{(2)}$.

The key element in our argument, is now to show that the $O(\mu/M)^0$ corrections to the gravitational waveforms depend on Eqs. 36 and 37, and therefore depend on $f_i^{(2)}$.

We can study the importance of the higher-order correction by considering two dimensionless quantities, $d\mathcal{N}_{\text{cyc}}/d(\ln f)$, the number of orbits spent in a logarithmic interval of frequencies, and $V/(r\omega)$. Then, we compare these quantities

between a theoretical template accurate to $O(\mu/M)^{-1}$, and a template accurate to $O(\mu/M)^0$. $V/(r\omega)$ is related to the rate of change of the envelope of the chirp wave. Notice that $d\mathcal{N}_{\text{cyc}}/d(\ln f) \equiv \omega^2/[(2)\pi\dot{\omega}]$ is gauge independent. The difference in these quantities between the $O(\mu/M)^0$ expressions and their $O(\mu/M)^{-1}$ counterparts can be expanded using the expressions above. We find that

$$\Delta \frac{d\mathcal{N}_{\text{cyc}}}{d(\ln f)} = -\frac{2}{3\pi} \sqrt{\frac{M}{r}} \left[\frac{3r^3\sigma_{(1)}}{2MV_{(1)}} + \frac{1r^4\sigma'_{(1)}}{3MV_{(1)}} - \frac{V_{(2)}}{V_{(1)}^2} \right], \quad (38)$$

and

$$\Delta \left(\frac{V}{r\omega} \right) / \left(\frac{V}{r\omega} \right)^2 = \sqrt{\frac{M}{r}} \left[\frac{1r^3\sigma_{(1)}}{2MV_{(1)}} - \frac{V_{(2)}}{V_{(1)}^2} \right]. \quad (39)$$

While the LHS of Eq. 38 is gauge invariant, the RHS is written as a sum of three terms, each of which is gauge dependent: it is only their sum which is gauge independent, in the same sense as the discussion following Eq. 33. Notice that the last two quantities are at $O(\mu/M)^0$. Notice also that Eqs. 38 and 39 do not depend on $\sigma_{(2)}$. All the second-order effects are included in $V_{(2)}$, and they all arise from $f_t^{(2)}$ and $f_r^{(2)}$. Recall that these components of the self-force are the dissipative ones. Therefore, in the circular orbit Schwarzschild case, the post-adiabatic effects originate in the first-order self-force in addition to the dissipative piece of the second-order self-force. We have no reason to expect that this division persists also in more general cases. The dependence of Eqs. 36 and 37 on $V_{(2)}$ implies that the second-order self-forces are necessary for the calculation of the $O(\mu/M)^0$ effects in the waveforms. Schematically, denoting the 4-acceleration by a , the dissipative part of $a^{(n)}$ with a subscript “ D ” and the conservative part with a subscript “ C ”, where n is the order of the self-force, we find that for circular Schwarzschild orbits the last terms that determine the self-force are given by

$$a = \underbrace{a^{(0)}}_{=0} + \left[\underbrace{a_D^{(1)}}_{\text{Waveform to } O(\mu/M)^{-1}} + \underbrace{a_C^{(1)}}_{\text{Waveform to } O(\mu/M)^0} \right] \left(\frac{\mu}{M} \right) + \left[\underbrace{a_D^{(2)}}_{\text{Waveform to } O(\mu/M)^0} + \underbrace{a_C^{(2)}}_{\text{Waveform to } O(\mu/M)} \right] \left(\frac{\mu}{M} \right)^2 + \dots \quad (40)$$

Here, we braced together the terms that contribute to the waveforms (specifically, the phase of the waveforms) at various orders in the mass ratio, specifically at order $O(\mu/M)^0$ which is important for parameter estimation. The first term on the RHS in Eq. 40 is the vanishing (in the absence of nongravitational forces) geodesic acceleration, that is included here for completeness.

A more general approach for the post-adiabatic EMRI inspiral was recently presented by Hinderer and Flanagan [30]. Hinderer and Flanagan wrote the EMRI equations of motion in action-angle variables and obtained the post-adiabatic corrections to the orbital motion. The method presented in [30] is more general than the argument we presented above, that was made simple by making the geodesic orbit circular. For that case, [30] found results in agreement with those of [14]. Specifically, for circular and for equatorial orbits, the leading order corrections

are suppressed by one power of the mass ratio, and give rise to phase errors of $O(\mu/M)^0$ over a complete inspiral through the relativistic regime. These post-1-adiabatic corrections are generated by the fluctuating, dissipative piece of the first-order self-force, by the conservative piece of the first-order self-force, and by the orbit-averaged, dissipative piece of the second-order self-force.

5.2 Conservative Self-Force Effects

We have argued that the $O(\mu/M)^0$ self-force effects (at least for the simple case of circular Schwarzschild orbits) include contributions from the dissipative second-order self-force and the conservative first-order self-force. What are these effects on the EMRI gravitational waveform?

The traditional approach to the self-force was to consider the adiabatic waveforms as a satisfactory approximation for the waveform, and to ignore conservative effects, mostly for the argument that they do not accumulate secularly. Conservative effects are in practice ignored when one is using the so-called radiation reaction without radiation-reaction forces. In that approach one uses the fluxes of fields associated with otherwise conserved quantities, and integrates them over a large sphere, taken in practice to be at infinity. As the fields associated with conservative effects drop off quicker at infinity than those associated with dissipative effects, one in practice discards all conservative effects when one integrate over a distant surface. The complementary viewpoint is that of the self-force. The adiabatic self-force is equivalent in practice to the force obtained from the “half-retarded minus half-advanced” potential. As both the advanced and the retarded potentials include the same divergent piece on the world line, their difference is finite, and is equal to the radiation-reaction potential. However, this difference does not capture the conservative piece of the self-force.

The “half-retarded minus half-advanced” approach, while equivalent to the adiabatic self-force [27, 43], suffers from a fundamental difficulty. Physical interaction is normally believed to be retarded, for reasons of causality. The explicit inclusion of the advanced potential in the physical self-force makes the self-force expression acausal, and in some sense also *anti-causal*. While the “half-retarded minus half-advanced” prescription undoubtedly leads to the correct adiabatic results (neglecting for the moment the discarded conservative contributions), the perception of physical reality it leads to is quite different from the accepted one, and required the reality of advanced interactions.

This question of the conservative contributions to the waveforms was considered first in [13]. The model used in [13] was a simple one, intended to only raise this question. Specifically, [13] argued that in addition to the dissipative forces, there is also a conservative force, that in the circular orbit Schwarzschild case is radial. This radial force depends only on the distance from the central black hole, and therefore affects the orbit like any other component to the total radial force that is different from the inverse-square law. Specifically, it causes precession of the

periastron. In the absence of specific information on the functional dependence of the radial component of the self-force, [13] instead considered the effect of various post-Newtonian orders of the radial self-force. It was shown that the effect would be a retrograde contribution to the periastron precession at $O(\mu/M)$, in addition the usual general relativistic periastron precession, which is a geodesic effect at $O(\mu/M)^0$.

The effect of such an excess periastron precession on the waveform is therefore dephasing of the latter: as the orbit is dephased by the periastron precession, the gravitational waveform suffers similar dephasing. As this dephasing is at $O(\mu/M)^0$ it may have an importance implications on EMRI gravitational wave detection and precise parameter estimation.

The simple argument of [13] was revisited in much greater depth in [41]. In [41] a simple toy problem of an electric charge moving slowly in the weak background of a central object, subjected to its *electromagnetic* self force. In this toy model the self-force is known analytically, and the dissipative and conservative contributions can be clearly separated. Pound, Poisson, and Nickel then made use of the conceptual clarity of their model, and eloquently showed how the conservative self-force effects contribute to the dephasing of the orbits, by changing the positional orbital elements while keeping the constant of the motion (the three principal orbital elements in the language of [41]) unchanged, precisely what one would expect of the effects of the conservative self-force effects on the orbit to be. The gravitational self-force case was considered in [40]. It was shown in [40] that the conservative effects cause significant long-term changes in the waveforms, by changing the phasing and the time dependence of the orbit and consequently also those of the waveform.

Another important conservative effect was studied recently in [9]. It was found that the innermost stable circular orbit (ISCO) in Schwarzschild is altered by the conservative self-force, so that $\Delta r_{\text{isco}} = -3.27 \mu$ in the Lorenz gauge, and $\Delta \Omega_{\text{isco}}/\Omega_{\text{isco}} = 0.487 \mu/M$ (gauge independent). These results provide an accurate strong-field benchmark against which various approximation methods can be tested.

Acknowledgements The author was supported in part by NSF grant No. PHY-0757344 and NASA/SSC grant No. NNX07AL52A.

References

1. T. Apostolatos, D. Kennefick, A. Ori, E. Poisson, Phys. Rev. D **47**, 5376 (1993)
2. J.G. Baker, W.D. Boggs, J. Centrella, B.J. Kelly, S.T. McWilliams, M. Coleman Miller, J.R. van Meter, Astrophys. J. **682**, L29 (2008)
3. J.G. Baker, J. Centrella, D.-I. Choi, M. Koppitz, J. van Meter, Phys. Rev. Lett. **96**, 111102 (2006)
4. L. Barack, L.M. Burko, Phys. Rev. D **62**, 084040 (2000)
5. L. Barack, D.A. Golbourn, Phys. Rev. D **76**, 044020 (2007)
6. L. Barack, D.A. Golbourn, N. Sago, Phys. Rev. D **76**, 124036 (2007)
7. L. Barack, C.O. Lousto, Phys. Rev. D **66**, 061502 (2002)

8. L. Barack, A. Ori, *Phys. Rev. Lett.* **90**, 111101 (2003)
9. L. Barack, N. Sago, *Phys. Rev. Lett.* **102**, 191101 (2009)
10. L. Barack, N. Sago, *Phys. Rev. D* **75**, 064021 (2007)
11. J.L. Barton, D.J. Lazar, D.J. Kennefick, G. Khanna, L.M. Burko, *Phys. Rev. D* **78**, 064042 (2008)
12. L.M. Burko, *Phys. Rev. Lett.* **84**, 4529 (2000)
13. L.M. Burko, *Int. J. Mod. Phys. A* **16**, 1471 (2001)
14. L.M. Burko, *Phys. Rev. D* **67**, 084001 (2003)
15. L.M. Burko, G. Khanna, *Europhys. Lett.* **78**, 60005 (2007)
16. M. Campanelli, C.O. Lousto, Y. Zlochower, *Phys. Rev. D* **73**, 061501 (2006)
17. M. Campanelli, C.O. Lousto, Y. Zlochower, D. Merritt, *Astrophys. J.* **659**, L5 (2007)
18. C. Cutler, L.S. Finn, E. Poisson, G.J. Sussman, *Phys. Rev. D* **47**, 1511 (1993)
19. M. Davis, R. Ruffini, W.H. Press, R.H. Price, *Phys. Rev. Lett.* **27**, 1466 (1971)
20. S.L. Detweiler, *Astrophys. J.* **225**, 687 (1978)
21. S. Detweiler, *Class. Q. Grav.* **22**, S681 (2005)
22. S. Detweiler, *Phys. Rev. D* **77**, 124026 (2008)
23. S. Detweiler, E. Messaritaki, B.F. Whiting, *Phys. Rev. D* **67**, 104016 (2003)
24. S. Detweiler, B.F. Whiting, *Phys. Rev. D* **67**, 024025 (2003)
25. S. Drasco, *Phys. Rev. D* **79**, 124016 (2009)
26. L.S. Finn, K.S. Thorne, *Phys. Rev. D* **62**, 124021 (2000)
27. D.V. Gal'tsov, *J. Math. Phys. A. Math. Gen.* **15**, 3737 (1982)
28. K. Glampedakis, S.A. Hughes, D. Kennefick, *Phys. Rev. D* **66**, 064005 (2002)
29. K. Glampedakis, D. Kennefick, *Phys. Rev. D* **66**, 044002 (2002)
30. T. Hinderer, É.É. Flanagan, *Phys. Rev. D* **78**, 064028 (2008)
31. C. Hopman, T. Alexander, *Astrophys. J.* **629**, 362 (2005)
32. S.A. Hughes, *Phys. Rev. D* **61**, 084004 (2000); Erratum, *ibidem* **63**, 049902(E) (2001)
33. S.A. Hughes, *Phys. Rev. D* **64**, 064004 (2001)
34. G. Khanna, *Phys. Rev. D* **69**, 024016 (2004)
35. K. Lackeos, G. Khanna, L. Barack, L.M. Burko, work in progress
36. K. Lackeos, G. Khanna, L.M. Burko, work in progress
37. R. Lopez-Aleman, G. Khanna, J. Pullin, *Class. Q. Grav.* **20**, 3259 (2003)
38. E. Poisson, *Phys. Rev. D* **47**, 1497 (1993); *ibidem* **48**, 1860 (1993); *ibidem* **52**, 5719 (1995); Erratum and Addendum, *ibidem* **55**, 7980 (1997)
39. E. Poisson, M. Sasaki, *Phys. Rev. D* **51**, 5753 (1995)
40. A. Pound, E. Poisson, *Phys. Rev. D* **77**, 044013 (2008)
41. A. Pound, E. Poisson, B.G. Nickel, *Phys. Rev. D* **72**, 124001 (2005)
42. F. Pretorius, *Phys. Rev. Lett.* **95**, 121101 (2005)
43. T.C. Quinn, R.M. Wald, *Phys. Rev. D* **60**, 064009 (1999)
44. E. Rosenthal, *Class. Q. Grav.* **22**, S859 (2005); *Phys. Rev. D* **74**, 084018 (2006)
45. N. Sago, L. Barack, S. Detweiler, *Phys. Rev. D* **78**, 124024 (2008)
46. M. Shibata, M. Sasaki, H. Tagoshi, T. Tanaka, *Phys. Rev. D* **51**, 1646 (1995)
47. P.A. Sundararajan, G. Khanna, S.A. Hughes, *Phys. Rev. D* **76**, 104005 (2007)
48. P.A. Sundararajan, G. Khanna, S.A. Hughes, S. Drasco, *Phys. Rev. D* **78**, 024022 (2008)
49. S. Suzuki, K. Maeda, *Phys. Rev. D* **58**, 023005 (1998)
50. T. Tanaka, Y. Mino, M. Sasaki, M. Shibata, *Phys. Rev. D* **54**, 3762 (1996)
51. T. Tanaka, M. Shibata, M. Sasaki, T. Nakamura, *Prog. Theor. Phys.* **90**, 65 (1993)
52. S.A. Teukolsky, *Astrophys. J.* **185**, 635 (1973)

High-Accuracy Comparison Between the Post-Newtonian and Self-Force Dynamics of Black-Hole Binaries

Luc Blanchet, Steven Detweiler, Alexandre Le Tiec, and Bernard F. Whiting

Abstract The relativistic motion of a compact binary system moving in circular orbit is investigated using the post-Newtonian (PN) approximation and the perturbative self-force (SF) formalism. A particular gauge-invariant observable quantity is computed as a function of the binary's orbital frequency. The conservative effect induced by the gravitational SF is obtained numerically with high precision, and compared to the PN prediction developed to high order. The PN calculation involves the computation of the 3PN regularized metric at the location of the particle. Its divergent self-field is regularized by means of dimensional regularization. The poles $\propto (d - 3)^{-1}$ that occur within dimensional regularization at the 3PN order disappear from the final gauge-invariant result. The leading 4PN and next-to-leading 5PN conservative logarithmic contributions originating from gravitational wave tails are also obtained. Making use of these exact PN results, some previously unknown PN coefficients are measured up to the very high 7PN order by fitting to the numerical SF data. Using just the 2PN and new logarithmic terms, the value of the 3PN coefficient is also confirmed numerically with very high precision. The consistency of this cross-cultural comparison provides a crucial test of the very different regularization methods used in both SF and PN formalisms, and illustrates the complementarity of these approximation schemes when modeling compact binary systems.

L. Blanchet and A. Le Tiec (✉)

*G*R \mathcal{E} C \mathcal{O} , Institut d'Astrophysique de Paris, C.N.R.S. & Université Pierre et Marie Curie, 98^{bis} Boulevard Arago, 75014 Paris, France
e-mail: blanchet@iap.fr; letiec@iap.fr

S. Detweiler and B.F. Whiting

Institute for Fundamental Theory, Department of Physics, University of Florida, Gainesville, FL 32611-8440, USA
e-mail: det@phys.ufl.edu; bernard@phys.ufl.edu

1 Introduction and Motivation

For the gravitational wave observatories LIGO/Virgo/GEO on Earth and LISA in space, routine identification of inspiralling compact binaries (binary systems composed of neutron stars and/or black holes) will require high-accuracy predictions from general relativity theory [24, 49]. Providing such predictions represents a formidable task that can be addressed using approximation schemes in general relativity. The two main approximation schemes available are: (i) the *post-Newtonian* expansion, well suited to describe the inspiralling phase of compact binaries in the slow motion and weak-field regime independently of the mass ratio; and (ii) the *self-force* approach, based on perturbation theory, which gives an accurate description of extreme mass ratio binaries even in the strong-field regime.

The post-Newtonian (PN) templates for compact binary inspiral have been developed to 3.5PN order in the phase [10, 14, 17] and 3PN order in the amplitude [15, 18] (see Blanchet's contribution in this volume).¹ These are suitable for the inspiral of two neutron stars in the frequency bandwidth of LIGO and Virgo detectors. For detection of black-hole binaries (with higher masses), the construction of template banks either requires the matching of the PN waveform with full numerical simulations for the merger phase and the ringdown of the final black-hole [1, 20], or using the effective-one-body formalism [21] (see also Damour's and Nagar's contribution in this volume).

In a completely different parameter regime, gravitational self-force (SF) analysis [30, 33, 36, 40, 43] (see also Poisson's contribution in this volume) is expected to provide templates for extreme mass ratio inspirals (EMRIs) anticipated to be present in the LISA frequency bandwidth. SF analysis is a natural extension of first-order perturbation theory, and the latter has a long history of comparisons with PN analysis [25, 37–39, 41, 45–47]. SF analysis itself, however, is just now mature enough to present some limited comparisons with PN analysis, although it is not yet ready for template generation.

In recent works [11, 12] (hereinafter referred to as Papers I and II, respectively) we performed a high-accuracy comparison between the PN and SF analyses in their common domain of validity, that of the slow motion weak-field regime of an extreme mass ratio binary (see illustration of various methods in Fig. 1). The problem was tackled previously by Detweiler [28], who computed numerically within the SF approach a certain gauge invariant quantity (called the redshift observable), and compared it with the 2PN prediction extracted from existing PN results [16]. We then extended this comparison in Papers I and II up to higher PN orders. This required an improvement in the numerical resolution of the SF calculation in order to distinguish more accurately the various contributions of very high PN order terms. However, our primary difficulty has been that the relevant PN results for the

¹ As usual the n PN order refers to terms equivalent to $(v/c)^{2n}$ beyond Newtonian theory, where v is a typical internal velocity of the material system and c is the speed of light.

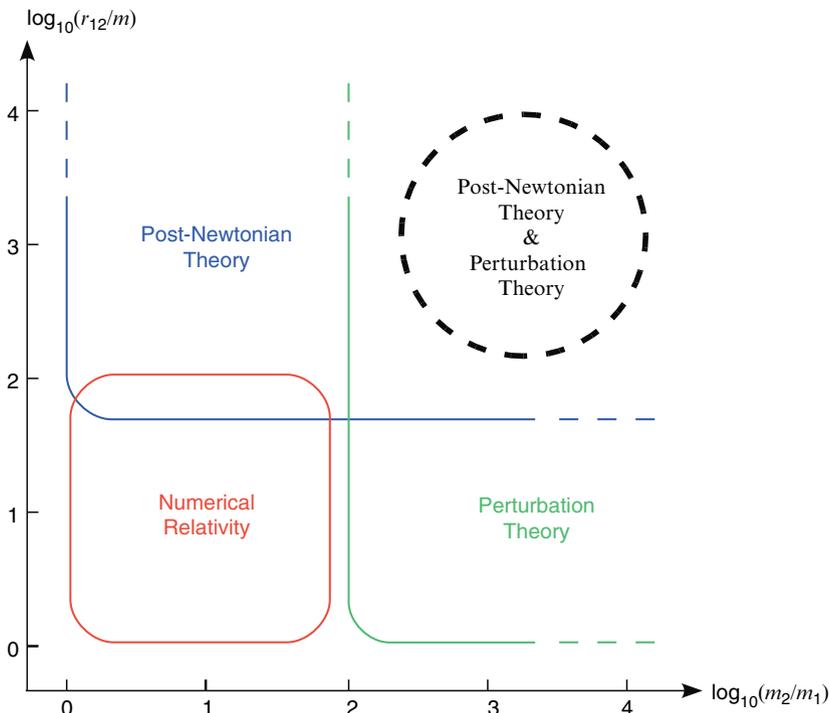


Fig. 1 Different analytical approximation schemes and numerical techniques are used to study black-hole binaries, depending on the mass ratio m_1/m_2 and the orbital velocity $v^2 \sim Gm/r_{12}$, where $m = m_1 + m_2$. The post-Newtonian theory and black hole perturbation theory can be compared in the slow motion regime ($v \ll c$ equivalent to $r_{12} \gg Gm/c^2$ for circular orbits) of an extreme mass ratio ($m_1 \ll m_2$) binary

metric were previously not available beyond the 2PN order, and had to be carefully derived. We have finally demonstrated an excellent agreement between the SF contribution to the analytical PN result (derived through 3PN order, with inclusion of specific logarithmic terms at 4PN and 5PN orders) and the exact numerical SF result.

In this article we present a summary of the Papers I and II. The plan is as follows: After having introduced in Section 2 the coordinate-invariant relation used to perform the comparison, we describe in Section 3 how the divergent self-field of point particles is regularized in both SF and PN formalisms. Section 4 provides a brief overview of how the SF computation proceeds. Sections 5 and 6 present the PN computations of the 3PN regularized metric, as well as of the 4PN and 5PN logarithmic contributions. The PN results for the gauge-invariant relation are discussed in Section 7. Finally, Section 8 is devoted to the comparison of these PN results with the SF numerical data, and the measurement of unknown high-order PN coefficients.

2 The Gauge-Invariant Redshift Observable

We consider a system of two (non-spinning) compact objects with masses m_1 and m_2 , and moving on slowly inspiralling quasi-circular orbits. In the PN analysis, let m_1 and m_2 be arbitrary; in the SF analysis, further assume that $m_1 \ll m_2$. We can then call m_1 the “particle,” and m_2 the “black hole.”

SF analysis shows that the dissipative parts of the SF for a circular orbit are the t and φ components. These result in a loss of energy and angular momentum from the small mass at the same precise rate as energy and angular momentum are radiated away [28]. In addition, earlier perturbative calculations of energy and angular momentum fluxes [25, 37–39, 41, 45–47] for this situation show them to be equivalent to the results of the PN analysis in their common domain of validity. Hence, by invoking an argument of energy and angular momentum balance, we know that the PN results also agree with the dissipative parts of the SF, and further comparison can reveal nothing new.

For our PN–SF comparison, we shall thus neglect the dissipative, radiation-reaction force responsible for the inspiral, and restrict ourselves to the conservative part of the dynamics. In PN theory this means neglecting the dissipative radiation-reaction force at 2.5PN and 3.5PN orders, and considering only the conservative dynamics at the even-parity 1PN, 2PN, and 3PN orders. This clean separation between conservative even-parity and dissipative odd-parity PN terms is correct up to 3.5PN order. However, such split breaks at 4PN order, since at that approximation arises a contribution of the radiation-reaction force, which originates from gravitational wave tails propagating to infinity [7] (this will be further discussed in Section 6). In SF theory, there is also a clean split between the dissipative and conservative parts of the SF (see e.g. [4]). This split is particularly transparent for a quasi-circular orbit, where the r component is the only nonvanishing component of the conservative SF.

Henceforth, the orbits of both masses are assumed to be and to remain circular, because we are ignoring the dissipative radiation-reaction effects. For our comparison we require two physical quantities which are precisely defined in the context of each of our approximation schemes. The orbital frequency Ω of the circular orbit as measured by a distant observer is one such quantity. The second quantity is defined as follows.

With circular orbits and no dissipation, the geometry has a helical Killing vector field k^α . A Killing vector is only defined up to an overall constant factor. In our case k^α extends out to a large distance where the geometry is essentially flat. There,

$$k^\alpha \partial_\alpha = \partial_t + \Omega \partial_\varphi, \quad (1)$$

in any natural coordinate system which respects the helical symmetry [44]. We let this equality define the overall constant factor, thereby specifying the Killing vector field uniquely.

An observer moving with the particle m_1 , while orbiting the black hole m_2 , would detect no change in the local geometry. Thus, the four-velocity u_1^α of the

particle is tangent to the Killing vector k^α evaluated at the location of the particle, which we denote by k_1^α . A second physical quantity is then defined as the constant of proportionality, call it u_1^T , between these two vectors, namely

$$u_1^\alpha = u_1^T k_1^\alpha. \tag{2}$$

The four-velocity of the particle is normalized so that $(g_{\alpha\beta})_1 u_1^\alpha u_1^\beta = -1$; $(g_{\alpha\beta})_1$ is the *regularized* metric at the particle’s location, whereas the metric itself is formally singular at the particle m_1 in both PN and SF approaches.

If we happen to choose a coordinate system such that (1) is satisfied everywhere, then in particular $k_1^t = 1$, and thus $u_1^T \equiv u_1^t$, the t component of the four-velocity of m_1 . The Killing vector on the particle is then $k_1^\alpha = u_1^\alpha / u_1^t$, and simply reduces to the particle’s ordinary PN coordinate velocity v_1^α / c . In such a coordinate system, the description of the invariant quantity we are thus considering is

$$u_1^T \equiv u_1^t = \left(-(g_{\alpha\beta})_1 \frac{v_1^\alpha v_1^\beta}{c^2} \right)^{-1/2}. \tag{3}$$

It is important to note that this quantity is precisely defined in both PN and SF frameworks, and it does not depend upon the choice of coordinates or upon the choice of perturbative gauge. The quantity u_1^T represents the redshift of light rays emitted from the particle and received on the helical symmetry axis perpendicular to the orbital plane [28]; we shall refer to it as the *redshift observable*.

3 Regularization Issues in the SF and PN Formalisms

The redshift observable (3) depends upon using a valid method of regularization. The regularized metric $(g_{\alpha\beta})_1$ is defined with very different prescriptions in the SF and PN approaches. Both analyses require subtle treatment of singular fields at the location of the masses. Subtracting away the infinite part of a field while carefully preserving the part which is desired is always a delicate task.

In the SF prescription, the regularized metric reads

$$g_{\alpha\beta}^{\text{SF}}(x) = \bar{g}_{\alpha\beta}(x) + h_{\alpha\beta}^{\text{R}}(x), \tag{4}$$

where $\bar{g}_{\alpha\beta}$ denotes the background Schwarzschild metric of the black hole, and where the “Regular” metric perturbation $h_{\alpha\beta}^{\text{R}}$ is smooth in a neighborhood of the particle, and given by the difference

$$h_{\alpha\beta}^{\text{R}} = h_{\alpha\beta}^{\text{ret}} - h_{\alpha\beta}^{\text{S}} \tag{5}$$

between the retarded metric perturbation $h_{\alpha\beta}^{\text{ret}}$ and the purely locally determined “Singular” field $h_{\alpha\beta}^{\text{S}}$. Following the Detweiler–Whiting prescription [30],

a Hadamard expansion of Green's functions in curved spacetime provides an expansion for $h_{\alpha\beta}^S$. In a neighborhood of the particle with a special, locally inertial coordinate system, $h_{\alpha\beta}^S$ appears as the m_1/r part of the particle's Schwarzschild metric along with its tidal distortion caused by the background geometry of the large black hole. Details of the expansion are given in Section 6.1 of [27]. The special locally inertial coordinates for a circular geodesic in the Schwarzschild metric are given as functions of the Schwarzschild coordinates in Appendix B of [29]. (See also Detweiler's and Barack's contributions in this volume.) Since the metric (4) is regular at the particle's position y_1^α , we simply have

$$\left(g_{\alpha\beta}^{\text{SF}}\right)_1 = g_{\alpha\beta}^{\text{SF}}(y_1). \quad (6)$$

In the perturbative SF analysis we are only working through first order in the mass ratio $q \equiv m_1/m_2$, and at that level of approximation $h_{\alpha\beta}^R = \mathcal{O}(q)$. Then u_1^T can be computed accurately to the same perturbative order and compares well with the PN result to 2PN order [28]. The regularized 2PN metric is known [16], and therefore the comparison is straightforward.

In the PN prescription on the other hand, one first computes the metric $g_{\alpha\beta}^{\text{PN}}(\mathbf{x}, t)$ using an iterative PN procedure at any field point outside the particle, in a coordinate system $x^\alpha = \{ct, x^i\}$. That metric is generated by the two particles, and includes both regular and singular contributions around each particle. Such iterative PN calculation is a very long and intricate procedure up to say 3PN, at which order it will be partly based on existing computations of the 3PN equations of motion using Hadamard [13] and dimensional [9] regularizations. Then we compute the regularized metric at the location of the particle by taking the limit when $\mathbf{x} \rightarrow \mathbf{y}_1(t)$, where $\mathbf{y}_1(t)$ is the particle's trajectory. In 3 spatial dimensions, that limit is singular. In order to treat the infinite part of the field, we extend the computation in d spatial dimensions, following the prescription of *dimensional regularization* [19, 48], which is based on an analytic continuation (AC) in the dimension d viewed as a complex number. We do not use Hadamard's regularization that found its limit at 3PN order. Considering the analytic continuation in a neighborhood of $\varepsilon \equiv d - 3 \rightarrow 0$, we define

$$\left(g_{\alpha\beta}^{\text{PN}}\right)_1 = \text{AC} \left[\lim_{\varepsilon \rightarrow 0} \lim_{\mathbf{x} \rightarrow \mathbf{y}_1} g_{\alpha\beta}^{\text{PN}}(\mathbf{x}, t) \right]. \quad (7)$$

The limit $\varepsilon \rightarrow 0$ does not exist in general due to the presence of poles $\propto \varepsilon^{-1}$ occurring at the 3PN order; we thus do not take the strict limit $\varepsilon \rightarrow 0$ but compute the singular Laurent expansion when $\varepsilon \rightarrow 0$. At 3PN order the result takes the schematic form

$$\left(g_{\alpha\beta}^{\text{PN}}\right)_1 = \frac{1}{\varepsilon} g_{\alpha\beta}^{(-1)}(\mathbf{y}_1, t) + g_{\alpha\beta}^{(0)}(\mathbf{y}_1, t) + \mathcal{O}(\varepsilon), \quad (8)$$

where $g_{\alpha\beta}^{(-1)}(\mathbf{y}_1, t)$ denotes the pole part, which is purely of 3PN order, and $g_{\alpha\beta}^{(0)}(\mathbf{y}_1, t)$ is the finite part. At higher PN orders we expect the presence of multipole

poles $\propto \varepsilon^{-n}$. As we shall see, the (simple) poles at 3PN order will disappear from the final gauge-invariant relationship $u_1^T(\Omega)$. In fact, the occurrence of poles at the 3PN order is specific to the use of the harmonic gauge condition. Previous work on the 3PN equations of motion of point particle binaries has shown that the poles can be absorbed into a renormalization of the world lines of the particles, so they should not appear in any physical coordinate invariant quantity. Thus, dimensional regularization is a powerful regularization method in the PN context. In particular this regularization is free of the ambiguities plaguing the Hadamard regularization at the 3PN order [9, 10, 26] (see the contributions by Schäfer and Blanchet in this volume).

Although the two regularizations in SF and PN analyses have been carefully designed, it appears to be nontrivial that they will yield results consistent up to a high level of approximation. Our cross-cultural comparison of the redshift observable u_1^T is a test of the equivalence of the SF and PN metrics (6) and (7) and is, thus, a test of the two independent (and very different) regularization procedures in use.

4 Circular Orbits in the Perturbed Schwarzschild Geometry

Previously, we described the truly coordinate and perturbative-gauge independent properties of Ω and the redshift observable u_1^T . In this section we use Schwarzschild coordinates, and we refer to “gauge invariance” as a property that holds within the restricted class of gauges for which (1) is a helical Killing vector. In all other respects, the gauge choice is arbitrary. With this assumption, no generality is lost, and a great deal of simplicity is gained.

The effect of the gravitational SF is most easily described as having m_1 move along a geodesic of the regularized metric (4). We are interested in circular orbits and let u^α be the four-velocity of m_1 .² This differs from the four-velocity \bar{u}^α of a geodesic of the straight Schwarzschild geometry at the same radial coordinate r by an amount of $\mathcal{O}(q)$. Recall that we are describing perturbation analysis with $q \ll 1$, therefore $h_{\alpha\beta}^R = \mathcal{O}(q)$, and all equations in this section necessarily hold only through first order in q .

It is straightforward to determine the components of the geodesic equation for the metric (4) [28], and then to find the components of the four-velocity u^α of m_1 when it is in a circular orbit at Schwarzschild radius r . We reiterate that the four-velocity is to be normalized with respect to $\bar{g}_{\alpha\beta} + h_{\alpha\beta}^R$ rather than $\bar{g}_{\alpha\beta}$, and that $h_{\alpha\beta}^R$ is assumed to respect the symmetry of the helical Killing vector. In this case we have³

$$(u^t)^2 = \frac{r}{r - 3m_2} \left[1 + \bar{u}^\alpha \bar{u}^\beta h_{\alpha\beta}^R - \frac{r}{2} \bar{u}^\alpha \bar{u}^\beta \partial_r h_{\alpha\beta}^R \right], \tag{9a}$$

² Since we are interested in the motion of the small particle m_1 , we remove the index 1 from u_1^α .

³ In all of this section we shall set $G = c = 1$.

$$(u^\varphi)^2 = \frac{r - 2m_2}{r(r - 3m_2)} \left[\frac{m_2(1 + \bar{u}^\alpha \bar{u}^\beta h_{\alpha\beta}^R)}{r(r - 2m_2)} - \frac{1}{2} \bar{u}^\alpha \bar{u}^\beta \partial_r h_{\alpha\beta}^R \right]. \quad (9b)$$

A consequence of these relations is that the orbital frequency of m_1 in a circular orbit about a perturbed Schwarzschild black hole of mass m_2 is, through first order in the perturbation, given by

$$\Omega^2 = \left(\frac{u^\varphi}{u^t} \right)^2 = \frac{m_2}{r^3} - \frac{r - 3m_2}{2r^2} \bar{u}^\alpha \bar{u}^\beta \partial_r h_{\alpha\beta}^R. \quad (10)$$

The angular frequency Ω is a physical observable, and is independent of the gauge choice. However, the perturbed Schwarzschild metric does not have spherical symmetry, and the radius of the orbit r is not an observable and does depend upon the gauge choice. That is to say, an infinitesimal coordinate transformation of $\mathcal{O}(q)$ might change $\bar{u}^\alpha \bar{u}^\beta \partial_r h_{\alpha\beta}^R$. But if it does, then it will also change the radius r of the orbit in just such a way that Ω^2 as determined from (10) remains unchanged. Both $u^t \equiv u^T$ and $u^\varphi \equiv \Omega u^T$ are gauge invariant as well.

Our principle interest is in the relationship between Ω and u^T , which we now establish directly using (9a) and (10). To this end, following [28], we introduce the gauge-invariant measure of the orbital radius

$$R_\Omega \equiv \left(\frac{m_2}{\Omega^2} \right)^{1/3}, \quad (11)$$

and readily establish a first-order, gauge-invariant, algebraic relationship between u^T (to which u^t evaluates in our gauge) and R_Ω (or equivalently Ω), namely:

$$(u^T)^2 = \left(1 - \frac{3m_2}{R_\Omega} \right)^{-1} \left(1 + \bar{u}^\alpha \bar{u}^\beta h_{\alpha\beta}^R \right). \quad (12)$$

See Paper I for a detailed derivation of this result. The lowest order term in q on the right-hand side is identical to what is obtained for a circular geodesic of the unperturbed Schwarzschild metric. Indeed, recall that the Schwarzschild part of u^T is known exactly as $u_{\text{Schw}}^T = (1 - 3m_2/R_\Omega)^{-1/2}$. Thus, if we write

$$u^T \equiv u_{\text{Schw}}^T + q u_{\text{SF}}^T + \mathcal{O}(q^2), \quad (13)$$

the first order term in (12) gives:

$$q u_{\text{SF}}^T = \frac{1}{2} \left(1 - \frac{3m_2}{R_\Omega} \right)^{-1/2} \bar{u}^\alpha \bar{u}^\beta h_{\alpha\beta}^R, \quad (14)$$

which is $\mathcal{O}(q)$, and contains the effect of the ‘‘gravitational self-force’’ on the relationship between u^T and Ω , even though it bears little resemblance to a force.

The numerical SF approach henceforth focuses attention on the calculation of the combination $\bar{u}^\alpha \bar{u}^\beta h_{\alpha\beta}^R$. See [28] and Paper I for more details on the implementation of the regularization of the perturbation, and on the numerical computation of the quantity $\bar{u}^\alpha \bar{u}^\beta h_{\alpha\beta}^R$.

5 Overview of the 3PN Calculation

Our aim is to compute the 3PN regularized metric (7) by direct PN iteration of the Einstein field equations in the case of singular point mass sources. In the dimensional regularization scheme, we look for the solution of the Einstein field equations in $d + 1$ space-time dimensions. We treat the space dimension as an arbitrary complex number, $d \in \mathbb{C}$, and interpret any intermediate formula in the PN iteration of those equations by analytic continuation in d . Then we analytically continue d down to the value of interest (namely 3), posing $d \equiv 3 + \varepsilon$. In most of the calculations we neglect terms of order ε or higher, that is, we retain the finite part and the eventual poles.

5.1 Iterative PN Computation of the Metric

Defining the gravitational field variable $h^{\alpha\beta} \equiv \sqrt{-g} g^{\alpha\beta} - \eta^{\alpha\beta}$,⁴ and adopting the harmonic coordinate condition $\partial_\beta h^{\alpha\beta} = 0$, we can write the “relaxed” Einstein field equations in the form of ordinary d’Alembert equations, namely

$$\square h^{\alpha\beta} = \frac{16\pi G^{(d)}}{c^4} |g| T^{\alpha\beta} + \Lambda^{\alpha\beta} [h, \partial h, \partial^2 h], \tag{15}$$

where $\square \equiv \eta^{\mu\nu} \partial_\mu \partial_\nu$ is the flat-space-time d’Alembertian operator in $d + 1$ dimensions. The gravitational source term $\Lambda^{\alpha\beta}$ in (15) is a functional of $h^{\mu\nu}$ and its first and second space-time derivatives; it depends explicitly on the dimension d . The matter stress-energy tensor $T^{\alpha\beta}$ is composed of Dirac delta-functions in d dimensions. Finally, the d -dimensional gravitational constant $G^{(d)}$ is related to the usual Newton constant G by

$$G^{(d)} = G \ell_0^\varepsilon, \tag{16}$$

where ℓ_0 denotes the characteristic length associated with dimensional regularization. We shall check that this length scale disappears from the final gauge-invariant three-dimensional result.

⁴ Here $g^{\alpha\beta}$ is the contravariant metric, inverse of the covariant metric $g_{\alpha\beta}$ of determinant $g = \det(g_{\alpha\beta})$, and $\eta^{\alpha\beta} = \text{diag}(-1, 1, 1, 1)$ represents an auxiliary Minkowski metric in Cartesian coordinates.

The 3PN metric is given in expanded form for general matter sources in terms of some “elementary” retarded potentials (sometimes called near-zone potentials) V , V_i , K , \hat{W}_{ij} , \hat{R}_i , \hat{X} , \hat{Z}_{ij} , \hat{Y}_i , and \hat{T} , which were introduced in Ref. [13] for three dimensions and generalized to d dimensions in Ref. [9]. All these potentials have a finite nonzero PN limit when $c \rightarrow +\infty$ and parameterize the successive PN approximations. Although this decomposition in terms of near-zone potentials is convenient, such potentials have no physical meaning by themselves.

Let us first define the combination

$$\mathcal{V} \equiv V - \frac{2}{c^2} \left(\frac{d-3}{d-2} \right) K + \frac{4\hat{X}}{c^4} + \frac{16\hat{T}}{c^6}. \quad (17)$$

Then the 3PN metric components can be written in the rather compact form [9]

$$g_{00}^{\text{PN}} = -e^{-2\mathcal{V}/c^2} \left(1 - \frac{8V_i V_i}{c^6} - \frac{32\hat{R}_i V_i}{c^8} \right) + \mathcal{O}(c^{-10}), \quad (18a)$$

$$g_{0i}^{\text{PN}} = -e^{-\frac{(d-3)\mathcal{V}}{(d-2)c^2}} \left(\frac{4V_i}{c^3} \left[1 + \frac{1}{2} \left(\frac{d-1}{d-2} \frac{V}{c^2} \right)^2 \right] + \frac{8\hat{R}_i}{c^5} + \frac{16}{c^7} \left[\hat{Y}_i + \frac{1}{2} \hat{W}_{ij} V_j \right] \right) + \mathcal{O}(c^{-9}), \quad (18b)$$

$$g_{ij}^{\text{PN}} = e^{\frac{2\mathcal{V}}{(d-2)c^2}} \left(\delta_{ij} + \frac{4}{c^4} \hat{W}_{ij} + \frac{16}{c^6} \left[\hat{Z}_{ij} - V_i V_j + \frac{1}{2(d-2)} \delta_{ij} V_k V_k \right] \right) + \mathcal{O}(c^{-8}), \quad (18c)$$

where the exponentials are to be expanded to the order required for practical calculations. The successive PN truncations of the field equations (15) give us the equations satisfied by all the above potentials up to 3PN order. We conveniently define from the components of the matter stress-energy tensor $T^{\alpha\beta}$ the following density, current density, and stress density:

$$\sigma \equiv \frac{2}{d-1} \frac{(d-2)T^{00} + T^{ii}}{c^2}, \quad (19a)$$

$$\sigma_i \equiv \frac{T^{0i}}{c}, \quad (19b)$$

$$\sigma_{ij} \equiv T^{ij}, \quad (19c)$$

where $T^{ii} \equiv \delta_{ij} T^{ij}$. As examples, the leading-order potentials in the metric obey

$$\square V = -4\pi G^{(d)} \sigma, \quad (20a)$$

$$\square V_i = -4\pi G^{(d)} \sigma_i, \quad (20b)$$

$$\square \hat{W}_{ij} = -4\pi G^{(d)} \left(\sigma_{ij} - \delta_{ij} \frac{\sigma_{kk}}{d-2} \right) - \frac{1}{2} \left(\frac{d-1}{d-2} \right) \partial_i V \partial_j V. \quad (20c)$$

All the potentials evidently include many PN corrections. The potentials V and V_i have a compact support (i.e., their source is localized on the isolated matter system) and will admit a finite limit when $\varepsilon \rightarrow 0$ without any pole. Most of the other potentials have, in addition to a compact-support part, a non-compact support contribution, such as that generated by the term $\propto \partial_i V \partial_j V$ in the source of \hat{W}_{ij} . These non-compact support pieces are the most delicate to compute, because they typically generate some poles $\propto 1/\varepsilon$ at the 3PN order. The d’Alembert equations satisfied by all higher-order PN potentials, whose sources are made of nonlinear combinations of lower-order potentials, can be found in Paper I. Clearly, the higher the PN order of a potential, the more complicated is its source, but it requires computations at a lower relative order.

Many of the latter potentials have already been computed for compact binary systems, and we have extensively used these results from [9, 13]. Notably, all the compact-support potentials such as V and V_i , and all the compact-support parts of other potentials, have been computed for any field point \mathbf{x} , and then at the source point \mathbf{y}_1 following the regularization. However, the most difficult non-compact support potentials such as \hat{X} and \hat{T} could not be computed at any field point \mathbf{x} , and were regularized directly on the particle’s world line. Since for the equations of motion we needed only the *gradients* of these potentials, only the gradients were regularized on the particle, yielding the results for $(\partial_i \hat{X})_1$ and $(\partial_i \hat{T})_1$ needed in the equations of motion. However, the 3PN metric requires the values of the potentials themselves regularized on the particles, that is, $(\hat{X})_1$ and $(\hat{T})_1$. For the present work we therefore computed, using the tools developed in [9, 13], the difficult nonlinear potentials $(\hat{X})_1$ and $(\hat{T})_1$, and especially the non-compact support parts therein. Unfortunately, the potential \hat{X} is always the most tricky to compute, because its source involves the cubically nonlinear and non-compact-support term $\hat{W}_{ij} \partial_{ij} V$, and it has to be evaluated at relative 1PN order.

In this calculation we also met a new difficulty with respect to the computation of the 3PN equations of motion. Indeed, we found that the potential \hat{X} is divergent because of the bound of the Poisson-like integral at *infinity*. Thus, the potential \hat{X} develops an IR divergence, in addition to the UV divergence due to the singular nature of the source and which is cured by dimensional regularization. The IR divergence is a particular case of the well-known divergence of Poisson integrals in the PN expansion for general (regular) sources, linked to the fact that the PN expansion is a singular perturbation expansion, with coefficients typically blowing up at spatial infinity. The IR divergence is discussed in Paper I, where we show how to resolve it by means of a finite part prescription.

The 3PN metric (18) is valid for a general isolated matter system, and we apply it to the case of a system of N point particles with “Schwarzschild” masses m_a and without spins (here $a = 1, \dots, N$). In this case we have

$$\sigma(\mathbf{x}, t) = \sum_a \tilde{\mu}_a \delta^{(d)}[\mathbf{x} - \mathbf{y}_a(t)], \tag{21a}$$

$$\sigma_i(\mathbf{x}, t) = \sum_a \mu_a v_a^i \delta^{(d)}[\mathbf{x} - \mathbf{y}_a(t)], \tag{21b}$$

$$\sigma_{ij}(\mathbf{x}, t) = \sum_a \mu_a v_a^i v_a^j \delta^{(d)}[\mathbf{x} - \mathbf{y}_a(t)], \tag{21c}$$

where $\delta^{(d)}$ denotes the Dirac density distribution in d spatial dimensions, such that $\int d^d \mathbf{x} \delta^{(d)}(\mathbf{x}) = 1$. We defined the effective masses of the particles by

$$\mu_a(t) = \frac{m_a}{\sqrt{(g g_{\alpha\beta})(\mathbf{y}_a, t) v_a^\alpha v_a^\beta / c^2}}, \tag{22a}$$

$$\tilde{\mu}_a(t) = \frac{2}{d-1} \left[d - 2 + \frac{\mathbf{v}_a^2}{c^2} \right] \mu_a(t). \tag{22b}$$

5.2 The Example of the Zeroth-Order Iteration

For illustration purposes, let us consider the simple case of the dimensional regularization of the Newtonian potential generated by two point particles of masses m_1 and m_2 . The PN metric (18) reduces to $g_{00}^{\text{PN}} = -1 + 2V/c^2 + \mathcal{O}(c^{-4})$, where V is solution of the Poisson equation $\Delta V = -4\pi G^{(d)}\sigma + \mathcal{O}(c^{-2})$, with source $\sigma(\mathbf{x}) = \frac{2(d-2)}{d-1} [m_1 \delta^{(d)}(\mathbf{x} - \mathbf{y}_1) + m_2 \delta^{(d)}(\mathbf{x} - \mathbf{y}_2)] + \mathcal{O}(c^{-2})$. Solving this Poisson equation in d space dimensions yields

$$V(\mathbf{x}) = \frac{2(d-2)}{d-1} k \left(\frac{G^{(d)}m_1}{|\mathbf{x} - \mathbf{y}_1|^{d-2}} + \frac{G^{(d)}m_2}{|\mathbf{x} - \mathbf{y}_2|^{d-2}} \right) + \mathcal{O}(c^{-2}), \tag{23}$$

where $k \equiv \Gamma(\frac{d-2}{2})/\pi^{\frac{d-2}{2}}$ tends to 1 when $\varepsilon \rightarrow 0$ (Γ is the usual Eulerian function). When $d = 3$, the Newtonian potential (23) is not defined in the limit $\mathbf{x} \rightarrow \mathbf{y}_1$ because of the divergent self-field of particle 1. By contrast, thanks to the analytic continuation in the space dimension, it is always possible to choose $\Re(d) < 2$ such that the d -dimensional potential (23) has a well-defined limit when $\mathbf{x} \rightarrow \mathbf{y}_1$, namely $(V)_1 \equiv V(\mathbf{y}_1)$ given by

$$(V)_1 = \frac{2(d-2)}{d-1} k \frac{G^{(d)}m_2}{r_{12}^{d-2}} + \mathcal{O}(c^{-2}), \tag{24}$$

where we have posed $r_{12} \equiv |\mathbf{y}_1 - \mathbf{y}_2|$. Relying on the unicity of the analytic continuation, we obtain the unique three-dimensional result

$$(V)_1 = \frac{Gm_2}{r_{12}} + \mathcal{O}(\varepsilon) + \mathcal{O}(c^{-2}). \tag{25}$$

This procedure is clear up to a high order, but let us mention a subtle point in the calculation of Paper I, namely that we had to systematically reintroduce the correction terms $\mathcal{O}(\varepsilon)$ in the *Newtonian* part of the metric and other quantities. Indeed, in

various operations such as replacing r_{12} in Eq. 25 by its 3PN expression in terms of the orbital frequency Ω , the corrections $\mathcal{O}(\varepsilon)$ will be multiplied by some poles at 3PN order, and they will therefore contribute *in fine* to the finite part at 3PN order. Such corrections are necessary only in the Newtonian results (since the poles arise only at 3PN order). For instance, the three-dimensional Newtonian potential (25) is to be replaced by its d -dimensional version valid up to terms $\mathcal{O}(\varepsilon^2)$, namely

$$(V)_1 = \frac{Gm_2}{r_{12}} \left\{ 1 + \varepsilon \left[\frac{1}{2} - \ln \left(\frac{r_{12} p}{\ell_0} \right) \right] + \mathcal{O}(\varepsilon^2) \right\} + \mathcal{O}(c^{-2}), \quad (26)$$

in which $p \equiv \sqrt{4\pi} e^{C/2}$ with $C = 0.5772 \dots$ being the Euler–Mascheroni constant.

This Newtonian calculation is generalized at higher orders in the iterative PN process described in the previous section. The result is the 3PN regularized metric (8) which is given in explicit form by Eqs. 4.2 of Paper I.

6 Logarithmic Terms at 4PN and 5PN Orders

We now discuss the logarithmic contributions in the near-zone metric of a generic isolated PN source, and then of a compact binary system. Our motivation is that knowing analytically determined *logarithmic* terms in the PN expansion is crucial for efficiently extracting from the SF data the numerical values of higher order PN coefficients. This will be further discussed in Section 8.

The occurrence of logarithmic terms in the PN expansion has been investigated in many previous works [2, 3, 5–7, 31, 32, 34, 35]. Notably Anderson et al. [3] found that the dominant logarithm arises at the 4PN order, and Blanchet and Damour [5, 7] showed that this logarithm is associated with gravitational wave tails modifying the usual 2.5PN radiation-reaction damping at the 4PN order. Furthermore, the general structure of the PN expansion is known [6]: it is of the type $\sum (v/c)^k [\ln(v/c)]^q$, where k and q are positive integers, involving only powers of logarithms; more exotic terms such as $[\ln(\ln(v/c))]^q$ cannot arise.

Following Paper II, we shall determine the leading 4PN logarithm and the next-to-leading 5PN logarithm in the conservative part of the dynamics of a compact binary system. The computation of such logarithmic contributions relies only very weakly on a regularization scheme. We shall thus work in three space dimensions; from now on we set $\varepsilon = 0$.

6.1 Physical Origin of Logarithmic Terms

Because of the nonlinearity of the field equations, the gravitational field at coordinate time t is in general not a function of the state of motion of the source at retarded time $t - r/c$, where $r = |\mathbf{x}|$ is the distance to the center of the source, but depends on the entire past “history” of the source. This means that the near-zone metric

depends on the source at all times before the current time t , say $t' \leq t$ (indeed in the near-zone one expands all retardations, i.e., $r/c \rightarrow 0$). This “hereditary” effect starts at 4PN order in the near-zone, and originates from gravitational wave tails, namely the scattering of gravitational radiation by the background curvature of the space-time generated by the mass M of the source. In a certain gauge where the hereditary terms are collected into the 00 component of the metric, we have

$$\delta g_{00}^{\text{tail}}(\mathbf{x}, t) = -\frac{8G^2 M}{5c^{10}} x^a x^b \int_{-\infty}^t dt' M_{ab}^{(7)}(t') \ln\left(\frac{c(t-t')}{2r}\right), \quad (27)$$

where M is the ADM mass of the source, and $M_{ab}^{(n)}$ is the n th time derivative of its quadrupole moment. This tail-induced contribution is of 4PN order.

The occurrence of the tail effect in the near-zone metric implies that the usual 2.5PN radiation-reaction force density in the matter source is corrected at the relative 1.5PN order as

$$F_{\text{rad}}^i(\mathbf{x}, t) = -\frac{2G}{5c^5} \rho x^a \left[M_{ia}^{(5)}(t) + \frac{4GM}{c^3} \int_{-\infty}^t dt' M_{ia}^{(7)}(t') \ln\left(\frac{c(t-t')}{2\lambda}\right) \right], \quad (28)$$

where ρ is the Newtonian mass density in the source, and λ is the typical wavelength of the radiation. The leading term in (28) is the standard Burke–Thorne [22, 23] radiation-reaction force density at the 2.5PN order, responsible for the leading radiation effect. The hereditary correction was obtained in [5, 7], and shown to be consistent with wave tails propagating at large distances from the source [8].

The radiation-reaction force (28) deserves its name because it is not invariant under a time reversal, and therefore gives rise to dissipative effects. A good way to see this is to change the condition of retarded potentials to advanced potentials, that is, to formally change c into $-c$. The first term in (28) is clearly non-invariant because it comes with an odd number of powers of $1/c$. The second term is also non-invariant, despite the fact that it comes with an even power of $1/c$ in front (i.e., $1/c^8$ in (28), corresponding to 4PN), because it is composed of an integral extending over the past rather than a time-symmetric integral.

However, thanks to the even power of $1/c$ carried by the hereditary integral (27), this means that there exists a *conservative* piece associated with it. Recall that in our calculation we neglect the dissipative radiation-reaction effects and are interested only in the conservative part of the dynamics; we have implemented this restriction by assuming the existence of the helical Killing vector (1), depending on the orbital frequency Ω of the circular motion. The presence of this scale Ω , imposed by the helical Killing symmetry, permits immediately to identify the conservative piece associated with the tail term in (27), through the decomposition

$$\ln\left(\frac{c(t-t')}{2r}\right) = \ln\left(\frac{c(t-t')}{2\lambda}\right) - \ln\left(\frac{r}{\lambda}\right), \quad (29)$$

where now the characteristic length scale λ is defined by $\lambda \equiv 2\pi c/\Omega$. The second term in (29), when inserted into the tail integral in (27), can be integrated out and gives rise to the conservative 4PN piece

$$\delta g_{00}^{\text{cons}}(\mathbf{x}, t) = \frac{8G^2 M}{5c^{10}} x^a x^b M_{ab}^{(6)}(t) \ln\left(\frac{r}{\lambda}\right). \quad (30)$$

This contribution to the metric has to be included in our study of the conservative part of the dynamics, while the purely dissipative piece in (27) enters the radiation reaction (28), and is excluded by our assumption of existence of the helical Killing vector. The term (30) is precisely the conservative 4PN logarithmic contribution in the near-zone metric that we want to compute, as well as its 5PN correction.

Notice that it is possible to find a gauge where all the logarithms are gathered in the 00 component of the metric, as we did in Eq. 30, only at 4PN order. When looking to subdominant logarithmic terms at 5PN order, we are obliged to include some vectorial $0i$ and tensorial ij components of the metric. But still it will be possible, and extremely convenient, to define a gauge in which the logarithms are “maximally” transferred to the 00 and $0i$ components (while the remaining contribution in the ij components is “minimal”). Using such a gauge saves a lot of calculations when performing the PN iteration; this was the strategy adopted in Paper II to compute the higher order 5PN logarithms beyond Eq. 30.

6.2 Expression of the Near-Zone Metric

To compute these 4PN and 5PN conservative logarithmic contributions, we make use of the multipolar post-Minkowskian wave-generation formalism [5–8]. We first identify these logarithmic contributions in the exterior of a generic isolated PN source. We then deduce the metric inside the matter source by a matching performed in the exterior part of the near-zone. We finally get the 4PN and 5PN logarithmic contributions in the near-zone metric, valid in a specific gauge defined in Paper II, as⁵

$$\begin{aligned} \delta g_{00} = & \frac{G^2 M}{c^{10}} \left[\frac{8}{5} \left(1 - \frac{2U}{c^2}\right) x^{ab} M_{ab}^{(6)} + \frac{4}{35c^2} r^2 x^{ab} M_{ab}^{(8)} - \frac{8}{189c^2} x^{abc} M_{abc}^{(8)} \right] \ln\left(\frac{r}{\lambda}\right) \\ & - \frac{8}{5} \frac{G^3 M}{c^{12}} x^a M_{ab}^{(6)} \int \frac{d^3 x'}{|\mathbf{x} - \mathbf{x}'|} \rho' x'^b \ln\left(\frac{r'}{\lambda}\right) + \mathcal{O}\left(\frac{1}{c^{14}}\right), \end{aligned} \quad (31a)$$

$$\delta g_{0i} = \frac{G^2 M}{c^{11}} \left[\frac{16}{21} \hat{x}^{iab} M_{ab}^{(7)} - \frac{64}{45} \varepsilon_{iab} x^{ac} S_{bc}^{(6)} \right] \ln\left(\frac{r}{\lambda}\right) + \mathcal{O}\left(\frac{1}{c^{13}}\right), \quad (31b)$$

⁵ We use shorthands such as $x^{ab} = x^a x^b$; $\hat{x}^{abc} = x^{abc} - \frac{1}{3}(\delta^{ab} x^c + \delta^{ac} x^b + \delta^{bc} x^a)r^2$ denotes the symmetric and trace-free part of x^{abc} ; ε_{abc} is the Levi-Civita antisymmetric symbol.

$$\delta g_{ij} = \frac{G^2 M}{c^{10}} \left[\frac{8}{5} x^{ab} M_{ab}^{(6)} \delta_{ij} \right] \ln \left(\frac{r}{\lambda} \right) + \mathcal{O} \left(\frac{1}{c^{12}} \right), \quad (31c)$$

where M_{ab} , M_{abc} and S_{ab} denote the mass quadrupole, mass octupole, and current quadrupole moments of the source, respectively, and where

$$U(\mathbf{x}, t) = G \int \frac{d^3 x'}{|\mathbf{x} - \mathbf{x}'|} \rho(\mathbf{x}', t) \quad (32)$$

is the Newtonian potential, sourced by the Newtonian mass density ρ of the source. We recall that $\lambda = 2\pi c/\Omega$, where Ω is the scale entering the Killing vector (1). Notice in particular the 5PN contribution in δg_{00} which involves the Poisson integral of a logarithmically modified source density, and which will contribute *in fine* to the 5PN logarithms in the case of binary systems.

We can then apply the result (31) to the case of a compact binary system moving on a circular orbit. In this case, Ω is the orbital frequency of the motion. We compute the metric at the location of one of the particles and obtain the result

$$\begin{aligned} (\delta g_{00})_1 = \frac{G^2 M}{c^{10}} \left[\frac{8}{5} \left(1 - \frac{2U_1}{c^2} \right) y_1^{ab} M_{ab}^{(6)} - \frac{8}{5c^2} U_1 y_1^a y_2^b M_{ab}^{(6)} \right. \\ \left. + \frac{4}{35c^2} y_1^2 y_1^{ab} M_{ab}^{(8)} - \frac{8}{189c^2} y_1^{abc} M_{abc}^{(8)} \right] \ln \left(\frac{r_{12}}{\lambda} \right) + \mathcal{O} \left(\frac{1}{c^{14}} \right), \end{aligned} \quad (33a)$$

$$(\delta g_{0i})_1 = \frac{G^2 M}{c^{11}} \left[\frac{16}{21} \hat{y}_1^{iab} M_{ab}^{(7)} - \frac{64}{45} \varepsilon_{iab} y_1^{ac} S_{bc}^{(6)} \right] \ln \left(\frac{r_{12}}{\lambda} \right) + \mathcal{O} \left(\frac{1}{c^{13}} \right), \quad (33b)$$

$$(\delta g_{ij})_1 = \frac{G^2 M}{c^{10}} \left[\frac{8}{5} y_1^{ab} M_{ab}^{(6)} \delta_{ij} \right] \ln \left(\frac{r_{12}}{\lambda} \right) + \mathcal{O} \left(\frac{1}{c^{12}} \right), \quad (33c)$$

where $U_1 = Gm_2/r_{12}$ is the Newtonian potential felt by particle 1. Finally, the last step is to replace the multipole moments M_{ab} , M_{abc} , and S_{ab} by the relevant PN expressions valid for circular-orbit compact binaries; we need also the ADM mass M , which reduces to $m = m_1 + m_2$ in first approximation. Note that the mass quadrupole moment M_{ab} (and also the ADM mass M) must crucially include a 1PN contribution. Again, we emphasize that for the 4PN and 5PN logarithms we do not need the full apparatus of dimensional regularization, in contrast to the fully fledged 3PN calculation sketched in Section 5.

7 Post-Newtonian Results for the Redshift Observable

To compute the gauge-invariant quantity u^T (associated with the particle 1 for helical symmetry, circular orbits), we adopt its coordinate form as given by (3), namely

$$u^t = \left(-(g_{\alpha\beta})_1 \frac{v_1^\alpha v_1^\beta}{c^2} \right)^{-1/2}, \tag{34}$$

and plug into it the 3PN regularized metric (8), explicitly computed from the 3PN near-zone expression (18) reduced to binary point masses, and including the 4PN and 5PN logarithmic corrections computed in Section 6. To begin with, this yields the expression of u^t for an arbitrary mass ratio $q = m_1/m_2$, and for a generic noncircular orbit in a general reference frame. We then choose the frame of the center of mass, which is consistently defined by the nullity of the center-of-mass integral of the motion, deduced from the equations of motion. Restricting ourselves to exactly circular orbits (consistently with the helical Killing symmetry we neglect radiation-reaction effects), the result is expressed by means of the convenient dimensionless gauge-invariant PN parameter

$$x \equiv \left(\frac{G m \Omega}{c^3} \right)^{2/3}, \tag{35}$$

which is directly related to the orbital frequency Ω of the circular orbit, and depends on the total mass $m = m_1 + m_2$ of the binary.

We discover most satisfactorily that all the poles $\propto 1/\varepsilon$ (as well as the associated constant ℓ_0) cancel out in the final expression for u^t . Our final result for a 3PN (plus 4PN and 5PN logarithmic terms), gauge-invariant, algebraic relationship between u^T (to which u^t now evaluates) and x (or equivalently Ω), is⁶

$$\begin{aligned} u^T(x) = & 1 + \left(\frac{3}{4} + \frac{3}{4}\Delta - \frac{\nu}{2} \right) x + \left(\frac{27}{16} + \frac{27}{16}\Delta - \frac{5}{2}\nu - \frac{5}{8}\Delta\nu + \frac{\nu^2}{24} \right) x^2 \\ & + \left(\frac{135}{32} + \frac{135}{32}\Delta - \frac{37}{4}\nu - \frac{67}{16}\Delta\nu + \frac{115}{32}\nu^2 + \frac{5}{32}\Delta\nu^2 + \frac{\nu^3}{48} \right) x^3 \\ & + \left(\frac{2835}{256} + \frac{2835}{256}\Delta - \left[\frac{2183}{48} - \frac{41}{64}\pi^2 \right] \nu - \left[\frac{12199}{384} - \frac{41}{64}\pi^2 \right] \Delta\nu \right. \\ & \left. + \left[\frac{17201}{576} - \frac{41}{192}\pi^2 \right] \nu^2 + \frac{795}{128}\Delta\nu^2 - \frac{2827}{864}\nu^3 + \frac{25}{1728}\Delta\nu^3 + \frac{35}{10368}\nu^4 \right) x^4 \\ & + \left(A_4(\nu) + \left[-\frac{32}{5} - \frac{32}{5}\Delta + \frac{64}{15}\nu \right] \nu \ln x \right) x^5 \\ & + \left(A_5(\nu) + \left[\frac{478}{105} + \frac{478}{105}\Delta + \frac{1684}{21}\nu + \frac{4388}{105}\Delta\nu - \frac{3664}{105}\nu^2 \right] \nu \ln x \right) x^6 \\ & + o(x^6). \end{aligned} \tag{36}$$

We introduced the notation $\Delta \equiv (m_2 - m_1)/m = \sqrt{1 - 4\nu}$, where $\nu = m_1 m_2 / m^2$ is the symmetric mass ratio. While it has been shown in [28] (see also Section 2 above)

⁶ The Landau o symbol for remainders takes its standard meaning.

that u^T is gauge invariant at any PN order in the extreme mass ratio limit $\nu \ll 1$, here we find that it is also gauge invariant for *any* mass ratio up to 3PN order (even up to 5PN order for the logarithmic terms). This result is expected from (2), according to which u^T is a scalar under our hypothesis of helical symmetry. Being proportional to the symmetric mass ratio ν , the 4PN and 5PN logarithmic contributions vanish in the test-mass limit – this is clear given that the Schwarzschild result for $u^T(\Omega)$ does not involve any logarithm. Notice that the functions $A_4(\nu)$ and $A_5(\nu)$ entering the expression of the non-logarithmic contribution to $u^T(\Omega)$ at the 4PN and 5PN orders are unknown, and would be very difficult to compute within standard PN theory. However, we know that they are polynomials in ν , with leading-order coefficient given by the Schwarzschild result (see Eq. 40).

We now investigate the small mass ratio regime $q \ll 1$, for comparison purposes with the perturbative SF calculation. We introduce a convenient PN parameter appropriate to the small mass limit of particle 1:

$$y \equiv \left(\frac{G m_2 \Omega}{c^3} \right)^{2/3}, \quad (37)$$

which is related to the usual PN parameter x by $x = y(1 + q)^{2/3}$, and to the gauge-invariant measure (11) of the orbital radius by $y = Gm_2/(R_\Omega c^2)$. We also use the expression of the symmetric mass ratio ν in terms of the (asymmetric) mass ratio $q = m_1/m_2$, namely $\nu = q/(1 + q)^2$. Our complete redshift observable, expanded through post-SF order, is of the type

$$u^T = u_{\text{Schw}}^T + q u_{\text{SF}}^T + q^2 u_{\text{PSF}}^T + \mathcal{O}(q^3), \quad (38)$$

where the Schwarzschild result is known in closed form as $u_{\text{Schw}}^T(y) = (1 - 3y)^{-1/2}$. By expanding the PN result (36) in powers of q , we find that the SF contribution reads

$$\begin{aligned} u_{\text{SF}}^T(y) = & -y - 2y^2 - 5y^3 + \left(-\frac{121}{3} + \frac{41}{32}\pi^2 \right) y^4 + \left(\alpha_4 - \frac{64}{5} \ln y \right) y^5 \\ & + \left(\alpha_5 + \frac{956}{105} \ln y \right) y^6 + o(y^6). \end{aligned} \quad (39)$$

The coefficients α_4 and α_5 are pure numbers that parametrize the small mass ratio expansions of the functions A_4 and A_5 through

$$A_4 = \frac{15309}{256} + \left(\alpha_4 - \frac{25515}{128} \right) q + \mathcal{O}(q^2), \quad (40a)$$

$$A_5 = \frac{168399}{1024} + \left(\alpha_5 - \frac{168399}{256} \right) q + \mathcal{O}(q^2). \quad (40b)$$

We also give the result for the combination $\bar{u}^\alpha \bar{u}^\beta h_{\alpha\beta}^R$ related to u_{SF}^T by Eq. 14, since this is the quantity primarily used in the numerical SF calculation:

$$\begin{aligned} \bar{u}^\alpha \bar{u}^\beta \hat{h}_{\alpha\beta}^R = & -2y - y^2 - \frac{7}{4}y^3 + \left(-\frac{1387}{24} + \frac{41}{16}\pi^2\right)y^4 + \left(a_4 - \frac{128}{5}\ln y\right)y^5 \\ & + \left(a_5 + \frac{5944}{105}\ln y\right)y^6 + o(y^6). \end{aligned} \quad (41)$$

We have conveniently rescaled the first-order perturbation $h_{\alpha\beta}^R$ by the mass ratio q , denoting $\hat{h}_{\alpha\beta}^R \equiv h_{\alpha\beta}^R/q$. Here a_4 and a_5 denote some unknown pure numbers related to α_4 and α_5 by

$$a_4 = 2\alpha_4 + \frac{9301}{64} - \frac{123}{32}\pi^2, \quad (42a)$$

$$a_5 = 2\alpha_5 - 3\alpha_4 + \frac{17097}{128} - \frac{369}{128}\pi^2. \quad (42b)$$

The expansions (39)–(41) were determined up to 2PN order $\propto y^3$ in [28], based on the Hadamard-regularized 2PN metric given in [16]. The result at 3PN order $\propto y^4$ was obtained in Paper I using the powerful dimensional regularization scheme. By comparing the expansion (39) with our accurate numerical SF data for $u_{\text{SF}}^T(\Omega)$, we shall be able to measure the coefficients α_4 and α_5 (or a_4 and a_5) with at least eight significant digits for the 4PN coefficient, and five significant digits for the 5PN coefficient. These results, as well as the estimation of even higher-order PN coefficients, will be detailed in the next section.

Similarly, from the PN result (36) valid for any mass ratio q , we get the post-SF contribution as

$$\begin{aligned} u_{\text{PSF}}^T(y) = & y + 3y^2 + \frac{97}{8}y^3 + \left(\frac{725}{12} - \frac{41}{64}\pi^2\right)y^4 + \varepsilon_4 y^5 \\ & + \left(\varepsilon_5 + \frac{4588}{35}\ln y\right)y^6 + o(y^6), \end{aligned} \quad (43)$$

which could in principle be compared to a future post-SF calculation making use of second-order black hole perturbation theory. Note that there is no logarithm at 4PN order in the post-SF term; the next 4PN logarithm would arise at cubic order q^3 , that is, at the post-post-SF level. The coefficients ε_4 and ε_5 in (43) are unknown, and unfortunately they are expected to be extremely difficult to obtain, not only analytically in the standard PN theory, but also numerically as they require a second-order perturbative SF scheme.

8 Numerical Evaluation of Post-Newtonian Coefficients

In the SF limit, the SF effect u_{SF}^T on the redshift observable u^T is related via (12) to the regularized metric perturbation $\hat{h}_{\alpha\beta}^R$ at the location of the particle through

$$u_{\text{SF}}^T = \frac{1}{2}(1-3y)^{-1/2} \bar{u}^\alpha \bar{u}^\beta \hat{h}_{\alpha\beta}^{\text{R}}, \quad (44)$$

where \bar{u}^α is the background four-velocity of the particle. Recall that here $\hat{h}_{\alpha\beta}^{\text{R}}$ stands for the perturbation *per unit mass ratio*, that is, $h_{\alpha\beta}^{\text{R}}/q$. In SF analysis, the combination $\bar{u}^\alpha \bar{u}^\beta \hat{h}_{\alpha\beta}^{\text{R}}$ arises more naturally than u_{SF}^T ; this is the quantity we shall be interested in fitting in this section. However, our final results in Table 5 will include the corresponding values of the coefficients for the redshift observable u_{SF}^T . We refer to Section II of Paper I for a discussion of the computation of the regularized metric perturbation $\hat{h}_{\alpha\beta}^{\text{R}}$, and the invariant properties of the combination $\bar{u}^\alpha \bar{u}^\beta \hat{h}_{\alpha\beta}^{\text{R}}$ with respect to the choice of perturbative gauge. In this section we often use $r \equiv 1/y$, a gauge-invariant measure of the orbital radius scaled by the black hole mass m_2 (see Eqs. 11 and 37).

Our earlier numerical work in [11, 12, 28] provided values of the function $\bar{u}^\alpha \bar{u}^\beta \hat{h}_{\alpha\beta}^{\text{R}}(r)$ which cover a range in r from 4 to 750. Following a procedure described in [29], we have used Monte Carlo analysis to estimate the accuracy of our values for $\bar{u}^\alpha \bar{u}^\beta \hat{h}_{\alpha\beta}^{\text{R}}$. As was reported in Paper I, this gives us confidence in these base numbers to better than one part in 10^{13} . We denote a standard error σ representing the numerical error in $\bar{u}^\alpha \bar{u}^\beta \hat{h}_{\alpha\beta}^{\text{R}}$ by

$$\sigma \simeq |\bar{u}^\alpha \bar{u}^\beta \hat{h}_{\alpha\beta}^{\text{R}}| \times \text{E} \times 10^{-13}, \quad (45)$$

where $\text{E} \simeq 1$ is being used as a placeholder to identify our estimate of the errors in our numerical results.

8.1 Overview

A common task in physics is creating a functional model for a set of data. In our problem we have a set of N data points f_i and associated uncertainties σ_i , with each pair evaluated at an abscissa r_i . We wish to represent this data as some model function $f(r)$ which consists of a linear sum of M basis functions $F_j(r)$ such that

$$f(r) = \sum_{j=1}^M c_j F_j(r). \quad (46)$$

The numerical goal is to determine the M coefficients c_j that yield the best fit in a least squares sense over the range of data. That is, the c_j are to be chosen such that

$$\chi^2 \equiv \sum_{i=1}^N \left[\frac{f_i - \sum_{j=1}^M c_j F_j(r_i)}{\sigma_i} \right]^2 \quad (47)$$

is a minimum under small changes in the c_j . For our application we choose the basis functions $F_j(r)$ to be a set of terms that are typical in PN expansions, such as r^{-1} , r^{-2} , ..., and also terms such as $r^{-5} \ln(r)$. We recognize that a solution to this extremum

problem is not guaranteed to provide an accurate representation of the data (r_i, f_i, σ_i) . The quality of the numerical fit is measured by χ^2 as defined in Eq. 47. If the model of the data is a good one, then the χ^2 statistic itself has an expectation value of the number of degrees of freedom in the problem, $N - M$, with an uncertainty (standard deviation) of $\sqrt{2(N - M)}$.

Our numerical work leans heavily upon Ref. [42] for solving the extremum problem for Eq. 47. The numerical evaluation of the fitting coefficient c_j includes a determination of its uncertainty Σ_j which depends upon (i) the actual values of r_i in use, (ii) all of the σ_i , and (iii) the set of basis functions $F_j(r)$. In fact, the estimates of the Σ_j do not depend at all on the data (or residuals) being fitted. As a consequence the estimates of the Σ_j are only valid if the data are well represented by the set of basis functions. For emphasis, the Σ_j depend upon $F_j(r_i)$ and upon σ_i but *are completely independent* of the f_i . Only if the fit is considered to be good, could the Σ_j give any kind of realistic estimate for the uncertainty in the coefficients c_j . If the fit is not of high quality (unacceptable χ^2), then the Σ_j bear no useful information [42]. We will come back to this point in the discussion below.

We also should remark that the task of determining coefficients in the $1/r$ characterization of our numerical data is almost incompatible with the task of determining an asymptotic expansion of $\bar{u}^\alpha \bar{u}^\beta \hat{h}_{\alpha\beta}^R$ from an analytic analysis. Analytically, the strict $r \rightarrow +\infty$ limit is always technically possible, whereas numerically, not only is that limit *never* attainable, but we must always contend with function evaluations at just a finite number of discrete points, obtained within a finite range of the independent variable, and computed with finite numerical precision. Nevertheless, this is what we have done below.

The numerical problem is even more constrained than we have just indicated. At large r , even though the data may still be computable there, the higher order terms for which we are interested in evaluating PN coefficients rapidly descend below the error level of our numerical data. This is clearly evident in Fig. 2 below. For small r , the introduction of so many PN coefficients is necessary that it becomes extremely difficult to characterize our numerical data accurately. Thus, in practice, we find ourselves actually working with less than the full range of our available data. At large r we could effectively drop points because they contribute so little to any fit we consider. At the other extreme, the advantage of adding more points in going to smaller r is rapidly outweighed by the increased uncertainty in every fitted coefficient. This results from the need to add more basis functions in an attempt to fit the data at small r . Further details will become evident in Section 8.4 below.

8.2 Framework for Evaluating PN Coefficients Numerically

In a generic fashion we describe an expansion of $\bar{u}^\alpha \bar{u}^\beta \hat{h}_{\alpha\beta}^R$ in terms of PN coefficients a_j and b_j with

$$\bar{u}^\alpha \bar{u}^\beta \hat{h}_{\alpha\beta}^R = \sum_{j \geq 0} \frac{a_j}{r^{j+1}} - \ln r \sum_{j \geq 4} \frac{b_j}{r^{j+1}}, \tag{48}$$

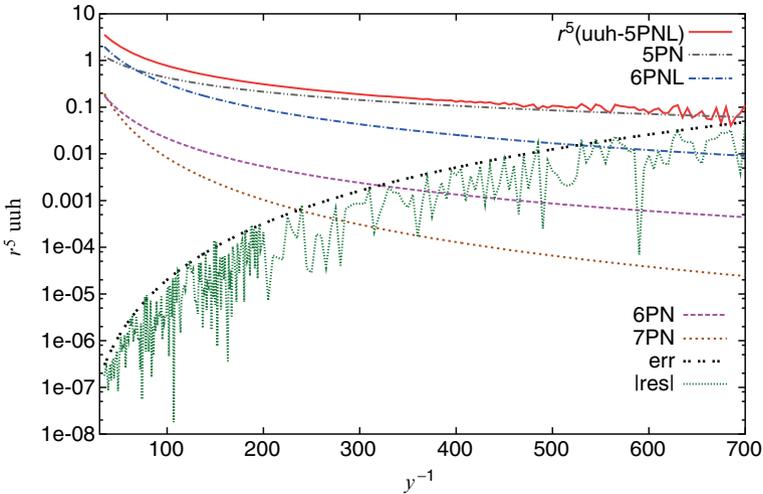


Fig. 2 The absolute value of the contributions of the numerically determined post-Newtonian terms to $r^5 \bar{u}^\alpha \bar{u}^\beta \hat{h}_{\alpha\beta}^R$. Here PNL refers to just the logarithm term at the specified order. The contribution of a_4 is not shown but would be a horizontal line (since the 4PN terms behaves like r^{-5}) at approximately 121.3. The remainder after a_4 and all the known coefficients are removed from $r^5 \bar{u}^\alpha \bar{u}^\beta \hat{h}_{\alpha\beta}^R$ is the top (red) continuous line. The lower (black) dotted line labelled “err” shows the uncertainty in $r^5 \bar{u}^\alpha \bar{u}^\beta \hat{h}_{\alpha\beta}^R$, namely $2Er^4 \times 10^{-13}$. The jagged (green) line labeled “|res|” is the absolute remainder after all of the fitted terms have been removed. The figure reveals that, with regard to the uncertainty of the calculated $\bar{u}^\alpha \bar{u}^\beta \hat{h}_{\alpha\beta}^R$, the choice $E \simeq 1$ was slightly too large

where a_0 is the Newtonian term, a_1 is the 1PN term, and so on. Similarly, for use in applications involving u^T we also introduce the coefficients α_j and β_j in the expansion of the SF contribution

$$u_{SF}^T = \sum_{j \geq 0} \frac{\alpha_j}{r^{j+1}} - \ln r \sum_{j \geq 4} \frac{\beta_j}{r^{j+1}}. \tag{49}$$

These series allow for the possibility of logarithmic terms, which are known not to start before the 4PN order. We also concluded in Paper II that $(\ln r)^2$ terms cannot arise before the 5.5PN order. Since we are computing a conservative effect, possible time-odd logarithmic squared contributions at the 5.5PN or 6.5PN orders do not contribute. But there is still the possibility for a conservative 7PN $(\ln r)^2$ effect, probably originating from a tail modification of the dissipative 5.5PN $(\ln r)^2$ term. However, we shall not permit for such a small effect in our fits. As discussed below in Section 8.4, we already have problems distinguishing the 7PN linear $\ln r$ term from the 7PN non-logarithmic contribution.

The analytically determined values of the coefficients $a_0, a_1, a_2, a_3, b_4, b_5$ and $\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_4, \beta_5$ computed in Ref. [28] and Papers I and II are reported in Table 1.

Table 1 The analytically determined PN coefficients for $\bar{u}^\alpha \bar{u}^\beta \hat{h}_{\alpha\beta}^R$ (left) and u_{SF}^T (right)

Coeff.	Value	Coeff.	Value
a_0	-2	α_0	-1
a_1	-1	α_1	-2
a_2	$-\frac{7}{4}$	α_2	-5
a_3	$-\frac{1387}{24} + \frac{41}{16}\pi^2$	α_3	$-\frac{121}{3} + \frac{41}{32}\pi^2$
b_4	$-\frac{128}{5}$	β_4	$-\frac{64}{5}$
b_5	$+\frac{5944}{105}$	β_5	$+\frac{956}{105}$

8.3 Consistency Between Analytically and Numerically Determined PN Coefficients

In this section, we investigate the use of our data for $\bar{u}^\alpha \bar{u}^\beta \hat{h}_{\alpha\beta}^R$ and the fitting procedures we have described above (and expanded upon in the beginning of Section 8.4). We will begin by fitting for enough of the other PN coefficients to be able to verify numerically the various coefficients a_3 , b_4 , and b_5 now known from PN analysis.

As a first step in this section, we will complete the task that was begun in [28], namely, the numerical determination of the coefficient a_3 (and α_3), this time taking fully into account the known logarithmic terms at 4PN and 5PN orders. For illustrative purposes only, these results are given in Table 2. We were able to obtain a fit with six undetermined parameters, and could include data from $r = 700$ down to $r = 35$. Note that, with the inclusion of the b_4 and b_5 coefficients, the precision of our tabulated value for a_3 has increased by more than four orders of magnitude from Paper I, although our accuracy is still no better than about 2Σ . Such a discrepancy is not uncommon. The uncertainty, Σ , reflects only how well the data in the given, finite range can be represented by a combination of the basis functions. It is not a measure of the quality of a coefficient when considered as a PN expansion parameter, which necessarily involves an $r \rightarrow +\infty$ limiting process.

Our next step is to include the known value for a_3 and to use our numerical data to estimate values for the b_4 and b_5 coefficients. Our best quality numerical result was obtained with five fitted parameters, over a range from $r = 700$ down to only $r = 65$, and is given in the first row of Table 3. Notice that while our b_4 is determined relatively precisely, it has only about 6Σ accuracy. The higher order coefficient b_5 is more difficult to obtain and, at this point, it is very poorly determined, but we can use the known value of b_4 in order to improve the accuracy for b_5 . These results are presented in Table 3, which again shows that we needed to fit for a total of six parameters to get a result of reasonable accuracy. With this, we have reached a limit for treating our data in this way, since adding further parameters and inner points does not result in any higher quality fit.

By now we have presented enough to show that we have data which allows high precision, with an accuracy that we now have some experience in relating to the computed error estimates. This experience will be valuable when we come to discuss further results in the next section. For convenience, we summarize the relevant information further, in Table 4, referring just to our estimates of known PN parameters, and relating our error estimates to the observed accuracy.

Table 2 The results of a numerical fit for a set of six coefficients that includes a_3 , which is now known analytically [11]. This fit uses the known results for b_4 and b_5 [12], but not the known value of a_3 . Thus, it is *not* the best fit of our data possible. The uncertainty in the last digit or two is in parentheses. The range runs from $r = 35$ to $r = 700$, with 266 data points and a respectable χ^2 of 264

3PN coeff.	Ref. [28]	Paper I	Paper II	PN (exact)
a_3	-32.34(6)	-32.479(10)	-32.5008069(7)	-32.50080538...
α_3	-27.61(3)	-27.677(5)	-27.6879035(4)	-27.68790269...

Table 3 The numerically determined PN coefficients for $\bar{u}^\alpha \bar{u}^\beta \hat{h}_{\alpha\beta}^R$. Each row represents a different fit. The first two columns give the starting point r_{\min} at the inner boundary of the fitting range, and the value of χ^2 statistic per degree of freedom (dof) for the chosen fit. The degrees of freedom, $N - M$, for the fit, range between 212 and 255, depending on r_{\min} . If a value for a coefficient is not shown, then either that parameter was not included in that particular fit (far right) or its analytically known value was used (e.g., b_4). The formal uncertainty of a coefficient in the last digit or two is in parentheses. The outer boundary is at 700 in each case

r_{\min}	χ^2/dof	a_4	b_4	a_5	b_5	a_6	b_6	a_7
65	0.961	-121.40(1)	-25.612(2)	-102(1)	45.5(3)	-2081(9)		
85	0.976	-121.3180(7)		-91.5(7)	48.5(2)	-2170(8)		
65	0.961	-121.313(1)		-79(2)	50.6(4)	-1868(44)	131(21)	
40	0.969	-121.3052(6)		-47(1)	55.7(2)	-359(41)	625(15)	-7722(162)

Table 4 Comparing the analytically known PN coefficients (column 5) with their numerically determined counterparts (column 3), and comparing the numerically determined error estimates (column 3) with the apparent accuracy (column 4). The source of the data is given in column 1

Source	Coeff.	Estimate	Accuracy	Exact result
Paper I	α_3	-27.677(5)	→ (11)	-27.6879...
Table 2	a_3	-32.5008069(7)	→ (15)	-32.50080538...
Table 3	b_4	-25.612(2)	→ (12)	-25.6
Table 3	b_5	+55.7(2)	→ (9)	+56.6095...

8.4 Determining Higher Order PN Terms Numerically

In this section we make maximum use of the coefficients which are already known. We find that in our *best-fit* analysis we can use a set of five basis functions corresponding to the unknown coefficients a_4, a_5, a_6, b_6 and a_7 .

In Table 5, we describe the numerical fit of our data over a range in r from 40 to 700. The χ^2 statistic is 259 and slightly larger than the degrees of freedom, 256, which denotes a good fit. Further, we expect that a good fit would be insensitive to changes in the boundaries of the range of data being fit, and we find, indeed, that if the outer boundary of the range decreases to 300 then essentially none of the data in the table changes, except for χ^2 and the degrees of freedom which decrease in a consistent fashion. Figure 2 shows that in the outer part of the range $\bar{u}^\alpha \bar{u}^\beta \hat{h}_{\alpha\beta}^R$ is heavily dominated

Table 5 The numerically determined values of higher-order PN coefficients for $\bar{u}^\alpha \bar{u}^\beta \hat{h}_{\alpha\beta}^R$ (left) and for u_{SF}^T (right). The uncertainty in the last digit or two is in parentheses. The range runs from $r = 40$ to $r = 700$, with 261 data points being fit. The χ^2 statistic is 259. We believe that a contribution from a b_7 confounds the a_7 coefficient. Both terms fall off rapidly and have influence over the fit only at small r . And the radial dependence of these two terms only differ by a factor of $\ln r$ [or possibly $(\ln r)^2$; see Paper II] which changes slowly over their limited range of significance

Coeff.	Value	Coeff.	Value
a_4	-121.30310(10)	α_4	-114.34747(5)
a_5	-42.89(2)	α_5	-245.53(1)
a_6	-215(4)	α_6	-695(2)
b_6	+680(1)	β_6	+339.3(5)
a_7	-8279(25)	α_7	-5837(16)

by only a few lower order terms in the PN expansion – those above the lower black double-dashed line in the figure.

The inner edge of the range is more troublesome. The importance of a given higher order PN term decreases rapidly with increasing r . Moving the inner boundary of the range outward might move a currently well determined term into insignificance. This could actually lead to a smaller χ^2 , but it would also lead to an increase in the Σ_j of every coefficient. Moving the inner edge of the range inward might require that an additional higher order term be added to the fit. This extra term loses significance quickly with increasing r so the new coefficient will be poorly determined and also result in an overall looser fit with an increase of Σ_j for all of the coefficients. If the inner boundary and the set of basis functions are chosen properly, then a robust fit is revealed when the parameters being fit are insensitive to modest changes in the boundaries of the range. The fit described in Table 5 appears to be robust. The parameters in this Table are consistent with all fits with the inner boundary of the range varying from 35 to 45 and the outer boundary varying from 300 to 700.

If an additional term, with coefficient b_7 , is added to the basis functions then, for identical ranges, each of the Σ_j increases by a factor of about 10, and the changes in a_4 and a_5 are within this uncertainty. The coefficient a_6 changes sign and b_6 and a_7 change by an amount significantly larger than the corresponding Σ_j . And the new coefficient b_7 is quite large. In the context of fitting data to a set of basis functions these are recognized symptoms of over-fitting and imply that the extra coefficient degrades the fit.

8.5 Summary

Our best fit can be visualized in Fig. 3, where we plot the SF effect u_{SF}^T on the redshift variable u^T as a function of $r = y^{-1}$, as well as several truncated PN series up to 7PN order, based on the analytically determined coefficients summarized in Table 1,

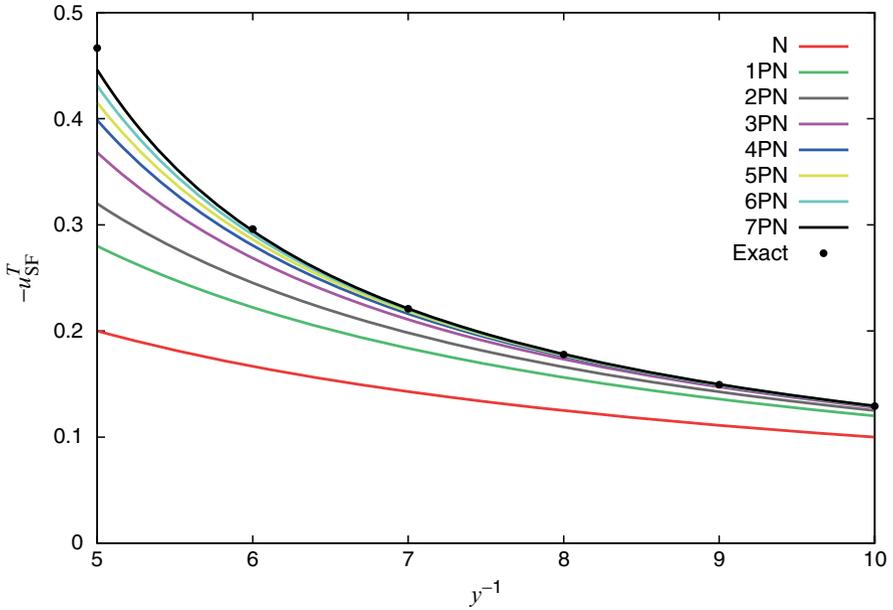


Fig. 3 The self-force contribution u_{SF}^T to the redshift observable u^T , plotted as a function of the gauge-invariant variable y . Note that y^{-1} is an invariant measure of the orbital radius scaled by the black hole mass m_2 (see Eqs. 11 and 37). The “exact” numerical points are taken from Ref. [28]. Here, PN refers to all terms, including logarithms, up to the specified order (however recall that we did not include in our fit a log-term at 7PN order)

as well as our best fit of the higher-order PN coefficients reported in Table 5. Observe in particular the smooth convergence of the successive PN approximations toward the exact SF results. Note, though, that there is still a small separation between the 7PN curve and the exact data in the very relativistic regime shown at the extreme left of Fig. 3.

We have found that our data in the limited range of $35 \leq r \leq 700$ can be extremely well characterized by a fit with five appropriately chosen (basis) functions. That is, the coefficients in Table 5 are well determined, with small uncertainties, and small changes in the actual details of the fit result in coefficients lying within their error estimates. Fewer coefficients would result in a very poor characterization of the same data while more coefficients result in large uncertainties in the estimated coefficients, which themselves become overly sensitive to small changes in specific details (such as the actual choice of points to be fitted). In practice, over the data range we finally choose, and with the five coefficients we fit for, we end up with exceedingly good results for the estimated coefficients, and with residuals which sink to the level of our noise. We have a very high quality fit which is quite insensitive to minor details. Nevertheless, as Table 4 hints, error estimates for these highest order coefficients should be regarded with an appropriate degree of caution.

Acknowledgements The authors acknowledge the 2008 Summer School on Mass and Motion, organized by A. Spallicci and supported by the University of Orléans and the CNRS, through which we experienced an extensive opportunity to understand each other's perspective and make rapid progress on this work. SD and BFW acknowledge support through grants PHY-0555484 and PHY-0855503 from the National Science Foundation. LB and ALT acknowledge support from the Programme International de Coopération Scientifique (CNRS-PICS).

References

1. P. Ajith, S. Babak, Y. Chen, M. Hewitson, B. Krishnan, A.M. Sintes, J.T. Whelan, B. Bruegmann, P. Diener, N. Dorband, J. Gonzalez, M. Hannam, S. Husa, D. Pollney, L. Rezzolla, L. Santamaria, U. Sperhake, J. Thornburg, Phys. Rev. D **77**, 104017 (2008); Erratum, *ibidem* **79**, 129901 (2009)
2. J.L. Anderson, in *Isolated Gravitating Systems in General Relativity*, ed. by J. Ehlers (North Holland, Amsterdam, 1979), p. 289
3. J.L. Anderson, R.E. Kates, L.S. Kegeles, R.G. Madonna, Phys. Rev. D **25**, 2038 (1982)
4. L. Barack, Class. Q. Grav. **26**, 213001 (2009)
5. L. Blanchet, Phys. Rev. D **47**, 4392 (1993)
6. L. Blanchet, T. Damour, Phil. Trans. R. Soc. Lond. A **320**, 379 (1986)
7. L. Blanchet, T. Damour, Phys. Rev. D **37**, 1410 (1988)
8. L. Blanchet, T. Damour, Phys. Rev. D **46**, 4304 (1992)
9. L. Blanchet, T. Damour, G. Esposito-Farèse, Phys. Rev. D **69**, 124007 (2004)
10. L. Blanchet, T. Damour, G. Esposito-Farèse, B.R. Iyer, Phys. Rev. Lett. **93**, 091101 (2004)
11. L. Blanchet, S. Detweiler, A. Le Tiec, B.F. Whiting, Phys. Rev. D **81**, 064004 (2010)
12. L. Blanchet, S. Detweiler, A. Le Tiec, B.F. Whiting, Phys. Rev. D **81**, 084033 (2010)
13. L. Blanchet, G. Faye, Phys. Rev. D **63**, 062005 (2001)
14. L. Blanchet, G. Faye, B.R. Iyer, B. Joguet, Phys. Rev. D **65**, 061501(R) (2002); Erratum, *ibidem* **71**, 129902 (E) (2005)
15. L. Blanchet, G. Faye, B.R. Iyer, S. Sinha, Class. Q. Grav. **25**, 165003 (2008)
16. L. Blanchet, G. Faye, B. Ponsot, Phys. Rev. D **58**, 124002 (1998)
17. L. Blanchet, B.R. Iyer, B. Joguet, Phys. Rev. D **65**, 064005 (2002); Erratum, *ibidem* **71**, 129903(E) (2005)
18. L. Blanchet, B.R. Iyer, C.M. Will, A.G. Wiseman, Class. Q. Grav. **13**, 575 (1996)
19. C.G. Bollini, J.J. Giambiagi, Phys. Lett. B **40**, 566 (1972)
20. M. Boyle, A. Buonanno, L.E. Kidder, A.H. Mroue, Y. Pan, H.P. Pfeiffer, M.A. Scheel, Phys. Rev. D **78**, 104020 (2008)
21. A. Buonanno, T. Damour, Phys. Rev. D **59**, 084006 (1999)
22. W.L. Burke, J. Math. Phys. **12**, 401 (1971)
23. W.L. Burke, K.S. Thorne, in *Relativity*, ed. by M. Carmeli, S.I. Fickler, L. Witten (Plenum, New York, 1970), p. 209
24. C. Cutler, T.A. Apostolatos, L. Bildsten, L.S. Finn, E.E. Flanagan, D. Kennefick, D.M. Markovic, A. Ori, E. Poisson, G.J. Sussman, K.S. Thorne, Phys. Rev. Lett. **70**, 2984 (1993)
25. C. Cutler, L.S. Finn, E. Poisson, G.J. Sussman, Phys. Rev. D **47**, 1511 (1993)
26. T. Damour, P. Jaranowski, G. Schäfer, Phys. Lett. B **513**, 147 (2001)
27. S. Detweiler, Class. Q. Grav. **22**, S681 (2005)
28. S. Detweiler, Phys. Rev. D **77**, 124026 (2008)
29. S. Detweiler, E. Messaritaki, B.F. Whiting, Phys. Rev. D **67**, 104016 (2003)
30. S. Detweiler, B.F. Whiting, Phys. Rev. D **67**, 024025 (2003)
31. T. Futamase, Phys. Rev. D **28**, 2373 (1983)
32. T. Futamase, B.F. Schutz, Phys. Rev. D **28**, 2363 (1983)
33. S.E. Gralla, R.M. Wald, Class. Q. Grav. **25**, 205009 (2008)

34. G.D. Kerlick, *Gen. Rel. Grav.* **12**, 467 (1980)
35. G.D. Kerlick, *Gen. Rel. Grav.* **12**, 521 (1980)
36. Y. Mino, M. Sasaki, T. Tanaka, *Phys. Rev. D* **55**, 3457 (1997)
37. E. Poisson, *Phys. Rev. D* **47**, 1497 (1993)
38. E. Poisson, *Phys. Rev. D* **48**, 1860 (1993)
39. E. Poisson, *Phys. Rev. D* **52**, 5719 (1995); Erratum and Addendum, *ibidem* **55**, 7980 (1997)
40. E. Poisson, *Living Rev. Rel.* **7**, URL (cited on 20 June 2010): <http://www.livingreviews.org/lrr-2004-6>
41. E. Poisson, M. Sasaki, *Phys. Rev. D* **51**, 5753 (1995)
42. W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, in *Numerical Recipes: The Art of Scientific Computing*, 3rd edn. (Cambridge University Press, Cambridge, 2007)
43. T.C. Quinn, R.M. Wald, *Phys. Rev. D* **56**, 3381 (1997)
44. N. Sago, L. Barack, S. Detweiler, *Phys. Rev. D* **78**, 124024 (2008)
45. H. Tagoshi, T. Nakamura, *Phys. Rev. D* **49**, 4016 (1994)
46. H. Tagoshi, M. Shibata, T. Tanaka, M. Sasaki, *Phys. Rev. D* **54**, 1439 (1996)
47. T. Tanaka, H. Tagoshi, M. Sasaki, *Prog. Theor. Phys.* **96**, 1087 (1996)
48. G. 't Hooft, M. Veltman, *Nucl. Phys. B* **44**, 189 (1972)
49. K.S. Thorne, in *Three Hundred Years of Gravitation*, ed. by S.W. Hawking, W. Israel (Cambridge University Press, Cambridge, 1987), p. 330

LISA and Capture Sources

Oliver Jennrich

Abstract LISA is a joint ESA/NASA mission to detect and observe gravitational waves. It is designed to register the change in distance between free-falling reference points to picometer accuracy, allowing to measure the effect of gravitational waves created by the coalescence of massive black holes almost anywhere in the universe, stellar mass black holes and neutron stars spiraling into massive black holes in other galaxies at intermediate distances, and tightly orbiting binary stars in our galaxy. LISA will be able to detect gravitational waves from coalescing massive black holes to redshifts of $z \sim 10$ and higher, allowing an unprecedented view into the early stages of galaxy formation. The signals from the many million binary stars in our galaxy yield information about the evolution and the morphology of our galaxy, giving a view of the population of binary stars unobstructed by dust. Among the most challenging, yet scientifically interesting sources are the captures of a small massive object by massive black holes where the mass ratio exceeds 1,000. Those events, named extreme mass ratio inspirals (EMRI), create very complex waveforms and allow to test general relativity to very high precision. LISA has been recently confirmed as a candidate for the L1 mission in ESA's Cosmic Vision program and is foreseen to be launched in the 2018 time frame.

1 LISA – A Mission to Detect and Observe Gravitational Waves

LISA is a spaceborne interferometric gravitational wave detector, jointly planned by ESA and NASA [16, 24, 48, 75] based on earlier ideas to build a gravitational wave detector in space [20, 30]. In contrast to ground-based gravitational wave detectors [1, 4, 33, 45, 66, 70, 71, 82] that have a typical sensitivity in the range from 1 Hz to 1 kHz, the sensitivity for LISA stretches between 0.1 mHz and 0.1 Hz, accessing a frequency window that is inaccessible to ground-based detectors due to seismic noise and gravity gradient noise.

O. Jennrich (✉)
ESA/ESTEC, Noordwijk, The Netherlands
e-mail: oliver.jennrich@esa.int

The sensitivity for low frequencies allows LISA to assess gravitational waves that are emitted by some of the most violent events in the universe, such as the coalescence of massive black holes ($m_{\text{BH}} \approx 10^5 M_{\odot} \cdots 10^7 M_{\odot}$) that occur during the formation and the growth of galaxies, signals from galactic binary systems, and the capture of stellar-size compact objects by massive black holes. It is even possible that LISA will be able to detect gravitational waves from times shortly after the Big Bang. The general science of LISA is discussed in detail in [3, 19, 25, 42, 43, 53, 54]; the focus of this paper is on the capture of compact objects by massive black holes.

1.1 Mission Concept

LISA comprises three spacecraft in a heliocentric orbit, forming an equilateral triangle with 5×10^6 km a side (the “armlength”). The plane of the triangle is inclined by 60° with respect to the ecliptic, causing the constellation to counter-rotate while orbiting around the Sun. This peculiar arrangement of the satellites results in a stable constellation, as orbital mechanics cause the constellation to rigid: it maintains its size and shape closely during the nominal mission time of 5 years [32] allowing to operate LISA without further station keeping maneuvers.

The rigidity of the constellation is only approximate, so that the distances between the satellites change by up to 1%, causing a differential velocity along the line of sight of up to 15 m/s, and resulting in the constellation to slightly change its form.

The distance to the Earth of about 50×10^6 km (Fig. 1), corresponding to a trailing (or leading) angle of 20° , has been chosen as a compromise between long-term stability of the constellation and communications requirements.

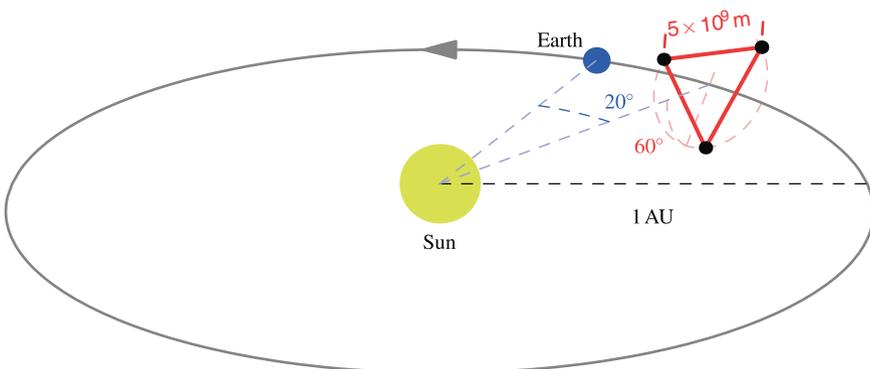


Fig. 1 Schematics of the LISA constellation orbiting the Sun. The constellation is inclined by 60° with respect to the ecliptic and trails the Earth by about 20° , resulting in a distance of about 50×10^6 km to Earth. The distance between the satellites (the armlength) is nominally 5×10^6 km

As gravitational waves cause a strain in the space-time [57], measuring the change in the distance of spatially separated objects is a common concept for measuring the effect of gravitational waves. Each spacecraft houses two test masses, kept in free fall, that form the reference points for an interferometric measurement of the inter-spacecraft distance.

1.2 Sensitivity

It is customary in the field of gravitational wave detection to state the sensitivity and noise figures in terms of a *linear spectral density* $S(f)$, that is, quantities are given in units¹ of $1/\sqrt{\text{Hz}}$; RMS values of the quantities can be retrieved by integrating the *spectral density* $S^2(f)$ over the frequency band Δf of interest and taking the square root:

$$\Delta x_{\text{rms}} = \sqrt{\int_{\Delta f} S_x^2(f) \frac{df}{2\pi}} \quad (1)$$

The interferometric measurement allows to assess the distance between the (almost) free-falling test masses to a level of $10 \text{ pm}/\sqrt{\text{Hz}}$, resulting in a strain sensitivity of about $e-20/\sqrt{\text{Hz}}$, enough to detect, for example, the coalescence of massive black holes even at redshifts of $z = 20$ with a signal-to-noise ratio of several hundreds [43, 49, 50]. The science requirements for LISA specify the sensitivity in a frequency window of 0.03 mHz to 0.1 Hz (Fig. 2) as this is where most of the sources for LISA are expected to emit.

The limitation to the sensitivity at frequencies below approximately 3 mHz is given by residual acceleration noise of the order of $3e-15 \text{ m/s}$ acting on the test masses. Any acceleration along the line connecting the test masses in two different satellites mimics the action of a gravitational wave and therefore limits LISA's sensitivity.

At higher frequencies, the noise associated with the position measurement limits the sensitivity to about $10 \text{ pm}/\sqrt{\text{Hz}}$. This noise is partly due to the unavoidable shot noise of the photons and partly due to imperfections and digitization noise in the phase meter.

The sensitivity is further reduced for frequencies above about 30 mHz by the transfer function of the detector: the effect of gravitational waves cancels out as the wavelength of the gravitational wave equals an integer multiple of the optical path in the detector [51].

The sensitivity of LISA to changes in the distance between the spacecraft is small, compared with ground-based detectors. However, as gravitational waves produce a *strain*, or a fractional change in distance, the large distance between the satellites provides a sensitivity to gravitational waves comparable to those of the much shorter ground-based detectors.

¹ Some authors use the *spectral density*, that is, $S^2(f)$ for the same purpose.

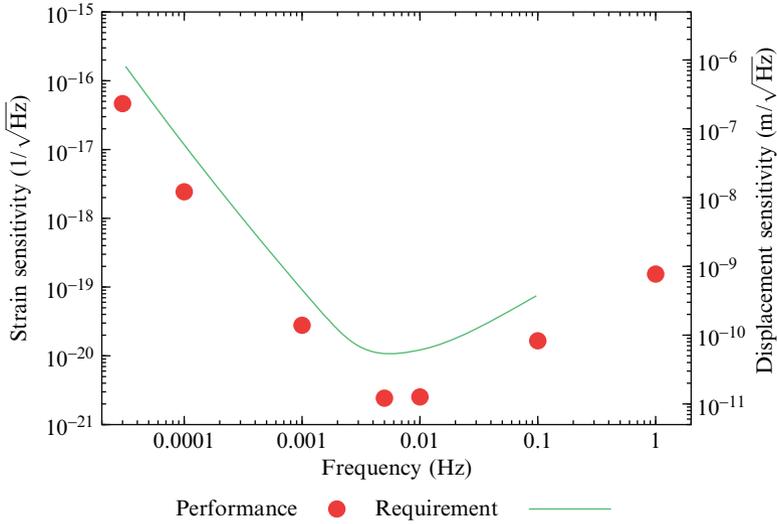


Fig. 2 Sensitivity requirements (*green solid*) and current best estimate of the performance (*red dots*). The strain sensitivity (*left axis*) is given for the full instrument, the displacement sensitivity (*right axis*) has been converted to a *single arm* link, that is represents the equivalent accuracy for the armlength measurement, combining all noise sources. The design of the instrument is such that a $\sim 35\%$ performance margin is included, resulting in the values for the current best estimate to be consistently lower than the requirement

1.3 Measurement Principle

For practical reasons, the interferometric measurement in LISA is broken up into three distinct parts: the measurement between the spacecraft, that is, between the optical benches that are fixed to the spacecraft, and the measurement between each of the test masses and its respective optical bench (see Fig. 4). While such a partition of the measurement is usually avoided as it increases the noise introduced by the detectors, it does little harm in the case of LISA, as the noise budget is dominated by the contribution of the shot noise in the measurement between the spacecraft.

The distance between the test mass and the optical bench is measured by reflecting light off the test mass and combining this measurement beam with a local oscillator on the optical bench (“test mass interferometer”).

To measure the distance between the spacecraft, about 2 W of infrared light (1.064 nm) is sent through a 40-cm telescope, used for receiving and transmitting, to the respective far spacecraft, that is, spacecraft A transmits to spacecraft B and C (nomenclature as in Fig. 3) etc., forming six single laser links. In an ordinary interferometer, the received light would be directly reflected back to the transmitting spacecraft A where the light would then be combined with a local oscillator, completing the measurement. However, due to the large distance between the two spacecrafts, a direct reflection of the light is not feasible. Diffraction widens

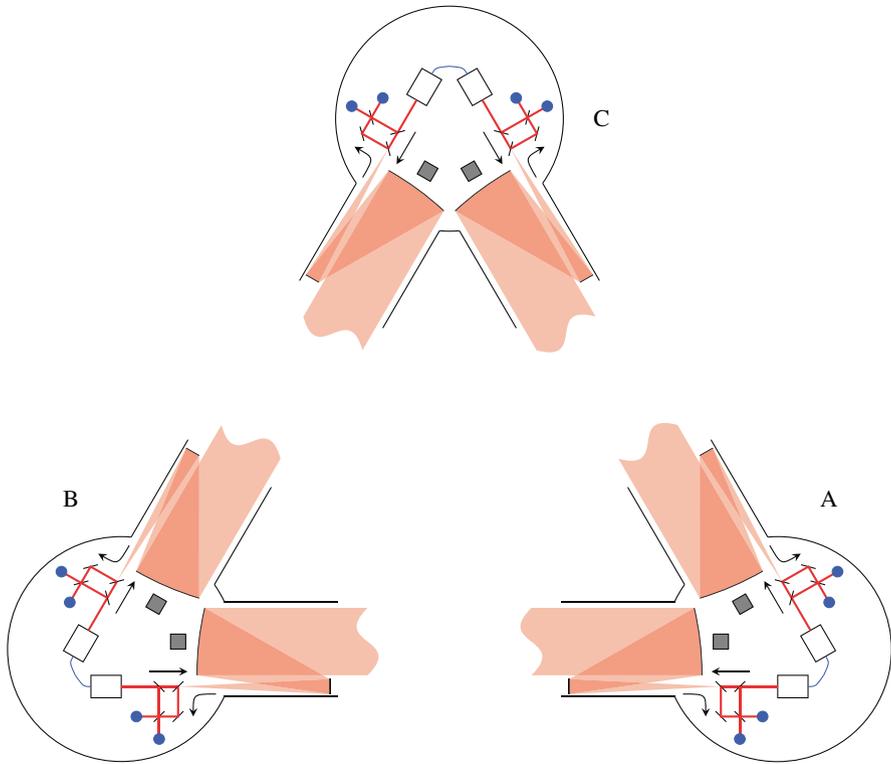


Fig. 3 Schematics of the LISA constellation. Each satellite transmits laser light to the two other satellites and receives laser light from the two other satellites. The laser used for transmitting is phase-locked to the received laser light, establishing a transponder scheme. As transmitted and received light share the same telescope, a polarization multiplexing scheme is used to separate the two beams. The use of linear polarisations requires slightly different optical benches on the sending and the receiving spacecraft

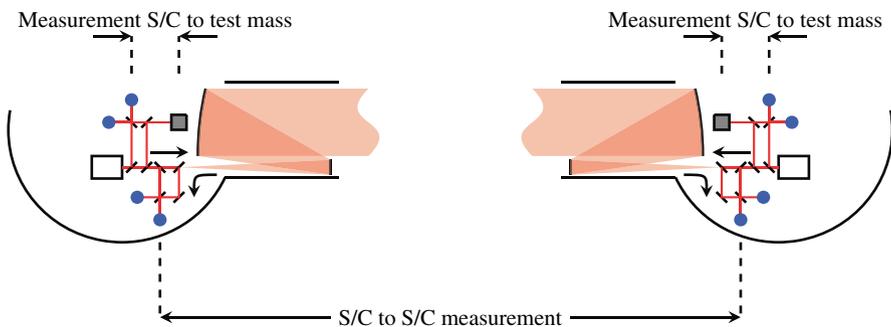


Fig. 4 The measurement of the distance between the test masses is broken up into three parts: The measurement between the two spacecraft, that is, the optical benches and one measurement each between the test mass and its optical bench

the transmitted laser beam to many kilometers at the receiving spacecrafts and the received power is of the order of 100 pW (or about 5×10^8 photons per second), resulting in a dilution factor of 10^{-10} , resulting in a photon rate of about 1 photon per second if direct reflection was attempted.

Therefore, a transponder scheme with offset phase-locking is implemented. The received light on spacecraft B is combined with a local oscillator derived from the transmitting laser (B) and the phase difference is measured. The frequency of the laser (B) can then be controlled so that the beat note corresponds precisely to an electronic offset frequency, rendering the phase of the transmitting laser a true copy of the phase of the received light. Recombining the light on spacecraft A then yields a beat note that contains the electronic offset frequency on spacecraft B, the Doppler shift (that can be as large as 15 m/s), and the signal due to the gravitational wave (“science interferometer”). To be able to distinguish a gravitational wave signal from a noise in the electronic offset frequency, the latter needs to be known precisely enough. This is achieved by multiplying the clock frequency on each spacecraft up to a few GHz, impose it as sidebands to the laser. On the other spacecraft, these sidebands can be compared to the local clock, allowing to assess and correct the noise of the local clock.

Extending the phase-locking scheme so that the two lasers on one spacecraft are phase-locked to each other as well, it is obvious that all six lasers can be phase-locked to one (arbitrarily chosen) master laser. Such a scheme requires then nine individual phase measurements, two for each arm and one each between the two lasers on a single spacecraft. In addition, each measurement of the test masses with respect to the optical bench requires another six phase measurements, so that a total of 15 phase measurements carry the complete information on the gravitational waves.

These signals will be sent to ground and combined on ground to form the LISA data streams that will be made available to the scientific community.

A more detailed overview of the instrumentation of LISA and the technology employed by LISA can be found in [47].

2 Capture Sources

The capture of compact objects by massive black holes is believed to occur regularly at the centers of galaxies where the central black hole is surrounded by a dense stellar environment. A “captured” object does not necessarily fall into the central black hole in the observation period, although eventually, most of them will. The term refers to an object that is in a bound orbit around the central black hole and comes close enough to it to emit a detectable amount of gravitational radiation. Such a scenario is possible only for compact objects, as non-compact objects (i.e., normal stars, molecular clouds) will be tidally disrupted long before they reach those orbits and will not emit gravitational waves in the LISA band.

Out of the stellar cloud surrounding the central black hole, compact stars, such as neutron stars and black holes, can be deflected toward the central black hole through interaction with other objects. If these “infalling” compact objects come

close enough to the central black hole, the energy loss through the emission of gravitational radiation is large enough to put them into a bound orbit around the central black hole [40, 44, 46, 63, 72]. Even low-mass main sequence stars can be captured [34], despite the fact that main sequence stars usually suffer from early tidal disruption.

Another mechanism that can bring compact objects close to the central black hole is based on tidal effects: Binary systems can be disrupted by tidal forces, leaving one constituent close to the central black hole [56]; compact cores of giant stars can be tidally stripped of their gas envelope by the central black hole and end up close to it [29].

Yet another class of compact objects might have formed in situ: black holes that have formed near the central black hole by the evolution of giant stars such as those seen near the central black hole in the Milky Way [52, 60, 62] or in the center of M31 [18].

The compact objects (neutron stars, black holes, and white dwarfs) resulting from those mechanisms would then survive as quasi point-like objects orbiting the central black hole until eventually their orbital energy is carried away through the emission of gravitational radiation and they vanish at the horizon. During the period of orbiting the central black hole, they would radiate waves in the LISA band for a long time, many of them emitting approximately 10^5 cycles or more of radiation before falling into the central black hole.

Two principal classes of systems can be distinguished according to their astrophysical characteristics and their mass ratio: extreme mass ratio inspirals (EMRI) and intermediate mass ratio inspirals (IMRI). For EMRIs, the ratio of the mass of the inspiraling object to that of the central black hole is 10^{-4} or smaller. This class comprises white dwarfs, neutron stars, or stellar-mass black holes. As black holes are the most massive objects in this class, they sink towards the denser regions of the galactic nucleus and will be therefore over-represented among the EMRI population. In addition, their higher mass makes for a stronger gravitational wave signal, allowing detection of them from larger distances, thus increasing their event rate further. The typical expected EMRI system is therefore a stellar-mass black hole with a mass of about $10 M_{\odot}$ falling into a $10^6 M_{\odot}$ massive black hole.

For IMRIs, the ratio between compact object and massive black hole is somewhat larger, between 10^{-4} and 10^{-2} . The inspiraling object here is expected to be an intermediate mass black hole, such as may have been formed in the first generation of star formation (Population III stars) [55]. The canonical expected source is an intermediate-mass black hole with a mass of $10^3 M_{\odot}$ falling into a $10^6 M_{\odot}$ massive black hole.

3 Science Return

Capture sources figure among the principal sources that allow to conduct fundamental tests of general relativity with LISA. Their signals contain rich information about the geometry around the central black holes and allow to test certain aspects of general relativity to unprecedented precision.

To emit gravitational radiation that can be detected with LISA, these objects are necessarily so close to their central black holes that it completely determines the orbital dynamics – surrounding stars or even accretion disks will have at most a small effect on the orbits. The compact object then acts as the textbook “test particle,” exploring the background space-time around the central black hole. The evolution of the signals, lasting for thousands or even hundreds of thousands of cycles, reflects in detail the near-geodesic orbits they follow around the central black hole; the orbits in turn map the space-time [67]. Comparing this phase evolution to the phase evolution of a signal from a compact object in a pure Kerr geometry allows to look for deviations from the geometry predicted by general relativity, thereby testing one of the most fundamental theorems of Einstein’s theory, that black holes are uniquely determined by their mass and their angular momentum [14].

An even more fundamental test of GR will be the direct observation of the event horizon. Direct evidence will come from seeing the signals changing characteristics as the compact object approaches the horizon and transfers from the inspiral phase to the coalescence or merger phase. If no such change will be observed, this would be strong evidence for the lack of an event horizon, and therefore ruling out a black hole as the central object.

Beyond fundamental physics, there are astrophysical payoffs for detecting these signals as well, as both EMRIs and IMRIs sample the stellar population near the central black holes. The different formation channels for EMRI lead to distinguishable properties of their orbits and consequently to different characteristics of the gravitational wave signals: direct captures of objects into close orbits are typically characterized by high eccentricities that can be clearly identified by the harmonic content of the gravitational wave signal; tidally captured objects will have nearly circular orbits with more or less random inclination, depending on their origin; and objects that evolved in the vicinity of the central black hole will occur in circular orbits that are most likely aligned with the accretion disk, and therefore with the spin of the black hole.

The observed rate of IMRI events will give a valuable handle on the demographics of the intermediate-mass black holes and their progenitors in the nucleus of galaxies. With additional information from simulations of mass segregation, stellar interactions and population evolution, the rates, and mass spectrum of the different classes of detected objects will provide unique insight into the coevolution of black holes and their host galaxies.

Capture sources can even play an important role in cosmology. The signals from EMRI and IMRI, depend only on a few parameters such as the masses of the central black hole and the compact object, the initial orbital angular momentum, the spin of the black hole and the initial separation, in particular, the frequency of the signals is determined by the mass of the central black hole, whereas the amplitude is determined by the mass of the infalling object.

As the signal shape and amplitude can be very precisely determined and as gravitational waves do not suffer from dispersion or absorption by the interstellar medium, the ratio of the observed signal strength to the calculated signal strength is a direct measure for the luminosity distance of the sources. If these measurements

can be combined with red-shift measurements of the host galaxy, an independent redshift–distance relationship can be given.

In the case of EMRIs, a direct detection of the host galaxy is unlikely, as EMRI are not expected to produce an electromagnetic signal. It is, however, possible to make a statistical argument: Provided LISA detects a suitable number of EMRIs (~ 20 to a redshift of $z = 0.5$), it is possible to establish such an independent redshift–distance relationship, thus determining Hubble’s constant to a precision of better than 1% [53].

A more detailed and very extensive review of the sources, the astrophysics, science, and detection of EMRIs using LISA can be found in [3].

It is worthwhile noting that the sensitivity of LISA is “just right” for the detection of EMRI in the sense that a significant degradation of the sensitivity in the frequency range of a few millihertz would diminish the science return for the EMRIs drastically. However, the cost benefit of giving up sensitivity in this frequency range as part of a descoping scenario is marginal and would at present not be justified [74].

4 Detection

Before exploiting the scientific potential of EMRI and IMRI observations, it is necessary to be able to detect those signals in the LISA data stream. Three main issues arise: the complexity of the signal, the algorithms used for detection, and the expected rate for EMRI and IMRI events.

Furthermore, the capture signals fall at least partly in the frequency range where the sensitivity is limited by the signal confusion of the galactic binary sources (see Section 4.3).

4.1 Capture Rates

Robust estimates for the rates of EMRI and IMRI require a model of the stellar distribution and dynamics in the galactic nucleus, taking into account the interaction between the stars, the relativistic effects close to the central black hole, and the relative abundance of the various stellar types.

Taking the relative abundance from stellar evolution simulations or observations, the two major approaches to the problem of stellar dynamics are based on direct N -body integration and on statistical methods.

The direct integration poses a significant numerical challenge, as a large number of objects has to be integrated and the equation of motion has to include post-Newtonian terms to accurately predict the number of captured objects [44]. Dedicated hardware and the general progress in computational speeds will allow to perform simulations yielding more accurate results.

Table 1 Number of EMRI detections for LISA, SNR, and distance thresholds. The predicted number of detectable EMRIs by LISA for a downgraded mission scenario (3-year mission, poor removal of confusion noise, see Section 4.3) and a nominal mission scenario (5-year mission, good removal of confusion noise). It should be noted that the estimates for the *event rates* are optimistic; true rates could be a factor of up to 100 lower, causing all detection numbers to scale accordingly. Numbers are taken from original data in [35]. The first two columns are the black hole mass M_\bullet and the compact body mass m_2 . The third column gives the SNR at a luminosity distance of 1 Gpc and the fourth column gives the maximum redshift at which such a source could be detected, given a detection threshold of $\text{SNR} = 30$. Values taken from [3]

M_\bullet ($10^6 M_\odot$)	m_2 (M_\odot)	SNR at 1 Gpc	z_{max}	Nominal #	Downgraded #
0.3	0.6	18	0.13	8	0.7
0.3	10	73	0.44	700	89
0.3	100	620	2.5	1	1
1	0.6	30	0.21	94	9
1	10	210	1.0	1,100	660
1	100	920	3.5	1	1
3	0.6	25	0.17	67	2
3	10	270	1.3	1,700	134
3	100	1,500	5.2	2	1

Statistical methods such as approaches based on the Fokker–Planck equation [26] or Monte Carlo methods are generally numerically cheaper. However, they have a harder time than direct integration methods in predicting EMRI rates, as the EMRI events are not only intrinsically rare, but also very sensitive to the details of the stellar dynamics around the central black hole as the capture “event” is rather gradual, as it requires a large number of orbits around the central black hole to allow the object to dissipate enough energy and bring it close enough to the black hole to be detectable in the LISA band[2].

The current implementations of statistical approaches have various shortcomings, so that a full understanding of the processes that lead to EMRIs and hence a robust prediction for their rate and their parameters, such as eccentricities, masses, stellar types, etc. are not yet available. Consequently, the rates for EMRIs assumed for LISA have quite a large spread, from a few to a few hundreds during the lifetime of the mission (see Table 1).

4.2 Signal Characteristics

The signals emitted from EMRI can be very complex due to the fact that the inspiraling object explores the metric close to the central black hole and experiences the full wealth of relativistic effects, which makes it difficult and numerically costly to create precise waveforms. It is, on the other hand, precisely this complexity that

makes the EMRI signals distinctive with respect to the precise geometry of the space-time around a black hole.

The standard approach, exploiting the post-Newtonian approximation of the fully relativistic equation of motions cannot work in the case of EMRI. Typically, the post-Newtonian expansion converges poorly when $v/c \gtrsim 0.3$ [21, 73]. As EMRI typically spend up to a few 10^6 orbits in this regime, using v/c as an expansion parameter is not useful.

Another technique that has seen dramatic and spectacular progress in the last few years, numerical relativity, is despite all this progress not very useful either. In contrast to (almost) equal mass mergers, where only a few orbits have to be calculated, the large number of orbits required to get a waveform for EMRI makes numerical relativity too slow for that purpose.

One of the methods that can be successfully employed to calculate at least approximate waveforms is based on perturbation theory. As the mass ratio in EMRIs is very small, the inspiralling object can be regarded as a small perturbation to the very well known background space-time of the central black hole. The computational complexity in this case stems from the accurate calculation of the gravitational self-force of the particle. While the mathematical theory of the self-force has been developed in the last decade ([64, 65] and references therein) a full, generic inspiral trajectory and a waveform from a particle on that trajectory is out of reach for the near future, as the calculations are computationally very expensive.

It is, however, much easier to compute adiabatic waveforms by calculating a sequence of geodesic trajectories and stitching together the waveforms resulting from these trajectories. These waveforms are called *adiabatic*, as this procedure only works as long as the timescale for the changes in the orbit are long compared to the orbital period itself, allowing to calculate average rates of change for the orbital constants. For a more detailed discussion see for example [3].

While the adiabatic waveforms are much easier to generate than the full waveform, their computation is still too expensive for the needs of data analysis in the current stage. In order to test algorithms and find the most suitable one for detecting EMRI in the LISA data stream, a very large number of signal templates has to be generated so that the whole parameter space is covered. Two families of approximate “kludge” waveforms have been developed, the analytic kludge and the numeric kludge. The purpose of these kludge waveforms is to allow their easy (and fast) calculation while at the same time being reasonably accurate, that is, yielding a good enough overlap (typically 95%) with the real waveform, so as not to lose too much signal to noise.

The analytic kludge is based on the gravitational wave emission from a particle on a Keplerian orbit and is then augmented by relativistic precession of the pericentre and the orbital plane and the inspiral, both taken from post-Newtonian calculations [12]. The fact that those waveforms can be expressed in an analytic way makes them very fast to calculate. However, their accuracy is limited in the later stages of the inspiral, as the approximation of the true orbits by Keplerian orbits becomes less and less accurate as the compact object comes close to the central black hole.

The numeric kludge waveforms are designed to overcome that problem by using the true geodesic orbit of the infalling particle, integrating the geodesic equation numerically. The building of the numerical kludge waveforms involves constructing an inspiral trajectory in phase space, for which a description that is accurate until the very end of the inspiral has been obtained [36] allowing to integrate the geodesic equation to get the trajectory of the inspiraling object in the Kerr background metric. The trajectory is then used to calculate the waveforms using standard radiation emission formulas. The numerical kludge waveforms show a very high overlap (95%) with the adiabatic waveforms for objects with a pericenter of larger than about $5M$ (where M is the mass of the black hole), making them a very useful tool in the LISA Data Analysis for detection and preliminary source identification, limiting the parameter range for a follow-up search with more accurate templates significantly.

4.3 Data Analysis

The LISA data analysis in general and for EMRI in particular makes significant use of template matching, where the noise-weighted inner product between the signal $h(t)$ and the template $r(t; \mathbf{q})$ is evaluated as

$$(h|r) = \int_0^{\infty} \frac{\tilde{h}(\omega)\tilde{r}^*(\omega; \mathbf{q}) + \tilde{h}^*(\omega)\tilde{r}(\omega; \mathbf{q})}{S(\omega)} d\omega, \quad (2)$$

where \tilde{h} denotes the Fourier transform of h (with respect to t), $S(\omega)$ is the power spectral density of the noise entering the data stream and \mathbf{q} stands for the parameters of the template.

The relevant noise for the detection of EMRI is not only the instrumental noise as given in Fig. 5 but includes also the so called *galactic binary foreground*. While reaching the instrumental sensitivity is a challenging problem, particularly in the face of laser frequency noise and relative spacecraft motion, it is well understood and techniques for dealing with the laser frequency noise have been developed [5, 23, 27, 28, 59, 68, 69, 76–80]. The galactic binary background is caused by millions of galactic binaries emitting gravitational waves [15, 17, 31, 41, 61], in the frequency band below a few millihertz, overlapping in time and frequency and creating a strongly coloured confusion noise.

It should be noted that depending on the event rates, signals from capture sources might contribute to the confusion noise [13].

Identifying the most likely parameters \mathbf{q}_0 of the template and determining the probability density function (pdf) $p(\mathbf{q})$ of the parameters \mathbf{q} , usually referred to as parameter estimation, requires one way or the other comparing the signal against a large number of templates. Currently, the data analysis for LISA is still in the exploratory phase, although quite a lot of progress has been made over the last few years, partly driven by the Mock LISA Data Challenge [6–9, 58].

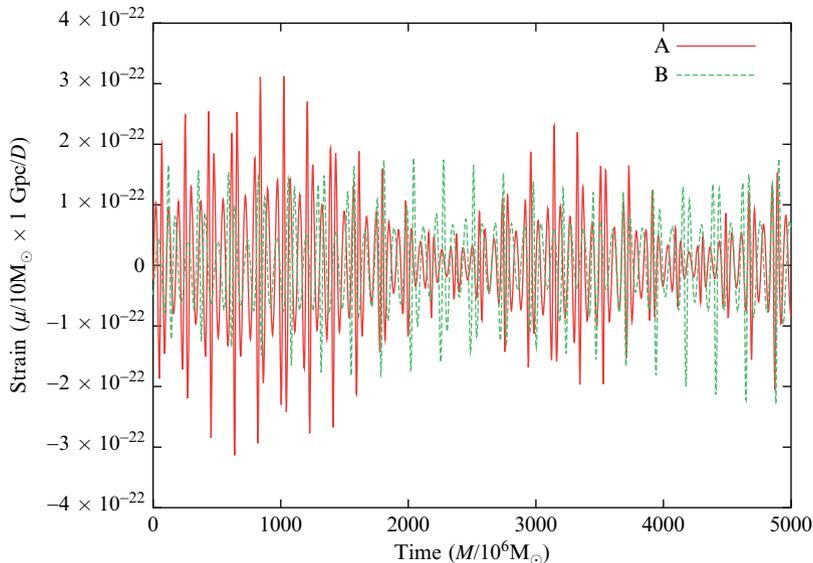


Fig. 5 Typical EMRI signals showing the complexity of the signals. Signals A (*red, solid*) and B (*green, dashed*) originate from identical systems, a $10 M_{\odot}$ black hole orbiting a $10^6 M_{\odot}$ central black hole, but differ in sky position and spin of the central black hole, given in the table below. The different initial conditions lead to a different time to plunge t_c , which for both systems lies far enough in the future to justify the use of analytic kludge waveforms. In the plot, time is given in units of the mass of the central object ($10^6 M_{\odot}$ equals about 5 s) and the strain in the ratio of the mass μ of the compact object to the luminosity distance D , normalized to 1 Gpc/ $10 M_{\odot}$. The signals are obtained through the `lisatools` software which is freely available at <http://code.google.com/p/lisatools/>

	Position		Spin		t_c (days)	
	Lat. ($^{\circ}$)	Long. ($^{\circ}$)	Mag. (S/M^2)	Polar ($^{\circ}$)		Azimuth ($^{\circ}$)
A	-15.4	77.6	0.52	95.1	8.1	384
B	-34.4	113.3	0.56	125.53	2.2	698

As the size of the template space will ultimately dominate the computational cost of the data analysis, a brief analysis of the expected size is helpful. Without any further assumptions, \mathcal{N} cycles of observed gravitational wave radiation and d parameters to be fit would lead to about \mathcal{N}^d templates for an optimal matched-filter search for EMRI waveforms. As noted before, EMRIs should emit about 10^5 cycles of radiation in the LISA band, resulting in 10^{5d} templates. Even for a small number of parameters as $d = 7$, this is an enormous number, exceeding the capabilities of any present computing architecture by many orders of magnitude.

Fortunately, the situation becomes a lot better, when the search is restricted to signals with a “high enough” SNR. For strong signals, hierarchical methods, such as ones that use crude templates [10,36,37], matching only short stretches of the signal,

or that sample the template space very coarsely, have been shown to work. The false alarm rate will be quite high for the initially estimated candidates, but drops off very quickly when follow-up searches around the candidates are performed.

For LISA, several detailed implementations of the first step of a hierarchical method have been investigated, such as *semi-coherent analysis* [35] and *time-frequency analysis* [37–39, 81].

Another approach is based on a stochastic search, usually conducted with some variant of the Metropolis–Hastings algorithm. Recent progress [11, 22] shows that stochastic searches are a promising method for tackling the problem of parameter estimation for EMRIs as well.

The actual performance of LISA with regard to the parameter estimation of EMRI depends to some degree on how well the signals from the galactic binaries can be identified, limiting the contamination of the EMRI signal by the signals of galactic binaries. The higher in frequency the EMRI are, the higher the “effective” SNR. Typical parameter accuracies for EMRI are very high [12]. Mass and spin of the central black hole, the mass of the inspiralling object, and the eccentricity of the orbit (at some time; eccentricity is a function of time) get estimated with a fractional accuracy of about 10^{-4} . The inclination $\cos(i)$ of the orbit as measured with respect to the spin of the central black hole can be determined to about 10^{-3} to 10^{-2} . The location of the EMRI in the sky can be constrained to a few square degrees and the luminosity distance to about 5%. The parameters for IMRI can be determined with even greater accuracy due to the higher mass of the IMRI that causes stronger gravitational waves.

5 Summary and Conclusions

Following [3], the science connected to EMRI and IMRI observations with LISA can be summarized in five categories: the characterization of the internal dynamics, the nature of the inspiralling objects, the characteristics of the central black hole, cosmology and early structure formation, and testing general relativity.

The characterization of the internal dynamics will yield information about the capture process itself, as different mechanisms lead to different eccentricities. Furthermore, different types of capture objects have different rates, lead to different orbits and different signal strengths, giving a handle on population studies. The characteristics of the central black hole can be assessed through the dependence of the signals on its spin and mass. EMRI observations can make a contribution to cosmology when used to establish an independent distance scale through either direct observation of the host galaxy or statistical methods whereas some information on structure formation and the growth of black holes can be obtained through the observation of EMRI rates. EMRI enable high-precision tests of general relativity, as they map out in detail the space-time around the central black hole.

EMRI pose the most difficult data analysis problem that LISA faces, because of the large number of potentially distinguishable signals and because of the challenges

in general relativity of computing reliable orbits over long durations. The numerical and conceptual problems of the data analysis are still a field of active research, however, the recent years have brought significant progress in algorithms and numerics so that the current situation allows to look forward in optimism regarding the remaining challenges.

References

1. F. Acernese, P. Amico, M. Al-Shourbagy, S. Aoudia, S. Avino, D. Babusci, G. Ballardin, F. Barone, L. Barsotti, M. Barsuglia, F. Beauville, M.A. Bizouard, C. Boccara, F. Bondu, L. Bosi, C. Bradaschia, S. Birindelli, S. Braccini, A. Brillet, V. Brisson, L. Brocco, D. Buskulic, E. Calloni, E. Campagna, F. Cavalier, R. Cavalieri, G. Cella, E. Chassande-Mottin, C. Corda, A.-C. Clapson, F. Cleva, J.-P. Coulon, E. Cuoco, V. Dattilo, M. Davier, R. De Rosa, L. Di Fiore, A. Di Virgilio, B. Dujardin, A. Eleuteri, D. Enard, I. Ferrante, F. Fidecaro, I. Fiori, R. Flaminio, J.-D. Fournier, O. Francois, S. Frasca, F. Frasconi, A. Freise, L. Gammaitoni, A. Gennai, A. Giazotto, G. Giordano, L. Giordano, R. Gouaty, D. Grosjean, G. Guidi, S. Hebri, H. Heitmann, P. Hello, L. Holloway, S. Karkar, S. Kreckelbergh, P. La Penna, N. Letendre, M. Lorenzini, V. Lorette, M. Loupias, G. Losurdo, J.M. Mackowski, E. Majorana, C.N. Man, M. Mantovani, F. Marchesoni, F. Marion, J. Marque, F. Martelli, A. Masserot, M. Mazzoni, L. Milano, C. Moins, J. Moreau, N. Morgado, B. Mours, A. Pai, C. Palomba, F. Paoletti, S. Pardi, A. Pasqualetti, R. Passaquieti, D. Passuello, B. Perniola, F. Piergiovanni, L. Pinard, R. Poggiani, M. Punturo, P. Puppò, K. Qipiani, P. Rapagnani, V. Reita, A. Remillieux, F. Ricci, I. Ricciardi, P. Ruggi, G. Russo, S. Solimeno, A. Spallicci, R. Stanga, R. Taddei, M. Tonelli, A. Toncelli, E. Tournefier, F. Travasso, G. Vajente, D. Verkindt, F. Vetrano, A. Viceré, J.-Y. Vinet, H. Vocca, M. Yvert, Z. Zhang, *Class. Q. Grav.* **23**, S63 (2006)
2. T. Alexander, C. Hopman, *Astrophys. J.* **590**, L29 (2003)
3. P. Amaro-Seoane, J.R. Gair, M. Freitag, M.C. Miller, I. Mandel, C.J. Cutler, S. Babak, *Class. Q. Grav.* **24**, 113 (2007)
4. M. Ando, TAMA Collabor. *Class. Q. Grav.* **19**, 1409 (2002)
5. J.W. Armstrong, F.B. Estabrook, M. Tinto, *Class. Q. Grav.* **20**, S283 (2003)
6. K.A. Arnaud, S. Babak, J.G. Baker, M.J. Benacquista, N.J. Cornish, C. Cutler, L.S. Finn, S.L. Larson, T. Littenberg, E.K. Porter, M. Vallisneri, A. Vecchio, J.Y. Vinet, T.M.L. Data Challenge Task Force, *Class. Q. Grav.* **24**, S551 (2007)
7. K.A. Arnaud, S. Babak, J.G. Baker, M.J. Benacquista, N.J. Cornish, C. Cutler, S.L. Larson, B.S. Sathyaprakash, M. Vallisneri, A. Vecchio, J.Y. Vinet, in *Laser Interferometer Space Antenna: 6th International LISA Symposium*, ed. by S.M. Merkowitz, J.C. Livas, AIP Conf. Proc. **873** (AIP, Melville, 2006) pp. 619–624
8. S. Babak, J.G. Baker, M.J. Benacquista, N.J. Cornish, J. Crowder, C. Cutler, S.L. Larson, T.B. Littenberg, E.K. Porter, M. Vallisneri, A. Vecchio, (the Mock LISA data challenge task force), G. Auger, L. Barack, A. Błaut, E. Bloomer, D.A. Brown, N. Christensen, J. Clark, S. Fairhurst, J.R. Gair, H. Halloin, M. Hendry, A. Jimenez, A. Królak, I. Mandel, C. Messenger, R. Meyer, S. Mohanty, R. Nayak, A. Petiteau, M. Pitkin, E. Plagnol, R. Prix, E.L. Robinson, C. Roever, P. Savov, A. Stroerer, J. Toher, J. Veitch, J.-Y. Vinet, L. Wen, J.T. Whelan, G. Woan, (the Challenge-2 participants), *Class. Q. Grav.* **25**, 114037 (2008)
9. S.L. Larson, E. Plagnol, E.K. Porter, M. Vallisneri, A. Vecchio (The Mock LISA Data Challenge Task Force), K. Arnaud, L. Barack, A. Błaut, C. Cutler, S. Fairhurst, J. Gair, X. Gong, I. Harry, D. Khurana, A. Królak, I. Mandel, R. Prix, B.S. Sathyaprakash, P. Savov, Y. Shang, M. Trias, J. Veitch, Y. Wang, L. Wen, J.T. Whelan (the Challenge-1B participants), *Class. Q. Grav.* **25**, 184026 (2008)
10. S. Babak, H. Fang, J.R. Gair, K. Glampedakis, S.A. Hughes, *Phys. Rev. D* **75**, 024005 (2007)
11. S. Babak, J.R. Gair, E.K. Porter, *Class. Q. Grav.* **26**, 135004 (2009)

12. L. Barack, C. Cutler, Phys. Rev. D **69**, 082005 (2004)
13. L. Barack, C. Cutler, Phys. Rev. D **70**, 122002 (2004)
14. L. Barack, C. Cutler, Phys. Rev. D **75**, 042003 (2007)
15. M.J. Benacquista, J. Phys. Conf. Ser. **32**, 147 (2006)
16. P. Bender, I. Ciufolini, K. Danzmann, W. Folkner, J. Hough, D. Robertson, A. Rüdiger, M. Sandford, R. Schilling, B. Schutz, R. Stebbins, T. Sumner, P. Touboul, S. Vitale, H. Ward, W. Winkler, J. Cornelisse, F. Hechler, Y. Jafry, R. Reinhard, *LISA - Laser Interferometer Space Antenna for the Detection and Observation of Gravitational Waves. Pre-Phase A Report*, MPQ 208 (Max Planck Institut für Quantenoptik, Garching, 1995), <http://www.mpg.de/cms/mpq/en/institute/service/library/reports/index.html>
17. P.L. Bender, D. Hils, Bull. Am. Astron. Soc. **30** (1988), <http://aas.org/archives/BAAS/v30n2/aas192/abs/S017009.html>
18. R. Bender, J. Kormendy, G. Bower, R. Green, J. Thomas, A.C. Danks, T. Gull, J.B. Hutchings, C.L. Joseph, M.E. Kaiser, T.R. Lauer, C.H. Nelson, D. Richstone, D. Weistrop, B. Woodgate, Astrophys. J. **631**, 280 (2005)
19. E. Berti, A. Buonanno, C.M. Will, Class. Q. Grav. **22**, S943 (2005)
20. B. Bertotti, in *Space Laser Applications and Technology*, Proc. ESA workshop, Les Diablerets 25–30 March 1984, ESA SP **202** (ESA, Noordwijk, 1984), p. 147
21. P.R. Brady, J.D.E. Creighton, K.S. Thorne, Phys. Rev. D **58**, 061501 (1998)
22. N.J. Cornish, arXiv e-prints (2008)
23. N.J. Cornish, R.W. Hellings, Class. Q. Grav. **20**, 4851 (2003)
24. K. Danzmann, Adv. Space Res. **32**, 1233 (2003)
25. C. Deffayet, K. Menou, Astrophys. J. **668**, L143 (2007)
26. J.A. de Freitas Pacheco, C. Filloux, T. Regimbau, Phys. Rev. D **74**, 023001 (2006)
27. S.V. Dhurandhar, J. Phys. Conf. S. **154**, 012047 (2009)
28. S.V. Dhurandhar, J.Y. Vinet, K.R. Nayak, Class. Q. Grav. **25**, 245002 (2008)
29. R. Di Stefano, J. Greiner, S. Murray, M. Garcia, Astrophys. J. **551**, L37 (2001)
30. J.E. Faller, P.L. Bender, J.L. Hall, D. Hils, M.A. Vincent, in *Kilometric Opt. Arrays in Space*, Coll. ESA, Cargèse 25–30 March 1984, ESA SP **226** (ESA, Noordwijk, 1985), p. 157
31. A.J. Farmer, E.S. Phinney, Mon. Not. R. Astron. Soc. **346**, 1197 (2003)
32. W.M. Folkner, F. Hechler, T.H. Sweetser, M.A. Vincent, P.L. Bender, Class. Q. Grav. **14**, 1405 (1997)
33. R.L. Forward, Phys. Rev. D **17**, 379 (1978)
34. M. Freitag, Astrophys. J. **583**, L21 (2003)
35. J.R. Gair, L. Barack, T. Creighton, C. Cutler, S.L. Larson, E.S. Phinney, M. Vallisneri, Class. Q. Grav. **21**, S1595 (2004)
36. J.R. Gair, K. Glampedakis, Phys. Rev. D **73**, 064037 (2006)
37. J. Gair, G. Jones, Class. Q. Grav. **24**, 1145 (2007)
38. J.R. Gair, I. Mandel, L. Wen, Class. Q. Grav. **25**, 184031 (2008)
39. J. Gair, L. Wen, Class. Q. Grav. **22**, S1359 (2005)
40. M.A. Gürkan, C. Hopman, Mon. Not. R. Astron. Soc. **379**, 1083 (2007)
41. D. Hils, P.L. Bender, R.F. Webbink, Astrophys. J. **360**, 75 (1990)
42. C.J. Hogan, in *Laser Interferometer Space Antenna: 6th International LISA Symposium*, ed. by S.M. Merkowitz, J.C. Livas, AIP Conf. Proc. **873** (AIP, Melville, 2006) pp. 30–40
43. D.E. Holz, S.A. Hughes, Astrophys. J. **629**, 15 (2005)
44. C. Hopman, T. Alexander, Astrophys. J. **645**, 1152 (2006)
45. J. Hough, J. Phys. Conf. Ser. **66**, 012002 (2007)
46. P.B. Ivanov, Mon. Not. R. Astron. Soc. **336**, 373 (2002)
47. O. Jennrich, arXiv e-prints (2009)
48. U.A. Johann, M. Ayre, P.F. Gath, W. Holota, P. Marenaci, H.R. Schulte, P. Weimer, D. Weise, J. Phys. Conf. S. **122**, 012005 (2008)
49. R.N. Lang, S.A. Hughes, Phys. Rev. D **74**, 122001 (2006)
50. R.N. Lang, S.A. Hughes, Astrophys. J. **677**, 1184 (2008)
51. S.L. Larson, W.A. Hiscock, R.W. Hellings, Phys. Rev. D **62**, 062001 (2000)

52. J.R. Lu, A.M. Ghez, S.D. Hornstein, M.R. Morris, E.E. Becklin, K. Matthews, *Astrophys. J.* **690**, 1463 (2009)
53. C.L. MacLeod, C.J. Hogan, *Phys. Rev. D* **77**, 043512 (2008)
54. K. Menou, Z. Haiman, B. Kocsis, *New Astron. Rev.* **51**, 884 (2008)
55. M.C. Miller, *Astrophys. J.* **618**, 426 (2005)
56. M.C. Miller, M. Freitag, D.P. Hamilton, V.M. Lauburg, *Astrophys. J.* **631**, L117 (2005)
57. C.W. Misner, K.S. Thorne, J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973)
58. Mock LISA Data Challenges (2006)
59. K.R. Nayak, J.Y. Vinet, *Phys. Rev. D* **70**, 102003 (2004)
60. S. Nayakshin, R. Sunyaev, *Mon. Not. R. Astron. Soc.* **364**, L23 (2005)
61. G. Nelemans, L.R. Yungelson, S.F. Portegies Zwart, F. Verbunt, *Astron. Astrophys.* **365**, 491 (2001)
62. T. Paumard, R. Genzel, F. Martins, S. Nayakshin, A.M. Beloborodov, Y. Levin, S. Trippe, F. Eisenhauer, T. Ott, S. Gillessen, R. Abuter, J. Cuadra, T. Alexander, A. Sternberg, *Astrophys. J.* **643**, 1011 (2006)
63. H.B. Perets, C. Hopman, T. Alexander, *Astrophys. J.* **656**, 709 (2007)
64. E. Poisson, *Class. Q. Grav.* **21**, 153 (2004)
65. E. Poisson, *Living Rev. Rel.* **7**, URL: <http://www.livingreviews.org/lrr-2004-6>
66. N.A. Robertson, *Class. Q. Grav.* **17**, R19 (2000)
67. F.D. Ryan, *Phys. Rev. D* **52**, 5707 (1995)
68. H.R. Schulte, P.F. Gath, M. Herz, **873**, 379 (2006)
69. D.A. Shaddock, B. Ware, R.E. Spero, M. Vallisneri, *Phys. Rev. D* **70**, 081101 (2004)
70. D. Shoemaker, *Class. Q. Grav.* **20**, 11 (2003)
71. D. Sigg, *LIGO Sci. Collabor.* **23**, S51 (2006)
72. S. Sigurdsson, M.J. Rees, *Mon. Not. R. Astron. Soc.* **284**, 318 (1997)
73. L.E. Simone, E. Poisson, C.M. Will, *Phys. Rev. D* **52**, 4481 (1995)
74. R.T. Stebbins, *Class. Q. Grav.* **26**, 094014 (2009)
75. R. Stebbins, S.M. Merkowitz, J.C. Livas, in *Laser Interferometer Space Antenna: 6th International LISA Symposium*, ed. by S.M. Merkowitz, J.C. Livas, AIP Conf. Proc. **873** (AIP, Melville, 2006) pp. 3–12
76. M. Tinto, S.V. Dhurandhar, *Living Rev. Rel.* **8**, URL: <http://www.livingreviews.org/lrr-2005-4>
77. M. Tinto, F.B. Estabrook, J.W. Armstrong, *Phys. Rev. D* **69**, 082001 (2004)
78. M. Tinto, S.L. Larson, *Class. Q. Grav.* **22**, S531 (2005)
79. M. Tinto, M. Vallisneri, J.W. Armstrong, *Phys. Rev. D* **71**, 041101 (2005)
80. M. Vallisneri, *Phys. Rev. D* **71**, 022001 (2005)
81. L.Q. Wen, J.R. Gair, *Class. Q. Grav.* **22**, S445 (2005)
82. B. Willke, *LIGO Sci. Collabor. Class. Q. Grav.* **24**, S389 (2007)

Motion in Alternative Theories of Gravity

Gilles Esposito-Farèse

Abstract Although general relativity (GR) passes all present experimental tests with flying colors, it remains important to study alternative theories of gravity for several theoretical and phenomenological reasons that we recall in these lecture notes. The various possible ways of modifying GR are presented, and we notably show that the motion of massive bodies may be changed even if one assumes that matter is minimally coupled to the metric as in GR. This is illustrated with the particular case of scalar-tensor theories of gravity, whose Fokker action is discussed, and we also mention the consequences of the no-hair theorem on the motion of black holes. The finite size of the bodies modifies their motion with respect to pointlike particles, and we give a simple argument showing that the corresponding effects are generically much larger in alternative theories than in GR. We also discuss possible modifications of Newtonian dynamics (MOND) at large distances, which have been proposed to avoid the dark matter hypothesis. We underline that all the previous classes of alternatives to GR may a priori be used to predict such a phenomenology, but that they generically involve several theoretical and experimental difficulties.

1 Introduction

Since general relativity (GR) is superbly consistent with all precision experimental tests – as we will see below several examples, one may naturally ask the question: Why should we consider alternative theories of gravity? The reason is actually three-fold. First, it is quite instructive to contrast GR's predictions with those of alternative models in order to understand better which features of the theory have been experimentally tested, and what new observations may allow us to test the remaining features [36]. Second, theoretical attempts at quantizing gravity or unifying it with other interactions generically predict the existence of partners to the graviton, that

G. Esposito-Farèse (✉)

IRéC, Institut d'Astrophysique de Paris, UMR 7095-CNRS, Université Pierre et Marie Curie-Paris 6, 98bis boulevard Arago, F-75014 Paris, France

e-mail: gef@iap.fr

is, extra fields contributing to the gravitational force. This is notably the case in all extra-dimensional (Kaluza–Klein) theories, where the components g_{ab} of the metric tensor [where a and b belong to the $D - 4$ extra dimensions] behave as $(D - 4)(D - 3)/2$ scalar fields (called moduli) in four dimensions. Independently of such moduli, supersymmetry (notably needed in string theory) also implies the existence of several fields in the graviton supermultiplet, in particular another scalar called the dilaton. The third reason why it remains important to study alternative theories of gravity is the existence of several puzzling experimental issues. Cosmological data are notably consistent with a Universe filled with about 72% of “dark energy” (a fluid with negative pressure opposite to its energy density) and 24% of “dark matter” (a fluid with negligible pressure and vanishingly small interaction with ordinary matter and itself) [101, 102]. Another strange phenomenon is the anomalous extra acceleration toward the Sun that the two Pioneer spacecrafts have undergone beyond 30 astronomical units [4, 85, 107] (see also [74]). Although such issues do not threaten directly GR itself, since they may be explained by the existence of unknown “dark” fluids (or by yet unmodeled sources of noise in the case of the Pioneer anomaly), they may nevertheless be a hint that something needs to be changed in the gravitational law at large distances.

Since many theoretical and experimental physicists devised their own gravity models, the field of alternative theories is much too wide for the present lecture notes. Detailed reviews may be found in Refs. [52, 108, 113, 114]. We shall focus here on the particular case of scalar-tensor theories and some of their generalizations. Before introducing them, let us recall that GR is based on two independent hypotheses, which can be most conveniently described by writing its action

$$S = \underbrace{\frac{c^3}{16\pi G} \int d^4x \sqrt{-g} R}_{\text{Einstein–Hilbert}} + \underbrace{S_{\text{matter}}[\text{matter}, g_{\mu\nu}]}_{\text{metric coupling}}, \quad (1)$$

where g denotes the determinant of the metric $g_{\mu\nu}$, R its scalar curvature, and we use the sign conventions of Ref. [82], notably the mostly plus signature. The first assumption of GR is that matter fields are universally and minimally coupled to one single metric tensor $g_{\mu\nu}$. This ensures the “Einstein equivalence principle,” whose consequences will be summarized in Section 2. The second hypothesis of GR is that this metric $g_{\mu\nu}$ propagates as a pure spin-2 field, i.e., that its kinetic term is given by the Einstein–Hilbert action. The core of the present lecture notes, Sections 3–6, will be devoted to the observational consequences of other possible kinetic terms.

2 Modifying the Matter Action

In the above action (1), square brackets in $S_{\text{matter}}[\text{matter}, g_{\mu\nu}]$ mean a functional dependence on the fields, that is, it also depends on their first derivatives. For instance, the action of a point particle,

$$S_{\text{point particle}} = - \int mc ds = - \int mc \sqrt{-g_{\mu\nu}(x) v^\mu v^\nu} dt, \quad (2)$$

depends not only on its spacetime position x but also on its 4-velocity $v^\mu \equiv dx^\mu/dt$. Since the matter action defines the *motion* of matter in a given metric $g_{\mu\nu}$, it is a priori what needs to be modified with respect to GR in order to predict different trajectories. This idea has been studied in depth by Milgrom in [80,81], where he assumed that the action of a point particle could also depend on its acceleration and even higher time derivatives: $S_{\text{pp}}(\mathbf{x}, \mathbf{v}, \mathbf{a}, \dot{\mathbf{a}}, \dots)$. However, any modification with respect to action (2) is tightly constrained experimentally for usual accelerations, notably by high-precision tests of special relativity. On the other hand, physics may happen to differ for tiny accelerations, much smaller than the Earth's gravitational attraction. In such a case, the mathematical consistency of the theory may be invoked to restrict the space of allowed theories. A theorem derived by Ostrogradski in 1850 [88,117] shows notably that the Hamiltonian is generically unbounded from below if $S_{\text{pp}}(\mathbf{x}, \mathbf{v}, \mathbf{a}, \dots, d^n \mathbf{x}/dt^n)$ depends on a *finite* number of time derivatives, and therefore that the theory is unstable. A possible solution would thus be to consider *nonlocal* theories, depending on an infinite number of time derivatives. This is actually what Milgrom found to be necessary in order to recover the Newtonian limit and satisfy Galileo invariance. Although nonlocal theories are worth studying, and are actually obtained as effective models of string theory, their phenomenology is quite difficult to analyze, and we will not consider them any longer in the present lecture notes. General discussions and specific models may be found for instance in [49, 50, 80, 99, 100].

Another possible modification of the matter action $S_{\text{matter}}[\text{matter}, g_{\mu\nu}]$ is actually predicted by string theory: Different matter fields are coupled to different metric tensors, and the action takes thus the form $S_{\text{matter}}[\text{matter}^{(i)}, g_{\mu\nu}^{(i)}]$. In other words, two different bodies a priori do not feel the same geometry, and their accelerations may differ both in norm and direction. However, the universality of free fall is extremely well tested experimentally, as well as the three other observational consequences of a metric coupling $S_{\text{matter}}[\text{matter}, g_{\mu\nu}]$, that we will recall below. The conclusion is that string theory must actually show that the different metrics $g_{\mu\nu}^{(i)}$ are almost equal to each other. One possible reason is that their differences may be mediated by massive fields, and would become thus exponentially small at large enough distances. But even in presence of massless fields contributing to the difference between the various $g_{\mu\nu}^{(i)}$, a generic mechanism has been shown to attract the theory toward GR during the cosmological expansion of the Universe [45, 47, 48].

Let us now recall the four observational consequences of a metric coupling $S_{\text{matter}}[\text{matter}, g_{\mu\nu}]$, as well as their best experimental verifications. If all matter fields feel the same metric $g_{\mu\nu}$, it is possible to define a ‘‘Fermi coordinate system’’ along any worldline, such that the metric takes the diagonal form $\text{diag}(-1, 1, 1, 1)$ and its first derivatives vanish. In other words, up to small tidal effects proportional to the spatial distance to the worldline, everything behaves as in special relativity. This is the mathematically well-defined notion of a freely falling elevator. The effacement of gravity in this coordinate system implies that (i) all coupling constants

and mass scales of the Standard Model of particle physics are indeed space and time independent. One of the best experimental confirmations is the time independence of the fine-structure constant, $|\dot{\alpha}/\alpha| < 7 \times 10^{-17} \text{year}^{-1}$, six orders of magnitude smaller than the inverse age of the Universe [38, 59, 98]. A second consequence of the validity of special relativity within the freely falling elevator is that (ii) local (nongravitational) experiments are Lorentz invariant. The isotropy of space has notably been tested at the 10^{-27} level in [34, 75, 89]. The third consequence of a metric coupling is the very existence of this freely falling elevator where gravity is effaced, that is, (iii) the universality of free fall: (non self-gravitating) bodies fall with the same acceleration in an external gravitational field. This has been tested at a few parts in 10^{13} both in laboratory experiments [2, 9], and by studying the relative acceleration of the Earth and the Moon toward the Sun [116]. The fourth consequence of a metric coupling is (iv) the universality of gravitational redshift. It may be understood intuitively by invoking the equivalence between the physics in a gravitational field and within an accelerated rocket: The classical Doppler effect suffices to show that clocks at the two ends of the rocket do not tick at the same rate, and one can immediately deduce that lower clocks are slower in a gravitational field. More precisely, a metric coupling implies that in a static Newtonian potential $g_{00} = -1 + 2U(\mathbf{x})/c^2 + \mathcal{O}(1/c^4)$, the proper times measured by two clocks is such that $\tau_1/\tau_2 = 1 + [U(\mathbf{x}_1) - U(\mathbf{x}_2)]/c^2 + \mathcal{O}(1/c^4)$. This has been tested at the 2×10^{-4} level 30 years ago by flying a hydrogen maser clock [109, 110], and the planned Pharo/Aces mission [92] should increase the precision by two orders of magnitude.

In conclusion, the four consequences of a metric coupling have been very well tested experimentally, notably the universality of free fall (i.e., the relative *motion of massive bodies* in a gravitational field). Therefore, although theoretical considerations let us expect that the Einstein equivalence principle is violated at a fundamental level, we do know that deviations from GR are beyond present experimental accuracy. In the following, we will thus restrict our discussion to theories that satisfy exactly this principle, that is, which assume the matter action takes the form $S_{\text{matter}}[\text{matter}, g_{\mu\nu}]$. On the other hand, we will now assume that the kinetic term of the gravitational field, say S_{gravity} , is not necessarily given by the Einstein–Hilbert action of Eq. 1.

3 Modified Motion in Metric Theories?

For a given background metric $g_{\mu\nu}$, the kinetic term S_{gravity} defines how gravitational waves propagate, and the matter action S_{matter} how massive bodies move in spacetime. If we assume a universal metric coupling $S_{\text{matter}}[\text{matter}, g_{\mu\nu}]$ as in GR, we are thus tempted to conclude that the motion of matter must be strictly the same as in GR, and that the present lecture notes should stop here. However, S_{gravity} also defines how $g_{\mu\nu}$ is *generated* by the matter distribution. Therefore, the motion of massive bodies within this metric does actually depend directly on the dynamics of gravity!

The clearest way to illustrate this conclusion is to integrate away the metric tensor, that is, to replace it in terms of its material sources, in order to construct the so-called Fokker action. We give below a schematic derivation of its expression, taken from [41]. Gauge-fixing subtleties are discussed notably in Appendix C of [46]. We start from an action of the form

$$S = S_\Phi[\Phi] + S_{\text{matter}}[\sigma, \Phi], \tag{3}$$

where Φ denotes globally all fields participating in the gravitational interaction, and σ denotes the matter sources. We also denote as $\bar{\Phi}[\sigma]$ a solution of the field equation $\delta S/\delta\Phi = 0$ for given sources σ . Let us now define the Fokker action

$$S_{\text{Fokker}}[\sigma] \equiv S_\Phi[\bar{\Phi}[\sigma]] + S_{\text{matter}}[\sigma, \bar{\Phi}[\sigma]], \tag{4}$$

and show that it gives the correct equations of motion for matter σ . Indeed, its variational derivative reads

$$\frac{\delta S_{\text{Fokker}}[\sigma]}{\delta\sigma} = \left(\frac{\delta S[\sigma, \Phi]}{\delta\sigma}\right)_{\Phi=\bar{\Phi}[\sigma]} + \left(\frac{\delta S[\sigma, \Phi]}{\delta\Phi}\right)_{\Phi=\bar{\Phi}[\sigma]} \frac{\delta\bar{\Phi}[\sigma]}{\delta\sigma}, \tag{5}$$

where the second term of the right-hand side vanishes because $\bar{\Phi}[\sigma]$ has been chosen as a solution of $\delta S/\delta\Phi = 0$. Therefore, $\delta S_{\text{Fokker}}[\sigma]/\delta\sigma = 0$ does yield the correct equations of motion $\delta S[\sigma, \Phi]/\delta\sigma = 0$ for matter within the background $\Phi = \bar{\Phi}[\sigma]$ it consistently generates. The most important point to notice here is that the Fokker action (4) is *not* simply given by the matter action $S_{\text{matter}}[\sigma, \Phi]$, computed in the consistent background $\Phi = \bar{\Phi}[\sigma]$. Not only the σ -dependence of this background must be taken into account when varying the Fokker action, but its definition (4) also depends crucially on the kinetic term (and the nonlinear dynamics) of the field, $S_\Phi[\bar{\Phi}[\sigma]]$.

To illustrate more vividly that the motion of massive bodies does depend on the dynamics of the gravitational field(s), let us give a diagrammatic representation of the above formal definition (4) of the Fokker action. We first introduce some symbols in Fig. 1, notably white blobs for matter sources and straight lines for field propagators. Using this notation, the original action (3) may be translated as in Fig. 2, which actually *defines* the various vertices. In this figure, the numerical factors have been chosen to simplify the field equation satisfied by $\bar{\Phi}[\sigma]$, which takes the diagrammatic form of Fig. 3. This figure tells us how to replace any black blob

- $\sigma =$ ○ Material sources
- $\square\Phi =$ ● Left-hand side of the field equations
- $\square^{-1} =$ — Propagator of the fields (Green function)
- $\Phi =$ —● Fields

Fig. 1 Diagrammatic representation of matter sources, fields, and their propagator

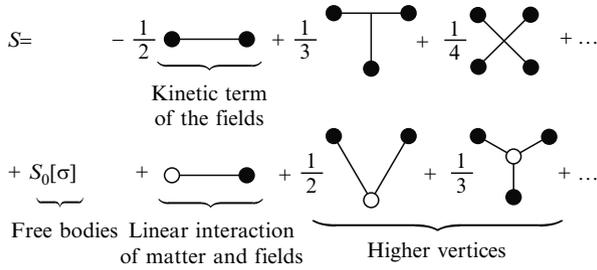


Fig. 2 Diagrammatic representation of the full action (3) of the theory, expanded in powers of Φ (black blobs). The first line corresponds to the field action $S_\phi[\Phi]$, and the second one to the matter action $S_{\text{matter}}[\sigma, \Phi]$ (describing notably the matter–field interaction)

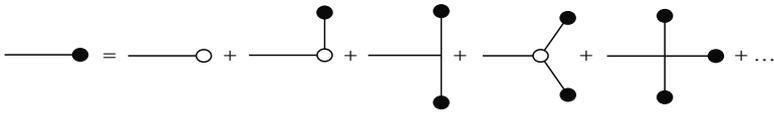


Fig. 3 Diagrammatic representation of the field equation $\delta S/\delta\Phi = 0$ satisfied by $\bar{\Phi}[\sigma]$

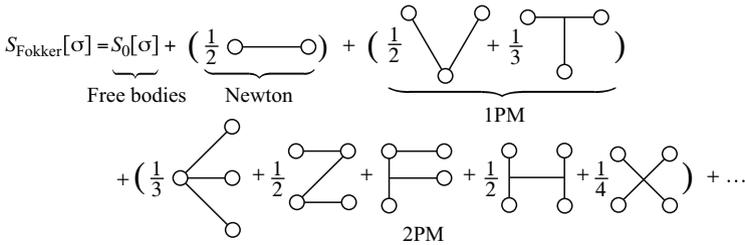


Fig. 4 Diagrammatic representation of the Fokker action (4), which depends only on matter sources σ (white blobs). The dumbbell diagram labeled “Newton” represents the Newtonian interaction $\propto G$, together with all velocity-dependent relativistic corrections. The 3-blob diagrams labelled “1PM” represent first post-Minkowskian corrections, that is, the $\mathcal{O}(G^2)$ post-Newtonian terms as well as their full velocity dependence. The 4-blob diagrams labeled “2PM” represent second post-Minkowskian corrections $\propto G^3$

(field Φ) by a white blob (source σ) plus higher corrections, in which one can again replace iteratively black blobs by white ones plus corrections. The Fokker action (4) is thus simply obtained by eliminating in such a way black blobs from Fig. 2, and the result is displayed in Fig. 4. This figure clearly shows that the dynamics of the field (i.e., the first line of Fig. 2) does contribute to that of massive bodies. Indeed, if it had not been taken into account, the Newtonian interaction would have been twice too large, no “T” diagram would have appeared in Fig. 4, and the numerical coefficients of all other diagrams would have also changed.

For any theory of gravity, whose field dynamics is imposed by S_{gravity} , one may now explicitly compute the diagrams entering Fig. 4. Of course, any gauge invariance must be fixed in order to define the field propagator as the inverse of the quadratic kinetic term (first dumbbell diagram of Fig. 2). In the case of GR, one may for instance fix the harmonic gauge, and the first line of Fig. 4 translates¹ as the well-known Einstein–Infeld–Hoffmann action describing the interaction of several massive bodies labelled A, B, \dots :

$$\begin{aligned}
 S_{\text{Fokker}} = & - \sum_A \int dt m_A c^2 \sqrt{1 - \mathbf{v}_A^2/c^2} \\
 & + \frac{1}{2} \sum_{A \neq B} \int dt \frac{G m_A m_B}{r_{AB}} \left[1 + \frac{1}{2c^2} (\mathbf{v}_A^2 + \mathbf{v}_B^2) - \frac{3}{2c^2} (\mathbf{v}_A \cdot \mathbf{v}_B) \right. \\
 & \left. - \frac{1}{2c^2} (\mathbf{n}_{AB} \cdot \mathbf{v}_A)(\mathbf{n}_{AB} \cdot \mathbf{v}_B) + \frac{\gamma^{\text{PPN}}}{c^2} (\mathbf{v}_A - \mathbf{v}_B)^2 \right] \\
 & - \frac{1}{2} \sum_{B \neq A \neq C} \int dt \frac{G^2 m_A m_B m_C}{r_{AB} r_{AC} c^2} (2\beta^{\text{PPN}} - 1) + \mathcal{O}\left(\frac{1}{c^4}\right). \tag{6}
 \end{aligned}$$

Here r_{AB} denotes the (instantaneous) distance between bodies A and B , \mathbf{n}_{AB} is the unit 3-vector pointing from B to A , \mathbf{v}_A is the 3-velocity of body A , and a sum over $B \neq A \neq C$ allows B and C to be the same body. The first line of Eq. 6, noted $S_0[\sigma]$ in Figs. 2 and 4, merely describes free bodies in special relativity. The second line of Eq. 6 describes the 2-body interaction, that is, the dumbbell diagram of Fig. 4 that we labeled “Newton.” Its lowest-order term is indeed the Newtonian gravitational potential, and we have also displayed its first post-Newtonian (1PN) corrections, of order $\mathcal{O}(v^2/c^2)$. Finally, the last line of Eq. 6 corresponds to the “V” and “T” diagrams labeled “1PM” in Fig. 4, computed here at their lowest (1PN) order.

The two coefficients β^{PPN} and γ^{PPN} entering Eq. 6 are simply equal to unity in GR. They were introduced by Eddington [51] to describe phenomenologically other possible theories of gravity, although he did not have any specific model in mind. It happens that the most natural alternatives to GR, scalar-tensor theories (that we will introduce in Section 4 below), do predict different values for these two parameters. This comes from the fact that massive bodies can exchange scalar particles in addition to the usual gravitons of GR. If we represent gravitons as curly lines and scalar fields as straight lines, the diagrams contributing to the Fokker action (6) are indeed displayed in Fig. 5. The four diagrams involving at least one scalar line contribute to change the values of the Eddington parameters β^{PPN} and γ^{PPN} . We will give their explicit values in Eq. 9, but we refer to Ref. [41] for their derivation from diagrammatic calculations.

Besides β^{PPN} and γ^{PPN} , many other parameters may actually be introduced to describe the most general behavior of massive bodies at the 1PN order. Under reasonable assumptions, notably that the matter action takes the metric form

¹ One needs to compute the integrals represented by the various diagrams to derive expression (6). See Ref. [41] for explicit diagrammatic calculations.

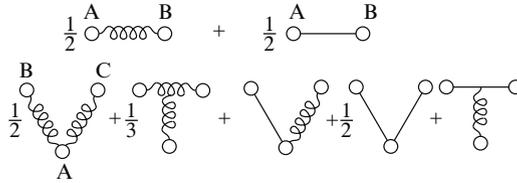


Fig. 5 Diagrams contributing to the N -body action (6) at 1PN order, in scalar-tensor theories of gravity. Graviton and scalar exchanges are represented, respectively, as curly and straight lines

$S_{\text{matter}}[\text{matter}, g_{\mu\nu}]$ and that the gravitational interaction does not involve any specific length scale, Nordtvedt and Will [115] showed that 8 extra parameters are a priori possible, in addition to Eddington’s β^{PPN} and γ^{PPN} . However, these 8 parameters vanish both in GR and in scalar-tensor gravity, therefore we will not introduce them in these lecture notes. A detailed presentation is available in the book [113].

4 Scalar-Tensor Theories of Gravity

Among alternative theories of gravity, those which involve scalar partners to the graviton are privileged for several reasons. Not only their existence is predicted in all extra-dimensional theories, but they also play a crucial role in modern cosmology (in particular during the accelerated expansion phases of the Universe). They are above all consistent field theories, with a well-posed Cauchy problem, and they respect most of GR’s symmetries (notably conservation laws and local Lorentz invariance even if a subsystem is influenced by external masses). To simplify the discussion, we will focus on models involving a single scalar field, although the study of tensor-multi-scalar theories can also be done in great detail [39]. We will thus consider the class of theories defined by the action [19, 86, 111]

$$S = \frac{c^3}{16\pi G_*} \int d^4x \sqrt{-g^*} (R^* - 2g_*^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi) + S_{\text{matter}}[\text{matter}; g_{\mu\nu} \equiv A^2(\varphi)g_{\mu\nu}^*]. \tag{7}$$

A potential $V(\varphi)$ may also be considered in this action, and is actually crucial in cosmology, but we will study here solar-system-size effects and assume that the scalar-field mass (and other self-interactions described by $V(\varphi)$) is small enough to be negligible at this scale. The physical metric $g_{\mu\nu}$, to which matter is universally coupled (and which defines thus the lengths and times measured by material rods and clocks), is the product of the Einstein metric $g_{\mu\nu}^*$ (whose kinetic term is the Einstein–Hilbert action) and a function $A^2(\varphi)$ characterizing how matter is coupled to the scalar field. It will be convenient to expand it around the background value φ_0 of the scalar field far from any massive body, as

$$\ln A(\varphi) = \ln A(\varphi_0) + \alpha_0(\varphi - \varphi_0) + \frac{1}{2}\beta_0(\varphi - \varphi_0)^2 + \mathcal{O}(\varphi - \varphi_0)^3, \tag{8}$$

where α_0 defines the linear coupling constant of matter to scalar excitations, β_0 its quadratic coupling to two scalar lines, etc.

4.1 Weak-Field Predictions

Newtonian and 1PN predictions depend only on these first two coupling constants, α_0 and β_0 . For instance, the effective gravitational constant between two bodies is not given by the bare constant G_* entering action (7), but by $G = G_*(1 + \alpha_0^2)$, in which a contribution G_* comes from the exchange of a (spin-2) graviton whereas $G_*\alpha_0^2$ is due to the exchange of a (spin-0) scalar field, each matter-scalar vertex bringing a factor α_0 . The first line of Fig. 5 gives a diagrammatic illustration of this sum. [Actually, the value of a gravitational constant depends on the chosen units, and the expression $G = G_*(1 + \alpha_0^2)$ corresponds to the ‘‘Einstein-frame’’ representation used to write action (7). An extra factor $A_0^2 = A(\varphi_0)^2$ enters when using the physical metric $g_{\mu\nu} = A^2(\varphi)g_{\mu\nu}^*$ to define observable quantities, and the actual gravitational constant that is measured reads $G_*A_0^2(1 + \alpha_0^2)$. No such extra factors A_0 enter the computation of dimensionless observable quantities, like the Eddington parameters β^{PPN} and γ^{PPN} .] Two kinds of 1PN corrections enter the Fokker action (6): velocity-dependent terms in the 2-body interaction (first line of Fig. 5), which involve the parameter γ^{PPN} , and the lowest-order 3-body interactions (second line of Fig. 5), which involve β^{PPN} . Diagrammatic calculations [43] or more standard techniques [39, 113] can be used to compute their expressions in scalar-tensor theories:

$$\gamma^{\text{PPN}} = 1 - \frac{2\alpha_0^2}{1 + \alpha_0^2}, \quad \beta^{\text{PPN}} = 1 + \frac{1}{2} \frac{\alpha_0\beta_0\alpha_0}{(1 + \alpha_0^2)^2}. \tag{9}$$

Here again, the factor α_0^2 comes from the exchange of a scalar particle between two bodies, whereas $\alpha_0\beta_0\alpha_0$ comes from a scalar exchange between three bodies (cf. the purely scalar ‘‘V’’ diagram of Fig. 5).

Several solar-system observations tightly constrain these 1PN parameters to be close to 1, i.e., their general relativistic values. The main ones are Mercury’s perihelion advance [96], Lunar Laser Ranging (which allows us to test the so-called Nordtvedt effect, i.e., whether there is a difference between the Earth’s and the Moon’s accelerations toward the Sun) [116], and experiments involving the propagation of light in the curved spacetime of the solar system (by order of increasing accuracy: radar echo delay between the Earth and Mars, light deflection measured by Very Long Baseline Interferometry over the whole celestial sphere [97], and time-delay variation to the Cassini spacecraft near solar conjunction [20]). These 1PN constraints are summarized in Fig. 6, and the conclusion is that GR is basically the only theory consistent with weak-field experiments. However, when translated in terms of the linear and quadratic coupling constants α_0 and β_0 of matter to the scalar field, the same solar-system constraints take the shape of Fig. 7. Therefore, the linear coupling constant $|\alpha_0|$ must be smaller than 3×10^{-3} , but we do not have any significant constraint on β_0 [nor any higher-order vertex entering expansion (8)].

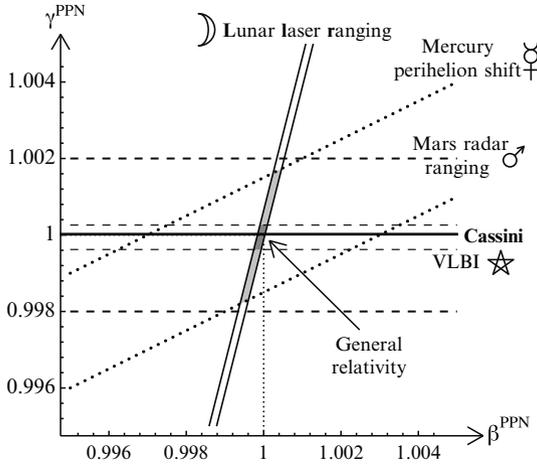


Fig. 6 Solar-system constraints on the post-Newtonian parameters β^{PPN} and γ^{PPN} . The allowed region is the tiny intersection of the “Lunar Laser Ranging” strip with the horizontal bold line labeled “Cassini.” General relativity, corresponding to $\beta^{\text{PPN}} = \gamma^{\text{PPN}} = 1$, is consistent with all tests

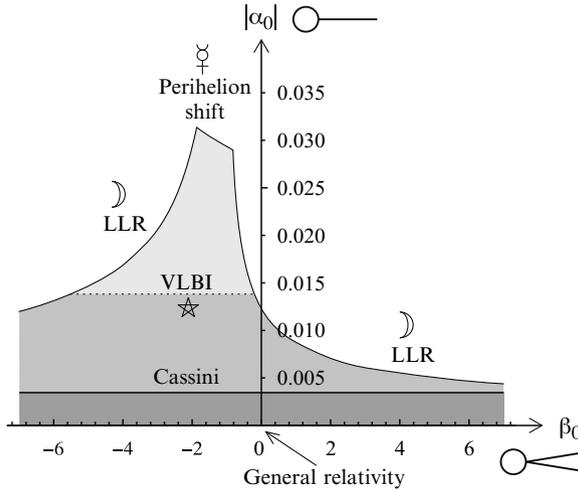


Fig. 7 Solar-system constraints on the matter-scalar coupling function $\ln A(\varphi)/A_0 = \alpha_0(\varphi - \varphi_0) + \frac{1}{2}\beta_0(\varphi - \varphi_0)^2 + \dots$. The allowed region is the dark grey horizontal strip. The vertical axis ($\beta_0 = 0$) corresponds to Brans–Dicke theory [27,56,71] with a parameter $2\omega_{\text{BD}} + 3 = 1/\alpha_0^2$. The horizontal axis ($\alpha_0 = 0$) corresponds to theories which are perturbatively equivalent to GR, that is, which predict strictly no deviation from it (at any order $1/c^n$) in the weak-field conditions of the solar system

4.2 Strong-Field Predictions

A qualitatively different class of constraints is obtained by studying scalar-tensor theories in the strong-field regime, that is, near compact bodies whose radius R is not extremely large with respect to their Schwarzschild radius $2Gm/c^2$. This is notably the case when considering neutron stars, whose ratio $Gm/Rc^2 \sim 0.2$ is not far from the theoretical maximum of 0.5 for black holes. Although the metric is very significantly different from the flat one inside such compact bodies and in their immediate vicinity, their orbital velocity in a binary system may nevertheless be small enough to perform a consistent expansion in powers of $v^2/c^2 \sim Gm/rc^2$, where r denotes the interbody distance (as opposed to their radius R). This is usually called a post-Keplerian expansion.² In such a case, one can show [39, 113] that the predictions of scalar-tensor theories are similar to those of weak-field conditions, with the only difference that the matter-scalar coupling constant α_0 and β_0 are replaced by body-dependent quantities, say α_A and β_A for a body labeled A . For instance, the effective gravitational constant describing the lowest-order attraction between two compact bodies A and B reads now $G_{AB} = G_*(1 + \alpha_A\alpha_B)$, instead of the weak-field expression $G = G_*(1 + \alpha_0^2)$ mentioned above. Similarly, the 1PN parameters β^{PPN} and γ^{PPN} are replaced by body-dependent ones β_{BC}^A and γ_{AB} , taking the same forms as in Eq. 9 but where α_0^2 is replaced by $\alpha_A\alpha_B$, and $\alpha_0\beta_0\alpha_0$ by $\alpha_B\beta_A\alpha_C$ (see Ref. [39] for precise expressions). All post-Keplerian effects can thus be derived straightforwardly, in a similar way as in the solar system. In addition to these predictions, one may also compute the energy loss due to the emission of gravitational waves by a binary system. It takes the schematic form³

$$\begin{aligned} \text{Energy flux} = & \left\{ \frac{\text{Quadrupole}}{c^5} + \mathcal{O}\left(\frac{1}{c^7}\right) \right\}_{\text{spin } 2} \\ & + \left\{ \frac{\text{Monopole}}{c} \left(0 + \frac{1}{c^2}\right)^2 + \frac{\text{Dipole}}{c^3} + \frac{\text{Quadrupole}}{c^5} + \mathcal{O}\left(\frac{1}{c^7}\right) \right\}_{\text{spin } 0}, \end{aligned} \quad (10)$$

where the first line comes from the emission of usual (spin-2) gravitons, and the second one from the emission of scalar (spin-0) waves. Note that the dipolar term is of order $\mathcal{O}(1/c^3)$, generically much larger than the standard $\mathcal{O}(1/c^5)$ quadrupole of GR. As expected for a dipole, it vanishes when considering a perfectly

² Two different (though related) meanings of “post-Keplerian” exist in the literature. We are here considering a post-Keplerian *expansion* in powers of v_{orbital}^2/c^2 , while keeping the full nonperturbative dependence in the gravitational self-energy Gm/Rc^2 . On the other hand, post-Keplerian *deviations* mean relativistic effects modifying the lowest-order Keplerian motion, like those described in Section 4.3. Only this latter meaning is used in GR, because its strong equivalence principle implies that the internal structure of a body does not influence its motion up to order $\mathcal{O}(1/c^{10})$, as recalled in Section 5.

³ The precise definitions of these multipoles and their explicit expressions may be found for instance in Section 6 of Ref. [39].

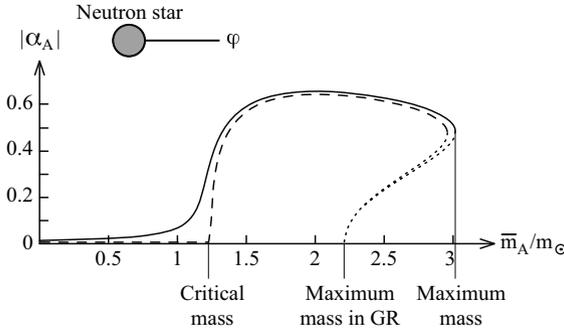


Fig. 8 Scalar charge α_A of a neutron star versus its baryonic mass \bar{m}_A , for the model $A(\varphi) = \exp(-3\varphi^2)$, that is, $\beta_0 = -6$. The solid line corresponds to a small value of α_0 (namely, the VLBI bound of Fig. 7), and the dashed line to $\alpha_0 = 0$. The dotted lines correspond to unstable configurations of the star

symmetrical binary system, because there is no longer any privileged spatial orientation. Its precise calculation [39, 113] shows indeed that it involves the *difference* of the scalar charges of the two bodies, and is actually proportional to $(\alpha_A - \alpha_B)^2$. Although the monopolar term is a priori of the even larger order $\mathcal{O}(1/c)$ for bodies which are not at equilibrium (e.g., collapsing or exploding stars), it reduces to order $\mathcal{O}(1/c^5)$ for usual bodies, as displayed in Eq. 10, because of the conservation of their scalar charge.

The only remaining difficulty, to derive the predictions of scalar-tensor theories in the strong-field regime, is to compute the body-dependent coupling constants α_A and β_A . This can be done thanks to numerical integrations of the field equations inside the bodies, as explained in [40, 42, 43]. For negative values of the parameter β_0 entering expansion (8), one shows, both analytically and numerically, that nonperturbative effects occur beyond a critical compactness Gm/Rc^2 depending on β_0 . For instance, for $\beta_0 = -6$, the linear coupling constant α_A of a neutron star to the scalar field takes the values displayed in Fig. 8.

One sees that even if α_0 is vanishingly small in the background (and thereby in the solar system), neutron stars can develop an order-one coupling constant to the scalar field. Their physics and their orbital motion can thus differ significantly from the predictions of GR, although the scalar field may have strictly no effect in the solar system.

4.3 Binary-Pulsar Tests

Now that we know how to compute the predictions of scalar-tensor theories even in strong-field conditions, how may we test them? It happens that nature has provided us with fantastic objects called pulsars. These are neutron stars (thereby very compact objects, $Gm/Rc^2 \sim 0.2$) which are rapidly rotating and highly magnetized, and which emit a beam of radio waves like lighthouses. They can thus be considered

as natural clocks, and the oldest pulsars are indeed very stable ones. Therefore, a pulsar A orbiting a companion B is a moving clock, the best tool that one could dream of to test a relativistic theory. Indeed, by precisely timing its pulse arrivals, one gets a stroboscopic information on its orbit, and one can measure several relativistic effects. Such effects do depend on the two masses m_A , m_B , which are not directly measurable. However, two different effects suffice to determine them, and a third relativistic observable then gives a test of the theory.

For instance, in the case of the famous Hulse–Taylor binary pulsar PSR B1913+16 [112], three relativistic parameters have been determined with great accuracy: (i) the Einstein time delay parameter γ_T , which combines the second-order Doppler effect ($\propto v_A^2/2c^2$) together with the redshift due to the companion ($\propto Gm_B/r_{ABC^2}$); (ii) the periastron advance $\dot{\omega}$ ($\propto v^2/c^2$); and (iii) the rate of change of the orbital period, \dot{P} , caused by gravitational radiation damping ($\propto v^5/c^5$ in GR, but of order v^3/c^3 in scalar-tensor theories; see Eq. 10). The same parameters have also been measured for the neutron star-white dwarf binary PSR J1141–6545, but with less accuracy [10, 11]. In addition to these three parameters, (iv) the “range” (global factor Gm_B/c^3) and (v) “shape” (time dependence) of the Shapiro time delay have also been determined for two other binary pulsars, PSR B1534+12 [103] and PSR J0737–3039 [31, 73, 76]. The latter system is particularly interesting because both bodies have been detected as pulsars. Since their independent timing gives us the (projected) size of their respective orbits, the ratio of these sizes provides a direct measure of (vi) the mass ratio $m_A/m_B \approx 1.07$. In other words, 6 relativistic parameters have been measured for the double pulsar PSR J0737–3039. After using two of them to determine the masses m_A and m_B , this system thereby provides $6 - 2 = 4$ tests of relativistic gravity in strong-field conditions.

The clearest way to illustrate these tests is to plot the various experimental constraints in the mass plane (m_A, m_B), for a given theory of gravity. Any theory indeed predicts the expressions of the various timing parameters in terms of these unknown masses and other Keplerian observables, such as the orbital period and the eccentricity. The equations $predictions(m_A, m_B) = observed\ values$ thereby define different curves in the mass plane, or rather different *strips* if one takes into account experimental errors. If these strips have a common intersection, there exists a pair of masses that is consistent with all observables, and the theory is confirmed. On the other hand, if the strips do not meet simultaneously, the theory is ruled out. Figure 9 displays this mass plane for the Hulse–Taylor binary pulsar. Its left panel shows that GR is superbly consistent with these data. Its right panel illustrates that the three strips can be significantly deformed in scalar-tensor theories, because scalar exchanges between the pulsar and its companion modify all theoretical predictions. In the displayed case, corresponding to a quadratic matter-scalar coupling constant $\beta_0 = -6$ (as in Fig. 8), the strips do not meet simultaneously and the theory is thus excluded. On the contrary, they may have a common intersection in other scalar-tensor theories, even if it does not correspond to the same values of the masses m_A and m_B that were consistent with GR. The allowed region of the theory space $(|\alpha_0|, \beta_0)$ is displayed in Fig. 10.

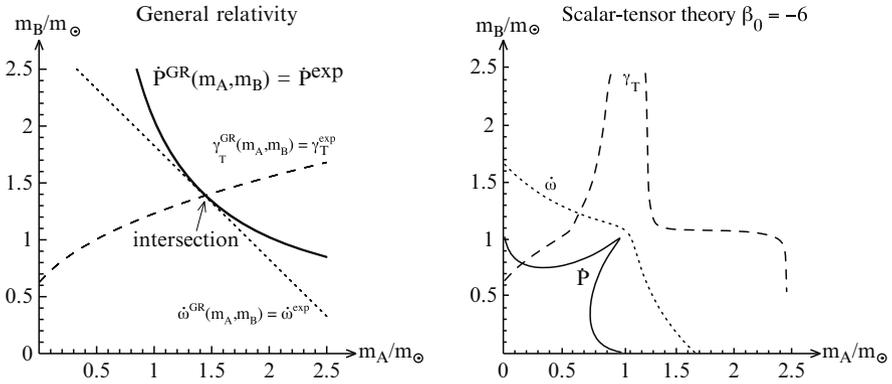


Fig. 9 Mass plane ($m_A =$ pulsar, $m_B =$ companion) of the Hulse–Taylor binary pulsar PSR B1913+16 in general relativity (*left panel*) and for a scalar-tensor theory with $\beta_0 = -6$ (*right panel*). The widths of the lines are larger than 1σ error bars. While GR passes the test with flying colors, the value $\beta_0 = -6$ is ruled out. [Actually, the right panel is plotted for a specific small value of $|\alpha_0| \approx 10^{-2}$, but the three curves keep similar shapes whatever α_0 , even vanishingly small, and they never have any common intersection for $\beta_0 = -6$]

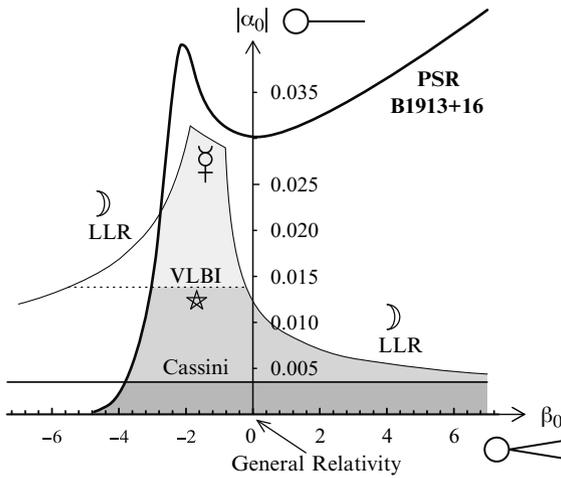


Fig. 10 Same theory plane ($|\alpha_0|, \beta_0$) as in Fig. 7, but taking now into account the constraints imposed by the Hulse–Taylor binary pulsar (*bold line*). LLR stands as before for “lunar laser ranging” and VLBI for “very long baseline interferometry.” The allowed region is the dark grey one. While solar system tests impose a small value of $|\alpha_0|$, binary pulsars impose the orthogonal constraint $\beta_0 > -4.5$

As mentioned above, several other relativistic binary pulsars are presently known, and Fig. 11 displays their simultaneous constraints on this theory plane [37,42–44,55]. To clarify this plot, we have used a logarithmic scale for the vertical ($|\alpha_0|$) axis. The drawback is that GR, corresponding to $\alpha_0 = \beta_0 = 0$, is sent down

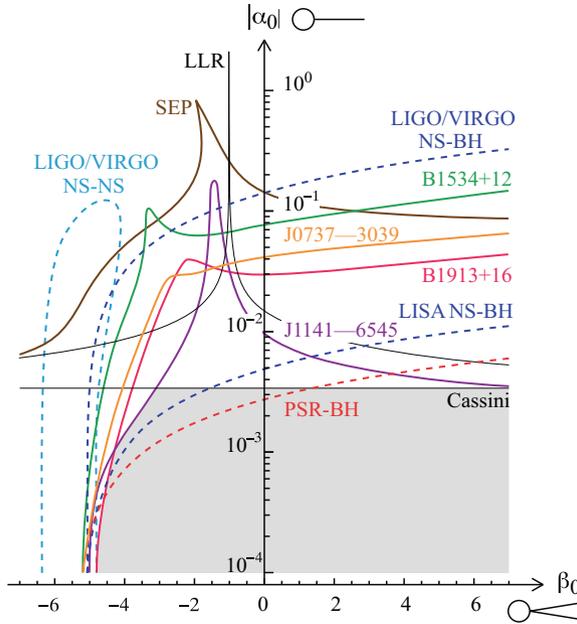


Fig. 11 Same theory plane as in Fig. 10, but now with a logarithmic scale for the linear matter-scalar coupling constant $|\alpha_0|$. This plot displays solar-system and several binary-pulsar constraints, using the latest published data. The allowed region is shaded. The curve labeled SEP corresponds to tests of the “strong equivalence principle” using a set of neutron star-white dwarf low-eccentricity binaries. The dashed lines corresponds to expected constraints if we find a relativistic pulsar-black hole (PSR-BH) binary, or when gravity-wave antennas detect the coalescence of double-neutron star (NS-NS) or neutron star-black hole (NS-BH) binaries

to infinity, but the important point to recall is that it does lie within the allowed (grey) region. Figures 10 and 11 illustrate vividly the *qualitative* difference between solar-system and binary-pulsar observations. Indeed, while weak-field tests only constrain the linear coupling constant $|\alpha_0|$ to be small without giving much information about the quadratic coupling constant β_0 , binary pulsars impose $\beta_0 > -4.5$ even for a vanishingly small α_0 . This constraint is due to the spontaneous scalarization of neutron stars which occurs when $-\beta_0$ is large enough, and which was illustrated in Fig. 8 above. Equations 9 allow us to rewrite this inequality in terms of the Eddington parameters β^{PPN} and γ^{PPN} , which are both consistent with 1 in the solar system. One finds

$$\frac{\beta^{\text{PPN}} - 1}{\gamma^{\text{PPN}} - 1} < 1.1. \tag{11}$$

The singular (0/0) nature of this ratio underlines why such a conclusion could not be obtained in weak-field experiments.

4.4 Black Holes in Scalar-Tensor Gravity

In Fig. 11 is displayed the order of magnitude of the tight constraint that can be expected if we are lucky enough to detect a relativistic pulsar-black hole binary. However, let us comment on the behavior of black holes in scalar-tensor theories. Since we saw in Fig. 8 above that nonperturbative effects can occur in strong-field conditions, one should a priori expect even larger deviations from GR for black holes (extreme compactness $Gm/Rc^2 = 0.5$) than for neutron stars (large compactness $Gm/Rc^2 \sim 0.2$). However, the so-called no-hair theorem [13, 28, 39, 60, 61, 63] shows that black holes must have a strictly vanishing scalar charge, $\alpha_{BH} = 0$. The basic idea is that otherwise the scalar field φ would diverge at the horizon, and this would be an unphysical solution. A first consequence is that a collapsing star must radiate away its scalar charge when forming a black hole. This is related to the generically large $\mathcal{O}(1/c)$ monopolar radiation of scalar waves predicted for nonequilibrium configurations, as discussed below Eq. 10. But the second crucial consequence is that black holes, once formed and stabilized, are not coupled at all to the scalar field, and therefore behave exactly as in GR: They generate the same solution for the Einstein metric $g_{\mu\nu}^*$, do not excite the scalar field φ , and move within the curved geometry of $g_{\mu\nu}^*$ as in GR. The conclusion is therefore that there is strictly no observable scalar-field effect in a binary black-hole system.

Of course, there do exist significant perturbations caused by the scalar field during the short time of the black-hole formation or when it captures a star (see e.g., [32]), because of the emission of a generically large amount of energy via scalar waves. Similarly, if one assumes that there exists a nonconstant background of scalar field $\varphi_0(x)$, then its own energy momentum tensor $T_{\mu\nu}$ contributes to the curvature of the Einstein metric $g_{\mu\nu}^*$, and even black holes would thus indirectly feel its presence. However, this would not be a consequence of the modification of gravity itself, but of the assumption of a nontrivial background $T_{\mu\nu}$. Even within pure GR, one could also have assumed that black holes move within a nontrivial background, caused for instance by the presence of dark matter or some large gravitational waves, and one would predict then a different motion than in a trivial background. One should thus qualify our conclusion above: Given some precise boundary conditions, for instance asymptotic flatness and no incoming radiation, black holes at equilibrium behave exactly as in GR.

It remains now to understand how a pulsar-black hole binary can allow us to constrain scalar-tensor theories as in Fig. 11. The reason is that one of the two bodies is *not* a black hole, but a neutron star with scalar charge $\alpha_A \neq 0$. Let us even recall that massive enough neutron stars can develop order-one scalar charges as in Fig. 8, even if matter did not feel at all the scalar field in the solar system ($\alpha_0 = 0$). A pulsar-black hole binary must therefore emit a large amount of dipolar waves $\propto (\alpha_A - \alpha_{BH})^2/c^3 = \alpha_A^2/c^3 = \mathcal{O}(1/c^3)$, as given by Eq. 10, and this can be several orders of magnitude larger than the usual $\mathcal{O}(1/c^5)$ quadrupolar radiation predicted by GR. Any pulsar-black hole binary whose variation of the orbital period, \dot{P} , is consistent with GR will thus tightly constrain scalar-tensor models.

5 Extended Bodies

In this section, we will discuss finite-size effects on the motion of massive bodies, in GR and in scalar-tensor theories [43, 62, 87]. It is still convenient to describe the position of such extended bodies by using one point in their interior, for instance their approximate center of mass. In other words, we will skeletonize the extended body’s worldtube as a unique worldline, say x_c . However, the action (2) describing the motion of a point particle cannot remain valid to all orders, because the metric $g_{\mu\nu}(x_{\text{left}})$ on one side of the body is not strictly the same as $g_{\mu\nu}(x_{\text{right}})$ on the other side. By expanding the metric around its value at x_c , it is thus clear that an effective action describing the motion of an extended body must depend on *derivatives* of the metric, and since we wish to construct a covariant expression, such derivatives must be built from contractions of the curvature tensor, its covariant derivatives, or any product of them. Moreover, if the body is nonspinning and spherical when isolated, it does not have any privileged direction, and the only 4-vector available to contract possible free indices is the 4-velocity $u^\mu = dx^\mu/ds$ of the point x_c . The first couplings to curvature that one may think of are thus of the form [62]

$$S_{\text{extended body}} = S_{\text{point particle}} + \int (k_1 R + k_2 R_{\mu\nu} u^\mu u^\nu + \dots) c ds, \quad (12)$$

where k_1 and k_2 denote body-dependent form factors (no term $R_{\mu\nu\rho\sigma} u^\mu u^\nu u^\rho u^\sigma$ is written since it vanishes because of the symmetry properties of the Riemann tensor). However, as recalled in [25, 43], any perturbative contribution to an action, say $S[\psi]$, which is proportional to its lowest-order field equations, $\delta S/\delta\psi$, is equivalent to a local field redefinition. Indeed, $S[\psi + \varepsilon] = S[\psi] + \varepsilon\delta S/\delta\psi + \mathcal{O}(\varepsilon^2)$, where the small quantity ε may be itself a functional of the field ψ . Therefore, up to *local* redefinitions of the worldline x_c and the metric $g_{\mu\nu}$ inside the body, which do not change the observable effects encoded in the metric $g_{\mu\nu}(x_{\text{obs}})$ at the observer’s location, one may replace R and $R_{\mu\nu}$ in action (12) above by their sources. Outside the extended body, such couplings to R or $R_{\mu\nu}$ have thus strictly no observable effect (see [25] for a discussion of the field redefinitions inside the body even when it is skeletonized as a point particle).

The first finite-size observable effects must therefore involve the Weyl tensor $C_{\mu\nu\rho\sigma}$, which does not vanish outside the body. Because of the symmetry properties of this tensor, the lowest-order terms one may construct from it take the form [62]

$$S_{\text{extended body}} = S_{\text{point particle}} + \int \left(k_3 C_{\mu\nu\rho\sigma}^2 + k_4 C_{\mu\nu\rho\alpha} C^{\mu\nu\rho\beta} u^\alpha u^\beta + k_5 C_{\mu\alpha\nu\beta} C^\mu{}_\gamma{}^\nu{}_\delta u^\alpha u^\beta u^\gamma u^\delta + \dots \right) c ds. \quad (13)$$

By comparing the dimensions of $S_{\text{point particle}} = -\int mc ds$ and these couplings to the Weyl tensor, we expect $k C^2 \sim m$ and therefore $k \sim m R_E^4$ (where k denotes any of the form factors k_3, k_4 , or k_5 , and R_E means the radius of the extended body, here written with an index to avoid a confusion with the curvature scalar).

In conclusion, for a compact body such that $Gm/R_E c^2 \sim 1$, we expect the form factors k to be of order $m(Gm/c^2)^4$, and the corrections to its motion will be proportional to $k c^2 C^2 = k c^2 \mathcal{O}(1/c^4) = \mathcal{O}(1/c^{10})$. Therefore, finite-size effects start at the fifth post-Newtonian (5PN) order, consistently with the Newtonian reasoning. Indeed, if one considers two bodies A and B that are spherical when isolated, one can show that B is deformed as an ellipsoid by the tidal forces caused by A , and this deformation induces an extra force $\sim (Gm_A/r_{AB}^2)(R_B/r_{AB})^5$ felt by A . If B is a compact body, that is, $Gm_B/R_B c^2 \sim 1$, one concludes that the extra force felt by body A is of order $\mathcal{O}(1/c^{10})$. We recovered above the same conclusion within GR thanks to a very simple dimensional argument. Note however that the explicit calculation is more involved than our estimate of the order of magnitude, because one must take into account the Weyl tensor generated by the *two* (or more) bodies of the system.

Let us now follow the same dimensional reasoning within scalar-tensor theories of gravity. We assumed in Eq. 7 that matter is coupled to the physical metric $g_{\mu\nu} = A^2(\varphi)g_{\mu\nu}^*$, therefore the action of a point particle reads

$$S_{\text{point particle}} = - \int mc \sqrt{-g_{\mu\nu} dx^\mu dx^\nu} = - \int A(\varphi) mc \sqrt{-g_{\mu\nu}^* dx^\mu dx^\nu}. \quad (14)$$

When studying the motion of such a point particle in the Einstein metric $g_{\mu\nu}^*$, it behaves thus as if it had a scalar-dependent mass $m_*(\varphi) \equiv A(\varphi)m$. [Incidentally, this provides us with another definition of the scalar charge, $\alpha = d \ln m_*(\varphi)/d\varphi$.] If we consider now an extended body in scalar-tensor gravity, the function $m_*(\varphi)$ must be replaced by a *functional*, that is, it can depend on the (multiple) derivatives of both the metric and the scalar field. Let us restrict again our discussion to nonspinning bodies that are spherical when isolated. Then the derivative expansion of this mass functional can be written as [43]

$$m[\varphi, g_{\mu\nu}^*] = m(\varphi) + I(\varphi)R^* + J(\varphi)R_{\mu\nu}^* u_*^\mu u_*^\nu + K(\varphi)\square^* \varphi + L(\varphi)\nabla_\mu^* \partial_\nu \varphi u_*^\mu u_*^\nu + M(\varphi)\partial_\mu \varphi \partial_\nu \varphi u_*^\mu u_*^\nu + N(\varphi)g_*^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi + \dots, \quad (15)$$

where star indices mean that we are using the Einstein metric $g_{\mu\nu}^*$ to compute curvature tensors, covariant derivatives, and the unit velocity $u_*^\mu \equiv dx^\mu/ds^*$, and where $I(\varphi), \dots, N(\varphi)$ are body-dependent form factors encoding how the extended body feels the various second derivatives of the fields φ and $g_{\mu\nu}^*$. Fortunately, by using as in GR the lowest-order field equations together with local field and worldline redefinitions, this long expression reduces to a mere

$$m[\varphi, g_{\mu\nu}^*] = m(\varphi) + N_{\text{new}}(\varphi)g_*^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi + \text{higher post-Keplerian orders}, \quad (16)$$

where $N_{\text{new}}(\varphi)$ is a linear combination of the previous $N(\varphi)$, $L(\varphi)$ and $I(\varphi)$. We already notice that the lowest-order finite-size effects involve only two derivatives of the scalar field, whereas action (13) in GR involved the square of a *second* derivative of the metric. As above, let us invoke dimensional analysis to deduce

that $N_{\text{new}} \sim mR_E^2$, where R_E still denotes the radius of the extended body. This is consistent with the actual calculation performed in [87] for weakly self-gravitating bodies, where it was proven that $N_{\text{new}} = \frac{1}{6}\beta_0 \times (\text{Inertia moment})$. For compact (therefore strongly self-gravitating) bodies, such that $Gm/R_E c^2 \sim 1$, one thus expects observable effects of order $Nc^2(\partial\varphi)^2 = \mathcal{O}(1/c^2)R_E^2 = \mathcal{O}(1/c^6)$, that is, of the third post-Newtonian (3PN) order. Finite-size effects are thus much larger in scalar-tensor theories⁴ than in GR, where they were of the 5PN order, $\mathcal{O}(1/c^{10})$. Moreover, when nonperturbative strong-field effects develop as in Fig. 8, one actually expect that $N_{\text{new}} \sim \beta_E \times (\text{Inertia moment})$ can be of order unity, because the coupling constant β_E of the body to two scalar lines can be extremely large with respect to the bare β_0 . Therefore, finite-size effects in scalar-tensor gravity should actually be considered as *first* post-Keplerian, $\mathcal{O}(v_{\text{orbital}}^2/c^2)$, perturbations of the motion, as compared to the 5PN order in GR.

6 Modified Newtonian Dynamics

The existence of dark matter (a pressureless and noninteracting fluid detected only by its gravitational influence) is suggested by several observations. For instance, type-Ia supernova data [7, 72, 90, 105] are consistent with a present acceleration of the expansion of the Universe, and tell us that the dark energy density should be of order $\Omega_\Lambda \approx 0.7$. On the other hand, the position of the first acoustic peak of the cosmic microwave background spectrum [101, 102] is consistent with a spatially flat Universe, that is, $\Omega_\Lambda + \Omega_{\text{matter}} \approx 1$. Combining these two pieces of information, one thus deduce that the matter density should be $\Omega_{\text{matter}} \approx 0.3$, a value at least one order of magnitude larger than all our estimates of baryonic matter in the Universe (for instance $\Omega_{\text{baryons}} \approx 0.04$ derived from Big Bang nucleosynthesis). Therefore, most of the cosmological matter should be nonbaryonic, that is, “dark” because it does not interact significantly with photons. Another, independent, evidence for the existence of dark matter is the flat rotation curves of clusters and galaxies [91]: The velocities of outer stars tend toward a constant value (depending on the galaxy or cluster), instead of going asymptotically to zero as expected in Newtonian theory (recall that Neptune is much slower than Mercury in the solar system). If Newton’s law is assumed to be valid, such nonvanishing asymptotic velocities imply the existence of much more matter than within the stars and the gas. Independently of

⁴These larger finite-size effects are due to the fact that a spin-0 scalar field can couple to the spherical inertia moment of a body, contrary to a spin-2 graviton. They should not be confused with the violation of the strong equivalence principle, which *also* occurs in scalar-tensor theories because all form factors $m(\varphi)$, $N(\varphi)$, ... depend on the body’s self-energy. Regardless of its finite size, the motion of a self-gravitating body in a uniform exterior gravitational field depends thus on its internal structure.

these experimental evidences, we also have many theoretical candidates for dark matter, notably the class of neutralinos occurring in supersymmetric theories (see e.g., [26]), and numerical simulations of structure formation have obtained great successes while incorporating dark matter (see e.g., [64]).

However, this unknown fluid might actually be an artifact of our interpretation of experimental data with a Newtonian viewpoint. It is thus worth examining whether the gravitational $1/r^2$ law could be modified at large distances, instead of invoking the existence of dark matter. In 1983, Milgrom realized that galaxy rotation curves could be fitted with a very simple recipe, that he called Modified Newtonian Dynamics (MOND) [79]. It does not involve any mass scale nor distance scale, but an *acceleration* scale denoted as a_0 (not to be confused with the linear matter-scalar coupling constant α_0 , defined in Eq. 8). Milgrom assumed that the acceleration a of a test particle caused by a mass M should read

$$\begin{aligned}
 a &= a_N = \frac{GM}{r^2} && \text{if } a > a_0, \\
 a &= \sqrt{a_0 a_N} = \frac{\sqrt{GM a_0}}{r} && \text{if } a < a_0,
 \end{aligned}
 \tag{17}$$

where a_N denotes the usual Newtonian expression. This phenomenological law happens to fit remarkably well galaxy rotation curves [95], for a universal constant $a_0 \approx 1.2 \times 10^{-10} \text{ m} \cdot \text{s}^{-2}$. [However, galaxy clusters require either another value of this constant, or some amount of dark matter, for instance in the form of massive neutrinos.] Moreover, it automatically recovers the Tully–Fisher law [106] $v_\infty^4 \propto M$, where M denotes the baryonic mass of a galaxy, and v_∞ the asymptotic velocity of visible matter in its outer region. The MOND assumption (17) would also explain in an obvious way why dark matter profiles seem to be tightly correlated to the baryonic ones [77].

However, reproducing the simple law (17) in a consistent relativistic field theory happens to be quite difficult. As mentioned in Section 2, Milgrom explored “modified inertia” models [80, 81], in which the action of a point particle is assumed to depend nonlocally on all the time derivatives $d^n \mathbf{x}/dt^n$ of its position \mathbf{x} . In these lecture notes, we focus on local field theories which “modify gravity”, that is, which assume that the kinetic term of the metric $g_{\mu\nu}$ (to which matter is universally coupled) is not the standard Einstein–Hilbert action (1).

6.1 Mass-Dependent Models?

It is actually very easy to devise a model predicting a force $\propto 1/r$. Indeed, let us consider a mere scalar field φ in flat spacetime, with a potential $V(\varphi) = -2a^2 e^{-b\varphi}$, where a and b are two constants. In a static and spherically symmetric situation, its field equation $\Delta\varphi = V'(\varphi)$ then gives the obvious solution $\varphi = (2/b) \ln(abr)$.

If we now assume that matter is linearly coupled to φ , it will feel a force $\propto \partial_i \varphi$, that is, the $1/r$ law we are looking for. However, this simple model presents two very serious problems. The first one is that the potential $V(\varphi) = -2a^2 e^{-b\varphi}$ is unbounded from below, and therefore that the theory is unstable. Actually, we will see that stability is indeed a generic difficulty of all models trying to reproduce the MOND dynamics. The second problem of this naive model is that it predicts a *constant* coefficient $2/b$ for the solution of the scalar field, instead of the factor \sqrt{M} entering the second line of (17). It happens that many models proposed in the literature do behave in the same way, although their more complicated writings hide the problem.⁵ The trick used by the corresponding authors is merely to set $b \propto 1/\sqrt{M}$ in the action of the theory. In other words, they are considering a *different* theory for each galaxy M ! It should be noted that the mass of an object is not a local quantity: It is the integral of the matter density over a particular region, whose symmetry plays also an important role. One of the most difficult steps in building a consistent theory of MOND is thus precisely to be able to predict this factor \sqrt{M} . Moreover, if one defined a potential $V(\varphi)$ as above, in terms of an integral of the matter density giving access to M , then it would mean that matter is coupled to φ in a highly nonlinear and nonlocal way, therefore the force it would feel would be much more complicated than the naive gradient $\partial_i \varphi$ assumed above. In the following, we will consider local field theories whose actions do not depend on M , but only on the constants G , c and a_0 .

6.2 Aquadratic Lagrangians or k-Essence

One of the most promising frameworks to reproduce the MOND dynamics is a generalization of the scalar-tensor theories we considered in Section 4. Their action takes the form [14–16, 18, 93, 94]

$$S = \frac{c^3}{16\pi G_*} \int d^4x \sqrt{-g^*} \{R^* - 2f(g_*^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi)\} + S_{\text{matter}} [\text{matter}; g_{\mu\nu} \equiv A^2(\varphi)g_{\mu\nu}^* + B(\varphi)U_\mu U_\nu]. \tag{18}$$

The first crucial difference with action (7) is that the kinetic term of the scalar field is now a *function* of the standard quadratic term $(\partial_\mu \varphi)^2$. Reference [17] showed that this suffices to reproduce the MOND law (17), including the important \sqrt{M} coefficient, provided $f(x) \propto x^{3/2}$ for small values of x (MOND regime) and $f(x) \rightarrow x$ for large x (Newtonian regime). Such aquadratic kinetic terms have also been analyzed later in the cosmological context, under the names of k-inflation [5] or k-essence [6, 33] (the letter k meaning that their dynamics is kinetic dominated).

⁵ See Section II.B of Ref. [30] for a critical discussion of various such mass-dependent models.

As in action (7), matter is assumed to be coupled to the scalar field via the function $A^2(\varphi)$ entering the definition of the physical metric $g_{\mu\nu}$. This ensures that test particles will undergo an extra acceleration caused by the scalar field. But the second difference with (7) is the presence of the non-conformal term $B(\varphi)U_\mu U_\nu$ entering the definition of $g_{\mu\nu}$. Here U_μ is a vector field, which can have either its own kinetic term, or can be simply chosen as $U_\mu = \partial_\mu\varphi$. This term is necessary to reproduce the light deflection caused by galaxies or clusters, which happens to be consistent with the prediction of GR in presence of a dark matter halo [12, 57, 78]. Indeed, light is totally insensitive to the conformal factor $A^2(\varphi)$ relating the physical and Einstein metrics in action (7). The simplest way to understand it is to recall that a photon's null geodesic satisfies $ds^2 = 0 \Leftrightarrow A^2(\varphi)ds_*^2 = 0 \Leftrightarrow ds_*^2 = 0$, therefore the photon propagates in the Einstein metric $g_{\mu\nu}^*$ without feeling the presence of the scalar field. (To be more precise, the photon does feel indirectly the presence of the scalar field via the extra curvature of $g_{\mu\nu}^*$ caused by its energy–momentum tensor, but this is a higher PN effect.) The only way to impose that the MOND potential (encoded in the scalar field φ) deflect light is thus to relate the physical metric $g_{\mu\nu}$ to the Einstein one in a non-conformal way, as in (18). This idea dates back to Ni's “stratified” theory of gravity [84, 113].

Obviously, several conditions must be imposed on the functions entering action (18) to warrant that the theory is stable and that it has a well-posed Cauchy problem. For instance, it is clear that $f(x) = x$ defines a standard kinetic term for the scalar field, whereas $f(x) = -x$ would define a negative-energy (ghost) mode. In order for the Hamiltonian to be bounded by below and for the scalar-field equations to be hyperbolic, one can actually show that the function f must satisfy the two conditions [3, 30]

$$\forall x, \quad f'(x) > 0, \quad \text{and} \quad \forall x, \quad 2x f''(x) + f'(x) > 0. \quad (19)$$

One also notices that gravitons are faster than scalar particles when $f''(x) < 0$, and slower when $f''(x) > 0$. It has been argued in [1] that gravitons should be the fastest modes, otherwise causal paradoxes can be constructed. However, Refs. [8, 29, 30] concluded that conditions (19) suffice for causality to be preserved, because they ensure that the widest causal cone always remains a cone, and never opens totally.

The analogues of conditions (19) become much more complicated within matter, where the two other functions $A(\varphi)$ and $B(\varphi)$ of action (18) also enter the game. Moreover, one also needs to ensure that the matter field equations always remain hyperbolic. We refer to [30] for a discussion of these issues.

6.3 Difficulties

Although the class of relativistic models (18) is the most promising one to reproduce the MOND phenomenology, it has anyway a long list of difficulties. Some of them can be solved, but at the price of complicated and unnatural Lagrangians.

For instance, we mentioned above the problem of light deflection, which is too small in conformally coupled scalar-tensor theories (7), therefore one needed to introduce a vector field U_μ in action (18). One may also notice that such models are not very predictive, since it would have been possible to predict fully different lensing and rotation curves. One may thus consider them as fine-tuned fits rather than fundamental theories imposed by some deep symmetry principles. The most famous model, called TeVeS (for Tensor–Vector–Scalar) [15, 16], also presents some discontinuous functions, and does not allow to pass smoothly from a time evolution (cosmology) to a spatial dependence (local physics in the vicinity of a galaxy). However, some cures are possible [30], although they again involve rather unnatural refinements.

One peculiarity of the TeVeS model is that its author imposed that gravitons and scalar particles are slower than photons (a priori to avoid causal paradoxes due to superluminal propagation, assuming that light is a more fundamental field than gravity). But Refs. [54, 83] proved that in such a case, high-energy cosmic rays would rapidly lose their energy by Cherenkov radiation of gravitational waves, and this would be inconsistent with their observation on Earth. However, a very simple solution exists to cure this problem: One just needs to flip a sign in one of the terms of the TeVeS model [15, 16], and merely accept that photons can be slower than gravitons [30]. If all the field equations are ensured to remain hyperbolic, with a common time direction, no causal paradox can be caused by such a situation.

The vector U_μ of action (18) is assumed to be timelike in the TeVeS model, therefore there a priori exists a preferred frame in which it takes the value $U_\mu = (1, 0, 0, 0)$ [58, 93]. References [15, 16] argue that this can anyway be consistent with the high-precision tests of local Lorentz invariance of gravity in the solar system if this vector is dynamical. However, Ref. [35] showed that the corresponding Hamiltonian is not bounded by below, precisely because of the kinetic term of this vector field. Therefore, the model is unstable. Other instabilities are also present in the slightly different model of Ref. [94].

A final difficulty is related to the PN tests of relativistic gravity, summarized in Section 4. In standard scalar-tensor theories (7), Fig. 7 shows that the linear matter-scalar coupling constant α_0 should be small. In the TeVeS model, of the form (18), the functions $A(\varphi)$ and $B(\varphi)$ have been *tuned* to mimic the Schwarzschild metric of GR up to the 1PN order, even for a large matter-scalar coupling constant α_0 . Therefore, the full plane $(|\alpha_0|, \beta_0)$ of Fig. 7 seems now allowed by experimental data. However, binary-pulsar tests do not depend on $A(\varphi)$ and $B(\varphi)$ in the same way, and one actually gets basically the same constraints as in Figs. 10 and 11, notably the tight bounds on α_0 imposed by the pulsar-white dwarf binary PSR J1141–6545 of Fig. 11. In conclusion, in spite of the tuning of $A(\varphi)$ and $B(\varphi)$ to mimic GR in the solar system, binary pulsars anyway impose that matter should be *weakly* coupled to the scalar field [30]. This is quite problematic because we wish the same scalar field to give rise, at large distances, to the MOND acceleration $\sqrt{GMa_0}/r$ whose magnitude is fixed. Figure 12 illustrates schematically the difficulty (more precise discussions are given in Ref. [30]): If we wish the function $f((\partial_\mu\varphi)^2)$ entering action (18) to have a natural enough shape, either scalar-field effects are too large to be consistent with binary-pulsar tests, or the solar-system size is already large

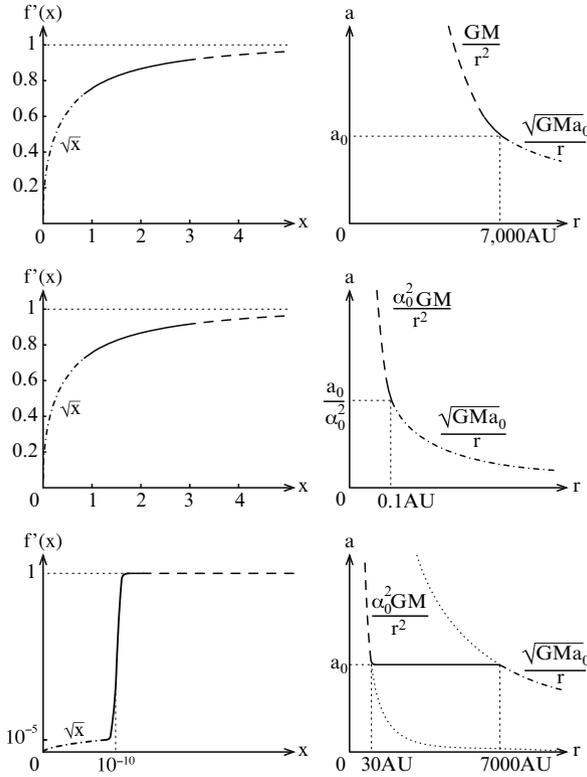


Fig. 12 Derivative of the function f defining the kinetic term of the scalar field in action (18), such that the MOND dynamics is predicted at large distances and Newtonian physics at small distances. In the first case, f' has a natural shape and the MOND acceleration is obtained at the expected distance $\sqrt{GM/a_0}$, but the scalar-field contribution to the acceleration is too large to be consistent with binary-pulsar tests. In the second case, f' has the same natural shape and the force caused by the scalar field is small enough when $r \rightarrow 0$, but all planets should undergo an extra MOND acceleration $\sqrt{GMa_0}/r$. In the third case, the scalar force is small enough in the solar system and the MOND law is predicted beyond the characteristic distance $\sqrt{GM/a_0}$, but the shape of f' is extremely unnatural

enough for all planets to be in the MOND regime and undergo a small $\sqrt{GMa_0}/r$ acceleration in addition to the Newtonian GM/r^2 . This would be ruled out by tests of Kepler’s third law [104]. The only remaining solution would be to impose a small enough $\alpha_0^2 GM/r^2$ acceleration caused by the scalar field within the solar system, and the MOND law $\sqrt{GMa_0}/r$ at large distances, but this would correspond to the quite unnatural shape of the function f' displayed in the third panel of Fig. 12. Although this is not yet excluded experimentally, it would suffice to improve binary-pulsar constraints by one order of magnitude to rule out this kind of fine-tuned model, because one would need a shape of function f' violating the consistency conditions (19).

6.4 Nonminimal Couplings

Another possible way to modify gravity is inspired by the behavior of extended bodies within GR itself, as illustrated in Section 5. One may assume that matter is not only coupled to the metric but also nonminimally to its curvature. Let us thus consider an action of the form

$$S = \frac{c^3}{16\pi G_*} \int d^4x \sqrt{-g^*} R^* + S_{\text{matter}} \left[\text{matter}; g_{\mu\nu} \equiv f(g_{\mu\nu}^*, R_{\lambda\mu\nu\rho}^*, \nabla_\sigma^* R_{\lambda\mu\nu\rho}^*, \dots) \right], \tag{20}$$

which is Lorentz invariant and satisfies the Einstein equivalence principle because all matter fields are coupled to the same tensor $g_{\mu\nu}$ (although it is built from the spin-2 metric $g_{\mu\nu}^*$ and its curvature tensor). An immediate bonus of this class of theories is that they reduce to GR in vacuum, therefore standard solutions for the metric $g_{\mu\nu}^*$ remain valid, notably the Schwarzschild solution for spherically symmetric configurations. The only difference with GR is that matter is no longer minimally coupled to $g_{\mu\nu}^*$, and therefore that its motion can be changed. Reference [30] showed that it is a priori possible to reproduce the MOND dynamics within this framework. However, the same theorem by Ostrogradski [88, 117] that we mentioned in Section 2 suffices to conclude that these models are generically unstable, and this can indeed be checked explicitly.

This instability can fortunately be avoided by slightly complicating the model in vacuum. Instead of pure GR, let us assume a scalar-tensor theory in vacuum, with a negligible scalar mass m_ϕ at galaxy scale. One may thus define an action

$$S = \frac{c^3}{16\pi G_*} \int d^4x \sqrt{-g^*} \{ R^* - 2g_*^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - 2m_\phi^2 \phi^2 \} + S_{\text{matter}} \left[\text{matter}; g_{\mu\nu} \equiv A^2[\phi] g_{\mu\nu}^* + B[\phi] \partial_\mu \phi \partial_\nu \phi \right], \tag{21}$$

and show that specific functionals $A[\phi] = A(\phi, \partial_\mu \phi)$ and $B[\phi] = B(\phi, \partial_\mu \phi)$ allow us to reproduce both the MOND dynamics and the right amount of light deflection by galaxies. The generic Ostrogradskian instability is avoided because the matter action involves second time derivatives of the scalar field only linearly [30]. Such a model is a priori as good as those which were previously proposed in the literature, and it is much simpler to analyze. Because of its simplicity, it has thus been possible to study its behavior *within* matter (an analysis which is usually too difficult to perform in other models). The bad news is that the scalar field equations do not remain hyperbolic within the dilute gas in outer regions of a galaxy. Therefore, this class of models is finally also ruled out to reproduce the MOND dynamics.

On the other hand, this class (21) is able to reproduce the Pioneer anomaly (if confirmed) in a consistent relativistic theory of gravity, without spoiling the other well-tested predictions of GR. This anomaly is actually a *simpler* problem than galaxy rotation curves. Indeed, one of the difficulties of MOND models was to predict a force $\propto \sqrt{M}$. In the case of the two Pioneer spacecrafts, we do not know

how their extra acceleration toward the Sun, $\delta a \approx 8.5 \times 10^{-10} \text{ m} \cdot \text{s}^{-2}$, is related (or not) to the mass M_\odot of the Sun. There are thus actually several stable and well-posed models able to reproduce this effect, that were constructed in Ref. [30] using the previous analyses of [68–70]. One of the simplest models reads schematically

$$S = \frac{c^3}{16\pi G_*} \int d^4x \sqrt{-g^*} \{R^* - 2g_*^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi - 2m_\varphi^2 \varphi^2\} + S_{\text{matter}} \left[\text{matter}; g_{\mu\nu} \equiv e^{2\alpha\varphi} g_{\mu\nu}^* - \lambda \frac{\partial_\mu \varphi \partial_\nu \varphi}{\varphi^5} \right], \quad (22)$$

where one needs to impose $\alpha^2 < 10^{-5}$ to pass solar-system and binary-pulsar tests, $\lambda \approx (\alpha G M_\odot / c^2)^3 (\delta a / v^2) \approx \alpha^3 (10^{-4} \text{ m})^2$ to fit the Pioneer anomaly, and where the scalar mass m_φ needs to be negligibly small at solar-system scale. Actually, refinements are necessary to define correctly this model when $\varphi \rightarrow 0$, but we refer to [30] for this discussion.

7 Conclusions

In these lecture notes, we have underlined that the study of contrasting alternatives to GR is useful to understand better which features of the theory have been tested, and to suggest new possible tests. One may modify either inertia (i.e., the matter action S_{matter}), gravity (i.e., the action defining its dynamics, S_{gravity}), or consider nonminimal couplings to curvature. We stressed that the dynamics of gravity directly influences the motion of massive bodies, and this was illustrated by constructing the Fokker action.

Our study of scalar-tensor theories of gravity exhibited a *qualitative* difference between solar-system experiments (weak fields) and binary-pulsar tests (strong fields). While the former tightly constrain the linear matter-scalar coupling constant, the latter are nonperturbatively sensitive to nonlinear couplings, and notably forbid that the quadratic matter-scalar coupling constant be large and negative. We mentioned that the no-hair theorem imposes that the motion of black holes in scalar-tensor theories is the same as in GR, contrary to other massive bodies. We also gave a simple dimensional argument showing that finite-size effects are much larger in alternative theories than in GR.

Finally, we illustrated that the MOND phenomenology may a priori be reproduced within any of the above classes of alternative theories (modified inertia, modified gravity, couplings to curvature), but that all the proposed models present several experimental and theoretical difficulties. However, many possible routes remain possible, and generalizations of Einstein-aether theories [53, 65–67, 118] able to reproduce MOND deserve to be studied, notably from the point of view of stability and well-posedness of the Cauchy problem. Of course, the simplest solution to account for galaxy rotation curves and cosmological data seems merely to accept the existence of dark matter. Instead of modifying gravity, one may then even devise some “modified dark matter” so that it reproduces the MOND successes while keeping the standard cold dark matter cosmological scenario [21–24].

References

1. A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis, R. Rattazzi, J. High Energy Phys. **10**, 014 (2006)
2. E.G. Adelberger, Class. Q. Grav. **18**, 2397 (2001)
3. Y. Aharonov, A. Komar, L. Susskind, Phys. Rev. **182**, 1400 (1969)
4. J.D. Anderson, P.A. Laing, E.L. Lau, A.S. Liu, M.M. Nieto, S.G. Turyshev, Phys. Rev. D **65**, 082004 (2002)
5. C. Armendáriz-Picón, T. Damour, V.F. Mukhanov, Phys. Lett. B **458**, 209 (1999)
6. C. Armendáriz-Picón, V.F. Mukhanov, P.J. Steinhardt, Phys. Rev. D **63**, 103510 (2001)
7. P. Astier et al. (The SNLS), Astron. Astrophys. **447**, 31 (2006)
8. E. Babichev, V. Mukhanov, A. Vikman, J. High Energy Phys. **02**, 101 (2008)
9. S. Baessler, B.R. Heckel, E.G. Adelberger, J.H. Gundlach, U. Schmidt, H.E. Swanson, Phys. Rev. Lett. **83**, 003585 (1999)
10. M. Bailes, S.M. Ord, H.S. Knight, A.W. Hotan, Astrophys. J. **595**, L49 (2003)
11. N.D.R. Bhat, M. Bailes, J.P.W. Verbiest, Phys. Rev. D **77**, 124017 (2008)
12. M. Bartelmann, P. Schneider, Phys. Rep. **340**, 291 (2001)
13. J.D. Bekenstein, Phys. Rev. Lett. **28**, 452 (1972); Phys. Rev. D **5**, 1239 (1972)
14. J.D. Bekenstein, Phys. Rev. D **48**, 3641 (1993)
15. J.D. Bekenstein, Phys. Rev. D **70**, 083509 (2004)
16. J.D. Bekenstein, PoS **JHW2004**, 012 (2005)
17. J. Bekenstein, M. Milgrom, Astrophys. J. **286**, 7 (1984)
18. J.D. Bekenstein, R.H. Sanders, Astrophys. J. **429**, 480 (1994)
19. P.G. Bergmann, Int. J. Theor. Phys. **1**, 25 (1968)
20. B. Bertotti, L. Iess, P. Tortora, Nature **425**, 374 (2003)
21. L. Blanchet, Class. Q. Grav. **24**, 3529 (2007)
22. L. Blanchet, Class. Q. Grav. **24**, 3541 (2007)
23. L. Blanchet, A. Le Tiec, Phys. Rev. D **78**, 024031 (2008)
24. L. Blanchet, A. Le Tiec, Phys. Rev. D **80**, 023524 (2009)
25. L. Blanchet, T. Damour, G. Esposito-Farèse, B.R. Iyer, Phys. Rev. D **71**, 124004 (2005)
26. C. Boehm, P. Fayet, J. Silk, Phys. Rev. D **69**, 101302(R) (2004)
27. C. Brans, R.H. Dicke, Phys. Rev. **124**, 925 (1961)
28. P. Breitenlohner, D. Maison, G.W. Gibbons, Commun. Math. Phys. **120**, 295 (1988)
29. J.-P. Bruneton, Phys. Rev. D **75**, 085013 (2007)
30. J.-P. Bruneton, G. Esposito-Farèse, Phys. Rev. D **76**, 124012 (2007)
31. M. Burgay, N. D'Amico, A. Possenti, R.N. Manchester, A.G. Lyne, B.C. Joshi, M.A. McLaughlin, M. Kramer, J.M. Sarkissian, F. Camilo, V. Kalogera, C. Kim, D.R. Lorimer, Nature **426**, 531 (2003)
32. B. Chauvineau, A.D.A.M. Spallicci, J.-D. Fournier, Class. Q. Grav. **22**, S457 (2005)
33. T. Chiba, T. Okabe, M. Yamaguchi, Phys. Rev. D **62**, 023511 (2000)
34. T.E. Chupp, R.J. Hoare, R.A. Loveman, E.R. Oteiza, J.M. Richardson, M.E. Wagshul, A.K. Thompson, Phys. Rev. Lett. **63**, 1541 (1989)
35. M.A. Clayton, arXiv:gr-qc/0104103v1 (2001)
36. T. Damour, in *Gravitation and Quantizations*, ed. by B. Julia, J. Zinn-Justin, Les Houches, Session LVII (Elsevier, Amsterdam, 1995), pp. 1–61
37. T. Damour, in *Physics of Relativistic Objects in Compact Binaries: from Birth to Coalescence* ed. by M. Colpi, P. Casella, V. Gorini, U. Moschella, A. Possenti, Astrophysics and Space Science Library **359** (Springer, Dordrecht and Canopus Publishing Lim., Bristol, 2009), pp. 1–41
38. T. Damour, F. Dyson, Nucl. Phys. B **480**, 37 (1996)
39. T. Damour, G. Esposito-Farèse, Class. Q. Grav. **9**, 2093 (1992)
40. T. Damour, G. Esposito-Farèse, Phys. Rev. Lett. **70**, 2220 (1993)
41. T. Damour, G. Esposito-Farèse, Phys. Rev. D **53**, 5541 (1996)
42. T. Damour, G. Esposito-Farèse, Phys. Rev. D **54**, 1474 (1996)
43. T. Damour, G. Esposito-Farèse, Phys. Rev. D **58**, 042001 (1998)

44. T. Damour, G. Esposito-Farèse (unpublished)
45. T. Damour, A.M. Polyakov, Nucl. Phys. B **423**, 532 (1994); Gen. Rel. Grav. **26**, 1171 (1994)
46. T. Damour, G. Schäfer, Gen. Rel. Grav. **17**, 879 (1985)
47. T. Damour, A. Vilenkin, Phys. Rev. D **53**, 2981 (1996)
48. T. Damour, F. Piazza, G. Veneziano, Phys. Rev. D **66**, 046007 (2002)
49. C. Deffayet, R.P. Woodard, J. Cosm. Astropart. Phys. **0908**, 23 (2009)
50. S. Deser, R.P. Woodard, Phys. Rev. Lett. **99**, 111301 (2007)
51. A. Eddington, *The Mathematical Theory of Relativity* (Cambridge University Press, London, 1923)
52. S. Eidelman et al. (Particle Data Group), Phys. Lett. B **592**, 1 (2004), contribution by T. Damour entitled *Experimental Tests of Gravitational Theories*, p. 186
53. C. Eling, T. Jacobson, D. Mattingly, arXiv:gr-qc/0410001v2 (2005)
54. J.W. Elliott, G.D. Moore, H. Stoica, J. High Energy Phys. **08**, 066 (2005)
55. G. Esposito-Farèse, in *Proc. 10th Marcel Grossmann Mtg*, Rio de Janeiro 20–26 July 2003, ed. by M. Novello, S. Perez Bergliaffa, R. Ruffini (World Scientific, Singapore, 2006), p. 647
56. M. Fierz, Helv. Phys. Acta **29**, 128 (1956)
57. B. Fort, Y. Mellier, Astron. Astrophys. Rev. **5**, 239 (1994)
58. B.Z. Foster, T. Jacobson, Phys. Rev. D **73**, 064015 (2006)
59. Y. Fujii, A. Iwamoto, T. Fukahori, T. Ohnuki, M. Nakagawa, H. Hidaka, Y. Oura, P. Möller, Nucl. Phys. B **573**, 377 (2000)
60. G.W. Gibbons, in *The Physical Universe: The Interface Between Cosmology, Astrophysics and Particle Physics*, ed. by J.D. Barrow, A.B. Henriques, M.T.V.T. Lago, M.S. Longair, Lecture Notes in Physics **383** (Berlin, Springer Verlag, 1991) p. 110
61. G.W. Gibbons, M.E. Ortiz, F. Ruiz Ruiz, Phys. Lett. B **240**, 50 (1990)
62. W.D. Goldberger, I.Z. Rothstein, Phys. Rev. D **73**, 104029 (2006)
63. S.W. Hawking, Commun. Math. Phys. **25**, 167 (1972)
64. E. Hayashi, J.F. Navarro, C. Power, A. Jenkins, C.S. Frenk, S.D.M. White, V. Springel, J. Stadel, T.R. Quinn, Mon. Not. R. Astron. Soc. **355**, 794 (2004)
65. T. Jacobson, PoS **QG-PH**, 020 (2007)
66. T. Jacobson, D. Mattingly, Phys. Rev. D **64**, 024028 (2001)
67. T. Jacobson, D. Mattingly, Phys. Rev. D **70**, 024003 (2004)
68. M.-T. Jaekel, S. Reynaud, Class. Q. Grav. **22**, 2135 (2005)
69. M.-T. Jaekel, S. Reynaud, Class. Q. Grav. **23**, 777 (2006)
70. M.-T. Jaekel, S. Reynaud, Class. Q. Grav. **23**, 7561 (2006)
71. P. Jordan, Z. Phys. **157**, 112 (1959)
72. R.A. Knop et al. (Supernova Cosmology Project), Astrophys. J. **598**, 102 (2003)
73. M. Kramer, I.H. Stairs, R.N. Manchester, M.A. McLaughlin, A.G. Lyne, R.D. Ferdman, M. Burgay, D.R. Lorimer, A. Possenti, N. D'Amico, J.M. Sarkissian, G.B. Hobbs, J.E. Reynolds, P.C.C. Freire, F. Camilo, Science **314**, 97 (2006)
74. C. Lämmerzahl, O. Preuss, H. Dittus, arXiv:gr-qc/0604052v1 (2006)
75. S.K. Lamoreaux, J.P. Jacobs, B.R. Heckel, F.J. Raab, E.N. Fortson, Phys. Rev. Lett. **57**, 3125 (1986)
76. A.G. Lyne, M. Burgay, M. Kramer, A. Possenti, R.N. Manchester, F. Camilo, M.A. McLaughlin, D.R. Lorimer, N. D'Amico, B.C. Joshi, J. Reynolds, P.C.C. Freire, Science **303**, 1153 (2004)
77. S.S. McGaugh, Phys. Rev. Lett. **95**, 171302 (2005)
78. Y. Mellier, Ann. Rev. Astron. Astrophys. **37**, 127 (1999)
79. M. Milgrom, Astrophys. J. **270**, 365 (1983)
80. M. Milgrom, Ann. Phys. (N.Y.) **229**, 384 (1994)
81. M. Milgrom, Phys. Lett. A **253**, 273 (1999)
82. C.W. Misner, K.S. Thorne, J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973)
83. G.D. Moore, A.E. Nelson, J. High Energy Phys. **09**, 023 (2001)
84. W.-T. Ni, Phys. Rev. D **7**, 2880 (1973)
85. M.M. Nieto, S.G. Turyshev, Class. Q. Grav. **21**, 4005 (2004)
86. K. Nordtvedt, Astrophys. J. **161**, 1059 (1970)

87. K. Nordtvedt, *Phys. Rev. D* **49**, 5165 (1994)
88. M. Ostrogradski, *Mem. Ac. St. Petersburg Rev.* **VI 4**, 385 (1850)
89. J.D. Prestage, J.J. Bollinger, W.M. Itano, D.J. Wineland, *Phys. Rev. Lett.* **54**, 2387 (1985)
90. A.G. Riess, L.-G. Strolger, J. Tonry, S. Casertano, H.C. Ferguson, B. Mobasher, P. Challis, A.V. Filippenko, S. Jha, W. Li, R. Chornock, R.P. Kirshner, B. Leibundgut, M. Dickinson, M. Livio, M. Giavalisco, C.C. Steidel, N. Benitez, Z. Tsvetanov, *Astrophys. J.* **607**, 665 (2004)
91. V.C. Rubin, N. Thonnard, W.K. Ford Jr., *Astrophys. J.* **225**, L107 (1978)
92. C. Salomon, C. Veillet, in *1st Symp. on the Utilization of the International Space Station*, Darmstadt 30 September - 2 October 1996, ESA SP **385** (ESA, Noordwijk, 1997), p. 397
93. R.H. Sanders, *Astrophys. J.* **480**, 492 (1997)
94. R.H. Sanders, *Mon. Not. R. Astron. Soc.* **363**, 459 (2005)
95. R.H. Sanders, S.S. McGaugh, *Ann. Rev. Astron. Astrophys.* **40**, 263 (2002)
96. I.I. Shapiro, in *Proc. 12th General Relativity and Gravitation Int. Conf.*, ed. by N. Ashby, D.F. Bartlett, W. Wyss (Cambridge University Press, Cambridge, 1990), p. 313
97. S.S. Shapiro, J.L. Davis, D.E. Lebach, J.S. Gregory, *Phys. Rev. Lett.* **92**, 121101 (2004)
98. A.I. Shlyakhter, *Nature* **264**, 340 (1976)
99. J.Z. Simon, *Phys. Rev. D* **41**, 3720 (1990)
100. M.E. Soussa, R.P. Woodard, *Class. Q. Grav.* **20**, 2737 (2003)
101. D.N. Spergel, L. Verde, H.V. Peiris, E. Komatsu, M.R. Nolta, C.L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S.S. Meyer, L. Page, G.S. Tucker, J.L. Weiland, E. Wollack, E.L. Wright, *Astrophys. J. Suppl.* **148**, 175 (2003)
102. D.N. Spergel, R. Bean, O. Doré, M.R. Nolta, C.L. Bennett, J. Dunkley, G. Hinshaw, N. Jarosik, E. Komatsu, L. Page, H.V. Peiris, L. Verde, M. Halpern, R.S. Hill, A. Kogut, M. Limon, S.S. Meyer, N. Odegard, G.S. Tucker, J.L. Weiland, E. Wollack, E.L. Wright, *Astrophys. J. Suppl.* **170**, 377 (2007)
103. I.H. Stairs, S.E. Thorsett, J.H. Taylor, A. Wolszczan, *Astrophys. J.* **581**, 501 (2002)
104. C. Talmadge, J.P. Berthias, R.W. Hellings, E.M. Standish, *Phys. Rev. Lett.* **61**, 1159 (1988)
105. J.L. Tonry et al. (Supernova Search Team), *Astrophys. J.* **594**, 1 (2003)
106. R.B. Tully, J.R. Fisher, *Astron. Astrophys.* **54**, 661 (1977)
107. S.G. Turyshev, V.T. Toth, L.R. Kellogg, E.L. Lau, K.J. Lee, *Int. J. Mod. Phys. D* **15**, 1 (2006)
108. J.-P. Uzan, *Rev. Mod. Phys.* **75**, 403 (2003)
109. R.F.C. Vessot, M.W. Levine, *Gen. Rel. Grav.* **10**, 181 (1978)
110. R.F.C. Vessot, M.W. Levine, E.M. Mattison, E.L. Blomberg, T.E. Hoffman, G.U. Nystrom, B.F. Farrel, R. Decher, P.B. Eby, C.R. Baugher, J.W. Watts, D.L. Teuber, F.D. Wills, *Phys. Rev. Lett.* **45**, 2081 (1980)
111. R. Wagoner, *Phys. Rev. D* **1**, 3209 (1970)
112. J.M. Weisberg, J.H. Taylor, *ASP Conf. Proc.* **302**, 93 (2003)
113. C.M. Will, *Theory and Experiment in Gravitational Physics*, revised edition (Cambridge University Press, Cambridge, 1993)
114. C.M. Will, *Living Rev. Rel.* **9**, URL: <http://www.livingreviews.org/lrr-2006-3>
115. C.M. Will, K. Nordtvedt, *Astrophys. J.* **177**, 757 (1972)
116. J.G. Williams, S.G. Turyshev, D.H. Boggs, *Phys. Rev. Lett.* **93**, 261101 (2004)
117. R.P. Woodward, in *The Invisible Universe: Dark Matter and Dark Energy*, ed. by L. Papantonopoulos, *Lect. Notes Phys.* **720** (Springer, Berlin, 2007), p. 403
118. T.G. Zlosnik, P.G. Ferreira, G.D. Starkman, *Phys. Rev. D* **75**, 044017 (2007)

Mass, Inertia, and Gravitation

Marc-Thierry Jaekel and Serge Reynaud

Abstract We discuss some effects induced by quantum field fluctuations on mass, inertia, and gravitation. Recalling the problem raised by vacuum field fluctuations with respect to inertia and gravitation, we show that vacuum energy differences, such as Casimir energy, do contribute to inertia. Mass behaves as a quantum observable and in particular possesses quantum fluctuations. We show that the compatibility of the quantum nature of mass with gravitation can be ensured by conformal symmetries, which allow one to formulate a quantum version of the equivalence principle. Finally, we consider some corrections to the coupling between metric fields and energy–momentum tensors induced by radiative corrections. Newton’s gravitation constant is replaced by two different running coupling constants in the sectors of traceless and traced tensors. There result metric extensions of general relativity (GR), which can be characterized by modified Ricci curvatures or by two gravitation potentials. The corresponding phenomenological framework extends the usual parametrized post-Newtonian (PPN) one, with the ability to remain compatible with classical tests of gravity while accounting for new features, such as Pioneer-like anomalies or anomalous light deflection.

1 Introduction

The relativistic conception of motion in space-time introduced by Einstein leads to consider inertial mass as a general property shared by all forms of energy [28, 29]. The law of inertial motion reflects the underlying symmetries of space-time [27].

M.-T. Jaekel (✉)

Laboratoire de Physique Théorique de l’ENS, 24 Rue Lhomond, F75231 Paris Cedex 05,
Centre National de la Recherche Scientifique (CNRS), Ecole Normale Supérieure (ENS),
Université Pierre et Marie Curie (UPMC)
e-mail: Marc.Jaekel@lpt.ens.fr

S. Reynaud

Laboratoire Kastler Brossel, Case 74, Campus Jussieu, F75252 Paris Cedex 05,
CNRS, ENS, UPMC
e-mail: Serge.Reynaud@spectro.jussieu.fr

Simultaneously to the development of relativity theory, Einstein laid the basis of a description of Brownian motion, that is, motion in a fluctuating environment [26]. A fundamental role is played by fluctuation-dissipation relations, which allow one to derive the force felt by a moving body from the fluctuating force exerted by its environment. Fluctuation-dissipation relations have been shown to admit a quantum extension [15, 67], which is well captured by the formalism of linear response theory [69]. It also appears that the vacuum state of quantum fields [83] is just a particular case of a thermal equilibrium state, in the limit of zero temperature. Motion in vacuum can either be considered from a relativistic point of view, using the space-time symmetries of empty space, or from a quantum point of view, using the response properties of vacuum with respect to motion. Then, the question naturally arises of testing the compatibility between these different approaches and of the consequences for the notion of inertia.

When treated in a naive way, the vacuum energy of quantum fields leads to difficulties at the cosmological level, due to the incompatibility of a very large value of vacuum energy with gravitational laws according to general relativity (GR). In this framework, the vacuum energy density is associated with the cosmological constant and observational evidence pleads for a small value of the latter [97, 101]. Astrophysical and cosmological observations suggest that, at least within GR, the energy of vacuum fields should not be treated as other forms of energy. A most radical position in this respect has been early advocated by Pauli [80] (see [32] for an English translation): “At this point it should be noted that it is more consistent here, in contrast to the material oscillator, not to introduce a zero-point energy of $\frac{1}{2}h\nu$ per degree of freedom. For, on one hand, the latter would give rise to an infinitely large energy per unit volume due to the infinite number of degrees of freedom, on the other hand, it would be principally unobservable since nor can it be emitted, absorbed or scattered and hence, cannot be contained within walls and, as is evident from experience, neither does it produce any gravitational field.” This position follows the remark that since all types of interactions, excepting gravitation, only involve differences of energy, the latter can be evaluated with respect to a minimum, for instance the vacuum, a prescription realized by normal ordering. Despite the fact that such a position cannot be sustained for gravitation, as no unambiguous frame-independent definition of the vacuum state is available in present quantum field theories [10, 35], it also appears to be wrong for other types of interaction, as we show in the following.

Vacuum fluctuations of the electromagnetic field are known to have observable consequences on atoms or microscopic scatterers [19, 44]. Spontaneous emission of atoms and Lamb shifts of frequencies have been accurately measured, in good agreement with theoretical predictions. Van der Waals forces, which result from the scattering of electromagnetic field vacuum fluctuations by atoms, play a crucial role in physicochemical processes and Casimir forces between two macroscopic electromagnetic plates, or mirrors, due to vacuum field fluctuations [18] have now been measured with good accuracy [12]. We show below that a consistent treatment of quantum field fluctuations and of the relativistic motion of a scatterer entails that

the energy of vacuum fluctuations must be taken into account in inertial effects. This result is obtained within the linear response formalism and follows from a consistent treatment of motion and of quantum field fluctuations [45, 51, 59].

Compatibility may be maintained between quantum and relativity theories, with the consequence of promoting mass to the status of a quantum observable. The quantum nature of the mass observable comes with a similar property for the observables describing positions in space and time. This also conflicts with the status given to space-time positions in Quantum Field Theory (QFT). The relativistic generalization of quantum theory has led to attribute to spatial positions the same representation as to time, that is, to represent space-time positions as underlying parameters. This position has also been enforced by arguments stating the inconsistency of defining a time operator [80, 102]. We shall also show that these arguments can be bypassed and that the relativistic requirement of dealing with observables and the quantum representation of the latter as operators can both be satisfied [56, 65]. As this property also ensures compatibility with the equivalence principle [58], this revives the question of the effects of quantum fluctuations on gravitation.

As previously remarked, GR, although in remarkable agreement with all tests of gravitation that have been performed, is challenged by observations at galactic and cosmological scales. Besides difficulties with the cosmological constant, anomalies affect the rotation curves of galaxies [3, 87] and, as observed more recently, the relation between redshifts and luminosities for type II supernovae [81]. These anomalies may be accounted for by keeping the theory of gravitation unchanged and by introducing unseen components under the form of dark matter and dark energy. As long as dark components remain unobserved by direct means, they are equivalent to deviations from GR occurring at large scales [17, 72, 90]. In this context, the anomaly that has been observed on Doppler tracking data registered on the Pioneer 10/11 probes [4], during their travel to the outer parts of the solar system, constitutes a further element of questioning.

Besides its geometric setting, gravitation may be treated within the framework of QFT as other fundamental interactions [33, 95, 99]. Then, its coupling to energy–momentum tensors leads to modifications that are induced by quantum field fluctuations [16, 23, 98]. These effects, or radiative corrections, induced on gravitation by quantum fluctuations of energy–momentum tensors are similar to those induced on motion and may similarly be treated within the linear response formalism [54]. As expected, these modifications affect the nature of gravitation at small length scales [2, 89, 93], but they may also affect its behavior at large length scales [22, 38]. GR must then be considered as an effective theory of gravity valid at the length scales for which it has been accurately tested [105] but not necessarily at other scales where deviations may occur [70]. The modified theory, while remaining in the vicinity of GR, entails deviations that may have already been observed, and may be responsible for the Pioneer anomaly [61]. This anomaly, which has escaped up to now all attempts of explanation based on the probes themselves or their spatial environment [77] may point at an anomalous behavior of gravity already occurring at scales of the order of the size of the solar system. As these scales cover the domain where GR has been tested most accurately, the confrontation of the Pio-

neer anomaly with classical tests of gravitation provides a favorable opportunity for constraining the possible extensions of gravitation theory [64].

This chapter contains three main sections. In Sect. 2, we review how inertial motion can be treated in a way that remains consistent both with the relativistic and quantum frameworks. We obtain that vacuum energy differences, such as Casimir energy, do contribute to inertia. We discuss in Sect. 3 the quantum properties of mass, in particular its fluctuations, induced by quantum field fluctuations, and its representation as an operator, belonging to an algebra which contains both quantum observables and the generators of space-time symmetries. Finally, in Sect. 4, we review the main features of metric extensions of GR that result from radiative corrections. We show that they lead to an extended framework, which has the ability to remain compatible with present gravity tests while accounting for Pioneer-like anomalies and predicting other related anomalies that could be tested.

2 Vacuum Fluctuations and Inertia

Most sources of conflicts between QFT and GR can be traced back to infinities arising when considering the contributions of quantum fields in vacuum [83]. As a result, although vacuum fields are responsible for now well-established mechanical effects, such as Casimir forces [18], the infinite value of vacuum energy has led to question its contribution to relativistic effects associated with the energy, including inertia and gravitation [80]. One usually admits that only energy differences are observable, which justifies to ignore the contribution of vacuum fields to energy, using a normal ordering prescription for all operators built on quantum fields, such as energy–momentum tensors. However, this prescription hardly applies to gravitation, due to nonlinearities of GR and to the impossibility to define vacuum in covariant way [10, 35]. Observations of the universe at very large scales also plead for considering the role played by vacuum energy [97]. In this section, we show that the energy corresponding to Casimir forces does indeed contribute to inertia, in full agreement with the relativistic law of inertia of energy [29]. There results that one cannot ignore the contribution of vacuum energy to the inertial effects associated with quantum fields. Furthermore, mass is affected by intrinsic quantum fluctuations and cannot be considered any more as a mere constant parameter, but should be represented as a quantum observable, that is, by an operator.

2.1 *Linear Response Formalism*

The appropriate formalism for dealing with the effects of vacuum fluctuations on inertia is available in QFT. A physical system in space-time can be considered in a general way as a scatterer of quantum fields. Scattering is completely determined by the interaction part of the Lagrangian or Hamiltonian of the whole system,

scatterer plus fields. A localized system may also be represented as a set of boundary conditions for fields propagating in space-time [36], but this simplified description appears to be too crude and a limiting case of the general description. Space-time evolution of the whole system is best described in the interaction picture [7, 44], where a unitary evolution operator provides the required quantum description of all observables and the large time limit gives the scattering matrix for fields. All space-time properties of the scatterer, and in particular its motion, may be obtained from general transformations of the interaction part of the Lagrangian or Hamiltonian.

To be more explicit, a physical system in space-time may be seen as performing a transformation of incoming free fields Φ^{in} into interacting fields Φ , which at large time become outgoing free fields again, that is, scattered fields Φ^{out} . Interacting fields Φ evolve according to a total Hamiltonian H that includes a free part H_0 and the interaction part H_I . The evolution in time t of the interacting fields $\Phi(t)$ is then described as an operator S_t acting on incoming fields $\Phi^{in}(t)$. Free fields evolve according to the free part H_0 of the Hamiltonian (\hbar is Planck constant, $[,]$ denotes the commutator of two operators)

$$\begin{aligned} \frac{d\Phi}{dt} &= -\frac{i}{\hbar}[H, \Phi], & H &\equiv H(\Phi), \\ \frac{d\Phi^{in}}{dt} &= -\frac{i}{\hbar}[H_0^{in}, \Phi^{in}], & H_0^{in} &\equiv H_0(\Phi^{in}), \\ \Phi(t) &\equiv S_t^{-1}\Phi^{in}(t)S_t. \end{aligned} \tag{1}$$

The evolution operator S_t is obtained from the interaction part H_I of the Hamiltonian by solving a differential equation resulting from the evolution of interacting and free fields (1) [44]

$$\begin{aligned} \frac{dS_t}{dt} &= \frac{i}{\hbar}H_I^{in}(t)S_t, & H &\equiv H_0 + H_I, & H_I^{in} &\equiv H_I(\Phi^{in}), \\ S_t &= \mathcal{T} \exp \frac{i}{\hbar} \int_{-\infty}^t H_I^{in}(t')dt'. \end{aligned} \tag{2}$$

\mathcal{T} denotes time ordering for the different terms building the exponential, while H_I^{in} denotes the interaction Hamiltonian written in terms of input fields Φ^{in} ($H_I(t) = S_t^{-1}H_I^{in}(t)S_t$ is the same expression, but written in terms of interacting fields Φ). Equations 2 determine the interacting fields in terms of input fields only. Similarly, output fields are determined by a scattering matrix S , which is the large time limit of the evolution operator S_t

$$\Phi^{out}(t) = S^{-1}\Phi^{in}(t)S, \quad S = S_\infty. \tag{3}$$

This description applies to a scatterer at rest or in motion indifferently, the only difference between the two cases lying in a modification of the interaction Hamiltonian describing the scatterer. All properties of the scatterer with respect to its localization or motion in space-time are encoded in the interaction term.

Motions correspond to modifications of the interaction Hamiltonian and, as long as they can be considered as small perturbations, may be treated using the formalism of linear response theory [67]. As can be seen from evolution equations (1, 2), a small perturbation of the interaction Hamiltonian (proportional to a time-dependent parameter $\delta\lambda(t)$) induces a perturbation of the evolution operator, hence of all observables that are built on interacting fields:

$$\begin{aligned}
 S_t^{-1}\delta S_t &= \frac{i}{\hbar} \int_{-\infty}^t \delta H_I(t') dt', & S_t^{-1}\delta H_I^{\text{in}}(t)S_t &= \delta H_I(t) \equiv B(t)\delta\lambda(t), \\
 \delta A(t) &= [A(t), S_t^{-1}\delta S_t] = \int_{-\infty}^{\infty} \frac{i}{\hbar} \theta(t-t') [A(t), B(t')] \delta\lambda(t') dt', \\
 \theta(t) &\equiv 0 \quad \text{for } t < 0, \quad \theta(t) \equiv 1 \quad \text{for } t > 0, \\
 \langle \delta A(t) \rangle &= \int_{-\infty}^{\infty} \chi_{AB}(t, t') \delta\lambda(t') dt', \\
 \chi_{AB}(t, t') &= \frac{i}{\hbar} \theta(t-t') \langle [A(t), B(t')] \rangle. \tag{4}
 \end{aligned}$$

The linear response of an observable A to a perturbation may be written as the action of a generator B and the response is captured in a susceptibility function χ_{AB} , which only depends on the commutator of B with A . Evolution equations (2) and hence the general form (4) of linear response hold independently of the particular form taken by the interaction Hamiltonian. Causality of responses follows from the time ordering prescription entering the definition of the evolution operator (2), as is made explicit in linear responses (4) by the presence of Heaviside step function θ in the time domain. When response functions are written in the frequency domain (after a Fourier transformation), causality equivalently corresponds to analyticity in the upper half of the complex plane. Kramers–Kronig relations then allow one to relate the real and imaginary parts of the susceptibility function

$$\begin{aligned}
 \chi(t) &\equiv \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} e^{-i\omega t} \chi[\omega], & \chi &\equiv \text{Re}(\chi) + i\text{Im}(\chi), \\
 \chi[\omega] &= \int_{-\infty}^{\infty} \frac{d\omega'}{\pi} \frac{\text{Im}(\chi)[\omega']}{\omega' - \omega - i\varepsilon}. \tag{5}
 \end{aligned}$$

For simplicity, invariance under time translation has been assumed; $\varepsilon \rightarrow 0^+$ defines the integration contour in the frequency complex plane.

In general, as can be seen on explicit cases dealing with quantum fields, short time singularities occur which generate infinities in time ordered products (2). These infinities take their origin in divergences occurring at high frequencies and are well known in QFT where they are treated by renormalization. Renormalization amounts to modify the interaction with additional counterterms to compensate the divergencies and by imposing renormalization conditions on final observables to fix the resulting ambiguities [44]. These counterterms modify the $\theta(t-t')$ factor

defining time ordered products (they are localized at equal times $t = t'$) and affect the reactive part of the susceptibility. They amount to modify dispersion relations (5) by subtractions (of a finite number of terms of a Laurent expansion of the response function) and the resulting ambiguities are raised with the renormalization prescriptions. Equations 4 then correspond to unambiguous relations, in the frequency domain, only between the dissipative part $Im(\chi_{AB})$ of the susceptibility and the commutator ξ_{AB}

$$\begin{aligned} \xi_{AB}(t-t') &\equiv \frac{1}{2\hbar} \langle [A(t), B(t')] \rangle \equiv \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} e^{-i\omega(t-t')} \xi_{AB}[\omega], \\ Im(\chi_{AB})[\omega] &= \xi_{AB}[\omega]. \end{aligned} \quad (6)$$

Determining the reactive part $Re(\chi_{AB})$ of the susceptibility requires either to compute the susceptibility or to use dispersion relations (5) with subtractions and additional constraints. Both ways must treat the same infinities and the additional constraints are provided by the prescriptions accompanying renormalization. Renormalization allows one to treat infinities at the expense of a detailed treatment of the interaction, at least at a perturbative level. In the following, we shall focus on characterizing responses in the most general way as possible, without entering a detailed description of interactions, and shall let aside the questions raised by criteria to be fulfilled for renormalizability.

Equations 4 and 6 show that the response of a system to a perturbation is strongly constrained by the quantum correlations of observables describing the unperturbed system. When the latter is in a vacuum state, quantum correlations may be further characterized, as vacuum can be seen as a thermal state at the limit of zero temperature. For a general thermal equilibrium, namely for a state described by a thermal density matrix, quantum correlations satisfy Kubo–Martin–Schwinger (KMS) relations [15, 44, 67] (T is the temperature, Boltzmann constant is set equal to 1, Tr is the trace in the Hilbert space providing mean values for observables)

$$\begin{aligned} C_{AB}(t-t') &\equiv \langle A(t)B(t') \rangle - \langle A(t) \rangle \langle B(t') \rangle, \\ \langle A \rangle &\equiv \text{Tr}(\rho A), \quad \rho = \frac{e^{-\frac{\hbar H}{T}}}{\text{Tr}\left(e^{-\frac{\hbar H}{T}}\right)}, \\ 2\hbar\xi_{AB}(t) &\equiv C_{AB}(t) - C_{BA}(-t), & 2\hbar\sigma_{AB}(t) &\equiv C_{AB}(t) + C_{BA}(-t), \\ 2\hbar\xi_{AB}[\omega] &= (1 - e^{-\frac{\hbar\omega}{T}})C_{AB}[\omega], & \sigma_{AB}[\omega] &= \coth\left(\frac{\hbar\omega}{2T}\right)\xi_{AB}[\omega]. \end{aligned} \quad (7)$$

In the zero temperature limit, these relations characterize the vacuum state and allow one to deduce all correlation functions from commutators (sgn is the sign function)

$$C_{AB}[\omega] = 2\hbar\theta(\omega)\xi_{AB}[\omega], \quad \sigma_{AB}[\omega] = \text{sgn}(\omega)\xi_{AB}[\omega]. \quad (8)$$

As expected for fluctuations in the ground state, correlations are stationary, with a spectrum limited to positive frequencies only. In particular, this entails

that the commutator in the time domain cannot vanish, showing the irreducible non-commuting character of quantum fluctuations. In the case of correlations of free quantum fields, commutators become pure numbers and mean values may be omitted

$$\begin{aligned} 2\hbar\xi_{\Phi_{\text{in}}(x)\Phi_{\text{in}}(x')}(t-t') &= [\Phi_{\text{in}}(t, x), \Phi_{\text{in}}(t', x')], \\ \chi_{\Phi\Phi}(t-t', x-x') &= 2i\theta(t-t')\xi_{\Phi_{\text{in}}(x)\Phi_{\text{in}}(x')}(t-t'). \end{aligned} \quad (9)$$

The response of a quantum field to its source, that is, the retarded propagator, is given by the correlation function associated with the field commutator.

2.2 Response to Motions

In this part, we shall only consider physical systems evolving in a flat space-time. Hence, displacements of these systems, including quantum fields, may be associated with symmetries of space-time. According to Noether theorem, when evolution, that is, propagation of fields and interaction, satisfies some symmetries, the corresponding generators are associated with conserved quantities [44]. Invariance under space-time translations implies conservation of energy–momentum, which corresponds to the generator of translation symmetry. In quantum theory, the generators of space-time symmetries acting on quantum observables and the corresponding conserved quantities are built from the energy–momentum tensor.

For a system made of quantum fields and a scatterer coupled to them, the generator of space translation for the whole system identifies with the total momentum, which is conserved when the coupling between fields and scatterer is a scalar, that is, a space-time invariant. This property may be stated in an equivalent way as the invariance under translation of the interaction part in the total Lagrangian or Hamiltonian. Hence, the momenta of field and scatterer, which generate their respective translations in space, have opposite actions. Then, according to the previous section, perturbation of the interaction Hamiltonian H_I under displacements of the scatterer δq is provided by the commutator of the field momentum P with H_I (4)

$$\delta H_I = -\frac{i}{\hbar}[H_I, P]\delta q = -\frac{i}{\hbar}[H, P]\delta q = F\delta q, \quad F \equiv \frac{dP}{dt} = -\frac{i}{\hbar}[H_I, P]. \quad (10)$$

Free field propagation being invariant under space-time translation, the field momentum P is conserved during propagation, hence commutes with H_0 . As shown by Eqs. 10, small motions δq of the scatterer and their effect on scattered quantum fields (4) are completely determined by a single observable F , the force felt by the scatterer. Then, perturbations of observables due to the scatterer's motions are obtained from their commutator with the force, that is, the radiation pressure, exerted by scattered fields.

In particular, when a scatterer is set into motion, the force exerted by scattered fields undergoes itself a perturbation δF which is related to the scatterer's displacement δq . According to the linear response equations (4, 6) and the general form (10) of the generator of displacements, this relation may be written as a motional susceptibility χ_{FF} that is determined by fluctuations of the force ξ_{FF} , exerted on the scatterer at rest

$$\begin{aligned} \langle \delta F(t) \rangle &= \int_{-\infty}^{\infty} \chi_{FF}(t-t') \delta q(t') dt', \\ \chi_{FF}(t) &= 2i\theta(t)\xi_{FF}(t), \quad \xi_{FF}(t-t') = \frac{1}{2\hbar} \langle [F(t), F(t')] \rangle. \end{aligned} \quad (11)$$

A warning remark must be made at this point. As underlined in previous section when discussing the causality properties of response functions (5), the relation (11) between motional susceptibilities and force correlations in fact suffers from ambiguities due to necessary subtractions in Kramers–Kronig relations. These are related to ambiguities affecting the time ordering defining the evolution operator (2) at the limit of equal times, due to a necessary treatment of divergences. The additional constraints allowing one to raise these ambiguities are equivalent to the renormalization prescriptions following the regularization of infinities. The same ambiguities affect the definition (10) of the generator of motional perturbations, built in terms of generators of space-time symmetries. Then, a complete and self-consistent treatment of motional perturbations would require a discussion in terms of renormalized quantities, including in particular the generator of motional perturbations, that is, the force F . As we shall not enter such a discussion in the present article, we must keep in mind that some defects might arise, which are due to an incomplete description and which can be cured by a more detailed treatment involving renormalization of physical quantities.

The force induced by quantum fields on a moving scatterer does not depend on the type of coupling describing the interaction between scatterer and fields, but only on the balance of energy–momentum resulting from overall energy–momentum conservation. This general property may be seen as a consequence of the fundamental connection between motions in space-time and symmetries which underlies relativity theory. Response functions to motion are then determined by correlation functions of the energy–momentum tensor of scattered quantum fields. One then expects that the description of forces induced by motion in vacuum fields remains consistent with the general principles of relativity theory and in particular with the law of inertia of energy [28]. In the following, we show that this law is indeed respected by vacuum energies, by studying exemplary cases built with mirrors and cavities.

For the sake of simplicity, and in order to discuss explicit expressions, we shall illustrate the following arguments on a simplified description of a scatterer [49]. Without essential reduction of generality, we shall consider quantum fields in a two-dimensional space-time, that is, fields propagating in a single spatial dimension ($\varphi(t-x)$ and $\psi(t+x)$ for the two different directions)

$$\begin{aligned}\phi(t, x) &\equiv \varphi(t - x) + \psi(t + x), & \Phi[\omega] &\equiv \begin{bmatrix} \varphi[\omega] \\ \psi[\omega] \end{bmatrix}, \\ \xi_{\Phi_{\text{in}}\Phi_{\text{in}}}[\omega] &= \begin{bmatrix} \xi_{\varphi_{\text{in}}\varphi_{\text{in}}}[\omega] & 0 \\ 0 & \xi_{\psi_{\text{in}}\psi_{\text{in}}}[\omega] \end{bmatrix} = \frac{1}{\omega}.\end{aligned}\quad (12)$$

The scatterer will also be represented by a scattering matrix S mixing the two counterpropagating directions and corresponding to a scatterer localized at a spatial position q . Considering scatterers with a mass that is large when compared with the field energy, we shall also neglect recoil effects, so that the scattering matrix (3) becomes a quadratic form of input fields and may be rewritten under the simple form of a 2×2 matrix acting on field components (written in the frequency domain)

$$\Phi_{\text{out}}[\omega] = S[\omega]\Phi_{\text{in}}[\omega], \quad S[\omega] \equiv \begin{bmatrix} s[\omega] & r[\omega]e^{-i\omega q} \\ r[\omega]e^{i\omega q} & s[\omega] \end{bmatrix}. \quad (13)$$

$r[\omega]$ and $s[\omega]$, respectively, describe the frequency-dependent reflection and transmission coefficients associated with the field scattering, the latter being defined in a reference frame that is comoving with the scatterer; q describes the spatial position of the scatterer in this frame. Besides its usual analyticity and unitarity properties ($|r|^2 + |s|^2 = 1$), the scattering matrix S will also be assumed to approach the identity matrix above some frequency cutoff, smaller than the scatterer's mass. This condition corresponds to transparency at high frequencies and, besides being indeed satisfied by real mirrors, allows one to neglect recoil effects. Furthermore, this property will provide a natural (physical) high frequency regulator, thus leading to finite results.

The energy–momentum tensor $(T^{\mu\nu})_{\mu, \nu=0,1}$ of propagating fields (12) is determined from their space-time derivatives

$$\begin{aligned}T^{00} = T^{11} &= \frac{1}{2}(\partial_t \phi^2 + \partial_x \phi^2) = \dot{\varphi}^2 + \dot{\psi}^2, & \dot{\varphi} &\equiv \partial_t \varphi, & \dot{\psi} &\equiv \partial_t \psi, \\ T^{01} = T^{10} &= -\partial_t \phi \partial_x \phi = \dot{\varphi}^2 - \dot{\psi}^2.\end{aligned}\quad (14)$$

Hence, the force F exerted on the scatterer is obtained from the balance of momenta of incoming and outgoing fields (of energy–momentum fluxes through the scatterer's surface)

$$F(t) = \dot{\varphi}_{\text{in}}^2(t - q) - \dot{\varphi}_{\text{out}}^2(t - q) - \dot{\psi}_{\text{in}}^2(t + q) + \dot{\psi}_{\text{out}}^2(t + q), \quad (15)$$

The force exerted on the scatterer in vacuum is then determined from the scattering matrix (3, 13), KMS relations (8) and commutators of incoming free fields (9, 12). Equations 15 and S-matrix unitarity imply that the mean force $\langle F(t) \rangle$ vanishes for a scatterer at rest in vacuum. But, force correlations $\xi_{FF}(t)$ do not vanish and remain finite due to the high frequency transparency of the scattering matrix (13).

Motions of the scatterer $\delta q(t)$ ($\delta q[\omega]$ in the frequency domain) modify the field scattering matrix S . The S-matrix of a moving scatterer can be directly obtained by applying a frame transformation to the S-matrix at rest. The corresponding frame transformation on fields (12), hence on scattering matrices (13), is provided by the change of coordinates that transforms the laboratory frame into a frame that is comoving with the scatterer. The transformation results in a modified scattering matrix $S + \delta S$ which may still be described by a 2×2 matrix, but with changes of frequency

$$\delta \Phi_{\text{out}}[\omega] = \int_{-\infty}^{\infty} \frac{d\omega'}{2\pi} \delta S[\omega, \omega'] \Phi_{\text{in}}[\omega'],$$

$$\delta S[\omega, \omega'] = i\omega' \delta q[\omega - \omega'] \left(S[\omega] \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} S[\omega'] \right). \quad (16)$$

As discussed in previous section, modifications of the scattering matrix (4) may also be derived from the perturbation induced by the generator associated with the force exerted on the scatterer (10). One checks that the latter, a quadratic form of input free fields (15) as the scattering matrix itself (13), induces a perturbation of the scattering matrix (4) which is identical to Eqs. 16. This property illustrates, on this simple scattering model, the general connection between motions and symmetry generators entailed by the principles of relativity theory.

In general, vacuum input fields are transformed by the moving scatterer into output fields that are no more in the vacuum state. In other words, general motions of the scatterer lead to radiation. In return, as a consequence of energy–momentum conservation, the moving scatterer feels a mean radiation reaction force $\langle \delta F \rangle$. For small perturbations, the reaction force is proportional to the scatterer’s motion δq . The perturbed force may be directly obtained from expression (15) and the perturbed scattering matrix (16), and may be expressed in terms of a motional susceptibility [49]

$$\langle \delta F[\omega] \rangle = \chi_{FF}[\omega] \delta q[\omega],$$

$$\chi_{FF}[\omega] = i\hbar \int_0^\omega \frac{d\omega'}{2\pi} \omega'(\omega - \omega') \{ 1 - s[\omega']s[\omega - \omega'] + r[\omega']r[\omega - \omega'] \}. \quad (17)$$

One verifies that the previously discussed relation between the dissipative part of the force susceptibility $Im(\chi_{FF})$ and force fluctuations ξ_{FF} (11) is satisfied by expression (17) [49]. One also notes that, as a result of the analyticity properties of the scattering matrix itself, expression (17) for the motional susceptibility satisfies the analyticity, that is, causality, properties which are characteristic of response functions. At perfect reflection ($r[\omega] \equiv -1, s[\omega] \equiv 0$), one also recovers from Eqs. 17 the known dissipative force for a perfect mirror moving in vacuum, when treated as a boundary condition for quantum fields [36]

$$\langle \delta F(t) \rangle = \frac{\hbar}{6\pi} \frac{d^3}{dt^3} \delta q(t). \quad (18)$$

One remarks that this result has been obtained in the present formalism without explicitly treating the infinite energy of vacuum. This is due to the frequency dependence of the scattering matrix (13), and more precisely to its high frequency transparency. The representation as a scatterer provides a more physical description than the treatment as a boundary condition. But one should remind that, as previously underlined, neither of these treatments really takes into account the fundamental divergences associated with space-time singularities and affecting the evolution of quantum fields. These divergences require a renormalization prescription to be managed in a self-consistent way. One also notes that the general expression (17) for the radiation reaction force in vacuum not only satisfies analyticity properties but also further positivity properties [74], entailed by the ground state nature of the vacuum state (8). As a result, it can be shown [48] that the radiation reaction force described by Eqs. 17 does not produce unstable motions, known as “runaway solutions” [88], as those which would be induced by expression (18).

The simple scattering model discussed here provides an explicit application of the linear response formalism to perturbations induced by motions in space-time. This formalism may further be used to analyze the Brownian motion undergone by a scatterer embedded in vacuum fields [50]. It allows one to obtain relations between fluctuations of positions in vacuum and the susceptibility describing the response of the system to an applied external force.

2.3 Relativity of Motion

The mechanical effects induced on a scatterer by quantum field fluctuations allow for a complete quantum description, using the linear response formalism. Their compatibility with relativity theory can also be checked explicitly. In particular, the energy–momentum balance responsible for the motional response of a scatterer leads to the same result when analyzed in different reference frames. This can be shown using the linear response formalism and the representation of motion as the action of space-time symmetries.

The force exerted by quantum fields on a scatterer (15) is obtained as a balance of momentum between outgoing and incoming fields, hence in terms of the scattering matrix and input field correlations only. For the simple scattering model introduced in previous section, this reads, using expressions (14, 15)

$$\begin{aligned} \Phi_{\text{out}}[\omega] &= S[\omega]\Phi_{\text{in}}[\omega], \\ \langle \Phi_{\text{in}}[\omega]\Phi_{\text{in}}[\omega'] \rangle &\equiv C_{\text{in}}[\omega, \omega'] \quad \rightarrow \quad \langle F[\omega] \rangle \equiv \mathcal{F}\{S, C_{\text{in}}\}[\omega]. \end{aligned} \quad (19)$$

The functional dependence $\mathcal{F}\{S, C_{\text{in}}\}$ of the force F in terms of the scattering matrix and incoming field correlations (19) does not depend on the choice of a

particular reference frame. Scattering may then equivalently be analyzed either in a laboratory frame, where the scatterer is moving, or in a comoving frame where the latter is at rest.

In the laboratory frame, the scatterer’s motion induces a perturbation δS of the scattering matrix. The latter is obtained by applying a transformation to the scattering matrix at rest (13), which corresponds either to a change of coordinates from the comoving frame to the laboratory, or to the action of a displacement generator (10), both transformations providing the same result $\delta S(\delta q)$ (16). When analyzed in the laboratory frame, incoming fields, hence their correlations C_{in} , remain unperturbed by the scatterer’s motion (12), so that the perturbed force reads

$$\begin{aligned} & \langle F[\omega] + \delta F[\omega] \rangle = \mathcal{F}\{S + \delta S(\delta q), C_{in}\}[\omega], \\ \rightarrow & \langle \delta F[\omega] \rangle = \chi_{FF}[\omega]\delta q[\omega]. \end{aligned} \tag{20}$$

In a comoving frame, the scattering matrix S , which describes the coupling between scatterer and fields, remains unchanged (13). But the space-time expression of incoming field correlations now undergoes a modification δC_{in} that can be obtained from field correlations in the laboratory (12) and a change of coordinates from the laboratory to a comoving frame $\delta C_{in}(\delta q)$, so that the perturbed force reads in that case

$$\begin{aligned} & \langle F[\omega] + \delta F[\omega] \rangle = \mathcal{F}\{S, C_{in} + \delta C_{in}(\delta q)\}[\omega], \\ \rightarrow & \langle \delta F[\omega] \rangle = \chi_{FF}[\omega]\delta q[\omega]. \end{aligned} \tag{21}$$

In either reference frame, the perturbation of the force exerted on the scatterer may be expressed as a motional susceptibility χ_{FF} . Both expressions (20) and (21) can be seen to lead to the same result (17) [49]. This property, here illustrated on a simple model, is in fact a general consequence of the principles ruling the evolution of quantum observables (2) and their space-time transformations (10). Interesting applications of this property may be envisaged. For instance, motion can be simulated by keeping the scatterer at rest but by modifying the field fluctuations reflected by the scatterer, using for instance optical devices acting on incoming fields, thus obtaining radiation and a reaction force equivalent to those induced on a scatterer by its motion [49].

When combined with symmetry properties of field fluctuations, relativity of motion leads to important consequences for the radiation reaction force. If the scatterer’s motion corresponds to a generator of a symmetry of field fluctuations, that is, to a frame transformation that leaves field fluctuations invariant, then the radiation reaction force is the same as for a scatterer at rest embedded in the same field fluctuations. This is in particular true for motions with uniform velocity in vacuum, due to the Lorentz invariance of vacuum field fluctuations. As a result, the mean radiation reaction force vanishes for uniform motions in vacuum. In contrast, uniform motions in a thermal bath induce a dissipative reaction force proportional to the scatterer’s velocity, in conformity with the transformation properties of the thermal

bath under a Lorentz transformation. These properties of the radiation reaction force may be checked by explicit computation in the previously discussed model [52].

Quantum field fluctuations in vacuum may be invariant under a larger group than the Poincaré group that is generated by translations and Lorentz symmetries. This is the case for the vacuum fluctuations of a scalar field in two-dimensional space-time or of electromagnetic fields in four-dimensional space-time. Both admit invariance under the action of the larger group of conformal symmetries, which include generators performing transformations to uniformly accelerated frames. As a consequence, the radiation reaction force in vacuum should be invariant under conformal transformations and uniformly accelerated motions should lead to a vanishing radiation reaction force. Indeed, scalar fields in a two-dimensional space-time lead to vacuum field correlations behaving as ω^3 at low frequency, corresponding to a radiation reaction force behaving as the third time derivative of the position (see Eq. 17)

$$\begin{aligned} C_{FF}[\omega] &\sim \theta(\omega)\omega^3, & \omega &\sim 0, \\ \langle \delta F \rangle &\sim \frac{d^3}{dt^3} \delta q(t). \end{aligned} \quad (22)$$

At perfect reflection, these relations hold at all frequencies, as shown by expression (18) [36]. The case of a scatterer embedded in electromagnetic fields $(A_\mu)_{\mu=0,1,2,3}$ in four-dimensional space-time similarly involves conformally invariant vacuum field fluctuations [55], which may be written, after an appropriate gauge transformation

$$C_{A_\mu A_\nu}(x, x') = \frac{\hbar}{\pi} \frac{\eta_{\mu\nu}}{(x-x')^2 - i\varepsilon(t-t')}, \quad \eta_{\mu\nu} \equiv \text{diag}(1, -1, -1, -1), \quad \varepsilon \rightarrow 0^+. \quad (23)$$

As a consequence, the radiation reaction force induced on the scatterer by its motion vanishes for uniform velocities and also for uniform accelerations (τ denotes proper time)

$$\langle \delta F^\mu \rangle \sim \frac{d^3 q^\mu}{d\tau^3} + \left(\frac{d^2 q}{d\tau^2} \right)^2 \frac{dq^\mu}{d\tau}. \quad (24)$$

It can be seen that the radiation reaction force in that case is proportional to a conformally invariant derivative of the scatterer's position, that is, to Abraham-Lorentz vector [55].

2.4 Inertia of Vacuum Fields

One should not conclude from the previous discussion that the vacuum of conformally invariant fields does not induce any motional effect on an accelerated scatterer. In fact, as previously remarked, only the dissipative part (imaginary part

in the frequency domain) of the motional susceptibility is unambiguously related to correlations of quantum fields (6). The reactive part (real part in the frequency domain) of the motional susceptibility is related to the latter through Kramers–Kronig relations (5) and requires more care to be determined. Although the general properties of the dissipative part of the motional force should remain unchanged by a renormalization of observables built on the energy–momentum tensor of quantum fields, those of the reactive part will in general be sensitive to this procedure. The model of a scatterer used in previous sections involves approximations that allow one to bypass the problems raised by coupling to high frequencies, but it appears insufficient to determine in a general way the relation between acceleration and motional forces. For that reason, we first extend the simple model of a single mirror to a model of a cavity, built with two mirrors still modeled in the same way. This model of a cavity will allow us to exhibit a fundamental relation between vacuum fields and the response of an embedded system to accelerated motion. Although still bypassing fundamental space-time singularities associated with quantum fields, the spatial extension introduced by a cavity will appear sufficient to exhibit this relation that was occulted by the approximations underlying the model of a single mirror.

Still considering fields propagating in a single spatial direction (12), each mirror building the cavity can be described by its scattering matrix ($S_i, i = 1, 2$), depending on its position (q_i) (13). As in the case of a single mirror, each scattering matrix determines outgoing fields in terms of incoming ones, the only difference being that intracavity fields are involved and play the role of incoming or outgoing fields, according to the mirror on which they reflect. The scattering matrix S for the total cavity may easily be obtained from the two mirrors’ scattering matrices. There results a relation between the determinants of the different scattering matrices that may be written

$$\det S[\omega] = \det S_1[\omega] \det S_2[\omega] e^{2i\Delta_q[\omega]},$$

$$\Delta_q[\omega] = \frac{i}{2} \text{Log} \frac{1 - r[\omega] e^{2i\omega q/c}}{1 - r[\omega]^* e^{-2i\omega q/c}}. \tag{25}$$

$q = q_2 - q_1$ denotes the spatial distance of the two mirrors, that is, the length of the cavity, and $r[\omega]$ is the product of the frequency-dependent reflection coefficients of the two mirrors. All dependence on the spatial extension of the system is then captured in a phase shift $\Delta_q[\omega]$. This corresponds to the phase shift undergone by incoming fields at frequency ω when scattered by a cavity of length q .

When applied to a cavity at rest, the balance of energy–momentum (14) between intracavity and exterior (input and output) fields leads to a difference between the two mean forces $\langle F_i \rangle, i = 1, 2$ acting on the two mirrors. The resulting force corresponds to the mean Casimir force F_C exerted by field fluctuations between the two sides of the cavity. When computed for input fields in vacuum, the mean Casimir force, and the corresponding Casimir energy, take a simple and suggestive form in terms of scattering matrices or of phase shifts (25) [46]

$$\begin{aligned} \langle F_2 \rangle - \langle F_1 \rangle &= F_C \equiv \partial_q E_C, \\ E_C &= \int_0^\infty \frac{d\omega}{2\pi} \hbar \omega \tau[\omega], \quad \tau[\omega] \equiv \frac{1}{2} \partial_\omega \Delta_q[\omega], \end{aligned} \quad (26)$$

The Casimir energy E_C may be seen as a sum over all field modes (two counterpropagating modes for each frequency ω) of a contribution built from two factors, $\hbar\omega/2$ the vacuum energy at the corresponding frequency and $\tau[\omega]$ the derivative of the phase shift at the same frequency. The factor $\tau[\omega]$ may also be seen as the time delay undergone by a field around frequency ω during its scattering by the cavity. Then, expression (26) for the Casimir energy may be given a simple interpretation: it corresponds to a part of the energy of vacuum fields that is stored inside the cavity during their scattering [46].

As in the case of a single mirror, the forces F_i , $i = 1, 2$ acting on the two mirrors of a cavity show fluctuations in vacuum which may be obtained from the mirrors' scattering matrices (13) and free field correlation functions (12), and which satisfy KMS relations in vacuum (8)

$$C_{F_i F_j}(t) = \langle F_i(t) F_j(0) \rangle - \langle F_i \rangle \langle F_j \rangle. \quad (27)$$

Correlation functions for fluctuations of the Casimir force follow from (27) [47].

Now, we consider that the two mirrors of the cavity are moving independently, and we directly determine the corresponding motional responses. The motions of the two mirrors correspond to perturbations (10) of the Hamiltonian describing their coupling to scattered fields

$$\delta H_I(t) = \sum_i F_i(t) \delta q_i(t). \quad (28)$$

Output and intracavity fields are then modified according to the perturbed scattering matrices (16) and the resulting perturbations $\langle \delta F_i \rangle$ of radiation pressure on the mirrors can be expressed at first order in the mirrors' displacements δq_i under the form of susceptibilities $\chi_{F_i F_j}$

$$\langle \delta F_i[\omega] \rangle = \sum_j \chi_{F_i F_j}[\omega] \delta q_j[\omega]. \quad (29)$$

Each mirror's motion induces a motional force on both mirrors and motional susceptibilities can be checked to satisfy fluctuation-dissipation relations [47] with Casimir force fluctuations

$$\chi_{F_i F_j}[\omega] - \chi_{F_j F_i}[-\omega] = \frac{i}{\hbar} \{C_{F_i F_j}[\omega] - C_{F_j F_i}[-\omega]\}. \quad (30)$$

As a result of their dependence on the energy density of intracavity fields, motional forces can be seen to be resonantly enhanced for motions at proper frequencies of

the cavity ($\omega_n = n \frac{\pi}{q}$). This resonance property has important consequences, as in particular it allows one to envisage experimental setups for exhibiting the motional effects of vacuum fields. Indeed, the relatively small magnitude of these effects in vacuum, as suggested by Casimir forces, can be compensated by the cavity finesse and lead to observable effects [68].

We now focus on global motions of the cavity. The total force induced by a global motion of the cavity may be evaluated in the quasistatic limit, that is, for slow motions corresponding to frequencies lower than the cavity modes. The motional force exerted by vacuum fields on the cavity may then be written as an expansion in the frequency

$$\begin{aligned} \langle \delta F[\omega] \rangle &= \{ \mu \omega^2 + \dots \} \delta q[\omega], \\ \langle \delta F(t) \rangle &= -\mu \frac{d^2}{dt^2} \delta q(t) + \dots, \quad \mu = \frac{1}{2} \sum_{ij} \partial_\omega^2 \chi_{F_i F_j}[0]. \end{aligned} \quad (31)$$

As expected, at zero frequency, the force (31) induced by a quasistatic motion of the mirrors corresponds to a variation of the mean Casimir force (26) due to the variation of the length of the cavity, so that it cancels in the total force. In conformity with Lorentz invariance of vacuum, the mirrors' velocities do not contribute so that the first contribution corresponds to the mirrors' accelerations. At lowest frequency order, the total motional force felt by the cavity (31) then depends on its global acceleration and may be seen as a correction to the mass of the cavity. This mass correction may be expressed in terms of the Casimir energy E_C and the mean Casimir force F_C [51]

$$\mu = \frac{E_C - F_C q}{c^2}. \quad (32)$$

Dependence on c , the light velocity, has been restored for illustrative purposes. One can see that the mass correction (32) precisely corresponds to the contribution of energy to inertia in the case of a stressed rigid body, as required by the law of inertia of energy according to relativity theory [29]. In fact, the law of inertia may be shown to express the conservation of the symmetry generator associated with Lorentz boosts [28]. The mass correction (32) generalizes this law to include vacuum energies.

Although performed on a model of a cavity that cannot be considered as complete (in particular, a complete description should be performed in terms of renormalized quantities), the previous discussion already allows one to conclude that field fluctuations modify the inertial response of a scatterer moving in vacuum. The resulting motional effects are compatible with the principles of relativity and may be consistently computed within the framework of linear response theory. Due to its relation with the fluctuations of quantum fields, the dissipative part of the motional response satisfies the same space-time symmetries as the vacuum state. Moreover, the reactive part of the motional response also conforms to these symmetries, so that the law of inertia of energy also applies to Casimir energies, a special case of the contribution of vacuum fields to energy (26).

3 Mass as a Quantum Observable

The study of the motion of a cavity embedded in quantum field fluctuations confirms the relativistic relation between mass and energy. It also shows that scattering of field fluctuations modifies the mass of a scatterer. This property follows from the principles of relativity theory relating motion with space-time symmetries, when applied within the linear response formalism (4, 10). There result fundamental modifications of the status of mass. Quantum fluctuations and symmetries impose to abandon the classical representation of mass as a mere characteristic parameter and to use the same representation as for all physical observables, that is, as a quantum operator.

3.1 Quantum Fluctuations of Mass

The mass induced by vacuum field fluctuations (31, 32) may be seen as the energy taken on field fluctuations due to time delays induced by scattering (25, 26). The mass correction not only depends on time delays associated with the scattering matrix but also on the energy density of incoming quantum field fluctuations. Following the same line of approach, motion in a thermal bath of quantum fields can be shown to induce, besides the well-known friction force, a mass correction that depends on the bath temperature [52].

The dependence of a scatterer's mass on the energy density of quantum field fluctuations shows that mass cannot be considered any more as a mere parameter characterizing the scatterer. The scatterer's mass inherits the quantum properties that are associated with the fluctuations of scattered fields and hence must be represented by a quantum operator. The contribution induced by scattered fields leads to mass quantum fluctuations which can also be characterized by correlation functions (7)

$$C_{MM}(t - t') \equiv \langle M(t)M(t') \rangle - \langle M(t) \rangle \langle M(t') \rangle,$$

$$C_{MM}(t) \equiv \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} e^{-i\omega t} C_{MM}[\omega]. \quad (33)$$

Correlation functions of the mass operator can be deduced from the linear response formalism used to obtain motional responses in vacuum. Mass correlation functions follow from the scattering matrix of quantum input fields (12, 13) and from correlation functions of their energy-momentum tensor (14).

In order to extend the analysis beyond the case of a cavity and to discuss simple explicit expressions, we consider again the model of a scatterer as in previous sections (12, 13), but now written under the form of an explicitly relativistic Lagrangian \mathcal{L} [53]

$$\mathcal{L} = \frac{1}{2} (\partial_t \phi^2 - \partial_x \phi^2) + \int ds \sqrt{1 - \dot{q}(s)^2} \delta(x - q(s)) M, \quad \dot{q} \equiv \frac{dq}{ds},$$

$$M \equiv M_0 + \Omega \phi^2(q). \quad (34)$$

ϕ is a scalar field propagating in a two-dimensional space-time and q the space-time trajectory, parametrized by s , of a point-like scatterer. The scatterer's mass M is the sum of a bare mass M_0 and of an interaction term localized on the scatterer's trajectory. The bare coupling Ω can also be seen to play the role of a frequency cutoff. Lagrangian (34) describes a general relativistic interaction between a scalar field and a point-like particle, with the further assumption of a quadratic interaction (the following arguments apply to general forms of interaction with minor complications). In the limit of a scatterer at rest, the Lagrangian (34) being quadratic in the fields, the scattering matrix (13) associated with the interaction is easily obtained, providing the corresponding phase shift or time delay

$$s[\omega] = 1 + r[\omega], \quad r[\omega] = -\frac{\Omega}{\Omega - i\omega},$$

$$2\tau[\omega] = \partial_\omega \Delta[\omega] = \frac{2\Omega}{\Omega^2 + \omega^2}. \tag{35}$$

Energy conservation for the system (34) shows that the scatterer's mass undergoes a correction μ , which is indeed related to the frequency-dependent time delay, as for a cavity

$$\mu \equiv \langle \Omega \phi(q)^2 \rangle = \int_0^\infty \frac{d\omega}{2\pi} \hbar \omega \tau[\omega]. \tag{36}$$

The mass correction (36), when written in terms of time delays (35) is infinite, due to an ultraviolet divergence. A counterterm must be added to the bare coupling to compensate this divergence. In fact, expression (35) is a crude approximation of the scattering matrix associated with the system (34), which is only valid when the scatterer remains approximately at rest during scattering, that is, for rather low field frequencies. For fields with an energy of the order of the scatterer's mass, recoil cannot be ignored and the scattering matrix, hence the time delay, differs significantly from (35). In a self-consistent treatment of infinities one must come back to the general definition (2,3) of the scattering matrix and consider a renormalization of physical observables. But the approximation (35) of the scattering matrix appears sufficient to exhibit the essential quantum properties of the mass observable.

The induced mass (36) depends on the energy density of vacuum fields, μ corresponding to its mean value, and exhibits quantum fluctuations. Using the simple model (34, 35), one derives the corresponding quantum fluctuations in the frequency domain (33), which can also be written in terms of time delays

$$C_{MM}[\omega] = 2\hbar^2 \theta(\omega) \int_0^\omega \frac{d\omega'}{2\pi} \omega'(\omega - \omega') \tau[\omega'] \tau[\omega - \omega']. \tag{37}$$

In spite of approximations that have been made, expression (37) for the correlations of mass quantum fluctuations remains valid for frequencies that are smaller than the frequency cutoff Ω . One recovers for the mass observable the characteristic spectrum of quantum fluctuations in vacuum: the factor proportional to $\theta(\omega)$ reduces excitations to positive frequencies only. This results, in the time domain,

into a nonvanishing commutator for the mass observable at different times, implying that mass must be represented by a quantum operator. The quantum properties of the mass observable manifest themselves at short timescales. Indeed, extrapolating (with the restrictions which have been formulated) this simple model to high frequencies, one observes that mass fluctuations cannot be neglected due contributions at short times

$$\langle M^2 \rangle - \langle M \rangle^2 = 2 \langle M \rangle^2. \quad (38)$$

However, the fluctuations of the mass observable become negligible in the low frequency domain. For the simple model (34), variations behave as the third power of frequency

$$C_{MM}[\omega] \sim \frac{\hbar^2}{6\pi} \theta(\omega) \frac{\omega^3}{\Omega^2} \quad \text{for} \quad \omega \ll \Omega. \quad (39)$$

This last property of the mass observable justifies its approximation by a constant parameter, as long as low frequency motions are considered. But fluctuations that come into play at higher frequencies limit the validity of this approximation [53]. This means that in a complete treatment, the renormalized mass of a scatterer is obtained under the form of a quantum operator. In other words, the renormalized mass of a scatterer follows from the renormalized energy–momentum tensor of scattered fields. The value of the parameter used in the renormalization prescription corresponds to the mean value of the renormalized mass observable.

3.2 Mass and Conformal Symmetries

The linear response formalism confirms, within the quantum framework, the relativistic relation between motion and space-time symmetries. Due to this relation, mass cannot be considered as an ordinary quantum observable. The quantum operator associated with mass must remain consistent with the general identification of generators of space-time symmetries with constants of motion. The relativistic relation between energy and inertial mass induced by acceleration has been seen in previous section to hold in a quantum framework and to include the energy due to vacuum fluctuations. This relation should also possess a general expression in terms of quantum operators associated with space-time symmetries. It appears that this property is ensured by the existence of a group of symmetries that extends the Poincaré group of translations and Lorentz transformations to include symmetries associated with accelerations.

In the following, we discuss the specific properties relating mass and acceleration in a quantum framework, in the case of a flat four-dimensional space-time. In a relativistic quantum theory, the generators of space-time symmetries describe changes of reference frame and also correspond to quantities that are preserved by the equations of motion. The transformations of quantum observables under translations

and Lorentz transformations are respectively given by their commutators with the energy–momentum $(P_\mu)_{\mu=0,1,2,3}$ and the angular momentum $(J_{\mu\nu})_{\mu,\nu=0,1,2,3}$. The actions of these generators on observables satisfy the Poincaré algebra, that is, the following commutation relations

$$\begin{aligned} [P_\mu, P_\nu] &= 0, \\ [J_{\mu\nu}, P_\rho] &= i\hbar (\eta_{\nu\rho} P_\mu - \eta_{\mu\rho} P_\nu), \\ [J_{\mu\nu}, J_{\rho\sigma}] &= i\hbar (\eta_{\nu\rho} J_{\mu\sigma} + \eta_{\mu\sigma} J_{\nu\rho} - \eta_{\mu\rho} J_{\nu\sigma} - \eta_{\nu\sigma} J_{\mu\rho}). \end{aligned} \quad (40)$$

$\eta_{\mu\nu} \equiv \text{diag}(1, -1, -1, -1)$ denotes Minkowski metric tensor and determines the light cones defining the causal structure of space-time. Propagations of electromagnetic fields and of gravitation fields, at the linearized level, follow these light cones so that the corresponding solutions of the equations of motion form a representation of the group of light cone symmetries [8, 21]. These correspond to conformal symmetries that include, besides the generators of Poincaré algebra (40), generators corresponding to a dilatation D and to transformations to accelerated frames $(C_\mu)_{\mu,\nu=0,1,2,3}$. The corresponding generators satisfy commutation rules that define the conformal algebra

$$\begin{aligned} [D, P_\mu] &= i\hbar P_\mu, & [D, J_{\mu\nu}] &= 0, \\ [P_\mu, C_\nu] &= -2i\hbar (\eta_{\mu\nu} D - J_{\mu\nu}), & [J_{\mu\nu}, C_\rho] &= i\hbar (\eta_{\nu\rho} C_\mu - \eta_{\mu\rho} C_\nu), \\ [D, C_\mu] &= -i\hbar C_\mu, \\ [C_\mu, C_\nu] &= 0. \end{aligned} \quad (41)$$

The generators of the conformal algebra (41) describe space-time symmetries associated with light cones and field propagation [13, 14]. Hence, they correspond to symmetries of the vacuum state and of motional responses in vacuum fields. They are satisfied in particular by the radiation reaction force in vacuum, as discussed in previous section. As shown in the following, they also allow to analyze the relation between conformal generators and motion in terms of quantum observables. For that purpose, one must first define the quantum observables that can be associated with positions in space-time.

According to the relativistic conception, positions in space and time should be defined as physical observables [27]. Time is delivered by special systems designed for that purpose, that is, clocks. A given set of synchronized clocks and emitters, disseminating time references along propagating signals (using electromagnetic fields for instance), builds a reference system which allows one to determine coordinates in space and time. The several time references received at a given location determine the positions of this location with respect to the reference system. This notion of space-time, based on a realization of positions by means of observables, is implemented nowadays in metrology [39] and in reference systems used for positioning and for navigation around the Earth and in the solar system [6, 84]. However, this implementation differs from the representations of space and time that are used in quantum field theories. In order to maintain the same status for space and time, positions are similarly represented in quantum field theories as mere

parameters. These parameters describe a classical manifold on which quantum fields and physical observables can be defined. In such a representation, positions in space-time have lost the nature of observables as required by relativity theory. But, nothing in principle prevents one from applying the relativistic definition of observables describing space-time positions within the context of quantum field theory. The main trade-off lies in the increased complexity of the observables representing positions in space-time, with the significant advantage of restoring a consistency between the principles of relativity theory and the formalism underlying quantum theory.

Following the relativistic approach, time references may be defined as observables built on the energy–momentum tensor of exchanged fields [57]. Then, localization in space-time by means of quantum fields leads to the definition of space-time positions as quantum observables $(X_\mu)_{\mu=0,1,2,3}$ built from quantities that are conserved by field propagation, that is, from the generators of space-time symmetries [56] (\cdot denotes the symmetrized product of operators)

$$X_\mu = \frac{1}{P^2} \cdot (P^\lambda \cdot J_{\lambda\mu} + P_\mu \cdot D). \quad (42)$$

Space-time positions are represented by operators (42) that belong to an extension of the enveloping algebra of the Lie algebra of conformal symmetries (41). In particular, the definition of space-time positions excludes massless field configurations ($P^2 \equiv M^2 = 0$), that is, its realization requires configurations involving fields propagating in different directions. Then, positions belong to an algebra which is generated by quantum fields and which contains quantum observables and relativistic frame transformations. Positions as defined by (42) are also conjugate to energy–momentum observables

$$[P_\mu, X_\nu] = -i\hbar\eta_{\mu\nu}. \quad (43)$$

One must note at this point that Eqs. 42 define quantum operators, which describe positions not only in space but also in time. Furthermore, the time operator thus defined is conjugate to the energy observable according to (43). As these operators are built on quantum fields that possess a state with minimal energy, the vacuum, all conditions seem satisfied to apply a well-known theorem [80], in its relativistic formulation [102]. This theorem states the impossibility to define a self-adjoint operator conjugate to the energy, when the latter is bounded from below. In fact, one can see that a condition assumed by the theorem, namely the self-adjointness of the time operator, is not fulfilled here. Although this property is often satisfied by quantum observables, it does not appear to be necessary in general [11]. Observables with real eigenvalues only require to be represented by hermitian operators and the definition domains of an operator and its adjoint may differ. This happens in particular when part of the Hilbert space must be excluded from the definition domain, as is the case for the time operator defined by (42). The exclusion of massless field configurations then allows one to escape the objection raised by the theorem and to define a time operator satisfying the required commutation rules with the generators of space-time symmetries [57].

It can be seen that positions thus defined (42) transform according to classical rules under rotations and dilatation

$$\frac{i}{\hbar}[J_{\mu\nu}, X_\rho] = \eta_{\mu\rho}X_\nu - \eta_{\nu\rho}X_\mu, \quad \frac{i}{\hbar}[D, X_\mu] = X_\mu. \quad (44)$$

Transformations to uniformly accelerated frames are then given by the conformal generators C_μ (41).

In conformity with relativity theory [27], the mass observable M is a Lorentz invariant (Poincaré invariant) built on energy–momentum P_μ

$$M^2 = P^\mu P_\mu, \quad \frac{i}{\hbar}[P_\mu, M^2] = \frac{i}{\hbar}[J_{\mu\nu}, M^2] = 0. \quad (45)$$

But the extension of space-time transformations to the conformal algebra, and in particular the action of the dilatation operator, shows that the mass observable cannot be considered as a mere parameter. Mass (45) and the conformal generators (41) are embedded within the same algebra of observables, with mass loosing its invariance property

$$\frac{i}{\hbar}[D, M] = -M. \quad (46)$$

The mass operator (45, 46) can also be seen to provide an extension to the quantum framework of the relativistic relation between positions (42) and the law of inertial motion. The transformation of mass (45) under a uniform acceleration is obtained from the conformal algebra (41) and is seen to involve the position observable

$$\frac{i}{\hbar}[C_\mu, M] = -2M \cdot X_\mu. \quad (47)$$

The transformation (47) of the mass observable takes the same form as the classical red-shift law describing the effect of acceleration on frequencies [29]. Rewriting Eq. 47 as the action of the generator Δ corresponding to an acceleration a^μ on the mass observable M , the result identifies with Einstein law, now written in terms of quantum positions X_μ

$$\Delta \equiv \frac{a^\mu}{2}C_\mu, \quad \frac{i}{\hbar}[\Delta, M] = -M \cdot \Phi, \quad \Phi = a^\mu X_\mu. \quad (48)$$

The transformation (48) of the mass is proportional to the mass itself and to a potential Φ which is given by the product of the acceleration with the quantum position. This quantum version of the classical red-shift law describes the effect on frequencies of an acceleration or an equivalent gravitational potential Φ , according to relativity theory [29]. Conversely, the transformation of mass under accelerations (47), given by the conformal algebra, can be used to define quantum positions, expressions (42) being then recovered [58].

The representation of frame transformations and motions as actions of generators of symmetries ensures that transformations of observables remain compatible

both with relativity and quantum theory. Expression (48) makes it explicit in the case of the transformation of mass under a uniform acceleration, which appears to be equivalent to a uniform gravitational potential. The consistency of both interpretations is ensured in this case by the relation made in the quantum framework between frequency and energy, more precisely, by the conformal invariance of Planck constant \hbar . This property holds, more generally, the particular form (48) of the gravitational potential in terms of quantum positions appearing as a particular case of a quantum generalization of the metric field. Using conformal symmetry, the covariance rules which, in classical theory, implement the equivalence between motion (or changes of reference frames) and gravitation (or metric fields) may be given a generalized form which applies to quantum observables [58]. Symmetries and their associated algebras then provide a way to extend to the quantum framework the equivalence principle which lies at the heart of GR [29].

4 Metric Extensions of GR

The previous part has shown that quantum fluctuations modify the relation of mass to inertial motion and gravitation. In this part, we discuss a similar modification of gravitation that is due to quantum fluctuations of stress tensors and leads to effects that might be observable at the macroscopic level.

In order to discuss this issue, one must first come back to the founding principles of GR. First, the equivalence principle, in its weak version, states the universality of free fall and gives GR its geometric nature. Violations of the equivalence principle are constrained by modern experiments to remain extremely small, below the 10^{-12} level, so that the equivalence principle is one of the best tested properties of nature [105]. The level of precision attained by tests, at least for scales ranging from the submillimeter to a few A.U., disfavors strong violations of the equivalence principle, so that modifications of GR should first be looked for among theories that still obey this principle.

Then, GR may be characterized, as a field theory, by the coupling it assumes between gravitation (or metric fields) and sources. This coupling is equivalent to the gravitation equations that determine the metric tensor from the distribution of energy–momentum in space-time. According to GR, the curvature tensor of the metric and the energy–momentum tensor of sources are in a simple relation [30, 31, 43]. Einstein curvature tensor $E_{\mu\nu}$ is simply proportional to the energy–momentum tensor $T_{\mu\nu}$ and a single constant, Newton gravitation constant G_N , describes the gravitational coupling

$$E_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G_N}{c^4}T_{\mu\nu}. \quad (49)$$

$R_{\mu\nu}$ and R denote the Ricci and scalar curvatures which are built on Riemann curvature tensor. The Einstein equations (49) can be derived from the Einstein–Hilbert Lagrangian simply equal to R . As a result of Bianchi identities, Einstein

curvature tensor has a null covariant divergence, like the energy–momentum tensor, so that Eq. 49 makes a simple connection between a geometric property of curvatures and the physical law of energy–momentum conservation. The latter property identifies with the geodesic motion describing free fall. But these properties could still be satisfied by fixing other relations between curvature and energy–momentum tensors, so that Einstein choice is the simplest but not the only physical possibility.

Indeed, as discussed in the following, even if gravitation is described by GR at the classical level, quantum fluctuations of stress tensors lead to modifications of gravitation equations (49), while preserving the metric nature of the theory.

4.1 Radiative Corrections

Form now on, we focus on gravitation theories that preserve the equivalence principle, that is, metric theories. Space-time is then represented by a four-dimensional manifold, endowed with a metric $(g_{\mu\nu})_{\mu,\nu=0,1,2,3}$, with Minkowskian signature, identifying with gravitation fields. Gravitation may be treated as a field theory that is characterized by its Lagrangian, giving equations of motions for the gravitation fields such as Einstein equations (49) for GR [100].

Einstein equations (49) describe the propagation of gravitation fields in empty space in the classical framework of GR. But, if treated on the same footing as fields corresponding to other fundamental interactions, gravitation fields must also possess quantum fluctuations which are induced by quantum fluctuations of sources, that is, of energy–momentum tensors [33, 95, 99]. Quantum fluctuations lead to effective equations for the propagation of gravitation fields that are modified, with consequences which may remain significant in the classical limit. To make this more explicit, it is convenient to first consider gravitational fluctuations in flat space. Fluctuations of gravitation fields are then represented as perturbations $((h_{\mu\nu})_{\mu,\nu=0,1,2,3})$ of Minkowski metric, which may equivalently be written as functions of position in spacetime or of a wavevector in Fourier space

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad \eta_{\mu\nu} = \text{diag}(1, -1, -1, -1), \quad |h_{\mu\nu}| \ll 1, \\ h_{\mu\nu}(x) \equiv \int \frac{d^4k}{(2\pi)^4} e^{-ikx} h_{\mu\nu}[k]. \tag{50}$$

The definition of metric fields suffers from ambiguities related to the choice of coordinates but, at the linearized level, gauge-invariant fields are provided by Riemann, Ricci, scalar, and Einstein curvatures

$$R_{\lambda\mu\nu\rho}[k] = \frac{1}{2} \{k_\lambda k_\nu h_{\mu\rho}[k] - k_\lambda k_\rho h_{\mu\nu}[k] - k_\mu k_\nu h_{\lambda\rho}[k] + k_\mu k_\rho h_{\lambda\nu}[k]\}, \\ R_{\mu\nu} = R^\lambda{}_{\mu\lambda\nu}, \quad R = R^\mu{}_\mu, \quad E_{\mu\nu} = R_{\mu\nu} - \eta_{\mu\nu} \frac{R}{2}, \tag{51}$$

Classically, metric fields are determined from energy–momentum sources by the Einstein equations of GR (49) which, at the linearized level and in the momentum domain, take a simple form

$$E_{\mu\nu}[k] = \frac{8\pi G_N}{c^4} T_{\mu\nu}[k]. \quad (52)$$

Equations of motion (52) in fact describe the coupling between metric fields and the total energy–momentum tensor of all fields, that is, the corresponding coupling terms in their common Lagrangian. Due to the nonlinear character of gravitation theory, these equations include the energy–momentum tensor of gravitation itself [100]. Equations 52 determine the metric fields that are generated by classical gravitation sources and may be seen as describing the response of metric fields to energy–momentum tensors, when quantum fluctuations are ignored. However, virtual processes associated with quantum fluctuations, that is, radiative corrections, must be taken into account when solving equations (52). There result modifications of the graviton propagator or of the effective coupling between gravitation and its sources [16, 23, 98]. It is well known that radiative corrections associated with Einstein equations (52) involve divergences which cannot be treated by usual means, due to the non-renormalizability of GR [96, 98]. However, these corrections result in embedding GR within a larger family of gravitation theories, with Lagrangians involving not only the scalar curvature but also quadratic forms in Riemann curvature. These theories appear to be renormalizable and to constitute reasonable extensions of GR. Furthermore, this enlarged family shows particular properties with respect to renormalization group trajectories [34, 40, 70], which hint at a consistent definition of a gravitation theory, with GR being a very good approximation within the range of length scales where it is effectively observed. Hence, we shall consider GR as an approximate effective theory and shall focus on the corrections to GR which remain to be taken into account in the range of length scales where GR is very close to the actual gravitation theory. It is also usually objected that gravitation theories with equations of motion involving higher derivatives of metric fields lead to violations of unitarity, or instability problems, which are revealed by the presence of ghosts. We shall note that arguments have been advanced for denying to consider these objections as real dead ends [41, 92]. Here, we shall just take the minimal position of restricting attention to a range of scales where both the gravitation theory remains close to GR and instabilities do not occur.

Keeping in mind the previous restrictions, one may see the modified gravitation propagator, including the effect of radiative corrections, as an effective response function of metric fields to energy–momentum tensors [54]. Gravitation equations then take the generalized form of a linear response relation between Einstein curvature and energy–momentum tensors

$$E_{\mu\nu}[k] = \chi_{\mu\nu}^{\lambda\rho}[k] T_{\lambda\rho}[k] = \left(\frac{8\pi G_N}{c^4} \delta_{\mu}^{\lambda} \delta_{\nu}^{\rho} + \delta\chi_{\mu\nu}^{\lambda\rho}[k] \right) T_{\lambda\rho}[k]. \quad (53)$$

The effective gravitation equations (53) take the same form as response functions induced by motion (4,10), with energy–momentum tensors playing the role of displacement generators. This property follows from the relation made in relativity theory between changes of coordinates and motion (this relation in particular provides the definition of a symmetric energy–momentum tensor [69]). Assuming that corrections induced by quantum fluctuations correspond to perturbations, the effective gravitation equations (53) should appear as a perturbation of Einstein equations (49). These perturbations may be captured in a function $\delta\chi_{\mu\nu}^{\lambda\rho}$, which may be seen as a momentum-dependent correction to the coupling constant G_N , or as a nonlocal correction to the gravitational coupling in the space-time domain.

Studying the coupling between gravitation and different fields [16, 54] allows one to derive some general properties of the modification brought by radiative corrections to Einstein equations, even if a complete determination of the modified response function $\chi_{\mu\nu}^{\lambda\rho}$ is not available. The different quantum fields coupling to gravitation lead to radiative corrections which differ in two sectors with different conformal weights. On one hand, conformally invariant fields such as electromagnetic fields, which correspond to traceless energy–momentum tensors, only affect the conformally invariant sector corresponding to Weyl curvatures. On the other hand, massive fields, which involve energy–momentum tensors with a non vanishing trace, modify both sectors. The two sectors then correspond to different modifications of the gravitational coupling constant G_N into running coupling constants, so that G_N should be replaced by two slightly different running coupling constants. In the linearized approximation, and in the case of a static pointlike source, the equations for metric fields (53) may be rewritten in terms of projectors on transverse components

$$\begin{aligned}
 T_{\mu\nu} &= \delta_{\mu 0}\delta_{\nu 0}T_{00}, & T_{00}[k] &= Mc^2\delta(k_0), \\
 E_{\mu\nu} &= E_{\mu\nu}^{(0)} + E_{\mu\nu}^{(1)}, & \pi_{\mu\nu}[k] &\equiv \eta_{\mu\nu} - \frac{k_\mu k_\nu}{k^2}, \\
 E_{\mu\nu}^{(0)} &= \left(\pi_\mu^0 \pi_\nu^0 - \frac{\pi_{\mu\nu} \pi^{00}}{3} \right) \frac{8\pi G^{(0)}}{c^4} T_{00}, & E_{\mu\nu}^{(1)} &= \frac{\pi_{\mu\nu} \pi^{00}}{3} \frac{8\pi G^{(1)}}{c^4} T_{00}, \\
 G^{(0)}[k] &= G_N + \delta G^{(0)}[k], & G^{(1)}[k] &= G_N + \delta G^{(1)}[k].
 \end{aligned}
 \tag{54}$$

At the linearized level, the decomposition on the two sectors with different conformal weights ($E^{(0)}$ and $E^{(1)}$) is easily performed in the momentum domain by means of projectors. The modified equations for gravitation then take the same form as Einstein equations (52), but in terms of two running coupling constants $G_N + \delta G^{(0)}$ and $G_N + \delta G^{(1)}$, which depend on the momentum or length scale and which slightly modify Newton gravitation constant G_N [60, 61]. These two scale-dependent couplings describe gravitation theories which remain close to GR within a certain range of scales along renormalization trajectories. They thus constitute a neighborhood of GR made of a large collection of theories labeled by two functions of an arbitrary scale parameter. As discussed above, GR just appears as a particular point in this neighborhood which is compatible with the observations made on our gravitational environment in an accessible range of scales.

4.2 Anomalous Curvatures

For the sake of simplicity, modifications of Einstein equations induced by radiative corrections have been presented, in the previous section, within the context of a linearized gravitation theory around a flat space-time (namely for weak gravitation fields and at first order in such fields). But the same mechanisms are easily seen to occur in the case of any space-time, endowed with an arbitrary background metric field, leading to modifications of Einstein equations (49) which may still be described as in Eqs. 53 by a general response of metric fields to energy–momentum tensors. Again, arguments derived from observations of our gravitational environment entail that gravitation equations should remain close to Einstein equations and take the following form

$$E_{\mu\nu} = \chi_{\mu\nu}^{\lambda\rho} \star T_{\lambda\rho} = \frac{8\pi G_N}{c^4} T_{\mu\nu} + \delta\chi_{\mu\nu}^{\lambda\rho} \star T_{\lambda\rho}. \quad (55)$$

The \star product denotes a convolution in space-time that replaces the ordinary product in the momentum domain. The effective response then introduces a nonlocal relation between Einstein curvature (with its full nonlinear dependence on metric fields) and energy–momentum tensors. The perturbed response function still differs in two sectors with different conformal weights and is thus equivalent to two different running coupling constants $G_N + \delta G^{(0)}$ and $G_N + \delta G^{(1)}$. Thus, also at the full nonlinear level, GR appears as embedded within a large family of gravitation theories labeled by two functions of a length scale parameter. Although the two running coupling constants remain close to Newton gravitation constant G_N , the relation they induce in general between curvatures and energy–momentum tensors is not explicit, due to an interplay between non linearity and non locality [62]. However, for discussing the observable consequences entailed by the modified gravitation equations (55), one only needs the solutions of these equations corresponding to given energy–momentum distributions. In that case, the modified equations (55) remaining close to Einstein equations (49), their solutions are small perturbations of the metric solutions satisfying Einstein equations.

The metric solution of generalized gravitation equations (55) lies in the vicinity of GR metric and satisfies perturbed equations (in the following, the notation $[\]_{\text{st}}$ stands for a GR solution of Einstein equations (49))

$$\begin{aligned} E_{\mu\nu}(x) &\equiv [E_{\mu\nu}]_{\text{st}}(x) + \delta E_{\mu\nu}(x), & [E_{\mu\nu}]_{\text{st}}(x) &= 0 \quad \text{when} \quad T_{\mu\nu}(x) = 0, \\ \delta E_{\mu\nu}(x) &\equiv \int d^4x' \delta\chi_{\mu\nu}^{\lambda\rho}(x, x') T_{\lambda\rho}(x'), \\ \delta E_{\mu\nu} &= \delta E_{\mu\nu}^{(0)} + \delta E_{\mu\nu}^{(1)}. \end{aligned} \quad (56)$$

Solutions of generalized equations (55) then identify with anomalous Einstein or Ricci curvatures, that is, with metrics leading to Ricci components which do not vanish in empty space (outside gravitational sources), contrarily to GR solutions. When considering solutions to the gravitation equations of motion (55), with the

latter corresponding to quantum perturbations of Einstein equations, the two running coupling constants $\delta G^{(0)}$ and $\delta G^{(1)}$ which modify Newton gravitation constant G_N are seen to be equivalent to anomalous Einstein curvatures (56). The latter possess two independent components only, which may be chosen as the components of Einstein curvature with conformal weight 0, $\delta E^{(0)}$ related to Weyl curvature, and with conformal weight 1, $\delta E^{(1)}$ equivalent to the scalar curvature. The relation between coupling constants and anomalous curvatures is obtained by extending the linearized limit (54) [62].

For applications which will be our main concern here, namely gravitation in the outer part of the solar system, it will be sufficient to consider the stationary and isotropic case. Using Schwarzschild coordinates, a stationary and isotropic metric $g_{\mu\nu}$ may be written (with t , r , and θ , φ denoting the time, radial, and angular coordinates, respectively)

$$g_{\mu\nu}dx^\mu dx^\nu \equiv g_{00}c^2dt^2 + g_{rr}dr^2 - r^2(d\theta^2 + \sin^2\theta d\varphi^2). \quad (57)$$

A stationary isotropic metric is then characterized by two functions of the radial coordinate r only, namely its temporal and radial components g_{00} and g_{rr} , respectively. For the sake of simplicity, we shall also ignore effects due to the size and rotation of the gravitational source (these can be introduced without qualitatively changing the following discussions) and consider a point-like gravitational source. The corresponding GR solution may be written in terms of a single Newtonian potential Φ_N

$$\begin{aligned} [E^\nu_\mu]_{\text{st}}(x) &= 8\pi\kappa\delta_\mu^0\delta_0^\nu\delta^{(3)}(x), \quad \kappa \equiv \frac{G_N M}{c^2}, \\ [g_{00}]_{\text{st}}(r) &= 1 + 2\Phi_N = -\frac{1}{[g_{rr}]_{\text{st}}(r)}, \quad \Phi_N \equiv -\kappa u, \quad u \equiv \frac{1}{r}. \end{aligned} \quad (58)$$

The metric given by (57) and solution of the generalized gravitation equations (55) lies in the vicinity of GR metric (58) and corresponds to two anomalous Einstein curvature components (56). In the stationary isotropic case, Ricci and Einstein curvatures have only two independent components, so that the anomalous components in the two different sectors, $\delta E^{(0)}$ and $\delta E^{(1)}$, may be replaced by the temporal and radial components of Einstein curvature δE^0_0 and δE^r_r (which may be rewritten in terms of Weyl and scalar curvatures). Then, solving for the metric field, the two anomalous curvature components become equivalent to anomalous parts in the two components of the isotropic metric

$$\begin{aligned} g_{00} &= [g_{00}]_{\text{st}} + \delta g_{00}, & \frac{\delta g_{00}}{[g_{00}]_{\text{st}}} &= \int \frac{du}{[g_{00}]_{\text{st}}^2} \int^u \frac{\delta E^0_0}{u'^4} du' + \int \frac{\delta E^r_r}{u^3} \frac{du}{[g_{00}]_{\text{st}}}, \\ g_{rr} &= [g_{rr}]_{\text{st}} + \delta g_{rr}, & \frac{\delta g_{rr}}{[g_{rr}]_{\text{st}}} &= -\frac{u}{[g_{00}]_{\text{st}}} \int \frac{\delta E^0_0}{u^4} du. \end{aligned} \quad (59)$$

In case of a stationary point-like source, the two running coupling constants characterizing the modified gravitation equations (55) are equivalent to two anomalous components of Ricci or Einstein curvature tensor (56), or else to two anomalous parts in the corresponding stationary isotropic metric (59). If the first representation better suits the framework of QFT, the last representation appears more appropriate to a study of the observable consequences of modified gravitation.

For a phenomenological analysis, it is even more convenient to rewrite the two independent degrees of freedom under the form of two gravitation potentials. In the stationary isotropic case, the two sectors of anomalous components (56) can be represented by two anomalous gravitational potentials, corresponding to the temporal and radial components of Einstein curvature

$$\begin{aligned}\delta E_0^0 &\equiv 2u^4(\delta\Phi_N - \delta\Phi_P)'', & (') &\equiv \partial_u, \\ \delta E_r^r &\equiv 2u^3\delta\Phi_P'.\end{aligned}\quad (60)$$

Solutions for the metric components then provide the perturbation (59) around GR solution (58) in terms of anomalous potentials

$$\begin{aligned}\delta g_{rr} &= \frac{2u}{(1 + 2\Phi_N)^2}(\delta\Phi_N - \delta\Phi_P)', \\ \delta g_{00} &= 2\delta\Phi_N + 4\kappa(1 + 2\Phi_N) \int \frac{u(\delta\Phi_N - \delta\Phi_P)' - \delta\Phi_N}{(1 + 2\Phi_N)^2} du.\end{aligned}\quad (61)$$

Equations 61 in terms of two gravitational potentials $\Phi_N + \delta\Phi_N$ and $\delta\Phi_P$ provide metric extensions which remain close to GR while accounting for nonlinearities in the metric. They correspond to a modification $\delta\Phi_N$ of Newton potential Φ_N and to the introduction of a second potential $\delta\Phi_P$. The two gravitational potentials describe, up to combinations, the two sectors introduced by the modified gravitation equations (55) and span the corresponding family of gravitation theories labeled by two functions of a length scale, which are equivalent to the two running coupling constants induced by radiative corrections [62].

4.3 Phenomenology in the Solar System

The family of extended metrics (58,61), obtained by solving the generalized gravitation equations (55), provides a basis for a phenomenological analysis of gravitation in the neighborhood of a stationary point-like source. Extended metrics (58,61) depend on two functions (of a single variable, the distance to the gravitational source) which parametrize a vicinity of GR. Observations performed in the gravitational field of the source should then allow one to characterize these two functions and thus to determine the nature of the theory describing gravitation in the neighborhood of a point-like source.

In the solar system, tests of gravity are usually performed by comparing observations with the predictions obtained from a family of parametrized post-Newtonian (PPN) metrics which extend GR metric by introducing additional parameters [25, 103]. In the approximation of a stationary point-like gravitational source, ignoring effects due to size and rotation, PPN metrics may be written in terms of a single potential ϕ (which takes in isotropic coordinates the same form as Newton potential Φ_N in Schwarzschild coordinates (58)) and of two constant parameters β and γ [104]

$$\begin{aligned}
 g_{\mu\nu}dx^\mu dx^\nu &\equiv g_{00}c^2dt^2 + g_{rr}\{dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2)\}, \\
 g_{00} &= 1 + 2\phi + 2\beta\phi^2 + \dots, \quad \phi = -\frac{G_N M}{c^2 r}, \\
 g_{rr} &= -1 + 2\gamma\phi + \dots.
 \end{aligned} \tag{62}$$

Eddington parameters γ and β , respectively, describe linear effects on light deflection and nonlinear effects on perihelia of planets. They are defined so that GR corresponds to the particular values $\beta = \gamma = 1$.

It is easily seen that PPN metrics (62) correspond to particular cases of the general metric extensions of GR that have been previously introduced

$$\begin{aligned}
 \delta\Phi_N &= (\beta - 1)\phi^2 + O(\phi^3), \quad \delta\Phi_P = -(\gamma - 1)\phi + O(\phi^2), \\
 \delta E_0^0 &= \frac{1}{r^2} O(\phi^2), \quad \delta E_r^r = \frac{1}{r^2} (2(\gamma - 1)\phi + O(\phi^2)) \quad [\text{PPN}], \tag{63}
 \end{aligned}$$

PPN metrics span a two-dimensional subspace, labeled by (β, γ) , of the neighborhood of GR corresponding to solutions of generalized gravitation equations (55), which are labeled by two functions $\delta\Phi_N$ and $\delta\Phi_P$. According to Eqs. 63, this subspace corresponds to metrics with Einstein curvatures which vanish at large distances of the gravitational source. Alternatively, the metric extensions of GR (58, 61) may be seen as generalizations of PPN metrics, where the two Eddington parameters β and γ are replaced by two arbitrary functions $\delta\Phi_N$ and $\delta\Phi_P$ of the radial distance to the gravitational source.

Assuming that the metric associated with the gravitational field of the Sun takes the form of a PPN metric, predictions can be made for motions in this gravitational field and compared with observations [105]. Observations in the solar system provide constraints on the form of the single potential ϕ and on the values of Eddington parameters β and γ . Modifications of Newton potential ϕ are usually parametrized in terms of an additional Yukawa potential depending on two parameters, a range λ and an amplitude α measured with respect to ϕ

$$\delta\phi(r) = \alpha e^{-\lambda r} \phi(r), \tag{64}$$

Corrections behaving as (64) have been looked for at various values of λ ranging from the millimeter scale [1] to the size of planetary orbits [20]. For ranges of

the order of the Earth–Moon [106] to Sun–Mars distances [42, 66, 85] tests have been performed by following the motions of planets and artificial probes. Although agreeing with GR, all these results still show [20, 86] that windows remain open for violations of the Newton force law at short ranges λ , below the millimeter, as well as long ones, of the order of or larger than the size of the solar system. Similarly, experiments performed up to now have confirmed that, assuming that β and γ take constant values, the latter should be close to 1. From Doppler ranging on Viking probes in the vicinity of Mars [42] to deflection measurements using VLBI astrometry [91] or radar ranging on the Cassini probe [9], the allowed values for $\gamma - 1$ have reduced with time. Precessions of planet perihelion [94] and polarization by the Sun of the Moon orbit around the Earth [78] constrain linear superpositions of β and γ . As a result, in order to remain compatible with gravity tests, PPN metrics must satisfy rather stringent constraints

$$|\gamma - 1| \leq 10^{-5}, \quad |\beta - 1| \leq 10^{-4}. \quad (65)$$

Obviously, gravity tests performed in the solar system provide evidence for a metric theory lying very close to GR. In case of a PPN metric (62), the latter should correspond to values of β and γ close to 1. These constraints however result from an analysis performed with the assumption of constant Eddington parameters, which only covers a small subspace of potential deviations from GR (63). Dependences of β and γ on the length scale are not excluded and deviations at very short or very large scale are loosely constrained. Gravity tests still leave room for alternative metric theories, such as metric extensions of GR (61), provided they satisfy these criteria. They should correspond to small anomalous potentials $\delta\Phi_N$ and $\delta\Phi_P$ so that to remain close to GR and thus compatible with present gravity tests.

As remarked in the introduction, deviations from GR may have already been observed in the very domain where classical tests have been performed, namely within the solar system. The Pioneer 10/11 probes, which were launched in the 1970s and tracked during their travel to the outer part of the solar system indeed showed anomalous behaviors. The trajectories of the probes were determined from the radio frequency signals sent to them, transponded on board and received by stations on Earth. The Doppler shifts affecting the signals received on Earth, when compared with the emitted ones, delivered the velocity and hence, by integration, the distance separating the probe from the Earth. In fact, Doppler data provide the trajectory once a modelization of all gravitation effects on light propagation and on geodesic trajectories has been made. It appeared that the modelization based on GR led to anomalies, taking the form of Doppler residuals, that is, differences between observed and modeled velocities $v_{\text{obs}} - v_{\text{model}}$, which could not be reduced [5]. The latter furthermore took the form of a linear dependence in time t , that is, of a roughly constant acceleration a_P , over distances ranging from 20 to 70 A.U., directed toward the Sun

$$v_{\text{obs}} - v_{\text{model}} \simeq -a_P(t - t_{\text{in}}), \quad a_P \simeq 0.8 \text{ nm s}^{-2}. \quad (66)$$

No satisfactory explanation in terms of systematic effects taking their origin on the probe itself or in its spatial environment has been found up to now [76]. Recently, a larger set of data, covering parts of the Pioneer 10/11 missions which had not been analyzed, has been recovered and put under scrutiny [82]. Several teams have engaged in an independent reanalysis of the Pioneer data [71, 73, 79]. Proposals have also been made for missions dedicated to a study of deep space gravity [24]. At present, the possibility that the Pioneer anomaly points at the necessity to change the theoretical framework cannot be ignored. In that case, the existence of an extended framework having the ability to account for the Pioneer anomaly while remaining consistent with all gravity tests constitutes a crucial element.

PPN metrics which are compatible with classical gravity tests practically reduce to GR metric (see Eq. 65). Furthermore, the radial dependence of their curvatures (63) does not allow PPN metrics to reproduce the properties of the Pioneer anomaly (66). On another hand, metric extensions of GR (58, 61) provide a phenomenological framework which enlarges the PPN neighborhood of GR, with corrections brought to curvatures which may remain significant at large distances from the gravitational source.

When performed within the framework of a general metric extension of GR, an explicit computation of the Doppler signal, in the physical configuration corresponding to the Earth and a remote probe, exhibits a discrepancy with the similar signal computed using GR [60, 62]. Computation is more easily achieved by using the time delay function, that is, a two-point function which describes the elapsed time between emission and reception of a lightlike signal propagating between two spatial positions. The time delay function is easily determined from the metric components by using a reference frame where the extended metric takes the form of a stationary isotropic metric [63]. The Doppler signal corresponding to a tracking of the probes by stations on Earth is then obtained as the time derivative of the time delay function evaluated on the trajectories of the Earth and probes. The time derivative of the Doppler signal itself may be written under the form of a time-dependent acceleration, as the observed Pioneer anomaly (66) [5]. For an extended metric (61) which is close to GR metric, dependences on metric components may be treated perturbatively and only the first order in metric perturbation may be kept. Metric perturbations may be seen to modify the expression for the Doppler signal in two ways, through a perturbation of the time delay two-point function itself, and through a perturbation of the trajectories on which this function is evaluated. A difference then emerges between the acceleration obtained for an extended metric and that entailed by GR metric, which corresponds to the discrepancy which would appear when comparing an observed signal with the similar signal predicted using GR. This difference can be considered as representing the anomaly which would be observed.

When expressed in terms of the anomalous potentials defining the extended metric (61), the anomalous acceleration contains several contributions due to different sources of perturbation, affecting in particular the propagation of lightlike signals and the trajectories of massive bodies. We shall consider the simplified case of a remote probe which is moving on an escape trajectory in the ecliptic plane and which is reaching the outer part of the solar system. Several contributions to the anomalous

acceleration may then be neglected and the latter takes a simplified expression [63]

$$\begin{aligned}\delta a &\simeq \delta a_{\text{sec}} + \delta a_{\text{ann}}, \\ \delta a_{\text{sec}} &\simeq -\frac{c^2}{2} \partial_r (\delta g_{00}) + [\ddot{r}]_{st} \left\{ \frac{\delta(g_{00} g_{rr})}{2} - \delta g_{00} \right\} - \frac{c^2}{2} \partial_r^2 [g_{00}]_{st} \delta r, \\ \delta a_{\text{ann}} &\simeq \frac{d}{dt} \{ [\dot{\phi}]_{st} \delta \rho \}.\end{aligned}\tag{67}$$

(r, ϕ) represents the position of the probe with respect to the Sun, in the ecliptic plane (r is the radial distance and ϕ the difference of angular positions between the probe and the Earth) and ρ denotes the impact parameter (with respect to the Sun) of the lightlike signal propagating between the probe and the Earth. The acceleration anomaly divides into two parts, a secular anomaly δa_{sec} which varies over large times only, typically for variations of the distance between the probe and the Earth of the order of several A.U., and a modulated anomaly δa_{ann} which describes variations with annual or semiannual periodicity. One notes that, due to several simplifications in the previous representation (neglecting in particular the motions of the stations, effects of the Earth atmosphere, deviations from the ecliptic plane, etc.) additional modulations, in particular with daily periodicity, have been ignored. Both parts of the anomalous acceleration are generated by anomalies in the two gravitational potentials ($\delta g_{00}, \delta g_{rr}$) and in the probe trajectory (δr). In particular, the latter cannot be ignored in the case of Pioneer 10/11 probes as no range capabilities (allowing one to directly determine the radial position of the probe) were available for them. Anomalous metrics (67) can be seen to lead to anomalous accelerations which exhibit the same qualitative features as the Pioneer anomaly [5, 71, 73, 79].

To obtain from the observed Pioneer anomaly quantitative constraints on the two gravitational potentials, one needs to enter a detailed analysis of navigation data, performing this analysis in the enlarged phenomenological framework provided by metric extensions of GR. As already shown in the above-simplified model (67), the acceleration anomaly exhibits a secular part and modulations which both depend on anomalies in the gravitational potentials and in the trajectory. Correlated anomalies should then play an essential role when confronting models pertaining to the extended phenomenological framework with a detailed analysis of Pioneer data [63].

Although a precise confrontation is not yet available, some consequences can nonetheless be drawn from the general form of the secular anomaly observed on the Pioneer probes. As shown by Eqs. 67 (anomaly modulations are used to eliminate the trajectory anomaly [63]), a constant anomalous acceleration $\delta a = -a_P \equiv -\frac{c^2}{l_H}$ over the distances covered by the Pioneer probes may follow from different forms of the anomalous gravitational potentials (61). On one hand, if only a perturbation of Newton potential is assumed, the latter should behave linearly with the heliocentric distance, that is, $\delta \Phi_N \simeq r/l_H$, to produce a constant acceleration. This form however conflicts with tests which have been performed between the Earth and Mars [5, 61, 85], so that an anomaly limited to the Newtonian sector is only allowed at large heliocentric distances [75], where it must still remain compatible with the

ephemeris of outer planets. On the other hand, a constant acceleration is also obtained for an anomaly in the second sector which behaves quadratically with the heliocentric distance, that is, $\delta\Phi_P \simeq \frac{c^2}{3G_{NM}} \frac{r^2}{I_H}$. A combination of anomalies in the two sectors may also lead to the form taken by Pioneer data [63]. One notes that these properties cannot be obtained from PPN metrics (63) and that the form just given for the anomalous potential in the second sector $\delta\Phi_P$ corresponds to a non vanishing constant curvature in the outer part of the solar system.

A remarkable feature of the phenomenological framework derived from metric extensions of GR (58, 61) is that, besides producing Pioneer-like anomalies, it also allows one to preserve the agreement with classical gravity tests. An important part of gravity tests is provided by ephemerides of inner planets, and in particular by perihelion precession anomalies [42]. Anomalies in perihelion precessions of planets (with respect to Newtonian gravitation) are well accounted for by the nonlinear dependence of GR metric on Newton potential (62). Metric extensions (58, 61) induce modifications which generalize the corrections obtained in the PPN framework (63). The anomalous gravitational potentials $\delta\Phi_N$ and $\delta\Phi_P$ may be seen as promoting the parameters β and γ to functions which depend on the heliocentric distance [62]. Observed anomalies in perihelion precessions constrain a particular combination of the two gravitational potentials for ranges around a few A.U. where planet ephemerides are known with precision. Provided the latter combination remains small within these ranges, compatibility of extended metrics with gravity tests is ensured.

Precise gravity tests are obtained by measurements of the deflection induced on a lightlike signal by the Sun gravitational field, as for instance those performed with Cassini probe [9]. Light deflection measurements can be seen to explore the behavior of gravitational potentials in the Sun vicinity, that is, at small heliocentric distances. The deflection angle is determined by the impact parameter of the lightlike signal and, within a PPN framework, is directly related to the value of the parameter γ . In the framework of metric extensions (58, 61), the effect of anomalous gravitational potentials on the deflection angle may be interpreted as due to a generalized parameter γ , with a functional dependence on the impact parameter [61]. Deviations of this generalized parameter from its constant value corresponding to GR are given by a combination of the two anomalous gravitational potentials (61), so that light deflection measurements provide constraints on the behavior of gravitational potentials in the range of a few solar radii. A precise analysis shows that these constraints still allow different behaviors of the two gravitational potentials with respect to the radial distance [60]. When compared with the value predicted by GR or even a PPN metric, the deflection angle might then show an anomalous behavior which becomes observable for large values of the impact parameter. Although light deflection is more important, within the PPN framework, for small impact parameters and is thus usually measured with lightlike signals grazing the Sun surface, a global mapping of light deflection over the sky, including large impact parameters, is programmed in the future mission GAIA [37]. A detected anomalous behavior of deflection angles could point at the necessity to use an enlarged phenomenological framework like that provided by metric extensions of GR.

5 Conclusion

Despite difficulties that emerge at their interface, quantum theory and relativity theory cannot be considered as independent frameworks, with quantum applications excluding motion and gravitation and relativity being limited to a representation of classical systems. Quantum fields possess fluctuations which result in mechanical effects and affect inertial responses. This entails a revision of our notions of mass and motion which becomes necessary when dealing with microscopic systems. Quantum field fluctuations also modify gravitation with the possibility of leading to observable effects at macroscopic scales. The effects of quantum fluctuations on relativistic systems cannot be ignored and it appears necessary to have a representation which is compatible with both frameworks.

We have shown that quantum field fluctuations, even in vacuum, must be taken into account when analyzing mechanical effects such as inertia. As exemplified on Casimir energy and contrarily to a common opinion, vacuum field fluctuations do contribute to inertia and the induced inertial mass complies both with quantum and with relativistic requirements. These properties are more easily put into evidence using the linear response formalism, which thus allows one to connect in a consistent way the quantum and relativistic frameworks. Infinities that are responsible for difficulties when interfacing quantum and relativity theories can be treated by renormalization techniques. The linear response formalism not only recovers the fundamental connection between interacting and free quantum fields lying at the basis of QFT, but also allows one to implement in a general way the fundamental relation made by relativity between motions and the symmetries of space-time. A treatment of motion in quantum vacuum may thus be given which remains consistent with the relativistic conception of space-time and which brings new light on the role played by quantum vacuum with respect to inertia [45, 59].

The effects of quantum fluctuations on inertia have important consequences for the notion of mass. There results that the mass parameter used for characterizing motion is a quasistatic approximation. Mass possesses quantum fluctuations which cannot be neglected in high frequency regimes, or equivalently at very short timescales, and these fluctuations should be described by the correlations of a quantum operator representing the mass observable [53]. This modification of the representation of mass is accompanied by similar extensions for positions and motions in space-time in order to comply with the requirements imposed on observables by quantum and relativity theories. Remarkably, these extensions are ensured in a consistent way by the existence of a single algebra where quantum observables and relativistic transformations are simultaneously represented. This unique algebra, built on space-time symmetries, provides relations between mass and motions, including uniformly accelerated motions, which hold in the quantum framework. This allows in particular to write a quantum version of Einstein effect and shows that the equivalence principle can be extended to the quantum framework [58].

According to the principles of relativity, quantum field fluctuations affect not only inertial but also gravitational masses, with the consequence that classical gravitation must be modified to account for radiative corrections. Presently, theoretical

analysis and experimental observations both support the equivalence principle as a faithful and well-tested basis for a gravitation theory. These arguments strongly favor metric theories but do not constrain the gravitational coupling. If classical tests performed up to now in the solar system tend to confirm Einstein–Hilbert Lagrangian, the corresponding coupling loses its simple form when corrections induced by quantum fluctuations are taken into account. Although applying renormalization techniques to gravitation theory appears as a formidable task, theoretical arguments and observations both point at the possibility of corrections to gravitation occurring at the classical level, that is, at large length scales. Radiative corrections induce modifications which amount to replace Newton gravitation constant by two gravitation running coupling constants corresponding to two sectors with different conformal weights. These may equivalently be represented by metric extensions of GR which can be characterized either by two nonvanishing Ricci curvatures or by two gravitational potentials. Besides modifications of Newton potential, a gravitational potential in a second sector opens new phenomenological possibilities. Metric extensions of GR appear as an efficient way to parametrize gravitation theories lying in the neighborhood of GR and offer larger possibilities than the usual PPN phenomenological framework. Having also the ability to remain compatible with classical gravity tests, they may account for Pioneer-like anomalies and may further lead to other correlated anomalies which could be observed in future experiments [64].

References

1. E.G. Adelberger, B.R. Heckel, A.E. Nelson, *Ann. Rev. Nucl. Part. Sci.* **53**, 77 (2003)
2. R.J. Adler, *Rev. Mod. Phys.* **54**, 729 (1982)
3. A. Aguirre, C.P. Burgess, A. Friedland, D. Nolte, *Class. Q. Grav.* **18**, R223 (2001)
4. J.D. Anderson, P.A. Laing, E.L. Lau, A.S. Liu, M.M. Nieto, S.G. Turyshev, *Phys. Rev. Lett.* **81**, 2858 (1998)
5. J.D. Anderson, P.A. Laing, E.L. Lau, A.S. Liu, M.M. Nieto, S.G. Turyshev, *Phys. Rev. D* **65**, 082004 (2002)
6. T.B. Bahder, *Phys. Rev. D* **68**, 063005 (2003)
7. G. Barton, *Introduction to Advanced Field Theory* (Interscience Publishers, New York, 1963)
8. H. Bateman, *Proc. Lond. Math. Soc.* **8**, 223 (1909)
9. B. Bertotti, L. Iess, P. Tortora, *Nature* **425**, 374 (2003)
10. N.D. Birrell, P.C.W. Davies, *Quantum Fields in Curved Space* (Cambridge University Press, New York, 1982)
11. N.N. Bogolubov, A.A. Logunov, I.T. Todorov, *Introduction to Axiomatic Quantum Field Theory* (Benjamin, Reading, 1975)
12. M. Bordag, U. Mohideen, V.M. Mostepanenko, *Phys. Rep.* **353**, 1 (2001)
13. A.J. Bracken, *Lett. Nuovo Cimento* **2**, 571 (1971)
14. A.J. Bracken, B. Jessup, *J. Math. Phys.* **23**, 1925 (1981)
15. H.B. Callen, T.A. Welton, *Phys. Rev.* **83**, 34 (1951)
16. D.M. Capper, M.J. Duff, L. Halpern, *Phys. Rev. D* **10**, 461 (1974)
17. S.M. Carroll, V. Duvvuri, M. Trodden, M.S. Turner, *Phys. Rev. D* **70**, 043528 (2004)
18. H.B.G. Casimir, *Proc. K. Ned. Akad. Wet.* **51**, 793 (1948)
19. C. Cohen-Tannoudji, J. Dupont-Roc, G. Grynberg, *Processus d'interaction entre photons et atomes* (InterEditions, Paris, 1988). English translation: *Atom-Photon Interactions* (Wiley, New York, 1992)

20. J. Coy, E. Fischbach, R. Hellings, C. Talmadge, E.M. Standish, Private communication (2003)
21. E. Cunningham, Proc. Lond. Math. Soc. **8**, 77 (1909)
22. C. Deffayet, G. Dvali, G. Gabadadze, A. Vainshtein, Phys. Rev. D **65**, 044026 (2002)
23. S. Deser, P. van Nieuwenhuizen, Phys. Rev. D **10**, 401 (1974)
24. H. Dittus, S.G. Turyshev, C. Lämmerzahl, S. Theil, R. Foerster, U. Johann, W. Ertmer, E. Rasel, B. Dachwald, W. Seboldt, F.W. Hehl, C. Kiefer, H.-J. Blome, J. Kunz, D. Giulini, R. Bingham, B. Kent, T.J. Sumner, O. Bertolami, J. Páramos, J.L. Rosales, B. Christophe, B. Foulon, P. Touboul, P. Bouyer, S. Reynaud, A. Brillet, F. Bondu, E. Samain, C.J. de Matos, C. Erd, J.C. Grenouilleau, D. Izzo, A. Rathke, J.D. Anderson, S.W. Asmar, E.L. Lau, M.M. Nieto, B. Mashhoon, in *1st Symp. on Environmental Testing for Space Programmes*, Noordwijk, 15–17 June 2004, ESA SP **558** (ESA, Noordwijk, 1997), p. 3
25. A.S. Eddington, *The Mathematical Theory of Relativity* (Cambridge University Press, Cambridge, 1957)
26. A. Einstein, Ann. Phys. (Leipzig) **17**, 549 (1905); Phys. Z. **18**, 121 (1917)
27. A. Einstein, Ann. Phys. (Leipzig) **18**, 639 (1905)
28. A. Einstein, Ann. Phys. (Leipzig) **20**, 627 (1906)
29. A. Einstein, Jahr. Radioakt. Elektron. **4**, 411 (1907); *ibidem*, **5**, 98 (1908)
30. A. Einstein, Sitzungsber. Preuss. Akad. Wiss., Phys. Math. Kl., 844 (1915)
31. A. Einstein, Ann. Phys. (Leipzig) **49**, 769 (1916)
32. C.P. Enz, in *Physical Reality and Mathematical Description*, ed. by C.P. Enz, J. Mehra (Reidel, Dordrecht, 1974), p. 124
33. R.P. Feynman, Acta Phys. Polon. **24**, 711 (1963)
34. E.S. Fradkin, A.A. Tseytlin, Nucl. Phys. B **201**, 469 (1982)
35. S.A. Fulling, *Aspects of QFT in Curved Spacetime* (Cambridge University Press, Cambridge, 1989)
36. S.A. Fulling, P.C.W. Davies, Proc. R. Soc. A **348**, 393 (1976)
37. GAIA, <http://www.rssd.esa.int/GAIA>
38. T. Goldman, J. Pérez-Mercader, F. Cooper, M.M. Nieto, Phys. Lett. B **281**, 219
39. B. Guinot, Metrologia **34**, 261 (1977)
40. H.W. Hamber, R.M. Williams, Phys. Rev. D **75**, 084014 (2007)
41. S.W. Hawking, T. Hertog, Phys. Rev. D **65**, 103515 (2002)
42. R.W. Hellings, P.J. Adams, J.D. Anderson, M.S. Keesey, E.L. Lau, E.M. Standish, Phys. Rev. Lett. **51**, 1609 (1983)
43. D. Hilbert, Nach. Ges. Wiss. Göttingen, 395 (1915)
44. C. Itzykson, J.-B. Zuber, *Quantum Field Theory* (McGraw Hill, New York, 1985)
45. M.-T. Jaekel, A. Lambrecht, S. Reynaud, New Astron. Rev. **46**, 727 (2002)
46. M.-T. Jaekel, S. Reynaud, J. Phys. I France **1**, 1395 (1991)
47. M.-T. Jaekel, S. Reynaud, J. Phys. I France **2**, 149 (1992)
48. M.-T. Jaekel, S. Reynaud, Phys. Lett. A **167**, 227 (1992)
49. M.-T. Jaekel, S. Reynaud, Q. Opt. **4**, 39 (1992)
50. M.-T. Jaekel, S. Reynaud, J. Phys. I France **3**, 1 (1993)
51. M.-T. Jaekel, S. Reynaud, J. Phys. I France **3**, 1093 (1993)
52. M.-T. Jaekel, S. Reynaud, Phys. Lett. A **172**, 319 (1993)
53. M.-T. Jaekel, S. Reynaud, Phys. Lett. A **180**, 9 (1993)
54. M.-T. Jaekel, S. Reynaud, Ann. Phys. (Leipzig) **4**, 68 (1995)
55. M.-T. Jaekel, S. Reynaud, Q. Semiclass. Opt. **7**, 499 (1995)
56. M.-T. Jaekel, S. Reynaud, Phys. Lett. A **220**, 10 (1996)
57. M.-T. Jaekel, S. Reynaud, Phys. Rev. Lett. **76**, 2407 (1996)
58. M.-T. Jaekel, S. Reynaud, Europhys. Lett. **38**, 1 (1997)
59. M.-T. Jaekel, S. Reynaud, Rep. Prog. Phys. **60**, 863 (1997)
60. M.-T. Jaekel, S. Reynaud, Class. Q. Grav. **22**, 2135 (2005)
61. M.-T. Jaekel, S. Reynaud, Mod. Phys. Lett. A **20**, 1047 (2005)
62. M.-T. Jaekel, S. Reynaud, Class. Q. Grav. **23**, 777 (2006)
63. M.-T. Jaekel, S. Reynaud, Class. Q. Grav. **23**, 7561 (2006)

64. M.-T. Jaekel, S. Reynaud, in *Gravitational Waves and Experimental Gravity*, XLII Rencontres de Moriond, ed. by J. Dumarchez, J. Trân Thanh Vân (Thé Gioi, Hanoi, 2007), p. 271
65. M.-T. Jaekel, S. Reynaud, in *XXVI Workshop on Geometric Methods in Physics*, ed. by P. Kielanowski, A. Odziejewicz, M. Sclichenmaier, T. Voronov (World Scientific, New York, 2007), p. 61
66. N.I. Kolosnitsyn, V.N. Melnikov, *Gen. Rel. Grav.* **36**, 1619 (2004)
67. R. Kubo, *Rep. Prog. Phys.* **29**, 255 (1966)
68. A. Lambrecht, M.-T. Jaekel, S. Reynaud, *Phys. Rev. Lett.* **77**, 615 (1996); *Eur. Phys. J. D* **3**, 95 (1998)
69. L.D. Landau, E.M. Lifschitz, *Cours de Physique Théorique, Physique Statistique*, Première partie, Ch. 12 (MIR, Moscou, 1984)
70. O. Lauscher, M. Reuter, *Class. Q. Grav.* **19**, 483 (2002)
71. A. Levy, B. Christophe, P. Berio, G. Metris, J.-M. Courty, S. Reynaud, *Adv. Space Res.* **43**, 1538 (2009)
72. A. Lue, R. Scoccimarro, G. Starkman, *Phys. Rev. D* **69**, 044005 (2004)
73. C.B. Markwardt, arXiv:gr-qc/0208046v1 (2002)
74. J. Meixner, in *Statistical Mechanics of Equilibrium and Non Equilibrium*, ed. by J. Meixner (North Holland, Amsterdam, 1965), p. 52
75. J.W. Moffat, *Class. Q. Grav.* **23**, 6767 (2006)
76. M.M. Nieto, J.D. Anderson, *Contemp. Phys.* **48**, 41 (2007)
77. M.M. Nieto, S.G. Turyshev, J.D. Anderson, *Phys. Lett. B* **613**, 11 (2005)
78. K. Nordtvedt, arXiv:gr-qc/0301024v1 (2003)
79. O. Olsen, *Astron. Astrophys.* **463**, 393 (2007)
80. W. Pauli, in *Die allgemeinen prinzipien der wellenmechanik*, ed. by H. Geiger, K. Scheel. *Handbuch der Physik*, vol. 24 (Springer, Berlin, 1933), p. 1
81. S. Perlmutter, G. Aldering, G. Goldhaber, R.A. Knop, P. Nugent, P.G. Castro, S. Deustua, S. Fabbro, A. Goobar, D.E. Groom, I.M. Hook, A.G. Kim, M.Y. Kim, J.C. Lee, N.J. Nunes, R. Pain, C.R. Pennypacker, R. Quimby, C. Lidman, R.S. Ellis, M. Irwin, R.G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B.J. Boyle, A.V. Filippenko, T. Matheson, A.S. Fruchter, N. Panagia, H.J.M. Newberg, W.J. Couch, *Astrophys. J.* **517**, 565 (1999); S. Perlmutter, M.S. Turner, M. White, *Phys. Rev. Lett.* **83**, 670 (1999)
82. Pioneer Anomaly Investigation Team Site, <http://www.issi.unibe.ch/teams/Pioneer>
83. M. Planck, *Verh. Deutsch. Phys. Ges.* **13**, 138 (1911)
84. *Proc. IEEE* **79**, *Special Issue on Time and Frequency*, 894–1079 (1991)
85. R.D. Reasenberg, I.I. Shapiro, P.E. MacNeil, R.B. Goldstein, J.C. Breidenthal, J.P. Brenkle, D.L. Cain, T.M. Kaufman, T.A. Komarek, A.I. Zygielbaum, *Astrophys. J. Lett.* **234**, L219 (1979)
86. S. Reynaud, M.-T. Jaekel, *Int. J. Mod. Phys. A* **20**, 2294 (2005)
87. A.G. Riess, A.V. Filippenko, P. Challis, A. Clocchiattia, A. Diercks, P.M. Garnavich, R.L. Gilliland, C.J. Hogan, S. Jha, R.P. Kirshner, B. Leibundgut, M.M. Phillips, D. Reiss, B.P. Schmidt, R.A. Schommer, R.C. Smith, J. Spyromilio, C. Stubbs, N.B. Suntzeff, J. Tonry, *Astron. J.* **116**, 1009 (1998)
88. F. Rohrlich, *Classical Charged Particles* (Addison Wesley, Reading, MA, 1965)
89. A.D. Sakharov, *Doklady Akad. Nauk SSSR* **177**, 70 (1967); *Sov. Phys. Doklady* **12**, 1040 (1968)
90. R.H. Sanders, S.S. McGaugh, *Annu. Rev. Astron. Astrophys.* **40**, 263 (2002)
91. S.S. Shapiro, J.L. Davis, D.E. Lebach, J.S. Gregory, *Phys. Rev. Lett.* **92**, 121101 (2004)
92. J.Z. Simon, *Phys. Rev. D* **41**, 3720 (1990)
93. K.S. Stelle, *Phys. Rev. D* **16**, 953 (1977); *Gen. Rel. Grav.* **9**, 353 (1978)
94. C. Talmadge, J.-P. Berthias, R.W. Hellings, E.M. Standish, *Phys. Rev. Lett.* **61**, 1159 (1988)
95. W.E. Thirring, *Ann. Phys. (N.Y.)* **16**, 96 (1961)
96. G. t'Hooft, M. Veltman, *Ann. Inst. H. Poincaré (A) Phys. Théor.* **20**, 69 (1974)
97. M.S. Turner, *Astrophys. J.* **576**, L101 (2002)
98. R. Utiyama, B. De Witt, *J. Math. Phys.* **3**, 608 (1962)

99. S. Weinberg, *Phys. Rev.* **138**, B988 (1965)
100. S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972)
101. S. Weinberg, arXiv:astro-ph/0005265v1 (2000)
102. A.S. Wightman, *Rev. Mod. Phys.* **34**, 845 (1962)
103. C.M. Will, K. Nordtvedt, *Astrophys. J.* **177**, 757 (1972); K. Nordtvedt, C.M. Will, *ibidem* **177**, 775 (1972)
104. C.M. Will, *Theory and Experiment in Gravitational Physics*, revised edition (Cambridge University Press, Cambridge, 1993); C.M. Will, *Living Rev. Rel.* **4**, URL: <http://www.livingreviews.org/lrr-2001-4>
105. C.M. Will, *Living Rev. Rel.* **9**, URL: <http://www.livingreviews.org/lrr-2006-3>
106. J.G. Williams, X.X. Newhall, J.O. Dickey, *Phys. Rev. D* **53**, 6730 (1996)

Motion in Quantum Gravity

Karim Noui

Abstract We tackle the question of motion in Quantum Gravity: what does motion mean at the Planck scale? Although we are still far from a complete answer we consider here a toy model in which the problem can be formulated and resolved precisely. The setting of the toy model is a three-dimensional Euclidean gravity. Before studying the model in detail, we argue that Loop Quantum Gravity may provide a very useful approach when discussing the question of motion in Quantum Gravity.

1 Introduction

1.1 *The Problem of Defining Motion in Quantum Gravity*

Motion is fundamentally a classical notion: “it refers to a change of position in space.” When we talk about quantum physics or relativity, the definition of motion has to be made more precise, for either the notion of position is not well-defined (in quantum physics) or the notion of space-time has to be rethought (in relativity). Indeed, when we turn on the Planck constant \hbar , matter is described in terms of wave functions that are not localized, so a point particle can, a priori, be everywhere at any time; one needs to introduce coherent states, for instance, to recover the reassuring notion of trajectory at the classical limit. When we turn on the light speed c , time is no longer absolute but becomes intimately mixed with spatial coordinates; we need to make precise what time means if we are to define the motion properly. When we turn on \hbar and c together, matter fields and their interactions are beautifully described within the Quantum Field Theory framework, which is rather nonintuitive, but one has to work quite hard to make a bridge to the classical world.

K. Noui (✉)

Laboratoire de Mathématiques et de Physique Théorique, UMR/CNRS 6083,
Fédération Denis Poisson, Faculté des Sciences et Techniques, Parc de Grandmont,
37200 Tours, France
e-mail: noui@lmpt.univ-tours.fr

What happens if we now introduce the gravitational constant G into the scenario? Newton gave us laws to explain the attraction between massive bodies and created tools for studying their trajectories, as long as the bodies are not too “small” and their velocities not too “high.” Turning on c and G together leads to general relativity, where space-time becomes a dynamical entity that interacts deeply with all types of matter and energy; we definitely lose the classical absolute background that is so necessary for defining the classical notion of motion. It is, all the same, possible to extend this notion and adapt it to general relativity; the geodesics for instance correspond to trajectories of infinitely light particles evolving in a space-time with which we assume they do not interact. The determination of trajectories without neglecting the self-force is a much more subtle, but more realistic and interesting, problem [6]. In particular, it puts forward the trivial but fundamental fact that the point-like description for massive matter-fields is completely meaningless in general relativity because it leads to black-hole singularities. Thus, to have a proper description of motion in general relativity, one needs to consider extended matter fields, which obviously makes the problem much more complicated at the technical and conceptual levels.

Defining motion in a theory where all the fundamental constants \hbar , G , and c are switched on is clearly too ambitious a problem. It is certainly too early to investigate it and we do not claim to solve it here. Rather, we would like to raise the preliminary questions that naturally arise while addressing such a concept, and to see if it is possible to answer some of them precisely. It goes without saying that the question of the fundamental structure of space-time comes first to mind. Even though it is commonly believed that space-time is no longer described in terms of a differential manifold at the Planck scale, it is also honest to claim that no one knows precisely how space-time appears at this scale. Nonetheless, there exist very fascinating proposals that one can take seriously when investigating the question of motion in quantum gravity.

1.2 *Quantum Gravity*

It is indeed openly recognized that a complete and consistent quantization of gravity that would give a precise description of space-time at the Planck scale is still missing. Many ways to attack this problem have been explored over the last 20 years, the two most popular surely being String Theory [24] and Loop Quantum Gravity [4]. While both these approaches aim to understand the deep and fundamental structure of space-time, they have developed very different strategies and achieved, so far, rather distinct results. For instance, String Theory proposes a version of quantum space-time with extra-dimensions whereas, in Loop Quantum Gravity, space-time is fundamentally four-dimensional with three-dimensional space slices that are discrete in some precise sense. The discreteness of space is, in fact, one of the most beautiful but intriguing achievements of Loop Quantum Gravity. Even if this result is controversial and unconfirmed, it makes Loop Quantum Gravity quite a

fascinating approach that certainly deserves to be investigated, to at least understand how far it can bring us toward the Planck regime.

To achieve this discreteness, Loop Quantum Gravity has adopted a very “conservative” point of view, namely, the canonical quantization of the Einstein–Hilbert theory reformulated in terms of Ashtekar variables [2] with no extra fields or extra-dimensions: only gravity and the laws of quantum physics. The basic idea is therefore very simple. One could naturally ask why such a simple idea has not been explored until recently, for gravity and quantum physics have existed for almost a century. In actuality, quantizing general relativity with the “standard” tools of quantum mechanics has been investigated from its inception, but it immediately faced huge problems: the canonical quantization à la ADM [1] leads to a system of highly nonlinear equations (the famous constraints) that are simply impossible to solve, whereas the perturbative path integral quantization makes no sense since gravity is non-renormalizable.

Does Loop Quantum Gravity overcome these fundamental difficulties? An honest answer would be: we still do not know. Why? Because, so far, Loop Quantum Gravity has “only” opened a new route toward the quantization of gravity, and we are still far from the end of the story. Nonetheless, the road is very fascinating. Among other things, it has allowed us to introduce very interesting new ideas, such as (so-called) background independence, and to formulate, for the first time, questions about the structure of space-time at the Planck scale, in a mathematically well-defined way. Loop Quantum Gravity is not (yet) a consistent theory of quantum gravity, but it has proposed very exciting preliminary results.

The starting point has been the discovery by Ashtekar of a new formulation of gravity. In the Ashtekar variables, gravity reveals strong similarities with $SU(2)$ Yang–Mills theory and, when starting to quantize general relativity, one makes use of the techniques developed for gauge theories. In particular, the physical states of quantum gravity are expected to be constructed from so-called spin-network states, which are a generalization of the Wilson loops and are associated to “colored three-dimensional topological graphs.” Thus, space slices are described in terms of graphs at the Planck regime and their geometrical content is encoded into the coloration of each graph. Roughly, colored graphs are for quantum gravity what the quantum numbers (n, ℓ, m) are for the hydrogen atom: (n, ℓ, m) characterize states of the electron in the hydrogen atom and a colored graph characterizes a state of quantum geometry. Spin-network states are shown to be eigenstates of certain geometrical operators, such as the area and the volume operators, with discrete eigenvalues, making quantum spaces discrete in Loop Quantum Gravity. The theoretical framework for describing these quantum geometries is mathematically very well defined and has already been exposed in several reference books and articles [4].

If we choose to view Loop Quantum Gravity as a starting point for understanding motion at the Planck scale, there comes the question of the description of the matter fields, and of their coupling to quantum gravity. Contrary to String Theory, Loop Quantum Gravity is, a priori, a quantization of pure gravity. A way to include matter in that scheme consists in first considering the classical coupling between the Einstein–Hilbert action with a (Klein–Gordon, Dirac, Maxwell, or Yang–Mills) field

and then quantizing the coupled system. Prior to quantization, one has to reformulate the coupled theory in terms of Ashtekar variables, which is in fact immediate. Thus, not only it is, in principle, possible to consider all the matter fields of the standard model but also one can directly include super-symmetry in that scheme.

As one could expect, in general, the presence of extra fields makes Loop Quantum Gravity much more complicated, but it has been shown that Loop Quantum Gravity techniques can be extended to these cases, and quantum gravity effects make the resulting Quantum Field Theories free of UV and IR divergences [28]. Thus, one concretely realizes the old idea that UV divergences in Quantum Field Theory are a reflection of our poor understanding of the physics at very short distances and quantum gravity should provide a regulator for Quantum Field Theory. However, we do not know how to solve the dynamics explicitly, that is, we do not have any ideas for the solution of the quantum equations of motion.

One idea for overcoming this difficulty is to assume that the matter field would be so “light” that it would not affect the (quantum) space-time structure and would follow the quantum analogue of a geodesic curve. This hypothesis appears immediately inconsistent, because there exists no regime in which space-time is quantized and the matter coupling to gravity can be neglected. A quantum gravity phenomenology has been developed to provide a more or less realistic picture of the effects of the quantized background on the motion. In that framework, many have predicted, for instance, a violation of Lorentz invariance, which manifests itself in the dispersion relation of some particles. These results have been discussed and criticized extensively in the literature. We will not continue this discussion here, but we do want to at least underline the fact that the discreteness of space is the one link that exists between this phenomenology and Loop Quantum Gravity. It is definitively clear that no one yet has a precise idea of what is motion in Loop Quantum Gravity.

1.3 Three-Dimensional Quantum Gravity Is a Fruitful Toy Model

One way to be more precise is to study simplified models of quantum gravity. Three-dimensional quantum gravity is such a toy model that has been explored considerably over the last 20 years, starting from the fundamental article of Witten who established an amazing relation between three-dimensional quantum gravity and the Jones polynomials [32]. Previously, three-dimensional gravity was supposed to be too trivial to deserve any attention: there are no local degrees of freedom, there is no gravitational attraction between massive particles – whose coupling to gravity creates “only” a conical singularity in the space-time at the location of the particle. This apparent simplicity hides not only incredibly rich mathematical structures but also a real physical interest in three-dimensional gravity, which may help us understand important conceptual issues concerning the problem of time, and how to deal with invariance under diffeomorphisms, for instance. The discovery of black holes in three-dimensional Lorentzian anti-de Sitter gravity has also greatly increased the interest in such a toy model [5]. Many quantization schemes have been developed, as evidenced in the book by Carlip [10]. The coupling to massive and spinning

particles has also been thoroughly studied at both the classical and quantum levels, and has revealed a close relationship between particle dynamics and knot invariants in three-dimensional manifolds [32].

Naively, it might seem to make no physical sense to quantize gravity coupled with point particles: in the regime where space-time becomes quantized, we expect the matter field to be quantized as well and then to be described in terms of fields instead of particles. In fact, the coupling to point particles is not completely devoid of physical interest because it appears to be a good starting point for understanding the coupling of quantum gravity to quantum fields. The first reason is that point particles do exist in three-dimensional general relativity contrary to the four-dimensional case. The second reason is simply to notice that if we do not know how to quantize gravity coupled to matter fields starting from the quantization of the matter fields in a given (flat) background and then perturbatively quantizing the geometry, we could try the other way around. Indeed, why not first try quantizing gravity non-perturbatively, keeping the matter classical, and then proceed to the quantization of the matter degrees of freedom in the quantum background? This point of view makes some sense, as pure quantum gravity is very well understood in three dimensions. Furthermore, it was very fruitful and led to the very first full quantization of a massive self-gravitating scalar field in the context of Euclidean Loop Quantum Gravity [12, 20].

The most important consequence of this study is certainly the fact that quantum gravity turns classical differential manifolds into noncommutative spaces where the noncommutativity is encoded into the Planck length $\ell_P = G\hbar/c^3$. More precisely, it has been argued that a quantum scalar field coupled to Euclidean three-dimensional gravity is equivalent to a sole quantum scalar field living in a non-dynamical but noncommutative space. The emerging noncommutative space appears to be a deformation of the standard three-dimensional Euclidean space and admits a quantum group, known as the Drinfeld (quantum) double $DSU(2)$, as “isometry group” [15]. As a consequence, the question of motion in three-dimensional quantum gravity turns into the question of motion in a noncommutative space. This problem is mathematically very well defined and admits a precise solution. In particular, the quantum space admits a fuzzy space formulation and a massive scalar field is described in terms of complex matrices. Equations of motion are finite difference equations involving the matrix coefficients and their solutions allow us to understand how the notion of motion is modified in quantum gravity. Once again, three-dimensional gravity appears as an incredibly good toy model to have for a first view of fundamental issues. This article is mainly devoted to explain how motion can be described in three-dimensional Euclidean quantum gravity.

1.4 Outline of the Article

This article is structured as follows. We start, in Section 2, with a very brief review of Loop Quantum Gravity: we focus on the aspects we think are the most important; details can be found in numerous informative references [4]. We present the main

lines of the quantization strategy, then describe the states of quantum geometry in terms of spin-networks and finally explain in which sense quantum geometries are discrete, presenting a computation of the spectrum of area operators. We also mention open issues concerning the problem of dynamics: how do we find solutions of the Hamiltonian constraint?

Section 3 is devoted to giving a precise answer to the question of motion in three-dimensional Euclidean quantum gravity. It is mainly based on the paper [21]. First, we explain why quantum gravity makes space-time noncommutative in that context. The emerging noncommutative geometry is a deformation of the classical three-dimensional Euclidean space whose “isometry” algebra is a deformation of the Euclidean symmetry algebra as well. We describe this noncommutative space and propose different equivalent formulations: of particular interest is its fuzzy space formulation where it appears as an union of concentric fuzzy spheres. Then, we show how to describe the dynamics of a massive scalar quantum field in such a noncommutative geometry: to be well defined, the scalar field must have different a priori independent components; its dynamics are governed by an action very similar to the classical one but nonlocal; equations of motion can be written as finite difference equations, which couple in general the different components of the field. We give the solutions when the field is free. When the field is not free, equations of motion do not admit generically explicit solutions. Faced with this technical difficulty, we perform a symmetry reduction to simplify the problem and propose a perturbative solution of the reduced system. The solution is interpreted as the motion of a particle in Euclidean quantum gravity.

We finish the paper with a section that contains our conclusion and outlook.

2 Casting an Eye Over Loop Quantum Gravity

Loop Quantum Gravity is a particularly intriguing candidate for a background independent non-perturbative Hamiltonian quantization of General Relativity. It is based on the Ashtekar formulation of gravity [2] that is (in a nutshell) a first order formulation where the fundamental variables are an $SU(2)$ connection A and its canonical variable, the electric field E .

2.1 The Classical Theory: Main Ingredients

The starting point is the classical canonical analysis of the Ashtekar formulation of gravity. In this framework, space-time is supposed to be (at least locally) of the form $\Sigma \times \mathbb{R}$ in order for the canonical theory to be well defined, in particular, for the Cauchy problem to be well posed. In terms of these variables, gravity offers interesting similarities with $SU(2)$ Yang–Mills theory that one can exploit to start quantizing the theory. The connection A is, strictly speaking, the analogue of the Yang–Mills gauge field. At this stage of our very brief description of the theory,

let us emphasize some aspects that are important for a good understanding of the hypotheses underlying the construction of Loop Quantum Gravity:

1. *The question of the covariance.* Due to the choice of a splitting $\Sigma \times \mathbb{R}$ of the space-time manifold, one is manifestly breaking the covariance of general relativity! It is the price to pay if one formulates a canonical description of General Relativity. In standard Quantum Field Theories (QFT), this aspect is not problematical, even if we make an explicit choice of a preferred time, because one recovers at the end of the quantization that the Quantum Theory is invariant under the Poincaré group. In General Relativity, the situation is more subtle because making a preferred time choice breaks a local symmetry whereas the Poincaré symmetry is a global one in standard QFT. The consequences of such a choice might be important in an eventual quantum theory of General Relativity. Spin-Foam models (Section 2.4) are introduced partly to circumvent this problem.
2. *Where does the group $SU(2)$ come from?* To answer this question, we briefly recall the construction of Ashtekar variables. The starting point is the first order formulation of Einstein–Hilbert action à la Palatini where the metric variables (described in terms of tetrads e) and the connection ω are considered as independent variables:

$$S[e, \omega] = \frac{1}{8\pi G} \int_{\mathcal{M}} \langle e \wedge e \wedge \star F(\omega) \rangle, \tag{1}$$

where $\langle \cdot \rangle$ holds for the trace in the fundamental representation of $sl(2, \mathbb{C})$ and \star is the hodge map in $sl(2, \mathbb{C})$. It becomes clear that the Palatini theory admits the Lorentz group $SL(2, \mathbb{C})$ as a local symmetry group. Then one performs a gauge fixing, known as the time gauge, which breaks the $SL(2, \mathbb{C})$ group into $SU(2)$, its subgroup of rotations. This is the origin of the symmetry group $SU(2)$ in Loop Quantum Gravity.

3. *The Barbero–Immirzi ambiguity.* In fact, there is a one parameter family of actions that are classically equivalent to the Palatini action. This remark has been observed first in the Hamiltonian context [7] before Holst [13] wrote the explicit form of the action:

$$S[e, \omega] = \frac{1}{8\pi G} \int_{\mathcal{M}} \left(\langle e \wedge e \wedge \star F(\omega) \rangle - \frac{1}{\gamma} \langle e \wedge e \wedge F(\omega) \rangle \right). \tag{2}$$

γ is the Barbero–Immirzi parameter. The canonical analysis of the Holst action leads to a set of canonical variables, which are a connection $A \equiv \frac{1}{2}(\omega - \gamma^{-1} \star \omega)$ and its conjugated variable E . The variable A is precisely the Ashtekar–Barbero connection. Historically, Ashtekar found this connection for $\gamma = i$: he noticed that the expression of the constraints of gravity simplify magically in that context but he had to deal with the so-called reality constraints to recover the real theory. So far, no one knows how to solve the reality constraint in the quantum theory. For that reason, the discovery of the Ashtekar–Barbero variable appeared as

a breakthrough, for the variables are no longer complex, but the price to pay is that some of the constraints (the Hamiltonian constraint) have a much more complicated expression than the complex ones. The parameter γ is not relevant in the classical theory but it leads to an ambiguity in the quantum theory that one can compare to the θ -ambiguity of QCD.

Contrary to Yang–Mills theory, gravity is not a gauge theory, it is a pure constraint system and admits as symmetry group the “huge” group of space-time diffeomorphisms supplemented with the $SU(2)$ gauge symmetries briefly described above. The symmetries are generated in the Hamiltonian sense by the constraints: the Gauss constraints $\mathcal{G}_a(x)$, the vectorial constraints $\mathcal{H}_i(x)$, and the famous Hamiltonian or scalar constraint $\mathcal{H}(x)$ where $a \in \{1, 2, 3\}$ are for internal or gauge indexes, $i \in \{1, 2, 3\}$ are for space indexes, and x denotes a space point. Of particular interest for what follows is the symmetry group $\mathcal{S} = \mathcal{G} \times \text{Diff}(\Sigma)$ where \mathcal{G} denotes the group of gauge transformations and $\text{Diff}(\Sigma)$ is the diffeomorphisms group on the hyperplane Σ . In principle, the physical phase space is obtained by first solving the constraints and second gauge-fixing the symmetries.

Currently, nobody knows how to construct the classical physical phase space, at least in four dimensions, and therefore it is nonsense to hope to quantize gravity after implementation of the constraints. In Loop Quantum Gravity, we proceed the other way around, namely, quantizing the nonphysical phase space before imposing the constraints. At this point, one could ask the question why solving the quantum constraints would be simpler than solving the classical ones. So far, we do not know any solution¹ of all the constraints even at the quantum level and hence no one knows precisely the physical degrees of freedom of Quantum Gravity. However, Loop Quantum Gravity provides very fascinating intermediate results that may give a glimpse of space-time at the Planck scale [26], a resolution of the initial singularity for the Big Bang model [8] and also a microscopic explanation of black-hole thermodynamics [25]. The problem of solving the Hamiltonian constraint is still open, but different strategies have been developed to attack it. Recently, new results [11] have opened a very promising way toward its resolution.

2.2 *The Route to the Quantization of Gravity*

This section is devoted to presenting the global strategy of Loop Quantum Gravity. We have adopted the point of view of [30], which seems to us very illuminating: we start with a general discussion on the quantization of constrained systems before discussing the case of Loop Quantum Gravity.

Starting from a symplectic (or a Poisson) manifold – the phase space \mathcal{P} – physicists know how to construct the associated quantum algebra. The basic idea is

¹ In fact, we know only one solution of all the constraints when there is a cosmological constant in the theory, known as the Kodama state [16]. This solution was discussed several years ago [27] but its physical interest remains minimal.

to promote the classical variables into quantum operators whose noncommutative product is constructed from the classical Poisson bracket. In that way, one constructs a quantum algebra \mathfrak{A} whose elements are identified with (smooth) functions on the classical phase space \mathcal{P} . The kinematical Hilbert space \mathcal{H} is the carrier space of an irreducible unitary representation of the algebra \mathfrak{A} . In the case of the quantization of a massive point particle evolving in a given potential, \mathfrak{A} is the Heisenberg algebra; the kinematical Hilbert space is unique due to the famous Stone–von Neumann theorem and the quantum states of the theory are very well understood if the dynamics are not too complicated. In general, \mathfrak{A} does not admit an unique unitary irreducible representation and one has to require some extra properties in order for \mathcal{H} to be unique. For instance, it is natural to ask that symmetries are unitarily represented on \mathcal{H} . Finally, the physical Hilbert space is obtained, directly or indirectly, from solving the constraints on the kinematical Hilbert space.

Loop Quantum Gravity is based on this program. One starts with the classical phase space \mathcal{P} which is the tangent bundle $T^*(\mathcal{C})$, where \mathcal{C} is the space of $SU(2)$ connections on the hypersurface Σ . A good “coordinate system” for \mathcal{P} is provided by the generators of the holonomy-flux algebra associated to edges e and surfaces S of Σ as follows:

$$A(e) \equiv P \exp\left(\int_e A\right) \quad \text{and} \quad E_f(S) \equiv \int_S \text{Tr}(f \star E), \tag{3}$$

where f is a Lie algebra valued function on Σ , \star is the Hodge star, Tr holds for the $SU(2)$ Killing form, and $P \exp$ is the notation for the path-ordered exponential. The symmetry group $\mathcal{S} = \mathcal{G} \times Diff(\Sigma)$ acts as an automorphism of the algebra of functions on the classical algebra. The quantization of the classical algebra is straightforward and leads to the quantum holonomy-flux algebra \mathfrak{A} . Many techniques have been used to study the representation theory of \mathfrak{A} and the Gelfand–Naimark–Segal construction is one of the most precise [30]. It consists in finding a positive state $\omega \in \mathfrak{A}^*$, which is central in the construction of the representation. Many such states exist but the requirement that ω is invariant under the action of \mathcal{S} makes the state unique [18]. Therefore, there is an unique representation π of the quantum holonomy-flux algebra \mathfrak{A} , which is invariant under the action of \mathcal{S} . This representation is the starting point of the construction of the physical states.

2.3 Spin-Networks Are States of Quantum Geometry

To make the representation π of \mathfrak{A} more concrete, let us describe its carrier space in terms of cylindrical functions. A cylindrical function $\Psi_{\Gamma, f}$ is defined from a graph $\Gamma \subset \Sigma$ with E edges and V vertices and a function $f \in C(SU(2))^{\otimes E}$. It is a complex valued function of the set of the holonomies $A = \{A(e_1), \dots, A(e_E)\}$ explicitly given by:

$$\Psi_{\Gamma, f}(A) = f(A(e_1), \dots, A(e_E)). \tag{4}$$

The set of cylindrical functions associated to the graph Γ is denoted Cyl_Γ . The carrier space of the representation π is given by the direct and non-countable sum $Cyl(\Sigma) \equiv \bigoplus_\Gamma Cyl_\Gamma$ over all graphs on Σ . Such a sum is mathematically well defined using the notion of a projective limit [3]. The vector space $Cyl(\Sigma)$ is endowed with a Hilbert space structure defined from the Ashtekar–Lewandowski measure

$$\langle \Psi_{\Gamma,f} | \Psi_{\Gamma',f'} \rangle = \delta_{\Gamma,\Gamma'} \int \left(\prod_e d\mu(A(e)) \right) \overline{f(A)} f'(A), \tag{5}$$

where $d\mu$ denotes the $SU(2)$ Haar measure. The delta symbol means that the scalar product between two states vanishes unless they are associated to exactly the same graph $\Gamma = \Gamma'$. This property makes the representation not weakly continuous. The completion of $Cyl(\Sigma)$ with respect to the Ashtekar–Lewandowski measure defines the kinematical Hilbert space \mathcal{H} . It remains necessary to impose the constraints in order to extract the physical states of quantum gravity from \mathcal{H} .

The Gauss constraint is quite easy to impose: a state $\Psi_{\Gamma,f}$ is invariant under the action of \mathcal{G} if the function f is unchanged by the action of the gauge group on the vertices of the graph. An immediate consequence is that the graph Γ has to be closed. The space of gauge invariant functions is denoted \mathcal{H}_0 and is endowed with an orthonormal basis: the basis of (gauge-invariant) spin-network states. A spin-network state $|S\rangle \equiv |\Gamma, j_e, \iota_v\rangle$ is associated to a graph Γ whose edges e are colored with $SU(2)$ unitary irreducible representations j_e and vertices v with intertwiners ι_v between representations of the edges meeting at v . Intertwiners are generalized Clebsh–Gordan coefficients. An example of a spin-network is given in Fig. 1 below.

Imposing the diffeomorphisms constraint is also relatively easy. Roughly, it consists in identifying states whose graphs are related by a diffeomorphism and that have the same colors once the graphs have been identified. The set of such conjugacy classes form the space \mathcal{H}_{diff} , which is endowed with a natural Hilbert structure inherited from the Ashtekar–Lewandowski measure. Elements of \mathcal{H}_{diff} are labeled by knots instead of graphs.

Before discussing the remaining constraint in the next section, let us give the physical interpretation of a spin-network state. To do so, we need to introduce some geometrical operators, such as those that relate to the area $a(S)$ of a surface S and

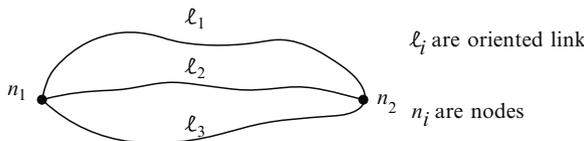


Fig. 1 The links are colored by representations of $SU(2)$ and the vertices by Clebsh–Gordan intertwiners

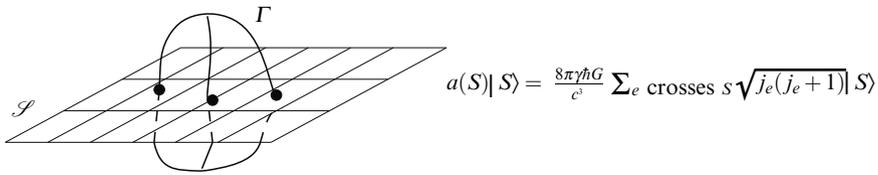


Fig. 2 Illustration of the action of the area operator on a given spin-network

the volume $v(R)$ of a domain R . The classical expressions of $a(S)$ and $v(R)$ are functions on the E -field given by [26]:

$$a(S) = \int_S d^2x \sqrt{E_i^a E_j^b n_a n_b} \quad \text{and} \quad v(R) = \int_R d^3x \sqrt{\frac{|\varepsilon_{abc} \varepsilon_{ijk} E^{ai} E^{bj} E^{ck}|}{3!}}, \quad (6)$$

where n_a denotes the normal of the surface S and ε_{abc} are the totally antisymmetric tensors. To promote these classical functions into quantum operators acting on the kinematical states, one has to introduce regularizations for the area or the volume due to the presence of the square roots in the previous classical definitions. There exist therefore some ambiguities in the definition of the quantum geometrical operators, above all in the case of the volume. For the area, the standard regularization leads to an operator $a(S)$ whose action on a spin-network state $|S\rangle$ is illustrated in Fig. 2, where the sum runs over all the edges of the graph Γ associated to $|S\rangle$ that cross the surface S . We have assumed that the edges always cross S transversely; the formula can be generalized for other, more general, cases [26]. We have explicitly introduced all the fundamental constants in order to show, in particular, the dependence on the Immirzi parameter γ [14]. We also see immediately that spin-network states are eigenstates of $a(S)$ with discrete eigenvalues. A similar but much more involved result exists for the volume operator $v(R)$: It acts on the nodes of the spin-network states and also has a discrete spectrum. As a result, at the kinematical level, space appears discrete in Loop Quantum Gravity.

2.4 The Problem of the Hamiltonian Constraint

Solving the Hamiltonian constraint is still an open issue. Two main roads have been developed to understand this constraint: the master program [29] and Spin-Foam models [22]. The master program, initiated and mainly developed by Thiemann, is an attempt to regularize the Hamiltonian constraint in order to find its kernel. Even if we still do not have a precise description of the physical Hilbert space, Thiemann proved an existence theorem that ensures physical states exist. We will not discuss this approach further here.

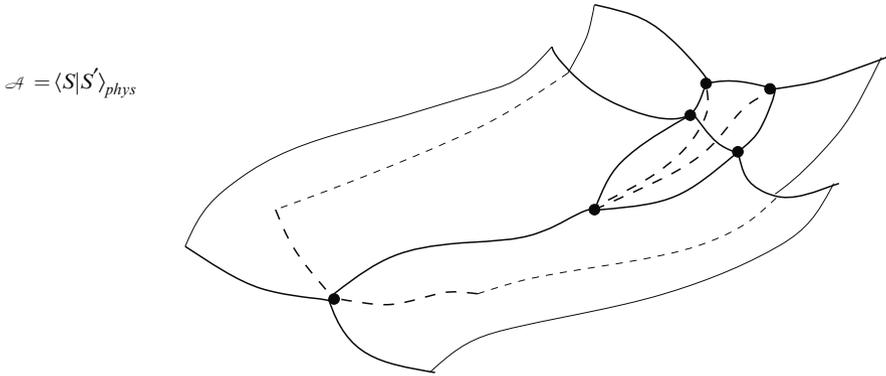


Fig. 3 Spin-Foam models propose an amplitude to each graph interpolating two given states. This amplitude is related to the physical scalar product between the two states

Spin-Foam models are an alternative attempt to solve the dynamics from a covariant point of view. The idea consists in finding the physical scalar product between spin-network states not necessarily solutions of the Hamiltonian constraint. Of course, the two problems are closely related. The physical scalar product should be given by the path integral of gravity, if one could give a meaning to this. Spin-Foam models are precisely proposals for the path integral of gravity. These proposals are based on the Plebanski formulation of gravity where gravity is described as a constrained BF theory. All BF theories are topological theories whose path integral can be easily (and formally) written in terms of combinatorial objects that we do not want to describe here. One starts with this path integral and tries to impose the constraints that make gravity a BF theory at the level of the path integral. For the moment, there is no precise implementation of the constraints, but there do exist proposals. Recently a promising new Spin-Foam model has been described [11].

In the context of Spin-Foam models, the physical scalar product between two spin-network states is given by a certain evaluation of topological graphs interpolating the two graphs defining the two spin-networks, as illustrated in Fig. 3. The rules for computing the amplitude of the graph are model dependent and can be viewed as generalizations of the Feynman rules for standard QFT. We may be far from having a clear and complete description of the physical Hilbert space of quantum gravity but the road proposed in LQG is very fascinating.

3 Three-Dimensional Euclidean Quantum Gravity

Let us underline two aspects, among the most important, concerning Loop Quantum Gravity in four dimensions. The first one is the possibility that space is discrete at the Planck scale. The second one is the difficulty in solving the dynamics of quantum gravity, and the subsequent impossibility of identifying the physical states.

Thus, to test the discreteness of space, one could use a simple toy model where the dynamics are easy to solve, and that already exhibits a discreteness of space. Three-dimensional Euclidean quantum gravity offers an ideal framework in this regard.

3.1 Construction of the Noncommutative Space

Anyone who would claim to have quantized gravity should, at the very least, be able to give a precise meaning to the formal expression for the path integral

$$\mathcal{Z} = \int [\mathcal{D}g][\mathcal{D}\varphi] e^{iS[g,\varphi]}, \quad (7)$$

where $S[g, \varphi]$ is the Einstein–Hilbert action for the metric g coupled to any matter field φ . The difficulty in performing such an integral is obviously hidden in the construction of a suitable measure $[\mathcal{D}g]$ for the space of metrics modulo diffeomorphisms.

If one uses standard perturbative techniques to compute (7), namely, one first writes $g = \eta + h$ as the sum of the flat metric η and a “fluctuation” h , then performs the integration over the variable φ on the flat metric and finally sums over all the fluctuations h , one gets into trouble because the theory is non-renormalizable. Furthermore, this method strongly breaks the covariance of the theory by specifying one background metric, and so appears not to be well adapted to general relativity.

As was mentioned in the introduction, in order to circumvent these difficulties, one could try the other way around, performing first the integration over the gravitational degrees of freedom. This idea makes sense for three-dimensional Euclidean gravity, which can be completely quantized by different techniques.

3.1.1 Quantum Gravity and Noncommutativity

Of particular interest is the Spin-Foam framework that gives tools for performing, at least formally, the integration over the metric variable in (7). Indeed, it has been argued that the path integral (7) reduces to a path integral of an effective quantum field theory $S_{eff}[\varphi]$ as follows [12]

$$\mathcal{Z} = \int [\mathcal{D}\varphi] e^{iS_{eff}[\varphi]}, \quad (8)$$

where $S_{eff}[\varphi]$ defines the action of a sole field φ on a fixed, but noncommutative, background. The noncommutative space is a deformation of the classical flat Euclidean space whose deformation parameter is the Planck length $\ell_P = \sqrt{G\hbar/c^3}$. Thus, quantum gravity would make “space-time” noncommutative, at least when space-time is three-dimensional and Euclidean. Now, we aim at giving a precise

definition of this noncommutative space. Before going into details of the definition, let us emphasize that this noncommutative space is unrelated to the particular Moyal noncommutative space [19] that appears within the String Theory framework.

The path integral approach to constructing the noncommutative geometry we have just outlined is certainly the most appealing at a conceptual level. Nonetheless, we will adopt here a more “canonical” way that is, at a technical level, simpler and also quite intuitive [15, 20]. Our starting point is the fact that the classical symmetry group of the theory is deformable into a quantum group. It is indeed well known that quantum groups play a crucial role in the quantization of three-dimensional gravity; the link between quantum gravity and knot invariants in three-dimensional manifolds [32] is certainly one of the most beautiful illustrations of this fact.

Three-dimensional gravity, for all values of the cosmological constant Λ and whatever the signature of space-time, is an exactly solvable system, as pointed out by Witten [31]. It can be reformulated as a Chern–Simons theory, which is a gauge theory whose gauge group is of the form $C^\infty(M, G)$, M being the space-time and G a Lie group. For $\Lambda = 0$ and Euclidean signature, the group $G = ISU(2) \equiv SU(2) \ltimes \mathbb{R}^3$ is the (universal cover of the) isometry group of the three-dimensional flat Euclidean space. This group gets deformed when the theory is quantized [17]. Only an idea of the deformation is given in the following section where we hope the reader gets at least the physical content of the deformation process. Mathematical and technical details can be found in [15] for instance.

3.1.2 The Quantum Double Plays the Role of the Isometry Algebra

In the combinatorial quantization scheme [9], the deformation of the isometry group is very clear. Classical groups are turned into quantum groups and the construction of the quantum physical states uses as a central tool the representation theory of these quantum groups. In the case we are interested in, the quantum group is the Drinfeld double of $SU(2)$, called also the quantum double or the double for short and denoted $DSU(2)$. The notion of quantum double is very general in the sense that it is possible to construct the quantum double DA for any Hopf algebra A . $DSU(2)$ is in fact the quantum double of the commutative algebra $C(SU(2))$ of smooth functions on $SU(2)$, which is endowed with a Hopf algebra structure: the algebra is defined by the pointwise product of functions and the co-algebra is determined by the standard coproduct $\Delta : C(SU(2)) \rightarrow C(SU(2)) \otimes C(SU(2))$ such that $\Delta(f)(a, b) = f(ab)$ for any group elements $a, b \in SU(2)$. The detailed definition of $DSU(2)$ can be found in several references [15, 20] but we only need to mention that, as a vector space, $DSU(2)$ is the tensor product $C(SU(2)) \otimes C[SU(2)]$, where $C[G]$ denotes the group algebra of G , that is, the algebra of formal linear combination of elements of G . In particular, G is a subset of $\mathbb{C}[G]$.

The double $DSU(2)$ is, precisely, a deformation of the algebra $\mathbb{C}[ISU(2)]$. In fact, the deformation concerns only the co-algebra structure, which is a central notion in constructing tensor products of representations. There exists an algebra morphism between $DSU(2)$ and $\mathbb{C}[ISU(2)] = C[\mathbb{R}^3] \otimes C[SU(2)]$, more precisely

$DSU(2)$ is included as an algebra into $\mathbb{C}[ISU(2)]$. The $\mathbb{C}[SU(2)]$ part of $DSU(2)$ is identified to $\mathbb{C}[SU(2)] \subset \mathbb{C}[ISU(2)]$ whereas $\mathbb{C}[\mathbb{R}^3]$ is sent to $C(SU(2))$. Thus, if one trivially identifies $\mathbb{C}[\mathbb{R}^3]$ with the algebra of functions $C(\mathbb{R}^3)$ on the Euclidean space \mathbb{R}^3 , then the deformation process transforms $C(\mathbb{R}^3)$ into $C(SU(2))$. Roughly, the deformation works as a compactification of the space \mathbb{R}^3 , which becomes the space $SU(2)$ that can be identified to the sphere S^3 . One understands that such a compactification needs a parameter with dimension of a length and here the Planck length ℓ_P enters. In other words, the Planck length is crucial for transforming momentum vectors \mathbf{v} in \mathbb{R}^3 into group elements $\nu(\mathbf{v})$ according to the formula $\nu(\mathbf{v}) = e^{i\ell_P \mathbf{v} \cdot \boldsymbol{\sigma}} \in SU(2)$, where the notation $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ holds for the generators of the Lie algebra $\mathfrak{su}(2)$.

In brief, the quantum double $DSU(2)$ is a deformation of the group algebra $\mathbb{C}[ISU(2)]$, where the rotational part is not affected by the deformation and the translational part is compactified in the sense described above. The group of translations becomes compact and also noncommutative. This is the fundamental reason why space-time becomes noncommutative.

3.1.3 The Quantum Geometry Defined by Its Momenta Space

The quantum geometry at the Planck scale is defined as the space that admits the quantum double as an isometry algebra. This definition is analogous to the classical one: indeed, the classical flat Euclidean space \mathbb{E}^3 admits the Euclidean group $ISU(2)$ as an isometry group and moreover can be described as the quotient $ISU(2)/SU(2)$. At the noncommutative level, one has to adapt such a construction (by quotient), for the noncommutative space is defined indirectly by its algebra of functions \mathcal{A} . However, the construction is quite easy to generalize and leads to the fact that \mathcal{A} is the convolution algebra of $SU(2)$ distributions [15], denoted

$$\mathcal{A} \equiv (C(SU(2))^*, \circ). \tag{9}$$

This algebra is trivially noncommutative. The algebra of functions $C(SU(2))$ endowed with the convolution product is a particular sub-algebra of \mathcal{A} and the product of two functions is explicitly given by:

$$(f_1 \circ f_2)(a) = \int d\mu(x) f_1(x) f_2(x^{-1}a), \tag{10}$$

where $d\mu(x)$ is the $SU(2)$ Haar measure.

The algebra \mathcal{A} admits different equivalent formulations, which have distinct physical interpretations. The formulation in (9) above is called the momentum representation: it is indeed a deformation of the commutative algebra $C(\mathbb{R}^3)$ of distributions on the tangent space \mathbb{R}^3 of \mathbb{E}^3 . At the Planck scale, the momenta become group-like.

By construction, \mathcal{A} provides a representation space of $DSU(2)$, which can be interpreted, in this way, as a symmetry algebra of \mathcal{A} whose action will be denoted \triangleright . More precisely, translation elements are functions on $SU(2)$ and act

by multiplication on \mathcal{A} whereas rotational elements are $SU(2)$ elements and act by the adjoint action:

$$\forall \phi \in \mathcal{A} \quad f \triangleright \phi = f\phi \quad \text{and} \quad u \triangleright \phi = \text{Ad}_u \phi. \tag{11}$$

The adjoint action is defined by the relation $\langle f, \text{Ad}_u \phi \rangle = \langle \text{Ad}_{u^{-1}} f, \phi \rangle$ with $\text{Ad}_u f(x)$ given by $f(u^{-1}xu)$ for any u, x in $SU(2)$.

3.1.4 The Fuzzy Space Formulation

Thus, we have a clear definition of the deformed space of momenta. To get the quantum analogue of the space $C(\mathbb{E}^3)$ itself, we need to introduce a Fourier transform on $C(SU(2))^*$. This is done by making use of harmonic analysis on the group $SU(2)$: the Fourier transform of a given $SU(2)$ -distribution is the decomposition of that distribution into (the whole set or a subset of) unitary irreducible representations (UIR) of $SU(2)$. These UIR are labeled by a spin j , they are finite dimensional of dimension $d_j = 2j + 1$. The Fourier transform is an algebra morphism that is explicitly defined by:

$$\mathcal{F} : C(SU(2))^* \longrightarrow \text{Mat}(\mathbb{C}) \equiv \bigoplus_{j=0}^{\infty} \text{Mat}_{d_j}(\mathbb{C}) \tag{12}$$

$$\phi \longmapsto \widehat{\Phi} \equiv \mathcal{F}[\phi] = \bigoplus_j \mathcal{F}[\phi]^j = \bigoplus_j (\phi \circ D^j)(e), \tag{13}$$

where $\text{Mat}_d(\mathbb{C})$ is the set of d -dimensional complex matrices, D_{mn}^j are the Wigner functions, and \circ is the convolution product. When ϕ is a function, its Fourier matrix components are obtained by performing the following integral

$$\mathcal{F}[\phi]_{mn}^j \equiv \int d\mu(u) \phi(u) D_{mn}^j(u^{-1}). \tag{14}$$

The inverse map $\mathcal{F}^{-1} : \text{Mat}(\mathbb{C}) \rightarrow C(SU(2))^*$ associates to any family of matrices $\widehat{\Phi} = \bigoplus_j \widehat{\Phi}^j$ a distribution according to the formula:

$$\langle f, \mathcal{F}^{-1}[\widehat{\Phi}] \rangle = \sum_j d_j \int d\mu(u) \overline{f(u)} \text{tr}(\widehat{\Phi}^j D^j(u)) \equiv \int d\mu(u) \overline{f(u)} \text{Tr}(\widehat{\Phi} D(u)) \tag{15}$$

for any function $f \in C(SU(2))$. We have introduced the notations $D = \bigoplus_j D^j$ and $\text{Tr} \widehat{\Phi} = \sum_j d_j \text{tr}(\widehat{\Phi}^j)$. Therefore, it is natural to interpret the algebra $\text{Mat}(\mathbb{C})$ as a deformation of the classical algebra $C(\mathbb{E}^3)$ and then three-dimensional Euclidean quantum geometry is fundamentally noncommutative and fuzzy.

3.1.5 Relation to the Classical Geometry

It is not completely trivial to show how the algebra of matrices $\text{Mat}(\mathbb{C})$ is a deformation of the classical algebra of functions on \mathbb{E}^3 .

To make it more concrete, it is necessary to construct a precise link between $C(SU(2))^*$ and $C(\mathbb{R}^3)^*$ for the former space is supposed to be a deformation of the latter. First, we remark that it is not possible to find a vector space isomorphism between them because $SU(2)$ and \mathbb{R}^3 are not homeomorphic: in more physical words, there is no way to establish a one-to-one mapping between distributions on $SU(2)$ and distributions on \mathbb{R}^3 , for $SU(2)$ and \mathbb{R}^3 have different topologies. Making an explicit link between these two spaces is in fact quite involved and one construction has been proposed in [15]. The aim of this section is to recall only the main lines of that construction; more details can be found in [15]. For pedagogical reasons, we also restrict the space $C(SU(2))^*$ to its subspace $C(SU(2))$ and then we are going to present the link between $C(SU(2))$ and $C(\mathbb{R}^3)$.

1. First, we need to introduce a parametrization of $SU(2)$: $SU(2)$ is identified with $S^3 = \{(\mathbf{y}, y_4) \in \mathbb{R}^4 | y^2 + y_4^2 = 1\}$ and any $u \in SU(2)$ is given by

$$u(\mathbf{y}, y_4) = y_4 - i\mathbf{y} \cdot \sigma \tag{16}$$

in the fundamental representation in terms of the Pauli matrices σ_i . For later convenience, we cut $SU(2)$ in two parts: the northern hemisphere U_+ ($y_4 > 0$) and the southern hemisphere U_- ($y_4 < 0$).

2. Then, we construct bijections between the spaces U_\pm and the open ball of \mathbb{R}^3 $B_{\ell_P} = \{\mathbf{p} \in \mathbb{R}^3 | p < \ell_P^{-1}\}$: to each element $u \in U_\pm$ we associate a vector $\mathbf{P}(u) = \ell_P^{-1}\mathbf{y}$. These bijections implicitly identify $\mathbf{P}(u)$ with the physical momenta of the theory. Note that this is a matter of choice: one could have chosen another expression for $\mathbf{P}(u)$ and there are no physical arguments to distinguish one from the other. We have made what seems to be, for various different reasons, the most natural and convenient choice.
3. As a consequence, any function $\phi \in C(SU(2))$ is associated to a pair of functions $\phi_\pm \in C(U_\pm)$, themselves being associated, using the previous bijections, to a pair of functions $\psi_\pm \in C_{B_{\ell_P}}(\mathbb{R}^3)$, which are functions on \mathbb{R}^3 with support on the ball B_{ℓ_P} . In that way, we construct two mappings $\alpha_\pm : C(U_\pm) \rightarrow C_{B_{\ell_P}}(\mathbb{R}^3)$ such that $\alpha_\pm(\phi_\pm) = \psi_\pm$ are explicitly given by:

$$\psi_\pm(\mathbf{p}) = \int d\mu(u) \delta^3(\mathbf{p} - \mathbf{P}(u)) \phi_\pm(u) = \frac{v_{\ell_P}}{\sqrt{1 - \ell_P^2 p^2}} \phi\left(u\left(\ell_P \mathbf{p}, \pm \sqrt{1 - \ell_P^2 p^2}\right)\right), \tag{17}$$

where $v_{\ell_P} = \ell_P^3 / 2\pi^2$. We have thus established a vector space isomorphism $\alpha = \alpha_+ \oplus \alpha_-$ between $C(SU(2))$ and $C_{B_{\ell_P}}(\mathbb{R}^3) \oplus C_{B_{\ell_P}}(\mathbb{R}^3)$. We need two functions on \mathbb{R}^3 to characterize one function of $C(SU(2))$. The mapping α_\pm satisfies the important following property: the action of the Poincaré group

$ISU(2) \subset DSU(2)$ on $C_{B_{\ell_P}}(\mathbb{R}^3)$ induced by the mappings α_{\pm} is the standard covariant one, namely,

$$\xi \triangleright \alpha_{\pm}(\phi_{\pm}) = \alpha_{\pm}(\xi \triangleright \phi_{\pm}) \quad \forall \xi \in ISU(2) \subset DSU(2). \quad (18)$$

In the r.h.s. (resp. l.h.s.), \triangleright denotes the action of $\xi \in ISU(2)$ (resp. ξ viewed as an element of $DSU(2)$) on $C(\mathbb{R}^3)$ (resp. $C(SU(2))$). This was, in fact, the defining property of the mappings α_{\pm} .

Now, we have a precise relation between $C(SU(2))$ and $C(\mathbb{R}^3)$. Using the standard Fourier transform $\mathfrak{F} : C(\mathbb{R}^3)^* \rightarrow C(\mathbb{E}^3)$ restricted to $C_{B_{\ell_P}}(\mathbb{R}^3)$, one obtains the following mapping:

$$\mathfrak{m} \equiv \mathfrak{F} \circ \alpha : C(SU(2)) \longrightarrow C_{\ell_P}(\mathbb{E}^3) \quad (19)$$

where $C_{\ell_P}(\mathbb{E}^3)$ is defined as the image of $C(SU(2))$ by \mathfrak{m} . It will be convenient to introduce the obvious notation $\mathfrak{m} = \mathfrak{m}_+ \oplus \mathfrak{m}_-$. We have the vector space isomorphism $C_{\ell_P}(\mathbb{E}^3) \simeq \widetilde{C}_{B_{\ell_P}}(\mathbb{R}^3) \oplus \widetilde{C}_{B_{\ell_P}}(\mathbb{R}^3)$, where $\widetilde{C}_{B_{\ell_P}}(\mathbb{R}^3)$ is the subspace of functions on \mathbb{E}^3 whose spectra are strictly contained in the open ball B_{ℓ_P} of radius ℓ_P^{-1} . Elements of $C_{\ell_P}(\mathbb{E}^3)$ are denoted $\Phi_+ \oplus \Phi_-$ where $\Phi_{\pm}(x) \in \widetilde{C}_{B_{\ell_P}}(\mathbb{R}^3)$. The explicit relation between $C(SU(2))$ and $C_{\ell_P}(\mathbb{E}^3)$ is

$$\Phi_{\pm}(x) \equiv \mathfrak{m}_{\pm}(\phi_{\pm})(x) = \int d\mu(u) \phi_{\pm}(u) \exp(iP(u) \cdot x). \quad (20)$$

This transform is clearly invertible.

It remains to establish the link between $C_{\ell_P}(\mathbb{E}^3)$ and the space of matrices $\text{Mat}(\mathbb{C})$. To do so, we make use of the mapping \mathcal{F} between $C(SU(2))$ and $\text{Mat}(\mathbb{C})$ and the mapping \mathfrak{m} between the same $C(SU(2))$ and $C_{\ell_P}(\mathbb{E}^3)$. If we denote by $\widehat{\Phi}_{\pm}$ the images of ϕ_{\pm} by \mathcal{F} , then we have:

$$\Phi_{\pm}(x) = \text{Tr}(K_{\pm}^{\dagger}(x) \widehat{\Phi}_{\pm}), \quad (21)$$

where K_{\pm} can be interpreted as the components of the element $K \equiv K_+ \oplus K_- \in \text{Mat}(\mathbb{C}) \otimes C_{\ell_P}(\mathbb{E}^3)$ defined by the integral:

$$K_{\pm}(x) \equiv \int_{U_{\pm}} d\mu(u) D(u) \exp(-iP(u) \cdot x). \quad (22)$$

The relation (21) is invertible. One can interpret the functions $\Phi_{\pm}(x)$ as a kind of continuation to the whole Euclidean space of the discrete functions $\widehat{\Phi}_{\pm mn}^j$, which are a priori defined only on an infinite but enumerable set of points. Given $x \in \mathbb{E}^3$, each matrix element $\widehat{\Phi}_{\pm mn}^j$ contributes to the definition of $\Phi_{\pm}(x)$ with a complex weight $\overline{K_{\pm nm}^j(x)}$.

For the moment, we have only described the vector space structure of $C_{\ell_P}(\mathbb{E}^3)$. However, this space inherits a noncommutative algebra structure when we ask the mapping \mathfrak{m} to be an algebra morphism. The product between two elements Φ_1 and

Φ_2 in $C_{\ell_P}(\mathbb{E}^3)$ is denoted $\Phi_1 \star \Phi_2$ and is induced from the convolution product \circ on $C(SU(2))$ as follows:

$$\Phi_1 \star \Phi_2 = m(m^{-1}(\Phi_1) \circ m^{-1}(\Phi_2)). \tag{23}$$

The \star -product is a deformation of the classical pointwise product.

In order to make the \star -product more intuitive, it might be useful to consider some examples of products of functions. The most interesting functions to consider first are surely the plane waves. Unfortunately, plane waves are not elements of $C(SU(2))$ but are pure distributions and hence, their study goes beyond what we have covered in this paper. Nevertheless, we will see that it is possible to extend the previously presented results to the case of the plane waves with some assumptions. Plane waves are defined as eigenstates of the generators P_a and then, as we have already underlined, a plane wave is represented by the distribution δ_u with eigenvalue $P_a(u)$, which is interpreted as the momentum of the plane wave. Plane waves are clearly degenerate, as $P_a(u)$ is not invertible in $SU(2)$: this result illustrates the fact that we need two functions $\Phi_+ \oplus \Phi_- \in C_{\ell_P}(\mathbb{E}^3)$ to characterize one function $\phi \in C(SU(2))$. The representations of the plane wave in the matrix space $\text{Mat}(\mathbb{C})$ and in the continuous space $C_{\ell_P}(\mathbb{E}^3)$ are, respectively, given by:

$$\mathcal{F}(\delta_u)^j = D^j(u)^{-1} \quad \text{and} \quad m(\delta_u)(x) \equiv w_u(x), \tag{24}$$

where $w_u(x) = \exp(iP_a(u)x^a) \oplus 0$ if $u \in U_+$ and $w_u(x) = 0 \oplus \exp(iP_a(u)x^a)$ if $u \in U_-$. The framework we have described does not include the case $u \in \partial U_+ = \partial U_-$, which is nonetheless completely considered in [15]. The \star -product between two plane waves reads:

$$w_u \star w_v = w_{uv} \tag{25}$$

if u, v , and uv belong to U_+ or U_- . This product can be trivially extended to the cases where the group elements belong to the boundary $\partial U_+ = \partial U_-$. As a result, one interprets $P_a(u) \boxplus P_a(v) \equiv P_a(uv)$ as the deformed addition rule for momenta in the noncommutative space.

Other interesting examples to consider are the coordinate functions. They are easily defined using the plane waves and their definition in the $C(SU(2))^*$ and $\text{Mat}(\mathbb{C})$ representations are:

$$\chi_a = 2i\ell_P \xi_a \delta_e \in C(SU(2))^* \quad \hat{x}_a = 2\ell_P D(J_a) \in \text{Mat}(\mathbb{C}), \tag{26}$$

where ξ_a is the $SU(2)$ left-invariant vector field and J_a the generators of the $\mathfrak{su}(2)$ Lie algebra satisfying $[J_a, J_b] = 2i\varepsilon_{ab}^c J_c$. In the $C_{\ell_P}(\mathbb{E}^3)$ representation, the coordinates are given by $X_a \equiv (x_a \oplus 0)$; only the first component is nontrivial. It becomes straightforward to show that the coordinates satisfy the relation

$$[X_a, X_b]_\star \equiv X_a \star X_b - X_b \star X_a = i\ell_P \varepsilon_{ab}^c X_c \tag{27}$$

and therefore do not commute, as expected.

3.2 Constructing the Quantum Dynamics

In this section, we introduce some mathematical tools for defining the dynamics in the noncommutative space – an integral in order to define an action, and a derivative operator in order to define the kinematical energy of the system.

3.2.1 An Integral on the Quantum Space to Define the Action

An important property is that the noncommutative space admits an invariant measure $h : C \rightarrow \mathbb{C}$. To be more precise, h is well defined on the restriction of $C \simeq C(SU(2))^*$ to $C(SU(2))$. The invariance is defined with respect to the symmetry action of the Hopf algebra $DSU(2)$. Let us give the expression of this invariant measure in the different formulations of the noncommutative space:

$$h(\phi) = \phi(e) = \text{Tr}(\widehat{\Phi}) = \int \frac{d^3x}{(2\pi)^3 v_{\ell_P}} \Phi_+(x), \tag{28}$$

where $\phi \in C(SU(2))$, $\widehat{\Phi} = \mathcal{F}[\phi]$, and $\Phi_+(x) = \mathfrak{m}_+[\phi](x)$. Note that $\int d^3x$ is the standard Lebesgue measure on the classical manifold \mathbb{E}^3 . Sometimes, such a measure is called a trace. It permits us to define a norm on the algebra C from the hermitian bilinear form

$$\langle \phi_1, \phi_2 \rangle \equiv h(\phi_1^b \phi_2) = \int d\mu(u) \overline{\phi_1(u)} \phi_2(u), \tag{29}$$

where $\phi^b(u) = \overline{\phi(u^{-1})}$.

3.2.2 Derivative Operators to Define the Dynamics

Derivative operators ∂_ξ can be deduced from the action of infinitesimal translations: given a vector $\xi \in \mathbb{E}^3$, we have $\partial_\xi = \xi^a \partial_a$ where $\partial_a = iP_a$ is the translation operator we have introduced in the previous section. When acting on the $C(SU(2))$ representation, ∂_ξ is the multiplication by the function $i\xi^a P_a$; it is the standard derivative when acting on the continuous $C_{\ell_P}(\mathbb{E}^3)$ representation (using the mapping \mathfrak{m}); finally it is a finite difference operator when acting on the fuzzy space representation $\text{Mat}(\mathbb{C})$ (using the Fourier transform \mathcal{F}). Its expression in the matrix representation is then given by:

$$(\partial_a \widehat{\Phi})_{st}^j = -\frac{1}{\ell_P d_j} D_{pq}^{1/2}(J_a) \left(\sqrt{(j+1+2qs)(j+1+2tp)} \widehat{\Phi}_{q+s, p+t}^{j+1/2} + (-1)^{q-p} \sqrt{(j-2qs)(j-2pt)} \widehat{\Phi}_{q+s, p+t}^{j-1/2} \right). \tag{30}$$

The interpretation of the formula (30) is clear. Note, however, an important point: The formula (30) defines a second-order operator in the sense that it involves $\widehat{\Phi}^{j-1/2}$ and $\widehat{\Phi}^{j+1/2}$ that are not nearest matrices but second nearest matrices.

The derivative operator is obviously necessary for defining a dynamics in the noncommutative fuzzy space. The ambiguity in the definition of P_a implies immediately an ambiguity in the dynamics. For instance, the fact that $C_a(j, k)$ relates matrices $\widehat{\Phi}^j$ with $\widehat{\Phi}^{j\pm 1/2}$ only is a consequence of the choice of P_a that is in fact a function whose non-vanishing Fourier modes are the matrix elements of a dimension 2 matrix: indeed, $P_a(u) = \ell_p^{-1} \text{tr}_{1/2}(J_a u)$. Another choice would lead to a different dynamics and then there is an ambiguity. Such ambiguities exist as well in full Loop Quantum Gravity [23].

3.2.3 Free Field: Solutions and Properties

Now, we have all the ingredients to study dynamics on the quantum space. Due to the fuzzyness of space, equations of motion will be discrete and therefore, there is in general no equivalence between the Lagrangian and Hamiltonian dynamics. Here, we choose to work in the Euler–Lagrange point of view, that is, the dynamics are governed by an action of the type:

$$S_\star[\Phi, J] = \frac{1}{2} \int \frac{d^3x}{(2\pi)^3 v_{\ell_p}} (\partial_\mu \Phi \star \partial_\mu \Phi + V(\Phi, J))_+(x), \quad (31)$$

where V is the potential that depends on the field Φ and eventually on some exterior fields J . The action has been written in the $C_{\ell_p}(\mathbb{E}^3)$ formulation to mimic easily the classical situation.

Obviously, finding the equations of motions reduces to extremizing the previous action, but with the constraint that Φ belongs to $C_{\ell_p}(\mathbb{E}^3)$: in particular, Φ (as well as the exterior field) admits two independent components Φ_\pm , which are classical functions on \mathbb{E}^3 whose spectra are bounded. The action (31) couples these two components generically. Even when one of the two fields vanishes, for instance $\Phi_- = 0$, it happens in general that the extrema of the functional $S[\Phi]$ differ from the ones that we obtain for a classical field Φ whose action would be formally the same functional, but defined with the pointwise product instead of the \star product. This makes the classical solutions in the deformed and nondeformed cases different in general. Let us state this point more precisely. When the field is free, in the sense that V is quadratic (with a mass term), deformed solutions are the same as classical ones. However, solutions are very different when the dynamics are nonlinear, and the differences are physically important.

First, let us consider the case of a free field: we assume that $V(\Phi) = \mu^2 \Phi \star \Phi$, where μ is a positive parameter. Equations of motion are:

$$\Delta \widehat{\Phi}^j + \mu^2 \widehat{\Phi}^j = 0 \quad \text{for all spin } j. \quad (32)$$

Due to the quite complicated expression of the derivative operator, it appears more convenient to solve this set of equations in the $C(SU(2))$ representation. Indeed, these equations are equivalent to the fact that $\phi = \mathcal{F}^{-1}[\hat{\Phi}]$ has support in the conjugacy classes $\theta \in [0, 2\pi[$ such that $\sin^2(\theta/2) = \ell_P^2 \mu^2$. Thus, a solution exists only if $\mu \leq \ell_P^{-1}$, in which case we write $\mu = \ell_P^{-1} \sin(m/2)$ with $0 < m < \pi$. Then the solutions of the previous system are given by $\hat{\Phi} = \hat{\Phi}_+ + \hat{\Phi}_-$ with:

$$\hat{\Phi}_\pm^j = \int d\mu(u) \mathbb{I}_m^\pm(u) \left(\alpha(u) D^j(u) + \beta(u) D^j(u)^\dagger \right), \tag{33}$$

where α and β are $SU(2)$ complex valued functions; the notation \mathbb{I}_m^\pm holds for the characteristic functions on the conjugacy class $\theta = m$ (for the $+$ sign) and $\theta = 2\pi - m$ (for the $-$ sign). These functions are normalized to one according to the relation $\int d\mu(u) \mathbb{I}_m^\pm(u) = 1$. If the fields $\Phi_\pm(x)$ are supposed to be real, the matrices $\hat{\Phi}^j$ are hermitian, and then α and β are complex conjugate functions. As a result, we obtain the general solution for the noncommutative free field written in the fuzzy space representation.

Using the mapping m , one can reformulate this solution in terms of functions on \mathbb{E}^3 . The components of Φ are given by:

$$\Phi_\pm(x) = \frac{\ell_P^2}{16\pi} \frac{\sin^2 \frac{m}{2}}{\cos \frac{m}{2}} \int_{B_{\ell_P}} d^3 p \delta(p - \mu) \left(\alpha_\pm(p) e^{ip \cdot x} + \beta_\pm(p) e^{-ip \cdot x} \right), \tag{34}$$

where B_{ℓ_P} is the Planck ball, $\alpha_\pm(p) = \alpha(u(p))$ where $u(p)$ is the inverse of $p(u)$ when u is restricted to the sets U_\pm ; a similar definition holds for β_\pm . We recover the usual solution for classical free scalar fields with the fact that the mass has an upper limit given by ℓ_P^{-1} . Therefore, the Planck mass appears to be a natural UV cutoff.

3.3 Particles Evolving in the Fuzzy Space

Important discrepancies between classical and fuzzy dynamics appear when one considers nonlinear interactions. In the case where we study the dynamics of a sole field ϕ , we may introduce self-interactions. However, even in the standard classical commutative space \mathbb{E}^3 , classical solutions of self-interacting field cannot be written in a closed form in general; and then one cannot expect to find explicit solutions for the self-interacting field evolving in the fuzzy background. Faced with such technical difficulties (which we postpone for future investigations), we will consider simpler models. We will perform symmetry reductions in order that the field ϕ depends only on one coordinate out of the three. We will interpret this model as describing one particle evolving in (Euclidean) fuzzy space-time.

3.4 Reduction to One Dimension

Let us define the algebra C^{1D} of symmetry reduced fields and its different representations: the group algebra, the matrix, and the continuous formulations.

First of all, C^{1D} can be identified to the convolution algebra $C(U(1))^*$ of $U(1)$ distributions. In particular, a function $\varphi \in C^{1D}$ is a function of $\theta \in [0, 2\pi]$. The matrix representation reduces to the Fourier representation of $C(U(1))^*$:

$$\mathcal{F}^{1D} : C(U(1))^* \longrightarrow \text{Diag}_\infty(\mathbb{C}), \quad \varphi \longmapsto \widehat{\Phi} \quad \text{with} \quad \widehat{\Phi}_a^a \equiv \varphi_a = \langle \varphi, e^{ia\theta} \rangle, \quad (35)$$

where \langle, \rangle is the duality bracket between $U(1)$ distributions and $U(1)$ functions and $\text{Diag}_\infty(\mathbb{C})$ is the algebra of infinite-dimensional diagonal complex matrices. This identity reduces to the following more concrete relation when φ is supposed to be a function:

$$\varphi_a = \frac{1}{2\pi} \int_0^{2\pi} d\theta \varphi(\theta) e^{ia\theta}. \quad (36)$$

The algebraic structure of $\text{Diag}_\infty(\mathbb{C})$ is induced from the convolution product \circ and is simply given by the commutative discrete pointwise product:

$$\forall \varphi, \varphi' \in C(U(1))^* \quad (\varphi \circ \varphi')_a = \varphi_a \varphi'_a. \quad (37)$$

Let us now construct the mapping between the convolution algebra $C(U(1))$ and the algebra $C_{\ell_P}(\mathbb{E}^1)$, which has to be understood for the moment as the one-dimensional analogue of $C_{\ell_P}(\mathbb{E}^3)$. We proceed in the same way as in the full theory:

1. first, we cut $U(1) \equiv [0, 2\pi]$ in two parts, $U_+ \equiv]-\frac{\pi}{2}, \frac{\pi}{2}[$ and $U_- \equiv]\frac{\pi}{2}, \frac{3\pi}{2}[$ where the symbol \equiv means equal modulo 2π ;
2. then, we construct two bijections between U_\pm and $B_{\ell_P}^{1D} \equiv]-\ell_P^{-1}; \ell_P^{-1}[$ by assigning to each $\theta \in U_\pm$ a momentum $P(\theta) = \ell_P^{-1} \sin \theta$;
3. the third step consists in associating to any function $\varphi \in C(U(1))$ a pair of functions $\varphi_\pm \in C(U_\pm)$, and a pair of functions $\psi_\pm \in C(\mathbb{R})$ induced by the previous bijections as follows

$$\alpha_\pm^{1D}(\varphi_\pm)(p) \equiv \psi_\pm(p) = \int \frac{d\theta}{2\pi} \delta(p - \ell_P^{-1} \sin \theta) \varphi_\pm(\theta); \quad (38)$$

4. finally, we make use of the standard one-dimensional Fourier transform \mathfrak{F}^{1D} to construct the mapping $\mathfrak{m}^{1D} = \mathfrak{m}_+^{1D} \oplus \mathfrak{m}_-^{1D} : C(U(1)) \rightarrow C_{\ell_P}(\mathbb{E}^1)$ where the components $\mathfrak{m}_\pm^{1D} = \mathfrak{F}^{1D} \circ \alpha_\pm^{1D}$ are given by:

$$\mathfrak{m}_\pm^{1D}(\varphi_\pm)(t) \equiv \Phi_\pm(t) = \int_0^{2\pi} \frac{d\theta}{2\pi} \varphi_\pm(\theta) \exp(iP(\theta)t). \quad (39)$$

The space $C_{\ell_P}(\mathbb{E}^1)$ is the image of $C(U(1))$ by \mathfrak{m} and therefore is defined by $\widetilde{C}(U_+) \oplus \widetilde{C}(U_-)$ where $\widetilde{C}(U_{\pm})$ are the images by \mathfrak{F}^{1D} of $C(U_{\pm})$. As in the full theory, this construction can be extended to the algebra $C(U(1))^*$ of distributions.

The link between the discrete and the continuous representations of C^{1D} is given by:

$$\Phi_{\pm}(t) = \sum_a \varphi_a K_{\pm}^a(t), \tag{40}$$

where the functions $K_{\pm}^a(t)$ are defined by the integrals

$$K_{\pm}^a(t) \equiv \int_{U_{\pm}} \frac{d\theta}{2\pi} e^{-ia\theta + iP(\theta)t} = (\pm 1)^a \int_0^{\frac{\pi}{2}} \frac{d\theta}{\pi} \cos\left(a\theta \mp \frac{t}{\ell_P} \sin\theta\right). \tag{41}$$

As in the general case, the relation (40) is invertible. The integral defining K_{\pm} is a simplified version of the general formula (22) and one can view these functions as the components of the element $K = K_+ \oplus K_- \in \text{Diag}_{\infty}(\mathbb{C}) \otimes C_{\ell_P}(\mathbb{E}^1)$. Furthermore, $K^a = K_+^a \oplus K_-^a$ is the image by \mathfrak{m}^{1D} of the (discrete) plane waves $\exp(-ia\theta)$. As a final remark, let us underline that K_+ and K_- are closely related by the property $K_-^a(-t) = (-1)^a K_+(t)$. This implies that the functions Φ_{\pm} are also closely related: if we assume for instance that $\varphi_{2n+1} = 0$ for any $n \in \mathbb{Z}$ then $\Phi_-(-t) = \Phi_+(t)$; if we assume on the contrary that $\varphi_{2n} = 0$ for any $n \in \mathbb{Z}$ then $\Phi_-(-t) = -\Phi_+(t)$. Such a property will have physical consequences as we will see in the sequel.

Let us give some physical interpretation of the formula (40). One can view it as a way to extend φ_a , considered as a function on \mathbb{Z} , into the whole real line \mathbb{R} . In that sense, this formula is a link between the discrete quantum description of a field and a continuous classical description. One sees that any microscopic time a contributes (positively or negatively) to the definition of a macroscopic time t with an amplitude precisely given by $K_{\pm}^a(t)$. At the classical limit $\ell_P \rightarrow 0$, $K_{\pm}^a(t)$ are maximal for values of the time $t = \pm \ell_P a$. In other words, the more the microscopic time $a\ell_P$ is close to the macroscopic time t , the more the amplitude $K_{\pm}^a(t)$ is important.

Concerning the reduced \star -product, it is completely determined by the algebra of the functions K^a viewed as elements of $C_{\ell_P}(\mathbb{E}^1)$ and a straightforward calculation leads to the following product between K^a type functions:

$$K^a \star K^b \equiv \mathfrak{m}^{1D}(\exp(-ia\theta) \circ \exp(-ib\theta)) = \delta^{ab} K_a. \tag{42}$$

This result clearly illustrates the non-locality of the \star -product.

Before going to the dynamics, let us give the expression of the derivative operator ∂_t . As for the general case, ∂_t is a finite difference operator whose action on $\text{Diag}_{\infty}(\mathbb{C})$ is given, as expected, by the following formula:

$$(\partial_t \varphi)_a = \frac{1}{2\ell_P} (\varphi_{a+1} - \varphi_{a-1}). \tag{43}$$

This expression is highly simplified compared to the more general one introduced in the previous section. However, we still have the property that ∂_t is in fact a second-order operator for it relates $a + 1$ and $a - 1$. An important consequence would be that the dynamics (of the free field) will decouple the odd components φ_{2n} and the even components φ_{2n+1} of the discrete field. Then, we will have two independent dynamics that could be interpreted as two independent particles evolving in the fuzzy space. In particular, one could associate the continuous fields $\Phi(t)^{\text{odd}}$ and $\Phi(t)^{\text{even}}$, respectively, to the families (φ_{2n}) and (φ_{2n+1}) . It is clear that $\Phi(t)^{\text{odd}}$ and $\Phi(t)^{\text{even}}$ are completely independent of one another and, using the basic properties of K_{\pm} , we find that the \pm components of each field are related by:

$$\Phi_{-}^{\text{odd}}(-t) = \Phi_{+}^{\text{odd}}(t) \quad \text{and} \quad \Phi_{-}^{\text{even}}(-t) = -\Phi_{+}^{\text{even}}(t). \tag{44}$$

Thus, Φ_{+} and Φ_{-} fundamentally describe two “mirror” particles.

3.4.1 Dynamics of a Particle: Linear versus Nonlinear

We have now assembled all the ingredients for studying the behavior of the one-dimensional field φ . When written in the continuous representation, its dynamics are governed by an action of type (31), but only one-dimensional, with no external field J , and the potential is supposed to be monomial, that is, of the form $V(\Phi) = \varepsilon/(\alpha + 1)\Phi^{*(\alpha+1)}$, with $\alpha + 1$ a non-null integer. The equations of motions are given by:

$$\Delta\Phi + \varepsilon\Phi^{*\alpha} = 0, \tag{45}$$

where $\Delta = \partial_t^2$. In the fuzzy space formulation, these equations read:

$$\frac{\varphi_{a+2} - 2\varphi_a + \varphi_{a-2}}{4\ell_P^2} = -\varepsilon\varphi_a^\alpha. \tag{46}$$

As was previously emphasized, we note that these equations do not couple odd and even integers a . For simplicity, we will consider only even spins, that is, we assume that $\varphi_{2n+1} = 0$ for all integer n .

The linear case has already been studied in the previous section. For present purposes, we consider the dynamics (46) with $\alpha \geq 2$ and we look for perturbative solutions in the parameter ε . The corresponding classical solution Φ_c reads at the first order

$$\Phi_c(t) = vt - \varepsilon \frac{v^\alpha t^{\alpha+2}}{(\alpha + 1)(\alpha + 2)} + \mathcal{O}(\varepsilon^2), \tag{47}$$

where we assume for simplicity that $\Phi_c(0) = 0$ and $\Phi'_c(0) = v$.

The perturbative expansion of the fuzzy solution is obtained using the same techniques. We look for solutions of the type $\varphi_a = \lambda a + \varepsilon \eta_a$ where $a = 2k$ by assumption, λ is a real number, and η must satisfy the following relation:

$$\begin{aligned} \eta_{2k} - \eta_{2k-2} &= -\frac{\ell_P^2}{4} \lambda^\alpha \sum_{n=1}^{k-1} (2n)^\alpha \\ &= -\frac{\ell_P^2}{4} (2\lambda)^\alpha \left[\frac{(k-1)^{\alpha+1}}{\alpha+1} + \frac{(k-1)^\alpha}{2} \right. \\ &\quad \left. + \frac{\alpha(k-1)^{\alpha-1}}{12} - \frac{\alpha(\alpha-1)(\alpha-2)}{720} (k-1)^{\alpha-3} \right. \\ &\quad \left. + \frac{\alpha(\alpha-1)(\alpha-2)(\alpha-3)(\alpha-4)}{30240} (k-1)^{\alpha-5} + \dots \right]. \end{aligned}$$

The solution is in general complicated. To be explicit, we will consider the case $\alpha = 2$. The formula simplifies considerably and, after some straightforward calculations, one can show that

$$\eta_{2k} = -\frac{\ell_P^2 \lambda^2}{12} k^2(k-1)(k+1) = -\frac{\ell_P^2 \lambda^2}{12} (k^4 - k^2). \tag{48}$$

In order to compute the $C_{\ell_P}(\mathbb{R}^1)$ representation of this solution, one uses the following relations for any integer n

$$S_{\pm}^{(n)}(t) \equiv \sum_{k=-\infty}^{+\infty} k^n K_{\pm}^{2k}(t) = \frac{1}{2(2i)^n} \frac{d^n}{d\theta^n} \exp(iP(\theta)t) \Big|_{\frac{1 \mp 1}{2} \pi}. \tag{49}$$

Applying this formula for $n = 1, 2$, and 4

$$S_{\pm}^{(1)}(t) = \pm \ell_P^{-1} t, \quad S_{\pm}^{(2)}(t) = 2\ell_P^{-2} t^2, \quad S_{\pm}^{(4)}(t) = 2(4\ell_P^{-4} t^4 + \ell_P^{-2} t^2) \tag{50}$$

one shows, after some simple calculations, that Φ_+ and Φ_- are simply related by $\Phi_+(t) = \Phi_-(-t)$ and Φ_+ is given by:

$$\Phi_+(t) = 2\lambda \ell_P^{-1} t - \varepsilon \frac{\ell_P^2 \lambda^2}{6} (2\ell_P^{-4} t^4 + \ell_P^{-2} t^2) + \mathcal{O}(\varepsilon^2). \tag{51}$$

To compare it with the classical solution Φ_c computed above (47), we impose the same initial conditions, which leads to $\lambda = v\ell_P/2$ and then the solution reads:

$$\Phi_+(t) = vt - \varepsilon \frac{v^2 t^4}{12} - \varepsilon \frac{\ell_P^2 v^2 t^2}{24} + \mathcal{O}(\varepsilon^2). \tag{52}$$

Let us interpret the solution. First, let us underline that Φ_+ and Φ_- are related by $\Phi_+(t) = \Phi_-(-t)$: thus, it seems that Φ_- corresponds to a particle evolving backward compared to Φ_+ . In that sense, the couple Φ_{\pm} behaves like a particle and a “mirror” particle: the presence of the mirror particle is due to quantum gravity effects. Second, we remark that the solution for Φ_+ differs from its classical counterpart at least order by order in the parameter ε . At the no-gravity limit $\ell_P \rightarrow 0$, Φ_+ tends to the classical solution (47). Therefore, we can interpret these discrepancies as an illustration of quantum gravity effects on the dynamics of a field.

3.4.2 Background Independent Motion

We finish this example with the question concerning the physical content of this solution. For the reasons given in the previous section, we concentrate only on the component Φ_+ . Can one interpret $\Phi_+(t)$ as the position $q(t)$ of a particle evolving in the fuzzy space? If the answer is positive, it is quite confusing because the position should be discrete valued whereas Φ_+ takes value in the whole real line a priori. In fact, we would like to interpret $\Phi_+(t) = Q(t) \in \mathbb{R}$ as the extension in the whole real line of a discrete position $q(t) \in \mathbb{Z}$. More precisely, we suppose that the space where the particle evolves is one-dimensional and discrete, and then its motion should be characterized by a \mathbb{Z} -valued function $q(t)$. If we restore the discreteness of the time variable, then the motion of the particle should, in fact, be characterized by a set of ordered integers $\{q(2k\ell_P), k \in \mathbb{Z}\}$. To make this description more concrete, we make use of the identity satisfied by $S_+^{(1)}$ (50), which implies that

$$Q(t) = \sum_{k=-\infty}^{+\infty} (2\ell_P k) K_+^{2k}(Q(t)). \tag{53}$$

This identity makes clear that $Q(t)$ can be interpreted as a kind of continuation in the whole real line of a set of discrete positions and $K_+^{2k}(Q(t))$ gives the (positive or negative) weight of the discrete point $2\ell_P k$ in the evaluation of the continuous point $Q(t)$. Therefore, one can associate an amplitude $\mathcal{P}(k|\tau)$ to the particle when it is at the discrete position $Q = 2\ell_P k$ and at the discrete time $t = 2\ell_P \tau$ (in Planck units) in the fuzzy space. This amplitude is given by:

$$\mathcal{P}(k|\tau) = \frac{K_+^{2k}(Q(2\ell_P \tau))}{\sum_{j=-\infty}^{+\infty} K_+^{2j}(Q(2\ell_P \tau))} = K_+^{2k}(Q(2\ell_P \tau)), \tag{54}$$

because the normalization factor equals one. These amplitudes cannot really be interpreted as statistical weights because they can be positive or negative. Nevertheless, they contain all the information of the dynamics of the particle in the sense that one can reconstruct the dynamics from these data. Therefore, we obtain a background independent description of the dynamics of the particle that can be a priori anywhere at any time: its position $2k\ell_P$ at a given time $2\tau\ell_P$ is characterized by the

amplitude previously defined. Furthermore, the amplitude is maximum around the classical trajectory, that is, when $Q(2\ell_P\tau) = 2\ell_P k$, and gives back the classical trajectory at the classical limit defined by $k, \tau \rightarrow \infty, \ell_P \rightarrow 0$ with the products $k\ell_P$ and $\tau\ell_P$, respectively, fixed to the values t (classical time) and Q (classical position).

4 Discussion

In this article, we have tackled the question of motion in Quantum Gravity. As we have already emphasized, it is certainly too early to discuss this question in detail. However it is at least possible to raise some preliminary problems that one needs to resolve if one aims at understanding what motion means at the Planck scale. Among the most fundamental problems are the questions of the deep structure of space-time and those of the description of matter fields in Quantum Gravity. Loop Quantum Gravity proposes a very clear answer to these questions (even if the specific viewpoint adopted here has warranted extensive discussion).

For this reason, we think that Loop Quantum Gravity presents a useful framework for discussing the question of motion at very short distances. We started with a very brief review of Loop Quantum Gravity, insisting on the kinematical aspects: the description of the kinematical states in terms of spin-networks and the computation of the spectrum of the so-called area and volume operators. We explained in what sense space appears discrete in Loop Quantum Gravity. We finished by mentioning the fundamental problem of the dynamics that, so far, no one knows how to solve, namely, the remaining Hamiltonian constraint. Nonetheless, different promising strategies have been developed to solve this issue. For the moment, this relative failure prevents us from discussing the question of the motion that is intimately linked to the question of the dynamics.

As a consequence, in a second part, we presented a toy model where the dynamics are very well understood: three-dimensional Euclidean quantum gravity with no cosmological constant. This model is exactly solvable and shares several characteristics with Loop Quantum Gravity, including the discreteness of space. Furthermore, the coupling to a matter field is very well understood and leads to a description of scalar fields in terms of complex matrices evolving in a noncommutative fuzzy geometry. Therefore, the question of motion at the Planck scale reduces in that case to the resolution of finite difference equations involving matrix coefficients, which are the quantum analogue of the equations of motions. We propose a solution of these equations in simple examples: the free field and the one-dimensional field. These examples are nice illustrations of what could represent motion more generally in Quantum Gravity.

It is nonetheless clear that we are far from a precise description of motion in full Loop Quantum Gravity. We hope that the examples we have developed in this paper will shed light on this problem.

References

1. R. Arnowitt, S. Deser, C. W. Misner, in *Gravitation: an Introduction to Current Research*, ed. by L. Witten (Wiley, New York, 1962), pp. 227–265
2. A. Ashtekar, Phys. Rev. Lett. **57**, 2244 (1986)
3. A. Ashtekar, J. Lewandowski, J. Math. Phys. **36**, 2170 (1995)
4. A. Ashtekar, J. Lewandowski, Class. Q. Grav. **21**, R23 (2004); C. Rovelli, *Quantum Gravity* (Cambridge University Press, Cambridge, 2004); T. Thiemann, *Introduction to Modern Canonical Quantum General Relativity* (Cambridge University Press, Cambridge, 2004)
5. M. Banados, J. Zanelli, C. Teitelboim, Phys. Rev. Lett. **69**, 1849 (1992)
6. L. Barack, Chapter 12 of this volume
7. J.F. Barbero, Phys. Rev. D **51**, 5507 (1995); G. Immirzi, Class. Q. Grav. **14**, 177 (1997)
8. M. Bojowald, Phys. Rev. Lett. **86**, 5227 (2001)
9. E. Buffenoir, K. Noui, P. Roche, Class. Q. Grav. **19**, 4953 (2002)
10. S. Carlip, *Quantum Gravity in 2+1 Dimensions* (Cambridge University Press, Cambridge, 1998)
11. J. Engle, R. Pereira, C. Rovelli, Phys. Rev. Lett. **99**, 161301 (2007)
12. L. Freidel, E. Livine, Class. Q. Grav. **23**, 2021 (2006); L. Freidel, E. Livine, Phys. Rev. Lett. **96**, 221301 (2006)
13. S. Holst, Phys. Rev. D **53**, 5966 (1996)
14. G. Immirzi, Nucl. Phys. Proc. Suppl. **57**, 65 (1997)
15. E. Joung, J. Mourad, K. Noui, J. Math. Phys. **50**, 052503 (2009)
16. H. Kodama, Phys. Rev. D **42**, 2548 (1990)
17. T.H. Koornwinder, N.M. Muller, J. Lie Theory **7**, 101 (1997); Erratum, *ibidem* **8**, 187 (1998); F.A. Bais, N.M. Muller, B.J. Schroers, Nucl. Phys. B **640**, 3 (2002)
18. J. Lewandowski, A. Okolow, H. Sahlmann, T. Thiemann, Commun. Math. Phys. **267**, 703 (2006)
19. J.E. Moyal, Proc. Camb. Phil. Soc. **45**, 99 (1949)
20. K. Noui, J. Math. Phys. **47**, 102501 (2006); K. Noui, Class. Q. Grav. **24**, 329 (2007)
21. K. Noui, Phys. Rev. D **78**, 105008 (2008)
22. A. Perez, Class. Q. Grav. **20**, R43 (2003)
23. A. Perez, Phys. Rev. D **73**, 044007 (2006)
24. J. Polchinski, *String Theory* (Cambridge University Press, Cambridge, 1998)
25. C. Rovelli, Phys. Rev. Lett. **14**, 3288 (1996); K. Krasnov, Phys. Rev. D **55**, 3505 (1997); A. Ashtekar, J. Baez, A. Corichi, K. Krasnov, Phys. Rev. Lett. **80**, 904 (1998)
26. C. Rovelli, L. Smolin, Nucl. Phys. B **442**, 593; Erratum, *ibidem* **456**, 753 (1995); A. Ashtekar, J. Lewandowski, Class. Q. Grav. **14**, A55 (1997); A. Ashtekar, J. Lewandowski, Adv. Theor. Math. Phys. **1**, 388 (1997)
27. L. Smolin, arXiv:hep-th/0209079v1 (2002); E. Witten, arXiv:gr-qc/0306083v2 (2003); L. Freidel, L. Smolin, Class. Q. Grav. **21**, 5685 (2004)
28. T. Thiemann, Class. Q. Grav. **15**, 1281 (1998)
29. T. Thiemann, Class. Q. Grav. **23**, 2249 (2006)
30. T. Thiemann, in *Approaches to Fundamental Physics*, ed. by I.-O. Stamatescu, E. Seiler, Lect. Notes Phys. **721** (Springer, Berlin, 2007), p. 185
31. E. Witten, Nucl. Phys. B **311**, 46 (1988)
32. E. Witten, Commun. Math. Phys. **121**, 351 (1989)

Free Fall and Self-Force: an Historical Perspective

Alessandro Spallicci

Abstract Free fall has signed the greatest markings in the history of physics through the leaning Pisa tower, the Woolsthorpe apple tree and the Einstein lift. The perspectives offered by the capture of stars by supermassive black holes are to be cherished, because the study of the motion of falling stars will constitute a giant step forward in the understanding of gravitation in the regime of strong field. After an account on the perception of free fall in ancient times and on the behaviour of a gravitating mass in Newtonian physics, this chapter deals with last century debate on the repulsion for a Schwarzschild–Droste black hole and mentions the issue of an infalling particle velocity at the horizon. Further, black hole perturbations and numerical methods are presented, paving the way to the introduction of the self-force and other back-action related methods. The impact of the perturbations on the motion of the falling particle is computed via the tail, the back-scattered part of the perturbations, or via a radiative Green function. In the former approach, the self-force acts upon the background geodesic; in the latter, the geodesic is conceived in the total (background plus perturbations) field. Regularisation techniques (mode-sum and Riemann–Hurwitz z function) intervene to cancel divergencies coming from the infinitesimal size of the particle. An account is given on the state of the art, including the last results obtained in this most classical problem, together with a perspective encompassing future space gravitational wave interferometry and head-on particle physics experiments. As free fall is patently non-adiabatic, it requires the most sophisticated techniques for studying the evolution of the motion. In this scenario, the potential of the self-consistent approach, by means of which the background geodesic is continuously corrected by the self-force contribution, is examined.

A. Spallicci (✉)

Observatoire des Sciences de l'Univers en région Centre, Université d'Orléans,
LPC2E, Campus CNRS, 3A Avenue de la Recherche Scientifique, 45071 Orléans, France
e-mail: spallicci@cnrs-orleans.fr

1 Introduction

The two-body problem in general relativity remains one of the most interesting problems, being still partially unsolved. Specifically, the free fall, one of the eldest and classical problems in physics, has characterised the thinking of the most genial developments and it is taken as reference to measure our progress in the knowledge of gravitation. Free fall contains some of the most fundamental questions on relativistic motion. The mathematical simplification, given by the reduction to a 2-dimensional case, and the non-likelihood of an astrophysical head-on collision should not throw a shadow on the merits of this problem. Instead, it may be seen as an arena where to explore part of the relevant features that occur to general orbits, e.g. the coupling between radial and time coordinates.

Although it is easily argued that radiation reaction has a modest impact on radial fall due to the feebleness of cumulative effects (anyhow, in case of high or even relativistic – a fraction of c – initial velocity of the falling particle, it is reasonable to suppose a non-modest impact on the waveform and possibly the existence of a signature), it would be presumptuous to consider free fall simpler than circular orbits, or even elliptic orbits if in the latter adiabaticity may be evoked. Adiabaticity has been variously defined in the literature, but on the common ground of the secular effects of radiation reaction occurring on a longer time scale than the orbital period. One definition refers to the particle moving anyhow, although radiating, on the background geodesic (local small deviations approximation), of obviously no-interest herein; another, currently debated for bound orbits, to the secular changes in the orbital motion being stemmed solely by the dissipative effects (radiative approximation); the third to the radiation reaction time scale being much longer than the orbital period (secular approximation), which is a rephrasing of the basic assumption.

But in radial fall such an orbital period does not exist. And as the particle falls in, the problem becomes more and more complex. In curved spacetime, at any time the emitted radiation may backscatter off the spacetime curvature, and interact back with the particle later on. Therefore, the instantaneous conservation of energy is not applicable and the momentary self-force acting on the particle depends on the particle's entire history. There is an escape route, though, for periodic motion. But energy–momentum balance cannot be evoked in radial fall, lacking the opportunity of any adiabatic averaging. The particle reaction to its radiation has thus to be computed and implemented immediately to determine the effects on the subsequent motion. It is a no-compromise analysis, without shortcuts. Thus, the computation and the application of the back-action all along the trajectory and the continuous correction of the background geodesic is the only semi-analytic way to determine motion in non-adiabatic cases. And once this self-consistent approach is mastered for radial infall, where simplification occurs for the two-dimensional nature of the problem, it shall be applicable to generic orbits.

It is worth reminding that the non-adiabatic gravitational waveforms are one of the original aims of the self-force community, since they express (i) the physics closer to the black hole horizon; (ii) the most complex trajectories; and (iii) the most tantalising theoretical questions.

The head-on collisions of black holes and the associated radiation reaction were evoked recently in the context of particle accelerators and thereby showing the richness of the applicability of the radial trajectory also beyond the astrophysical realm. As gravity is claimed by some authors to be the dominant force in the transplanckian region, the use of general relativity is adopted for their analysis.

This chapter reviews the problem of free fall of a small mass into a large one, from the beginning of science, whatever this may mean, to the application of the self-force and of a concurring approach, in the last 14 years. There is no pretension of exhaustiveness and, furthermore, justifiably or not, for this review some topics have been disregarded, namely: any orbit different from radial fall; radiation reaction in electromagnetism; but also the head-on of comparable masses and Kerr geometry, post-Newtonian (pN) and effective one-body (EOB) methods; and quantum corrections to motion.

Herein, the terms of self-force and radiation reaction are used rather loosely, though the latter does not include non-radiative modes. Thus, the self-force describes any of the effects upon an object's motion which are proportional to its own mass. Nevertheless, to the term self-force is often associated a specific method and it is preferable to adopt the term back-action whenever such association is not meant.

Geometric units ($G = c = 1$) and the convention $(-, +, +, +)$ are adopted, unless otherwise stated. The full metric is given by $\bar{g}_{\alpha\beta}(t, r) = g_{\alpha\beta}(r) + h_{\alpha\beta}(t, r)$ where $g_{\alpha\beta}$ is the background metric of a black hole of mass M and $h_{\alpha\beta}$ is the perturbation caused by a test particle of mass m .

2 The Historical Heritage

The analysis of the problem of motion certainly did not start with a refereed publication and it is arduous to identify individual contributions. Therefore, an arbitrary and convenient choice has led to select only renowned names.

Aristotélēs in the fourth century BC analysed motion qualitatively rather than quantitatively, but he was certainly more geared to a physical language than his predecessors. His views are scattered through his works, though mainly exposed in the *Corpus Aristotelicum*, collection of the works of Aristotélēs, that has survived from antiquity through Medieval manuscript transmission [11]. He held that there are two kinds of motion for inanimate matter, natural and unnatural. Unnatural motion is when something is being pushed: in this case the speed of motion is proportional to the force of the push. Natural motion is when something is seeking its natural place in the universe, such as a stone falling, or fire rising. For the natural motion of objects falling to Earth, Aristotélēs asserted that the speed of fall is proportional to the weight, and inversely proportional to the density of the medium the body is falling through. He added, though, that there is some acceleration as the body approaches more closely its own element; the body increases its weight and speeds up.

The more tenuous a medium is, the faster the motion. If an object is moving in void, Aristotélēs believed that it would be moving infinitely fast.

After two centuries, Hipparkhos said, through Simplikios [186], that bodies falling from high do experience a restraining factor which accounts for the slower movement at the start of the fall.

Gravitation was a domain of concern in the flourishing Islamic world between the ninth and the thirteenth century for ibn Shākir, al-Bīrūnī, al-Haytham, al-Khazini. It is doubtful whether gravitation was in their minds in the form of a mutual attraction of all existing bodies, but the debate acquired significant depth, although not benefiting of any experimental input. Conversely, in Islamic countries, experiments were performed as deemed necessary for the development of science. In this sense, there was a large paradigmic shift with respect to Greek philosophers, more oriented to abstract speculations.

Leonardo da Vinci¹ stated that each object does not move by its own, and when it moves, it moves under an unequal weight (for a higher cause); and when the wish of the first engine stops, immediately the second stops [121]. Further, in the context of fifteenth century gravitation, da Vinci compared planets to magnets for their mutual attraction.

Perception of the beginning of modern science in the early seventeenth century is connected on one hand to a popular legend, according to which Galilei dropped balls of various densities from the Tower of Pisa, and found that lighter and heavier ones fell at the same speed (in fact, he did quantitative experiments with balls rolling down an inclined plane, a form of falling that is slow enough to be measured without advanced instruments); on the other hand, modern science developed when the natural philosophers abandoned the search for a cause of the motion, in favour of the search for a law describing such motion. The law of fall, stating that distances from rest are as the squares of the elapsed times, appeared already in 1604 [86] and further developed in two famous essays [87, 88].

After another century, another legend is connected to Newton [213] who indeed himself told that he was inspired to formulate his theory of gravitation [156] by watching the fall of an apple from a tree, as reported by W. Stukeley and J. Conduit. The fatherhood of the inverse square law, though, was claimed by R. Hooke and it can be traced back even further in the history of physics.

The pre-Galilean physics, see Drake [63],² had an insight on phenomena that is not to be dismissed at once. For instance, the widespread belief that fall is unaffected by the mass of the falling object shall be examined throughout the chapter, through the concept of Newtonian back-action and through the general relativistic analysis of the capture of stars by supermassive black holes.

¹ Leonardo spent his final years at Amboise, nowadays part of the French Région Centre, under invitation of François I, King of France and Duke of Orléans.

² His book presents the contributions by several less-known researchers in the flow of time, being a well argued and historical – but rather uncritical – account. An other limitation is the neglect of non-Western contributions to the development of physics.

3 Uniqueness of Acceleration and the Newtonian Back-Action

One of the most mysterious and sacred laws in general relativity is the equivalence principle (EP). Confronted with ‘the happiest thought’ of Einstein’s life, it is a relief, for those who adventure into its questioning, to find out that notable relativists share this humble opinion.³ This principle is variously defined and here below some most popular versions are listed:

1. All bodies equally accelerate under inertial or gravitational forces.
 2. All bodies equally accelerate independently from their internal composition.
- In general relativity, the language style gets more sophisticated:
3. At every spacetime point of an arbitrary gravitational field, it is possible to choose a locally inertial coordinate system such that the laws of nature take the same form as in an unaccelerated coordinate system. The laws of nature concerned might be all laws (strong EP), or solely those dealing with inertial motion (weak EP) or all laws but those dealing with inertial motion (semi-strong EP).
 4. A freely moving particle follows a geodesic of spacetime.

It is evident that both conceptually and experimentally, the above different statements are not necessarily equivalent,⁴ although they can be connected to each other (e.g. the EP states that the ratio of gravitational mass to inertial mass is identical for all bodies and convenience suggests that this ratio is posed equal to unity). In this chapter, only the fourth definition will be dealt with⁵ and interestingly, it can

³ Indeed, it has been stated by Sygne [196] ‘...Perhaps they speak of the principle of equivalence. If so, it is my turn to have a blank mind, for I have never been able to understand this principle...’

⁴ For a review on experimental status of these fundamental laws, see Will’s classical references [211, 212], or else Lämmerzahl’s alternative view [113], while the relation to energy conservation is analysed by Haugan [99].

⁵ For the first definition, it is worth mentioning the following observation [196] ‘...Does it mean that the effects of a gravitational field are indistinguishable from the effects of an observer’s acceleration? If so, it is false. In Einstein’s theory, either there is a gravitational field or there is none, according as the Riemann tensor does not or does vanish. This is an absolute property; it has nothing to do with any observer’s world-line. Space-time is either flat or curved...’ Patently, the converse is also far reaching: if an inertial acceleration was strictly equivalent to one produced by a gravitational field, curvature would be then associated to inertial accelerations. Rohrlich [173] stresses that the gravitational field must be static and homogeneous and thus in absence of tidal forces. But no such a gravitational field exists or even may be conceived! Furthermore, the particle internal structure has to be neglected.

The second definition is under scrutiny by numerous experimental tests compelled by modern theories as pointed out by Damour [46] and Fayet [79].

First and last two definitions are correct in the limit of a point mass. An interesting discussion is offered by Ciufolini and Wheeler [38] on the non-applicability of the concept of a locally inertial frame (indeed a spherical drop of liquid in a gravity field would be deformed by tidal forces after some time, and a state-of-the-art gradiometer may reach sensitivities such as to detect the tidal forces of a weak gravitational field in a freely falling cabin). Mathematically, locality, for which the metric tensor $g_{\mu\nu}$ reduces to the Minkowski metric and the first derivatives of the metric tensor are zero, is limited by the non-vanishing of the Riemann curvature tensor, as in general certain

be reformulated, see Detweiler and Whiting [59, 209], in terms of geodesic motion in the perturbed field. Then, the back-action results into the geodesic motion of the particle in the metric $g_{\alpha\beta}(r) + h_{\alpha\beta}^R(t, r)$ where $h_{\alpha\beta}^R$ is the regular part of the perturbation caused by a test particle of mass m . Thus, the concept of geodesic motion is adapted to include the influence of m through $h_{\alpha\beta}^R$.

A teasing paradox concerning radiation has been conceived relative to a charge located in an Earth orbiting spacecraft. Circularly moving charges do radiate, but relative to the freely falling space cabin the charge is at rest and thus not radiating. Ehlers [171] solves the paradox by proposing that ‘It is necessary to restrict the class of experiments covered by the EP to those that are isolated from bodies of fields outside the cabin’. The transfer of this paradox to the gravitational case, including the case of radial fall, is immediate.

The EP is receptive of another criticism directed at the relation between the foundations of relativity and their implementation: it is somehow confined to the introduction of general relativity, while, for the development of the theory, a student of general relativity may be rather unaware of it.⁶

A popular but wrong interpretation of the EP states that all bodies fall with the same acceleration independently from the value of their mass (sometimes referred as the uniqueness of acceleration). This view is portrayed or vaguely referred to in some undergraduate textbooks, and anyhow largely present in various websites. Concerning the uniqueness of acceleration, non-radiative relativistic modes in a circular orbit were analysed by Detweiler and Poisson [58], who showed how the low multipole contributions to the gravitational self-acceleration may produce physical effects, within gauge arbitrariness ($l = 0$ determines a mass shift, $l = 1$ a centre of mass shift). In [58], the stage is set by a discussion on the gravitational self-force in Newtonian theory for a circular orbit. Herein, exactly the same pedagogical demonstration of theirs is applied to free fall.

A small particle of mass m is in the gravitational field of a much larger mass M . The origin of the coordinate system coincides with the centre of mass. The positions of M , m and a field point P are given by $\vec{\rho}$, \vec{R} and \vec{r} , respectively (the absolute value of \vec{r} is r and $m\vec{R} + M\vec{\rho} = 0$). In case of the sole presence of M , the potential and the acceleration at P are given by:

$$\Phi_0(r) = -\frac{M}{r}, \quad \vec{g}_0(r) = -\nabla\Phi_0(r) = -\frac{M}{r^3}\vec{r}. \tag{1}$$

If m is also present, M is displaced from the origin and the potential is:

$$\Phi(r) = -\frac{M}{|\vec{r} - \vec{\rho}|} - \frac{m}{|\vec{r} - \vec{R}|}. \tag{2}$$

combinations of the second derivatives of $g_{\mu\nu}$ cannot be removed. Pragmatically, it may be concluded that violating effects on the EP may be negligible in a sufficiently small spacetime region, close to a given event.

⁶ Again, this opinion is comforted [196] ‘...the principle of equivalence performed the essential office of midwife at the birth of general relativity...I suggest that the midwife be now buried with appropriate honours...’.

Since $m \ll M$, Eq. 2 is rewritten in the form of a small variation, that is $\Phi(r) = \Phi_0(r) + \delta\Phi(r)$ or else $\delta\Phi(r) = \Phi(r) - \Phi_0(r)$; thus:

$$\delta\Phi(r) = \underbrace{-\frac{m}{|\vec{r} - \vec{R}|}}_{\Phi_S} - \underbrace{\frac{M}{|\vec{r} - \vec{\rho}|}}_{\Phi_R} + \frac{M}{r}. \quad (3)$$

The potential $\delta\Phi(r)$ determines a field that exerts a force on m , that is the back-action of the particle. The singular term Φ_S diverges, but isotropically around the particle position and thus not contributing to the particle motion. Instead, the remaining regular part acts on the particle. Since $m \ll M$, the regular parts of the potential and of the acceleration, being $\nabla = \partial_r(\vec{r}/r)$, are:

$$\Phi_R(r) \simeq -\frac{M}{r} \left[1 - \frac{m}{M} \frac{\vec{R} \cdot \vec{r}}{r^2} \right] + \frac{M}{r} = m \frac{\vec{R} \cdot \vec{r}}{r^3}, \quad (4)$$

$$\vec{g}_R(r) = -\nabla\Phi_R(r) = m \frac{3(\vec{R} \cdot \vec{r})\vec{r} - r^2\vec{R}}{r^5}. \quad (5)$$

At the particle position, the two components of the acceleration are:

$$\vec{g}_R(R) = 2m \frac{\vec{R}}{R^3}, \quad \vec{g}_0(R) = -M \frac{\vec{R}}{R^3}, \quad (6)$$

and finally the total acceleration is given by (in vector and scalar form):

$$\vec{g}(R) = \vec{g}_0(R) + \vec{g}_R(R) = -\frac{M - 2m}{R^3} \vec{R}, \quad (7)$$

$$g(R) = -\frac{M}{R^2} \left(1 - 2\frac{m}{M} \right). \quad (8)$$

The Newtonian back-action of a particle of mass m falling into a much larger mass M is expressed as a correction to the classical value. This result is more easily derived if both the partition between singular and regular parts and the vectorial notation are left aside. The force exerted on m is (the origin of coordinate system is made coincident with the centre of mass for simplicity, so that $mR = M\rho$):

$$m\ddot{R} = -\frac{Mm}{(\rho + R)^2}, \quad (9)$$

and thus

$$\ddot{R} = -\frac{M}{R^2 \left(1 + \frac{\rho}{R} \right)^2} \simeq -\frac{M}{R^2} \left(1 - 2\frac{m}{M} \right), \quad (10)$$

but also

$$\ddot{R} = -\frac{M}{R(\rho + R)\left(1 + \frac{\rho}{R}\right)} \simeq -\frac{M}{R(\rho + R)}\left(1 - \frac{m}{M}\right). \quad (11)$$

It appears from the preceding computations that the falling mass is slowed down by a factor (1 or 2) proportional to its own mass and dependent upon the measurement approach adopted. It may be argued that the m/M term arises because the computation is referred to the centre of mass, but to shift the centre of mass to the centre of M is equivalent to deny the influence of m .

But instead, what about the popular belief that heavier objects fall faster? Let us consider the mass m at height h from the soil and the Earth radius R_{\oplus} ; since $\rho + R = h + R_{\oplus}$, it is found that:

$$\ddot{h} = -\frac{M}{(h + R_{\oplus})^2}\left(1 + \frac{m}{M}\right). \quad (12)$$

The mass is now falling faster thanks to a different observer system and the popular belief appears being confirmed. On the other hand, in a coordinate system whose origin is coincident and comoving with the centre of mass of the larger body M , any back-action effect disappears. For d , the distance between the two bodies, it is well known that:

$$m\ddot{d} = \frac{mM}{d^2}. \quad (13)$$

Nevertheless, the translational speed of the moving centre of mass of M (if the latter is fixed, any influence of m is automatically ruled out) is depending upon the value of m and the same applies to Eq. 9: it is not possible to find an universal reference frame in which the centre of the main mass moves equally for all various falling masses. Thus, the uniqueness of acceleration is result of an approximation, although often portrayed as an exact statement,⁷ or else consequence of gauge choice.⁸ The uniqueness of acceleration holds as long as the values of the masses of the falling

⁷ The difference between fall in vacuum and in the air has been the subject of a polemic between the former French Minister of Higher Education and Research Claude Allègre and the Physics Nobel Prize Georges Charpak, solicited by the satirical weekly 'Le Canard Enchaîné' [120]. The Minister affirmed on French television in 1999 'Pick a student, ask him a simple question in physics: take a petanque and a tennis ball, release them; which one arrives first? The student would tell you 'the petanque'. Hey no, they arrive together; and it is a fundamental problem, for which 2000 years were necessary to understand it. These are the basis that everyone should know'. The humourists wisecracked that the presence of air would indeed prove the student being right and tested their claim by means of filled and empty plastic water bottles being released from the second floor of their editorial offices. . . and asked the Nobel winner to compute the difference due to the air, whose influence was denied by the Minister. But in this polemic, no one drew the attention to the Newtonian back-action, also during the polemic revamped in 2003 by Allègre [1] who compared this time a heavy object and a paper ball. Such forgetfulness or misconception is best represented by the Apollo 15 display of the simultaneous fall of a feather and a hammer [4].

⁸ During the Bloomington 2009 Capra meeting, this state of affairs was presented as 'the confusion gauge'.

bodies are negligible. Correctly stated, the principle hardly sounds like a principle: all bodies fall with the same acceleration independently from their mass, if... we neglect their mass. Although the preceding is elementary, misconceptions tend to persist in colleges and higher education.

It is concluded that the Newtonian back-action manifests itself with different numerical factors possibly carrying opposite sign (from the Pisa tower – 100 pisan arms tall – for an observer situated at its feet, Newtonian back-action shows roughly as proportional to $1.7 \times 10^{-24} \text{ m/s}^2$ for each falling kilogram). This feature corresponds to the gauge freedom in general relativity.

For the latter, when considering perturbations, the energy radiated through gravitational waves is proportional to m^2/M and thus the energy leaking from the nominal motion. Therefore, the concept of uniqueness of acceleration is further affected, as it will be shown further.

Finally, it is quoted [59] that with only local measurements, the observer has no means of distinguishing the perturbations from the background metric. In the next section, it is shown that the concept of locality or non-locality of measurements associated to free fall, even without taking into account radiation reaction, is far from being evident and has fueled a controversy for more than 90 years.

4 The Controversy on the Repulsion and on the Particle Velocity at the Horizon

The concept of light being trapped in a star was presented in 1783 by Michell [144] in front of the Royal Society audience and later by Laplace [115, 116]. Preti [163] describes the close resemblance between the algebraic formulation of Laplace [116] and the concept of a black hole, term coined in 1967 by Wheeler [208]. In the last century, the Earth, once the attracting mass of reference, was silently replaced by the black hole. But, as many centuries before and after Newtonian gravity were necessary to formulate motion on the Earth, it should not be a surprise that it is taking more than a century to resolve the same Newtonian questions, in the more complex Einsteinian general relativity, on a black hole.

The existence and the detectability of gravitational waves, the validity of the quadrupole formula are among the notorious debates that have characterised general relativity, as described by Kennefick [111]. But closer to the topic of this chapter is the, surprisingly since almost endless, controversy on the radial motion in the unperturbed Schwarzschild or properly Schwarzschild–Droste (henceforth SD) metric [65, 66, 181],⁹ intertwined with the early debates on the apparent singularity at $2M$

⁹ Rothman [174] gives a brief historical account on Droste's independent derivation of the same metric published by Schwarzschild, in the same year 1916. Eisenstaedt [76] mentions previous attempts by Droste [64] on the basis of the preliminary versions of general relativity by Einstein and Grossmann [73], later followed by Einstein's works (general relativity was completed in 1915 and first systematically presented in 1916 [72]) and Hilbert's [101]. Antoci [2] and Liebscher [3]

and on the belief of the impenetrability of this singularity due to the infinite value of precession¹⁰ at $9/4M$.

Most references for the analysis of orbital motion, e.g. the first comprehensive analysis by Hagihara [98] or the later and popular book by Chandrasekhar [34], do not address this debate, that has invested names of the first rank in the specialised early literature.

An historically oriented essay by Eisenstaedt [77] critically scrutinises the relation that relativists have with free fall.¹¹ This section does not have any pretension of topical (e.g. photons in free fall are not dealt with) or bibliographical completeness. The questions posed in this debate concern the radial fall of a particle into an SD black hole and may be summarised as:

- Is there an effect of repulsion such that masses are bounced back from the black hole? Or more mildly, does the particle speed, although always inward, reaches a maximal value and then slows down? And if so, at which speed or at which radial coordinate?
- Does the particle reaches the speed of light at the horizon?

The discussion is largely a reflection of coordinate arbitrariness (and unawareness of its consequences), but the debaters showed sometimes a passionate affection to a coordinate frame they considered more suitable for a “real physical” measurement than other gauges. Further, ill-defined initial conditions at infinity, inaccurate wording (approaching rather than equalling the speed of light), sometimes tortuous reasonings despite the great mathematical simplicity, scarce propension to bibliographic research with consequent claim of historical findings [125], they all contributed to the duration of this debate. The approach of this section is to cut through any tortuous reasoning [160] and show the essence of the debate by means of a clean and simple presentation, thereby paying the price of oversimplification.

emphasise Hilbert’s [102] and Weyl’s [207] later derivations of solutions for spherically symmetric non-rotating bodies. Incidentally, Ferraris, Francaviglia and Reina [80] point to the contributions of Einstein and Grossmann [74], Lorentz [126] and obviously Hilbert [101] to the variational formulation.

¹⁰ Earman and Eisenstaedt [69] describe the lack of interest of Einstein for singularities in general relativity. The debate at the Collège de France during Einstein’s visit in Paris in 1922 included a witty exchange on precession (the Hadamard ‘disaster’), see Biezunski [26].

¹¹ The translation of the title and of the introduction to Section 5 of [77] serves best this paragraph ‘The impasse (or have the relativists fear of the free fall?) [...] the problem of the free fall of bodies in the frame of [...] the Schwarzschild solution. More than any other, this question gathers the optimal conditions of interest, on the technical and epistemological levels, without inducing nevertheless a focused concern by the experts. Though, is it necessary to emphasise that it is a first class problem to which classical mechanics has always showed great concern ... from Galileo; which more is the reference model expressing technically the paradigm of the lift in free fall dear to Einstein? The matter is such that the case is the most elementary, most natural, an extremely simple problem ... apparently, but which raises extremely delicate questions to which only the less conscious relativists believe to reply with answers [...]. Exactly the type of naive question that best experts prefer to leave in the shadow, in absence of an answer that has to be patently clear to be an answer. Without doubts, it is also the reason for which this question induces a very moderate interest among the relativists. ...’

Four types of measurements can be envisaged: local measurement of time dT , non-local measurement of time dt , local measurement of length dR , non-local measurement of length dr . Locality is somewhat a loose definition, but it hints at those measurements by rules and clocks affected by gravity (of the SD black hole) and noted by capital letters T, R , while non-locality hints at measurements by rules and clocks not affected by gravity (of the SD black hole) and noted by small letters t, r .¹² Therefore, for determining (velocities and) accelerations, four possible combinations do exist:

- Unrenormalised acceleration d^2r/dt^2
- Semi-renormalised acceleration d^2R/dt^2
- Renormalised acceleration d^2R/dT^2
- Semi-renormalised acceleration d^2r/dT^2

The latter has not been proposed in the literature and discussion will be limited to the first three types. The former two present repulsion at different conditions, while the third one never presents repulsion.

The first to introduce the idea of gravitational repulsion was Droste [65,66] himself. He defines:

$$dR = \frac{dr}{\sqrt{1 - \frac{2M}{r}}}, \tag{14}$$

which, after integration, Droste called the distance δ from the horizon. This quantity is derived from the SD metric posing $dt = 0$, delicate operation since the relation between proper and coordinate times varies in space as explained by Landau and Lifshits [114]; thus it may be accepted only for a static observer (obviously the notion of static observer raises in itself a series of questions; see e.g. Doughty [62], Taylor and Wheeler [198]). Through a Lagrangian and the relation of Eq. 14, for radial trajectories Droste derives that the semi-renormalised velocity and acceleration are given by (A is a constant of motion, equal to unity for a particle falling with zero velocity at infinity):

$$\frac{dR}{dt} = -\sqrt{\left(1 - \frac{2M}{r}\right) \left(1 - A + \frac{2AM}{r}\right)}, \tag{15}$$

$$\frac{d^2R}{dt^2} = -\frac{M}{r^2} \left[\sqrt{1 - \frac{2M}{r}} - \frac{2(dR/dt)^2}{\sqrt{1 - \frac{2M}{r}}} \right] = \frac{M}{r^2} \left(1 - 2A + \frac{4AM}{r}\right) \sqrt{1 - \frac{2M}{r}}, \tag{16}$$

¹² This definition is not faultless (there is no shield to gravity), but it is the most suitable to describe the debate, following Cavalieri and Spinelli [31, 32, 193] and Thirring [199].

where the constant of motion A is given by:

$$A = \left(1 - \frac{2M}{r}\right)^{-1} - (dr/dt)^2 \left(1 - \frac{2M}{r}\right)^{-3}.$$

From Eq. 16, two conditions may be derived for the semi-renormalised acceleration, for either of which the repulsion (the acceleration is positive) occurs for $A = 1$ if $r < 4M$ or else $dR/dt > \sqrt{1/2}\sqrt{1 - 2M/r}$.

Instead in his thesis [65], Droste investigated the unrenormalised velocity acceleration and for zero velocity at infinity, they are:

$$\frac{dr}{dt} = - \left(1 - \frac{2M}{r}\right) \sqrt{\frac{2M}{r}}, \quad (17)$$

$$\frac{d^2r}{dt^2} = -\frac{M}{r^2} \left[1 - \frac{2M}{r} - \frac{3(dr/dt)^2}{1 - \frac{2M}{r}} \right] = -\frac{M}{r^2} \left(1 - \frac{2M}{r}\right) \left(1 - \frac{6M}{r}\right), \quad (18)$$

for which repulsion occurs if, still for a particle falling from infinity with zero initial velocity, $r < 6M$ or else $dr/dt > 1/\sqrt{3}(1 - 2M/r)$.

The impact of the choice of coordinates on generating repulsion was not well perceived in the early days of general relativity. Further, many notable authors as Hilbert [102, 103], Page [157], Eddington [70], von Laue [205] in the German original version of his book, Bauer [24], de Jans [52–54] although indirectly by referring to the German version of [205], arrive independently and largely ignoring the existence of Droste's work, to the same conclusions in semi-renormalised or unrenormalised coordinates.

The initial conditions¹³ may astray the particle from being attracted by the gravitating mass. Indeed, Droste [66] and Page [157] refer to particles having velocities at infinity equal or larger than $1/\sqrt{2}$ for the semi-renormalised coordinates and equal or larger than $1/\sqrt{3}$ for the unrenormalised coordinates. These conditions dictate to Droste and Page that the particle is constantly slowed down when approaching the black hole and therefore impose to gravitation an endless repulsive action.

¹³ Generally, the setting of the proper initial conditions may be a delicate issue e.g. when associated with an initial radiation content expressing the previous history of the motion as it will be later discussed; or, in absence of radiation, when an external (sort of third body) mechanism prompting the motion to the two body system is to be taken into account. The latter case is represented by the thought experiment conceived by Copperstock [39] aiming to criticise the quadrupole formula. The experiment consisted in two fluid balls assumed to be in static equilibrium and held apart by a strut, with membranes to contain the fluid, until time $t = 0$. Between $t = 0$ and $t = t_1$, the strut and the membranes are dissolved and afterwards the balls fall freely. Due to the static initial conditions, there is a clear absence of incident radiation, but the behaviour of the fluid balls in the free fall phase depends on how the transition from the equilibrium to the free fall takes place. This initial dependence obscured the debate on the quadrupole formula.

In the later French editions of his book, von Laue [205] writes the radial geodesic in proper time, but it is only in 1936 that Drumaux [67] fully exploits it. Drumaux criticises the use of the semi-renormalised velocity and considers Eq. 14 as defining the physical measurement of length dR . Similarly, the relation between coordinate and proper times (for $dr = 0$) provides the physical measurement of time dT :

$$dT = \sqrt{1 - \frac{2M}{r}} dt. \quad (19)$$

Thereby, Drumaux derives the renormalised velocity and acceleration in proper time:

$$\frac{dR}{dT} = \sqrt{\frac{2M}{r}}, \quad (20)$$

$$\frac{d^2R}{dT^2} = -\frac{M}{r^2} \sqrt{1 - \frac{2M}{r}}, \quad (21)$$

for which no repulsion occurs. This approach is followed by von Rabe [206], Whittaker [210], Srinivasa Rao [194], Zel'dovich and Novikov [216]. Nevertheless, McVittie, almost 30 years after Drumaux [143], still reaffirms that the particle is pushed away by the central body as do Treder [201], also in cooperation with Fritze [202], Markley [138], Arifov [9, 10], McGruder [142]. A discussion on radar and Doppler measurements with semi-renormalised measurements was offered by Jaffe and Shapiro [107, 108]. The controversy seems to be extinguished in the 1980s, although recent research papers still refer to it, e.g. Kutschera and Zajiczek [112].

For the particle's velocity at the horizon, another, though related, debate has taken place in some of the above mentioned references as well as in Landau and Lifshits [114], Baierlein [12], Janis [109, 110], Rindler [170], Shapiro and Teukolsky [182], Frolov and Novikov [85], Mitra [151], Crawford and Tereno [40], Müller [153] the last ones being recently published. Whether the velocity is c or less, it is still the question posed by these papers.

The further step forward in the analysis of a freely falling mass into an SD black hole has taken place in the period from 1957 to 1997. In these 40 years,¹⁴ the falling mass finally radiates energy (the radiated gravitational power is proportional to the square of the third time derivative of the quadrupole moment which is different than zero), but its motion is still unaffected by the radiation emitted. The influence of the radiation on the motion of a particle of infinitesimal size was not dealt with until 1997.

¹⁴ Free fall has also been studied in other contexts. Synge [196] undertakes a detailed investigation of the problem and shows that, actually, the gravitational field (i.e. the Riemann tensor) plays an extremely small role in the phenomenon of free fall and the acceleration of 980 cm/s is, in fact, due to the curvature of the world line of the tree branch. The apple is accelerated until the stem breaks, then the world line of the apple becomes inertial until the ground collides with it.

5 Black Hole Perturbations

Perturbations were first dealt with by Regge and Wheeler [167, 168], where an SD black hole was shown to regain stability after undergoing small vibrations about its spherical form, if subjected to a small perturbation.¹⁵ The analysis was carried out thanks to the first application to a black hole of the Einstein equation at higher order.

The SD metric describes the background field $g_{\mu\nu}$ on which the perturbations $h_{\mu\nu}$ arise. It is given by:

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\phi^2). \quad (22)$$

Equation 22 originates from the Einstein field equation in vacuum, consisting in the vanishing of the Ricci tensor $R_{\mu\nu} = 0$.

Instead, the Regge–Wheeler equation derives from the vacuum condition, but this time posed on the first order variation of the Ricci tensor $\delta R_{\mu\nu} = 0$. The generic form of the variation of the Ricci tensor was found by Eisenhart [75] and it is given by $\delta R_{\mu\nu} = -\delta\Gamma_{\mu\nu;\beta}^{\beta} + \delta\Gamma_{\mu\beta;\nu}^{\beta}$ where the tensor $\delta\Gamma_{\beta\gamma}^{\alpha}$, variation of the Christoffel symbol (a pseudo-tensor), is: $\delta\Gamma_{\beta\gamma}^{\alpha} = 1/2 g^{\alpha\nu}(h_{\beta\nu;\gamma} + h_{\gamma\nu;\beta} - h_{\beta\gamma;\nu})$, being the perturbation $h_{\mu\nu} = \delta g_{\mu\nu}$. Replacing the latter in the vanishing variation of the Ricci tensor, a system of ten second order differential equations in $h_{\mu\nu}$ was obtained. Exploiting spherical symmetry, finally Regge and Wheeler got a vacuum wave equation out of the three odd-parity equations giving birth to a field that has grown immensely from the end of the 1950s.¹⁶

Zel'dovich and Novikov [215] first considered the problem of gravitational waves emitted by bodies moving in the field of a star, on the basis of the quadrupole formula, thus at large distances from the horizon, where only a minimal part of the radiation is emitted.

While a less known semi-relativistic work by Ruffini and Wheeler [178, 179] appeared in the transition from the 1960s to the 1970s, it was the work by Zerilli [218–220], where the source of perturbations was considered in the form of a radially falling particle, that opened the way to study free fall in a fully, although linearised, relativistic regime at first order. The Zerilli equation rules even-parity waves in the presence of a source, i.e. a freely falling point particle,

¹⁵ For a critical assessment of black hole stability, see Dafermos and Rodnianski [44].

¹⁶ A well-organised introduction, largely based on works by Friedman [84] and Chandrasekhar [33], is presented in the already mentioned book by the latter [34]. Some selected publications geared to the finalities of this chapter are to be listed: earlier works by Mathews [141], Stachel [195], Vishveshvara [204]; the relation between odd and parity perturbations [35]; the search for a gauge invariant formalism by Martel and Poisson [140] complements a recent review on gauge invariant non-spherical metric perturbations of the SD black hole spacetimes by Nagar and Rezzolla [154]; a classic reference on multiple expansion of gravitational radiation by Thorne [200]; the derivation by computer algebra by Cruciani [41, 42] of the wave equation governing black hole perturbations; the numerical hyperboloidal approach by Zenginoglu [217].

generating a perturbation for which the difference from the SD geometry is small. The energy–momentum tensor $T_{\mu\nu}$ is given by the integral of the world-line of the particle, the integrand containing a four-dimensional invariant δ Dirac distribution for the representation of the point particle trajectory. The vanishing of the covariant divergence of $T_{\mu\nu}$ is guaranteed by the world-line being a geodesic in the background SD geometry; in this way, the problem of the linearised theory on flat spacetime (for which the particle moves on a geodesic of flat space that determines uniform motion and thereby without emission of radiation) is avoided. Finally, the complete description of the gravitational waves emitted is given by the symmetric tensor $h_{\mu\nu}$, function of r, θ, ϕ and t .

The formalism can be summarised as follows [218–220].¹⁷ Due to the spherical symmetry of the SD field, the linearised field equations for the perturbation $h_{\mu\nu}$ are in the form of a rotationally invariant operator on $h_{\mu\nu}$, set equal to the energy–momentum tensor also expressed in spherical tensorial harmonics:

$$Q[h_{\mu\nu}] \propto T_{\mu\nu}[\delta(z_u)], \tag{23}$$

where the $\delta(z_u)$ Dirac distribution represents the point particle on the unperturbed trajectory z_u .

The rotational invariance is used to separate out the angular variables in the field equations. For the spherical symmetry on the 2-dimensional manifold on which t, r are constants under rotation in the θ, ϕ sphere, the ten components of the perturbing symmetric tensor transform like three scalars, two vectors and one tensor:

$$h_{tt}, h_{tr}, h_{rr} \quad (h_{t\theta}; h_{t\phi}), (h_{r\theta}; h_{r\phi}) \quad \begin{pmatrix} h_{\theta\theta} & h_{\theta\phi} \\ h_{\phi\theta} & h_{\phi\phi} \end{pmatrix}.$$

In the Regge–Wheeler–Zerilli formalism, the even perturbations (the source term for the odd perturbations vanishes for the radial trajectory, and given the rotational invariance through the azimuthal angle, only the index referring to the polar or latitude angle survives), going as $(-1)^l$, are expressed by the following matrix:

$$h_{\mu\nu} = \begin{pmatrix} \left(1 - \frac{2M}{r}\right) H_0 Y & H_1 Y & h_0 Y_{,\theta} & h_0 Y_{,\phi} \\ \text{sym} & \left(1 - \frac{2M}{r}\right)^{-1} H_2 Y & h_1 Y_{,\theta} & h_1 Y_{,\phi} \\ \text{sym} & \text{sym} & r^2 [KY + GY_{,\theta\theta}] & r^2 G (Y_{,\theta\phi} - \cot \theta Y_{,\phi}) \\ \text{sym} & \text{sym} & \text{sym} & r^2 \sin^2 \theta \left[K + G \left(\frac{Y_{,\phi\phi}}{\sin^2 \theta} + \cot \theta Y_{,\theta} \right) \right] \end{pmatrix}, \tag{24}$$

¹⁷ Two warnings: the literature on perturbations and numerical methods is rather plagued by editorial errors (likely herein too...) and different terminologies for the same families of perturbations. Even parity waves have been named also polar or electric or magnetic, generating some confusion (see the correlation table, Table II, in [220]). Sago, Nakano and Sasaki [180] have corrected the Zerilli equations (a minus sign missing in all right-hand side terms) but introduced a wrong definition of the scalar product leading to errors in the coefficients of the energy-momentum tensor.

where $H_0, H_1, H_2, h_0, h_1, K, G$ are functions of (t, r) and the l multipole index is not displayed. After angular dependence separation, the seven functions of (t, r) are reduced to four due to a gauge transformation for which $G = h_0 = h_1 = 0$, i.e. the Regge–Wheeler gauge.

For a point particle of proper mass m , represented by a Dirac delta distribution, the stress-energy tensor is given by:

$$T^{\alpha\beta} = m \frac{u^\alpha u^\beta}{u^0 r^2} \delta[r - z_u(t)] \delta^2[\Omega], \tag{25}$$

where $z_u(t)$ is the trajectory in coordinate time and u^α is the 4-velocity.

Any symmetric covariant tensor can be expanded in spherical harmonics [218]. For radial fall it has been shown that only three even source terms do not vanish and that two functions of (t, r) become identical. Finally, six equations are left with three unknown functions $H_0 = H_2, K, H_1$. After considerable manipulation, the following wave equation is obtained:

$$\frac{\partial^2 \Psi_l(t, r)}{\partial r^{*2}} - \frac{\partial^2 \Psi_l(t, r)}{\partial t^2} - V_l(r) \Psi_l(t, r) = S_l(t, r), \tag{26}$$

where $r^* = r + 2M \ln(r/2M - 1)$ is the tortoise coordinate; the potential $V_l(r)$ is given by:

$$V_l(r) = \left(1 - \frac{2M}{r}\right) \frac{2\lambda^2(\lambda + 1)r^3 + 6\lambda^2 M r^2 + 18\lambda M^2 r + 18M^3}{r^3(\lambda r + 3M)^2},$$

being $\lambda = 1/2(l - 1)(l + 2)$. The source $S_l(t, r)$ includes the derivative of the Dirac distribution (denoted δ'), coming from the combination of the $h_{\mu\nu}$ and their derivatives¹⁸:

$$S_l = \frac{2(r - 2M)\kappa}{r^2(\lambda + 1)(\lambda r + 3M)} \times \left\{ \frac{r(r - 2M)}{2u^0} \delta'[r - z_u(t)] - \left[\frac{r(\lambda + 1) - 3M}{2u^0} - \frac{3Mu^0(r - 2M)^2}{r(\lambda r + 3M)} \right] \delta[r - z_u(t)] \right\}, \tag{27}$$

for $u^0 = 1/(1 - 2M/z_u)$ being the time component of the 4-velocity and $\kappa = 4m\sqrt{2l + 1}\pi$. The geodesic in the unperturbed SD metric $z_u(t)$ assumes different forms according to the initial conditions¹⁹; herein, the simplest form is

¹⁸ There is an editorial error, a numerical coefficient, in the corresponding expressions (2.16) in [133] and (2.8) in [134], which the footnote 1 at page 3 in [139] does not address.

¹⁹ For a starting point different from infinity or a non-null starting velocity, but not their combination, see Lousto and Price [133–135], Martel and Poisson [139].

given, namely zero velocity at infinity. Then, $z_u(t)$ is the – numerical – inverse function of:

$$t = -4M \left(\frac{z_u}{2M} \right)^{1/2} - \frac{4M}{3} \left(\frac{z_u}{2M} \right)^{3/2} - 2M \ln \left[\frac{\sqrt{\frac{z_u}{2M}} - 1}{\sqrt{\frac{z_u}{2M}} + 1} \right]. \quad (28)$$

The coordinate velocity \dot{z}_u of the particle may be given in terms of its position z_u :

$$\dot{z}_u = - \left(1 - \frac{2M}{z_u} \right) \left(\frac{2M}{z_u} \right)^{1/2}. \quad (29)$$

The dimension of the wavefunction Ψ is such that the energy is proportional to $\int_0^\infty \dot{\Psi}^2 dt$. The wavefunction, in the Moncrief form [152] for its gauge invariance, is related to the perturbations via:

$$\Psi_l(t, r) = \frac{r}{\lambda + 1} \left[K^l + \frac{r - 2M}{\lambda r + 3M} \left(H_2^l - r \frac{\partial K^l}{\partial r} \right) \right], \quad (30)$$

where the Zerilli [219] normalisation is used for Ψ_l . For computations, this allows the choice of a convenient gauge, like the Regge–Wheeler gauge. The inverse relations for the perturbation functions K , H_2 , H_1 are given by Lousto [129, 132]:

$$K = \frac{6M^2 + 3M\lambda r + \lambda(\lambda + 1)r^2}{r^2(\lambda r + 3M)} \Psi + \left(1 - \frac{2M}{r} \right) \Psi_{,r} - \frac{\kappa u^0 (r - 2M)^2}{(\lambda + 1)(\lambda r + 3M)r} \delta, \quad (31)$$

$$\begin{aligned} H_2 = & - \frac{9M^3 + 9\lambda M^2 r + 3\lambda^2 M r^2 + \lambda^2(\lambda + 1)r^3}{r^2(\lambda r + 3M)^2} \Psi + \frac{3M^2 - \lambda M r + \lambda r^2}{r(\lambda r + 3M)} \Psi_{,r} \\ & + (r - 2M) \Psi_{,rr} + \frac{\kappa u^0 (r - 2M) [\lambda^2 r^2 + 2\lambda M r - 3M r + 3M^2]}{r(\lambda + 1)(\lambda r + 3M)^2} \delta \\ & - \frac{\kappa u^0 (r - 2M)^2}{(\lambda + 1)(\lambda r + 3M)} \delta', \end{aligned} \quad (32)$$

$$H_1 = \frac{\lambda r^2 - 3M\lambda r - 3M^2}{(r - 2M)(\lambda r + 3M)} \Psi_{,t} + r \Psi_{,tr} - \frac{\kappa u^0 \dot{z}_u (\lambda r + M)}{(\lambda + 1)(\lambda r + 3M)} \delta + \frac{\kappa u^0 \dot{z}_u r (r - 2M)}{(\lambda + 1)(\lambda r + 3M)} \delta'. \quad (33)$$

Several works by Davis, Press, Price, Ruffini and Tiomno [47–49, 175–177], but also by individual scholars like Chung [36], Dymnikova [68] and the forerunners of the Japanese school as Tashiro and Ezawa [197], Nakamura with Oohara and Kojima [155] or with Shibata [184], appeared in the frequency domain in the 1970s

and fewer later on, analysing especially the amplitude and the spectrum of the radiation emitted. Haugan, Petrich, Shapiro and Wasserman [100, 158, 183] modeled the source as a finite-sized star of dust.

For an infalling mass from infinity at zero velocity, the energy radiated to infinity for all modes [48] and the energy absorbed by the black hole [49] for each single mode, and for all modes²⁰ are given by respectively (beware, in physical units):

$$\sum_l E_l^r = 0.0104 \frac{m^2 c^2}{M}, \quad E_l^a = 0.25 \frac{m^2 c^2}{M}, \quad \sum_l E_l^a = \frac{\pi}{8} m c^2, \quad (34)$$

while most of the energy is emitted below the frequency:

$$f_m = 0.08 \frac{c^3}{GM}. \quad (35)$$

Up to 94% of the energy is radiated between $8M$ and $2M$ and 90% of it in the quadrupole mode.

Unfortunately, the analysis in the frequency domain does not contribute much to the understanding of the particle motion, the limitation having origin in the absence of exact solutions. A Fourier anti-transform of an approximate solution, for instance valid at high frequencies, does not reveal which effect on the motion has the neglect of lower frequencies. Thus, the lack of availability of any time domain solution has impeded progress in the comprehension of motion in the perturbative two-body problem. Although studies on analytic solutions were attempted throughout the years, e.g. Fackerell [78], Zhdanov [221], Leaver [117–119], Mano, Suzuki and Takasugi [137] and Fiziev [82], they were limited to the homogeneous equation.

6 Numerical Solution

The breakthrough arrived thanks to a specifically tailored finite differences method. It consists of the numerical integration of the inhomogeneous wave equation in time domain, proposed by Lousto and Price [134, 135] and based on the mathematical formalism of the particle limit approximation developed in [133] in the Eddington–Finkelstein coordinates [71, 81]. A parametric analysis of the initial data by Martel and Poisson has later appeared [139]. Confirmation of the results, among which the waveforms at infinity, is contained in [5].

The grid cells are separated in two categories, according to whether the cell is crossed or not by the particle. The latter category, Fig. 1, is then formed by the cells for which $r \neq z_u(t)$ and the $\Psi(t, r)$ evolution is not affected by the source.

²⁰ The divergence in summing over all l modes is said to be taken away by considering a finite size particle [49].

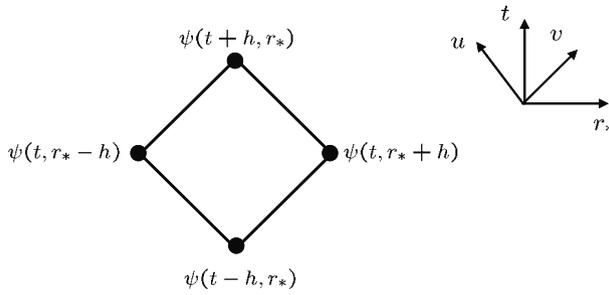


Fig. 1 An empty cell never crossed by the particle world-line

It is then sufficient to integrate each term of the homogeneous wave equation. The wave operator allows an exact integration (passing to the r^* tortoise coordinate):

$$\iint_{Cell} (\partial_{r^*}^2 - \partial_t^2) \Psi dA = -4[\Psi(t+h, r^*) + \Psi(t-h, r^*) - \Psi(t, r^*+h) - \Psi(t, r^*-h)]. \tag{36}$$

Instead, the product potential-wavefunction is given by:

$$\iint_{Cell} V(r)\Psi dA = V(r)h^2[\Psi(t+h, r^*) + \Psi(t-h, r^*) + \Psi(t, r^*+h) + \Psi(t, r^*-h) + \mathcal{O}(h^3)]. \tag{37}$$

The evolution algorithm defines Ψ at the upper cell corner as computed out of the three preceding values:

$$\Psi(t+h, r^*) = -\Psi(t-h, r^*) + [\Psi(t, r^*+h) + \Psi(t, r^*-h)] \left[1 - \frac{h^2}{2} V(r) \right]. \tag{38}$$

For the cells crossed by the particle, a different integration scheme is imposed (Fig. 2). The product potential-wavefunction is given by:

$$\iint_{Cell} V(r)\Psi dA = V(r) \times [A_3\Psi(t+h, r^*) + A_2\Psi(t-h, r^*) + A_4\Psi(t, r^*+h) + A_1\Psi(t, r^*-h) + \mathcal{O}(h^3)], \tag{39}$$

where A_1, A_2, A_3, A_4 are the sub-surfaces of the cell. The integration of the source term is given by²¹:

²¹ There are editorial errors in the corresponding expressions (3.6) in [134] and (3.4) in [139].

$$\begin{aligned}
 \iint_{Cell} SdA = & - \int_{t_i}^{t_o} \frac{2\kappa(r-2M)}{E(2\lambda+1)(\lambda z_{u0} + 3M)^2} \times \left[\frac{6M}{z_u} (1-E^2) + \lambda(\lambda+1) \right. \\
 & \left. - \frac{3M^2}{z_u^2} + \frac{4\lambda M}{z_u} \right] dt + \frac{2\kappa(r-2M)}{E(2\lambda+1)(\lambda z_u + 3M)} \times \left\{ Sign_i \left[1 + \frac{Sign_i}{E} \right. \right. \\
 & \left. \left. \times \sqrt{\frac{2M}{z_u} - \frac{2M}{z_{u0}}} \right]^{-1} + Sign_o \left[1 - \frac{Sign_o}{E} \sqrt{\frac{2M}{z_u} - \frac{2M}{z_{u0}}} \right]^{-1} \right\},
 \end{aligned}
 \tag{40}$$

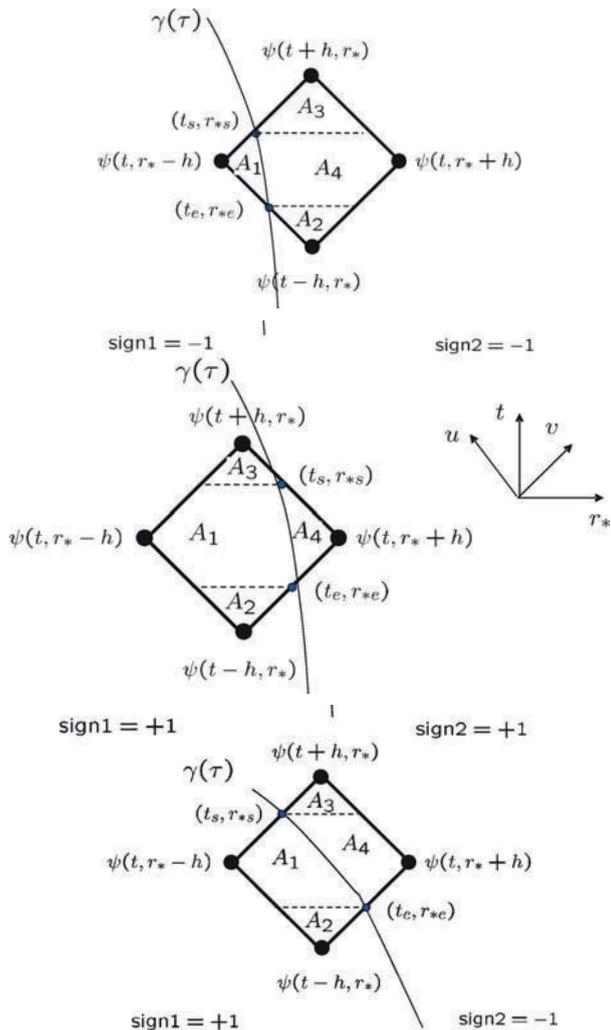


Fig. 2 There are only three physical cases for the particle crossing the (t, r^*) cell

where t_i corresponds to the time of entry of the particle in the cell and t_o the time of departure from the cell; z_{i0} is the initial position of the particle; $E = \sqrt{1 - 2M/z_{i0}}$; $Sign_i = +1$ if the particle enters the cell on the right, -1 if on the left; $Sign_o = +1$ if the particle leaves the cell on the right, -1 if on the left (Fig. 2). Through the evolution algorithm, the value of Ψ at the upper cell corner is given by²²:

$$\begin{aligned} \Psi(t+h, r^*) &= -\Psi(t-h, r^*) \left[1 + \frac{V(r)}{4}(A_2 - A_3) \right] \\ &+ \Psi(t, r^*+h) \left[1 - \frac{V(r)}{4}(A_4 + A_3) \right] \\ &+ \Psi(t, r^*-h) \left[1 - \frac{V(r)}{4}(A_1 + A_3) \right] \\ &- \frac{1}{4} \left(1 - \frac{V(r)}{4}A_3 \right) \iint_{Cell} S(t, r) dA. \end{aligned} \quad (41)$$

For the value of Ψ at $t = h$, the unavailability of Ψ at $t = -h$ is circumvented by using a Taylor expansion of $\Psi(r^*, -h)$ for the initial conditions $t = 0$:

$$\Psi(r^*, -h) = \Psi(r^*, 0) - h \frac{\partial \psi}{\partial t} \Big|_{t=0}. \quad (42)$$

The setting of initial conditions constitutes a delicate, technical and largely debated issue. Apart from the technical difficulty in the numerical implementation, it suffices to state how much it is crucial to match the initial radiation conditions, that represent the earlier history of the particle, with its position and velocity. For those z_{i0} starting points that are sufficiently far from the horizon, the errors on the initial conditions are fortunately not relevant at later times.

Another numerical issue is the evaluation of the wavefunction and the perturbations at the position of the particle, but unfortunately not described in the literature and too technical for this book. The wavefunction belongs to the C^{-1} continuity class²³ and the values before and after the particle position are computed and compared to the jump conditions posed on the wavefunction and its derivatives [172]. Further, it is necessary to obtain the third derivatives of the wavefunction to determine the correction to the geodesic background motion of the particle. Given the second order convergence of the above described algorithm, this is not easily achieved without recurring to a fourth order scheme [131].

²² There are editorial errors in the corresponding expressions (3.9) in [134] and (3.5) in [139].

²³ The Heaviside or step distribution, like the wavefunction of the Zerilli equation, belongs to the C^{-1} continuity class; the Dirac delta distribution and its derivative belong to the C^{-2} and C^{-3} continuity class, respectively.

7 Relativistic Radial Fall Affected by the Falling Mass

7.1 The Self-Force

It has been addressed in the previous section that the perturbative two-body problem involving a black hole and a particle with radiation emission has been tackled 40 years ago. For computation of radiation reaction, it may be worth recalling that before 1997, only pN methods existed in the weak field regime. Indeed, it is only slightly more than a decade that we possess methods [150, 165] for the evaluation of the self-force²⁴ in strong field for point particles thanks to concurring situations. On one hand, theorists progressed in understanding radiation reaction and obtained formal prescriptions for its determination, and on the other hand, the appearance of requirements from the LISA (Laser Interferometer Space Antenna) project [124] for the detection of captures of stars by supermassive black holes (EMRI, Extreme Mass Ratio Inspiral), notoriously affected by radiation reaction.

Such factors, theoretical progress and experiment requirements, have pushed the researchers to turn their efforts in finding an efficient and clear implementation of the theorists prescriptions²⁵ by tackling the problem in the context of perturbation theory, for which the small mass m corrects the geodesic equation of motion on a fixed background via a factor $\mathcal{O}(m)$ (for a review, see Poisson [58] and Barack [15]).

Before the appearance of the self-force equation and of the regularisation methods, the main theoretical unsolved problem was represented by the infinities of the perturbations at the particle's position. After determination of the perturbations through Eqs. 26 and 31–33, the trajectory of the particle could be corrected simply by requiring it to be a geodesic of the total (background plus perturbations) metric (the Christoffel connection $\bar{\Gamma}_{\alpha\beta}^{\mu}$ refers to the full metric):

$$\frac{d^2 x^\alpha}{d\tau^2} + \bar{\Gamma}_{\beta\gamma}^{\alpha} \frac{dx^\beta}{d\tau} \frac{dx^\gamma}{d\tau} = 0, \quad (43)$$

²⁴ A point-like mass m moves along a geodesic of the background spacetime if $m \rightarrow 0$; if not, the motion is no longer geodesic. It is sometimes stated that the interaction of the particle with its own gravitational field gives rise to the self-force. It should be added, though, that such interaction is due to an external factor like a background curved spacetime or a force imposing an acceleration on the mass. In other words, a single and unique mass in an otherwise empty universe cannot experience any self-force. Conceptually, the self-force is thus a manifestation of non-locality in the sense of Mach's inertia [136].

²⁵ It is currently believed that the core of most galaxies host supermassive black holes on which stars and compact objects in the neighbourhood inspiral-down and plunge-in. Gravitational waves might also be detected when radiated by the Milky Way Sgr*A, the central black hole of more than 3 million solar masses [30, 83]. The EMRIs are further characterised by a huge number of parameters that, when spanned over a large period, produce a yet unmanageable number of templates. Thus, in alternative to matched filtering, other methods based on covariance or on time and frequency analysis are investigated. If the signal from a capture is not individually detectable, it still may contribute to the statistical background [17].

but the perturbation behaves as:

$$h_{\alpha\beta} \sim \frac{1}{\sqrt{(g_{\gamma\delta} + u_\gamma u_\delta) [x^\gamma - z_u^\gamma(\tau)] [x^\delta - z_u^\delta(\tau)]}}, \quad (44)$$

thus diverging as the inverse of the distance to the particle and imposing a singular behaviour to $\bar{\Gamma}_{\alpha\beta}^\mu$ on the trajectory of the particle. Thus, the small perturbations assumption breaks down near the particle, exactly where the radiation reaction should be computed.

The solution was brought by the self-force equation, formulated in 1997 and baptised MiSaTaQuWa,²⁶ from the surname first two initials of its discoverers, who determined it using various approaches, all yielding the same formal expression. In the MiSaTaQuWa prescription, the self-force is only well defined in the harmonic (de Donder) [50, 51] gauge (stemmed from the Lorenz gauge [127]) and any departure from it – its relaxation – undermines the validity of the equation of motion.

Mino, Sasaki and Tanaka [150] used two methods, namely the conservation of the total stress-energy tensor and the matched asymptotic expansion. The former generalises the analysis of DeWitt and Brehme [60] and Hobbs [105], consisting in the calculation of the electromagnetic self-force in curved spacetime previously performed in flat space by Dirac [61]. It evaluates the perturbation near the world-line using the Hadamard expansion [97] of the retarded Green function [93–95]. Then, it deduces the equation of motion by imposing the conservation of the rank-two symmetric total stress-energy tensor, via integration of its divergence over the interior of a thin world-tube around the particle's world-line.

The latter, reformulated by Poisson [159], in a buffer zone matches asymptotically the expansion of the black hole perturbed background by the particle with the expansion around the particle distorted by the black hole.

Also in 1997, the axiomatic approach by Quinn and Wald [165] was presented. To them, the self-force is identified by comparison of the perturbation in curved spacetime with the perturbation in flat spacetime. The procedure allows elimination of the divergent part and extraction of the finite part of the force.

On the footsteps of Dirac's definition of radiation reaction, in 2003 Detweiler and Whiting [59], see also Poisson [159], offered a novel approach. In flat spacetime, the radiative Green function is obtained by subtracting the singular contribution, half-advanced plus half-retarded, from the retarded Green function. In curved spacetime, and in the gravitational case, the attainment of the radiative Green function passes through the inclusion of an additional, purposely built, function. The singular part does not exert any force on the particle, upon which only the regular field acts [55]. The latter, solely responsible of the self-force, satisfies the homogeneous wave equation and may be considered a radiative field in interaction with the particle.

²⁶ In 2002 at the Capra Penn State meeting by Eric Poisson.

This approach emphasises that the motion is a geodesic of the full metric and it implies two notable features: the regularity of the radiative field and the avoidance of any non-causal behaviour.²⁷

Gralla and Wald have attempted a more rigorous way of deriving a gravitational self-force [92]. Their final prescription, namely self-consistency versus the first order perturbative correction to the geodesic of the background spacetime, shall be addressed later in this chapter. On the same track of improving rigour, an alternative approach and a new derivation of the self-force have been proposed by Gal'tsov and coworkers [91] and by Pound [161], respectively.

The determination of the self-force has allowed not only targeted applications geared to more and more complex astrophysical scenarios, but also fundamental investigations: on the role of passive, active and inertial mass by Burko [29]; the already quoted papers on the Newtonian self-force [58], on the EP [59, 209], on the relation to energy conservation by Quinn and Wald [166]; on the relation between self-force and radiation reaction examined through gauge dependence and adiabaticity by Mino [145–148]; the differentiation between adiabatic, secular and radiative approximations as well as the relevance of the conservative effects by Pound and Poisson [162]; on the relation between h^{tail} and h^R , tail and regular parts of the field by Detweiler [56].

The following wishes to be constrained to a physical and sketchy picture of the self-force. For this purpose, the original MiSaTaQuWa approach – the force acts on the background geodesic – is more intuitive. One pictorial description refers to a particle that crosses the curved spacetime and thus generates gravitational waves. These waves are partly radiated to infinity (the instantaneous part) and partly scattered back by the black hole potential (the non-local part), thus forming tails which impinge on the particle and give origin to the self-force. Alternatively, the same phenomenon is described by an interaction particle-black hole generating a field which behaves as outgoing radiation in the wave-zone and thereby extracts energy from the particle. In the near-zone, the field acts on the particle and determines the self-force which impedes the particle to move on the geodesic of the background metric. The total force is thus written as:

$$F_{full}^{\alpha}(x) = F_{inst.}^{\alpha} + F_{tail}^{\mu}, \quad (45)$$

where $F_{inst.}^{\alpha}$ is computed from the contributions that propagate along the past light cone and F_{tail}^{α} has the contributions from inside the past light cone, product of the scattering of perturbations due to the motion of the particle in the curved spacetime created by the black hole,²⁸ (Fig. 3) (pictorially, in a curved spacetime, the radiation is not solely confined to the wave front). The self-force is then computed by taking

²⁷ Given the elegance of this classic approach, the self-force expression should be rebaptised as MiSaTaQuWa-DeWh.

²⁸ Detweiler and Whiting [59] refer to the contribution inside the light cone via the Hadamard expression [97] of the Green function.

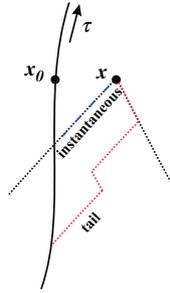


Fig. 3 The radiation is going to infinity (instantaneous) or is scattered back (tail). The latter part determines the self-force. The self-force is defined for x to $\rightarrow z_u$ where z_u is the position of the particle on the world-line τ , while x is the evaluation point

the limit $F_{self}^\mu = F_{tail}^\mu [x \rightarrow z_u(t)]$. Thus, it is conceived as force acting on the background geodesic [150, 165], wherein $\Gamma_{\beta\gamma}^\alpha$ refers to the background metric:

$$F_{self}^\alpha = m \frac{Du^\alpha}{d\tau} = \frac{d^2x^\alpha}{d\tau^2} + \Gamma_{\beta\gamma}^\alpha u^\beta u^\gamma. \tag{46}$$

All MiSaTaQuWa-DeWh approaches produce the same equation for the self-acceleration, given by:

$$a_{self}^\alpha = -(g^{\alpha\beta} + u^\alpha u^\beta) \left(\nabla_\delta h_{\beta\gamma}^* - \frac{1}{2} \nabla_\beta h_{\gamma\delta}^* \right) u^\gamma u^\delta, \tag{47}$$

where the star indicates the tail (MiSaTaQuWa) or radiative (DeWh) component. Equation 47 is not gauge invariant and depends upon the de Donder gauge condition:

$$\tilde{h}_{\nu}^{\mu\nu*} = 0, \tag{48}$$

where $\tilde{h}_{\gamma\delta}^* = h_{\gamma\delta}^* - \frac{1}{2} g_{\gamma\delta} h^*$ and $h^* = g^{\mu\nu} h_{\mu\nu}^*$. Barack and Ori have shown [21] that under a coordinate transformation of the form $x^\alpha \rightarrow x^\alpha - \xi^\alpha$, under which the perturbation transforms according to:

$$h_{\mu\nu} \rightarrow h_{\mu\nu} + \xi_{\mu;\nu} + \xi_{\nu;\mu}, \tag{49}$$

the self-force acceleration transforms as:

$$a_{self}^\alpha \rightarrow a_{self}^\alpha - \left[(g^{\alpha\lambda} + u^\alpha u^\lambda) \frac{d^2\xi_\lambda}{d\tau^2} + R^\alpha_{\mu\lambda\nu} u^\mu \xi^\lambda u^\nu \right], \tag{50}$$

where the terms are evaluated at the particle and $R^\alpha_{\mu\lambda\nu}$ is the Riemann tensor of the background geometry. Thus, for a given two-body system, the MiSaTaQuWa-DeWh acceleration is to be mentioned together with the chosen gauge.²⁹

²⁹ The self-force being affected by the gauge choice, the EP allows to find a gauge where the self-force disappears. Again, as in Newtonian physics, such gauge will be dependent of the mass m , impeding the uniqueness of acceleration.

The identification of the tail and instantaneous parts was not accompanied by a prescription of the cancellation of divergencies, which indeed arrived 3 years later thanks to the mode-sum method by Barack and coworkers [13, 14, 16, 20]. The mode-sum method relies on solutions to interwoven difficulties, mostly related to the divergent nature of the problem, but tentatively presented as separate hereafter.

Spherical symmetry allows the force to be expanded into spherical harmonics and turns out to be once more the key factor for black hole physics, after having been the expedient for the determination of the wave equation. The divergent nature of the problem is then transformed into a summation problem. For each multipole, the full force is finite and the divergence appears only upon infinite summing over l, m .

Furthermore, the tail component cannot be calculated directly, but solely as difference between the full force and the instantaneous part; thus, the self-force is computed as:

$$F_{self}^{\alpha} = \lim_{x \rightarrow z_u} \sum_l \left[F_{full}^{\alpha l \pm}(x) - F_{inst.}^{\alpha l \pm}(x) \right]. \quad (51)$$

Each of the two quantities $F_{full}^{\alpha l}(x)$ and $F_{inst.}^{\alpha l}$ is discontinuous through the particle location and the superscript \pm indicates the two (different) values obtained by taking the particle limit from outside ($x \rightarrow z_u^+$) and inside ($x \rightarrow z_u^-$). However, the difference in Eq. 51 does not depend upon the direction from which the limit is taken.

The full and the instantaneous parts have the same singular behaviour at large l and close to the particle; their difference should be sufficient to ensure a regular behaviour at each l . Unfortunately, another obstacle arises from the difficulty of calculating the instantaneous part mode by mode. Therefore, the divergence is dealt with by seeking a function $H^{\alpha l}$, such that the series $\sum_l \left[F_{full}^{\alpha l \pm} - H^{\alpha l \pm} \right]$ is convergent.

The function $H^{\alpha l}$ mimics the instantaneous component at large l and close to the particle. Once such condition is ensured, Eq. 51 is rewritten as:

$$F_{self}^{\alpha} = \sum_l \left[F_{full}^{\alpha l \pm} - H^{\alpha l \pm} \right] - D^{\alpha l \pm}, \quad (52)$$

where

$$D^{\alpha l \pm}(x) = \lim_{x \rightarrow z_u} \sum_l \left[F_{inst.}^{\alpha l \pm}(x) - H^{\alpha l \pm}(x) \right]. \quad (53)$$

The addition and subtraction of the function $H^{\alpha l}$ guarantees the pristine value of the computation. In general, for $L = l + 1/2$:

$$H^{\alpha l} = A^{\alpha} L + B^{\alpha} + C^{\alpha}/L. \quad (54)$$

Thus, the mode-sum amounts to [14]: (i) numerical computation of full modes; (ii) derivation of the regularisation parameters A, B, C , and D (obtained on a local analysis of the Green's function near coincidence, $x \rightarrow z_u$, at large l); (iii) computation of Eq. 52 whose behaviour has to show a $1/L^2$ fall off if previous steps are correctly carried out.

7.2 The Pragmatic Approach

The straightforward pragmatic approach by Lousto, Spallicci and Aoudia [129, 130, 189, 190] is the direct implementation of the geodesic in the full metric (background + perturbations) and it is coupled to the renormalisation by the Riemann–Hurwitz ζ function. These two features justify the pragmatic adjective. Though the application of the ζ function is somewhat artificial and the pragmatic method is somewhat naive, the latter has the merit of: (i) a clear identification of the different factors participating in the motion; (ii) potential applicability to any gauge and to higher orders of the ζ function renormalisation.

Dealing only with time and radial components, two geodesic equations can be written and then combined into a single one, after elimination of the geodesic parameter. Thus, for radial fall the coordinate acceleration is given by the sole radial component:

$$\ddot{z}_p = \bar{\Gamma}_{rr}^t \dot{z}_p^3 + 2\bar{\Gamma}_{tr}^t \dot{z}_p^2 - \bar{\Gamma}_{rr}^r \dot{z}_p^2 + \bar{\Gamma}_{tt}^t \dot{z}_p - 2\bar{\Gamma}_{tr}^r \dot{z}_p - \bar{\Gamma}^t t^r, \quad (55)$$

where $\bar{\Gamma}_{\beta\gamma}^\alpha$ refers to the full metric and z_p is given by Eq. 56. Equation 55 refers to:

- The full metric field $\bar{g}_{\mu\nu}(t, r)$ previously defined
- The displacement Δz , difference between the perturbed $z_p(t)$ and the unperturbed $z_u(t)$ positions, and the coordinate time derivatives:

$$z_p = z_u + \Delta z, \quad \dot{z}_p = \dot{z}_u + \Delta \dot{z}, \quad \ddot{z}_p = \ddot{z}_u + \Delta \ddot{z}; \quad (56)$$

- The Taylor expansion of the field and its spatial derivative:

$$\begin{aligned} \bar{g}_{\mu\nu} |_{z_p} &= \bar{g}_{\mu\nu} |_{z_u(t)} + \Delta z \bar{g}_{\mu\nu,r} |_{r=z_u(t)}, \\ \bar{g}_{\mu\nu,r} |_{z_p} &= \bar{g}_{\mu\nu,r} |_{z_u(t)} + \Delta z \bar{g}_{\mu\nu,rr} |_{r=z_u(t)}. \end{aligned} \quad (57)$$

The unperturbed trajectory of the particle $z_u(t)$ is given by the inverse of the relations $T(r)$, e.g. Eq. 28. Supposing that the relative strengths of the perturbations and the deviations behave as:

$$\frac{[h^{(1)}]^2}{g} \simeq \frac{h^{(2)}}{g} \ll \frac{h^{(1)}}{g} \simeq \frac{\Delta \dot{z}}{\dot{z}_p} \simeq \frac{\Delta z}{z_p}. \quad (58)$$

Then, the coordinate acceleration correction is given by an expansion up to first order for all quantities, which corresponds to the expression in [129, 130]³⁰:

$$\Delta \ddot{z} = \alpha_1 (g, \dot{z}_u) \Delta z + \alpha_2 (g, \dot{z}_u) \Delta \dot{z} + \alpha_6 (h, \dot{z}_u). \quad (59)$$

³⁰ Apart from some editorial errors therein, $\alpha_{1,2,6}$ correspond to the A, B, C coefficients in [129, 130], which are not to be confused with the A, B, C coefficients of the mode-sum!

Table 1 Representation of the terms of Eq. 59 in gauge-independent form. The elements in each column are produced by the terms of Eq. 55, in the first row. From the algebraic sum of the elements between horizontal lines, the terms of Eq. 59 are derived

	$\bar{\Gamma}_{rr}^t \dot{z}_p^3$	$2\bar{\Gamma}_{tr}^t \dot{z}_p^2$	$-\bar{\Gamma}_{rr}^r \dot{z}_p^2$	$\bar{\Gamma}_{tt}^t \dot{z}_p$	$-2\bar{\Gamma}_{tr}^r \dot{z}_p$	$-\bar{\Gamma}_{tt}^r$
$\alpha_1 \Delta z$		$g_{,r}^{tt} g_{tt,r} \dot{z}_u^2 \Delta z$	$-\frac{1}{2} g_{,r}^{rr} g_{rr,r} \dot{z}_u^2 \Delta z$			$\frac{1}{2} g_{,r}^{rr} g_{tt,r} \Delta z$
		$g^{tt} g_{tt,rr} \dot{z}_u^2 \Delta z$	$-\frac{1}{2} g^{rr} g_{rr,rr} \dot{z}_u^2 \Delta z$			$\frac{1}{2} g^{rr} g_{tt,rr} \Delta z$
$\alpha_2 \Delta \dot{z}$		$2g^{tt} g_{tt,r} \dot{z}_u \Delta \dot{z}$	$-g^{rr} g_{rr,r} \dot{z}_u \Delta \dot{z}$			
α_6	$g^{tt} h_{tr,r} \dot{z}_u^3$	$-h^{tt} g_{tt,r} \dot{z}_u^2$	$\frac{1}{2} h^{rr} g_{rr,r} \dot{z}_u^2$	$\frac{1}{2} g^{tt} h_{tt,t} \dot{z}_u$	$-g^{rr} h_{rr,t} \dot{z}_u$	$-g^{rr} h_{tr,t}$
	$-\frac{1}{2} g^{tt} h_{rr,t} \dot{z}_u^3$	$g^{tt} h_{tt,r} \dot{z}_u^2$	$-\frac{1}{2} g^{rr} h_{rr,r} \dot{z}_u^2$	$\frac{1}{2} h^{tr} g_{tt,r} \dot{z}_u$	$h^{tr} g_{tt,r} \dot{z}_u$	$-\frac{1}{2} h^{rr} g_{tt,r}$
	$-\frac{1}{2} h^{tr} g_{rr,r} \dot{z}_u^3$					$\frac{1}{2} g^{rr} h_{tt,r}$

The particle determines in first instance the emission of radiation $h_{\alpha\beta}$, which after backscattering by the black hole potential, interacts with the particle itself resulting into a change in acceleration (the coefficient α_6 depending on $h_{\mu\nu}$ and derivatives). The latter places the particle elsewhere from where it should have been, that is $z_u(t)$. The field is thus to be evaluated at this new position resulting into a further variation in acceleration (the terms $\alpha_1 \Delta z$ and $\alpha_2 \Delta \dot{z}$ depending on $g_{\mu\nu}$ and derivatives).

All terms in Eqs. 59 and 62 are of $1/M$ order; the terms $\alpha_1 \Delta z$ and $\alpha_2 \Delta \dot{z}$ represent the background field evaluated on the perturbed trajectory; α_6 represents the perturbed field on the background trajectory. The expressions in Table 1 are gauge independent, while in Table 2 they are shown in the Regge–Wheeler gauge ($H_0 = H_2$ and $K = 0$ as in head-on geodesics). Finally, the coefficient α_0 is the lowest order term corresponding to a particle radially falling into the SD black hole and not affected by the perturbations. It corresponds to the unrenormalised acceleration and it is to be added to the terms of Eqs. 59 and 62 to compute the total acceleration:

$$\alpha_0 = g^{tt} g_{tt,r} \dot{z}_u^2 - \frac{1}{2} g^{rr} g_{rr,r} \dot{z}_u^2 + \frac{1}{2} g^{rr} g_{tt,r} = -\frac{M}{r^2} \left[1 - \frac{2M}{r} - 3 \left(1 - \frac{2M}{r} \right)^{-1} \dot{z}_u^2 \right]. \quad (60)$$

If $\Delta \ddot{z}$ receives its main contribution from the background metric $g_{\mu\nu}$ or else cumulative effects are let to grow, a different expansion may be considered.³¹

³¹ Supposing that the relative strengths of the perturbations and the deviations behave as:

$$\frac{[h^{(1)}]^2}{g} \simeq \frac{h^{(2)}}{g} \ll \frac{h^{(1)}}{g} < \frac{\Delta \dot{z}}{\dot{z}_p} \simeq \frac{\Delta z}{z_p}, \quad (61)$$

then, the coordinate acceleration correction would be given by an expansion up to first order in perturbations and second order in deviation [190]:

$$\Delta \ddot{z} = \alpha_1 (g, \dot{z}_u) \Delta z + \alpha_2 (g, \dot{z}_u) \Delta \dot{z} + \alpha_3 (g, \dot{z}_u) \Delta z^2 + \alpha_4 (g) \Delta \dot{z}^2 + \alpha_5 (g, \dot{z}_u) \Delta z \Delta \dot{z} + \alpha_6 (h, \dot{z}_u) + \alpha_7 (h, \dot{z}_u) \Delta z + \alpha_8 (h, \dot{z}_u) \Delta \dot{z}. \quad (62)$$

Table 2 Representation of the terms of Eq. 59 in Regge–Wheeler gauge

$\alpha_1 \Delta z$	$-\frac{M}{r^2} \left[\frac{6M}{r^2} - \frac{2}{r} + \frac{6(r-M)}{r^2} \left(1 - \frac{2M}{r} \right)^{-2} \dot{z}_u^2 \right] \Delta z$
$\alpha_2 \Delta \dot{z}$	$\frac{6M}{r^2} \left(1 - \frac{2M}{r} \right)^{-1} \dot{z}_u \Delta \dot{z}$
α_6	$\frac{1}{r-2M} \left[\frac{r^2 H_{0,t}}{2(r-2M)} - \frac{MH_1}{r-2M} - rH_{1,r} \right] \dot{z}_u^3 - \frac{3}{2} H_{0,r} \dot{z}_u^2 - 3 \left(\frac{H_{0,t}}{2} - \frac{MH_1}{r^2} \right) \dot{z}_u$ $+ \frac{r-2M}{r} \left[\frac{2MH_0}{r^2} + \frac{(r-2M)H_{0,r}}{2r} - H_{1,t} \right]$

In radial fall, it has been indicated by two different heuristic arguments [129, 132] that the metric perturbations should be of C^0 continuity class at the location of the particle.³² One argument [129] is based on the integration over r of the Hamiltonian constraint, which is the tt component of the Einstein equations (Eq. C7a in [220]); the other [132] on the structure of selected even perturbation equations. In [6], a stringent analysis on the C^0 continuity is pursued in terms of the jump conditions that the wavefunctions and derivatives have to satisfy for guaranteeing the continuity of the perturbations.³³ Anyhow, the connection coefficients and the metric perturbation derivatives have a finite jump and they can be computed as the average of their values at $z_u \pm \varepsilon$ with $\varepsilon \rightarrow 0$.

In Eq. 62: (i) solely second order terms in perturbations are not considered; (ii) the terms $\alpha_2(g, \dot{z}_u) \Delta \dot{z}$, $\alpha_3(g, \dot{z}_u) \Delta z^2$, $\alpha_4(g) \Delta \dot{z}^2$, $\alpha_5(g, \dot{z}_u) \Delta z \Delta \dot{z}$ represent the background field evaluated on the perturbed trajectory at second order in deviation; (iii) α_{3-5} tend to infinity close to the horizon, conversely to the $\alpha_{1,2}$ coefficients; (iv) $\alpha_7(h, \dot{z}_u) \Delta z$, $\alpha_8(h, \dot{z}_u) \Delta \dot{z}$ represent the perturbed field on the perturbed trajectory, and the α_{7-8} coefficients are larger near the horizon. These last two coefficients may be regularised in l by the Riemann–Hurwitz ζ function as shown in [190].

³² The jump conditions were also dealt with by Sopuerta and Laguna [188].

³³ Having suppressed the l index for clarity of notation, after visual inspection of Eq. 26, containing a derivative of the Dirac delta distribution, it is evinced that the wavefunction Ψ is of C^{-1} continuity class and thus can be written as:

$$\Psi(t, r) = \Psi^+(t, r) \Theta_1 + \Psi^-(t, r) \Theta_2, \tag{63}$$

where $\Theta_1 = \Theta[r - z_u(t)]$ and $\Theta_2 = \Theta[z_u(t) - r]$ are two Heaviside step distributions. Computing the first and second, space and time and mixed derivatives, Dirac delta distributions and derivatives are obtained of the type $\delta[r - z_u(t)]$ and $\delta'[r - z_u(t)]$, respectively. It is wished that the discontinuities of Ψ and its derivatives are such that they are canceled when combined in K , H_2 , and H_1 . After replacing Ψ and its derivatives in Eqs. 31–33, continuity requires that the coefficients of Θ_1 must be equal to the coefficients of Θ_2 , while the coefficients of δ and δ' must vanish separately. After some tedious computing and making use of one of the Dirac delta distribution properties, $f(r)\delta'[r - z_u(t)] = f[z_u(t)]\delta'[r - z_u(t)] - f'(z_u(t))\delta[r - z_u(t)]$, at the position of the particle, the jump conditions for Ψ and its derivatives are found. Furthermore, the jump conditions allow a new method of integration, as shown by Aoudia and Spallicci [6].

The supposed C^0 continuity class of the metric perturbations allows to deal with the divergence with l of the α_6 coefficient [129, 130]. The divergence originates from the infinite sum over the finite multipole component contributions. One way of regularising this sum is to subtract to each mode precisely the $l \rightarrow \infty$ contribution, since for ever larger l the metric perturbations tend to some finite asymptotic behaviour. Thus, the subtraction from each mode of the $l \rightarrow \infty$ part leads to a convergent series. The renormalisation by the Riemann–Hurwitz ζ function was proposed first in [129, 130] and then extended to higher orders in [190]. For $L = l + 1/2$, it can be shown that:

$$\alpha_6 = \sum_{l=0}^{\infty} \alpha_6^l, \quad \alpha_6^l = \alpha_{6\pm}^a L + \alpha_6^b L^0 + \alpha_{6\pm}^c L^{-1} + \alpha_6^d L^{-2} + \mathcal{O}(L^{-3}). \quad (64)$$

Equation 64 is cast to have a similar form to the mode-sum expression. The average of $\alpha_{6\pm}^a$ and $\alpha_{6\pm}^c$ vanish at the position of the particle, whereas $\sum_{l=0}^{\infty} \alpha_6^b = \infty$ determines the divergence.

The Riemann ζ function [169] and its generalisation, the Hurwitz ζ function [106], are defined by:

$$\zeta(s) = \sum_{l=1}^{\infty} (l)^{-s}, \quad \zeta(s, a) = \sum_{l=0}^{\infty} (l + a)^{-s}, \quad (65)$$

where in our case $a = 1/2$. Two special values of the Hurwitz function, namely $\zeta(0, 1/2) = 0$ and $\zeta(2, 1/2) = 1/2 \pi^2$, cancel the divergent term and determine that the term $\sum_{l=0}^{\infty} \alpha_6^d L^{-2}$ gets a finite value, respectively. Barack and Lousto [18] have shown the concordance of the mode-sum and the ζ regularisations for radial fall.

8 The State of the Art

It is now time to discuss the state of the art of the radial fall affected by its mass and the emitted radiation. As shown in the introduction, the adiabatic approximation requires that a given orbital parameter q changes slowly over time scales comparable to the orbital period P (this is somewhat a coarse definition since the small mass always “reacts” immediately): $\Delta q = \dot{q}P \ll q$. For circular and moderately elliptic orbits, the above condition, where q is function of the semi-latus rectum p and eccentricity e , is transformed into a condition on the m/M ratio [43]. In radial fall, though, it is far from being evident, and even possible, to identify a condition on adiabaticity within which any simplification may occur. The febleness of cumulative effects for radiation reaction does not imply their non-existence. On the contrary, this is the case where most care and sophisticated techniques are demanded for the computation of the motion affected by the back-action, even if the latter has moderate effects. Therefore it is not surprising, thanks to the febleness and to the difficulties, that solely two studies (one based on the pragmatic method, the other on the self-force) exist.

8.1 Trajectory

The perennial question on the behaviour of the infalling mass reflects itself in the determination along which direction the back-action is exerted.

Lousto (Fig. 2a in [129]) suggests that the α_6 term, denoted therein C (the variation of the coordinate acceleration of the particle due solely to perturbations; it corresponds to the self-force when referred to coordinate time, see next section), increases approaching the Zerilli potential at $3.1M$ and reaches its peak value around $2.4M$. The same reference (Fig. 2b in [129]) shows that the coordinate acceleration, thus including $\alpha_1 \Delta z$ and $\alpha_2 \Delta \dot{z}$ terms, is slowed down³⁴ and mostly until around $3.1M$. The two statements are not contradictory if the discrepancy is attributed to $\alpha_1 \Delta z$ and $\alpha_2 \Delta \dot{z}$. In the same reference [129], deceleration is expected as the system is losing energy and momentum. This repulsive behaviour – before the Zerilli potential peak – is confirmed in the abstract of [130].

Conversely, for Barack and Lousto [18] the radial component of the self-force is found to point inward (i.e. towards the black hole) throughout the entire plunge, independently on the starting point z_{i0} . The work done by the self-force is considered positive, resulting in an increase of the energy parameter E throughout the plunge. To these results, it is attached a specific gauge choice (as opposed to the energy flux at infinity, which is gauge invariant) [18].

The upward or inward direction impressed on the particle by its mass and the emitted radiation and whether this direction is maintained throughout the plunge or part of it is a fundamental, if not the main one, feature to acquire in such analysis. Again here, the statements from the three papers might not be contradictory as Lousto [129, 130] describes motion in coordinate time, and includes geodesic deviations. Conversely, Barack and Lousto [18] describe motion in proper time and apply only the self-force without geodesic deviations. Nevertheless, it is of interest to remark that the concept of repulsion resurfaces again solely in coordinate time as in the elder debate. The coefficient of the geodesic deviation coefficient α_1 changes sign during fall, while it does not occur to α_2 [7] which remains negative throughout.

For a particle starting from rest at a finite distance from the black hole, an analytic approximation of the self-force for the $l \geq 2$ modes (while for $l = 0, 1$ the solutions in [220] are mentioned) is given by Barack and Lousto[18]:

$$F_{self}^r = F_{self}^{r\ l=0} + F_{self}^{r\ l=1} + \sum_{l=2}^{\infty} \left\{ -\frac{15}{16} m^2 \frac{E^2}{r^2} \left(E^2 + \frac{4M}{r} - 1 \right) \frac{1}{L^2} + \mathcal{O}(L^{-4}) \right\}, \quad (66)$$

where E is the orbital energy of the particle. The force has only negative and even powers of l , which makes the sum quickly convergent and provides an excellent

³⁴ Lousto [129] comments only this former part and not the acceleration boost taking place after the Zerilli potential peak.

approximation for the numerical evaluation of the first few lower multipoles. The derivation of such net expression for the self-force is not described in [18].³⁵

8.2 Regularisation Parameters

Regularisation parameters of the mode-sum method have been confirmed independently by different papers [14, 22] and they are consistent with the results obtained by the application of the ζ function [18]. In radial fall, there is a regular gauge transformation between the de Donder and Regge–Wheeler gauges [21], and thus the regularisation parameters were also determined in the latter gauge [5, 18]. The results are:

$$A_{\pm}^r = \mp \frac{m^2}{z_u^2} E, \quad A_{\pm}^t = \mp \frac{m^2}{z_u^2} \frac{\dot{z}_u}{f}, \quad A_{\pm}^{\theta} = A_{\pm}^{\varphi} = 0, \quad (67)$$

$$B^r = -\frac{m^2}{2z_u^2} E^2, \quad B^t = -\frac{m^2}{2z_u^2} \frac{E \dot{z}_u}{f}, \quad B^{\theta} = B^{\varphi} = 0, \quad (68)$$

$$C^{\alpha} = D^{\alpha} = 0, \quad (69)$$

where $f \equiv 1 - 2M/z_u$ and $\dot{z}_u = -(E^2 - f)^{1/2}$.

8.3 Effect of Radiation Reaction on the Waveforms During Plunge

Thanks to a suggestion of B. Whiting, preliminary indications were found [5]. The waveforms shifts are of the order of (tens of) seconds for a particle sensibly radiating for few thousands of seconds, having started at rest from a finite distance between $4M$ and $40M$. The assumption used therein is energy–momentum balance (the energy radiated to infinity and absorbed by the black hole is imposed to be equal to the energy change in the particle fall). This assumption is likely jeopardised by the lack of instantaneous energy conservation [166].

The correct alternative is the computation of the self-force and its continuous implementation all along the trajectory. It is thus mandatory to consider the application of the recently proposed self-consistent prescription [92], unfortunately not yet part of the state of the art in terms of its application.

³⁵ In the Rapid Communication [18] there are seven citations of a yet unpublished material containing mathematical and numerical justifications of the results therein. The author acknowledges private communications by L. Barack.

9 Beyond the State of the Art: the Self-Consistent Prescription

In [92] a rigorous derivation and application of the self-force equation is proposed. In the derivation, emphasis is put upon the de Donder gauge and the consequences of relaxation, that is not enforcing this gauge. On one hand, the de Donder gauge is imposed by the nature of the self-force, solely defined in this gauge. On the other hand, the relaxation of the de Donder gauge stems from the need of departing from the background geodesic to find the self-force that causes such departure. Previous derivations were based on the assumption of deviations from geodesic motion expected to be small; by consequence, the de Donder gauge violation should likewise be small. Instead, the new derivation by Gralla and Wald is a rigorous, perturbative, result, obtained without the step of de Donder gauge relaxation and containing the geodesic deviation terms.

But Gralla and Wald go a step further [92]. For the evolution of an orbit, rather than a first order perturbation equation containing geodesic deviation terms, they recommend a self-consistent approach. Such prescription basically affirms the greater accuracy of a first order perturbation expansion along a continuously corrected trajectory as opposed to a higher order perturbation expansion made on the background geodesic.³⁶ Self-consistency bypasses the issue of relaxation, since at each integration step a new geodesic is found.³⁷

The “classic” first order perturbative expansion for the motion of a small body determines that the first order metric perturbations satisfy:

$$\nabla^\nu \nabla_\gamma \tilde{h}_{\alpha\beta} - 2R^\gamma{}_{\alpha\beta}{}^\delta \tilde{h}_{\gamma\delta} = -16\pi M u_\alpha u_\beta \frac{\delta^{(3)}(x^\mu)}{\sqrt{-g}} \frac{d\tau}{dt}, \tag{70}$$

where $x^\mu = 0$ corresponds to a geodesic γ of the background spacetime, and u^α is the tangent to γ . It is reminded that $\tilde{h}_{\alpha\beta} \equiv h_{\alpha\beta} - \frac{1}{2}h g_{\alpha\beta}$ with $h = h_{\alpha\beta} g^{\alpha\beta}$. For a retarded solution to this equation (thus satisfying the de Donder gauge condition) of the type:

$$h_{\alpha\beta}^{\text{tail}}(x) = M \int_{-\infty}^{\tau_{\text{ret}}} \left(G_{\alpha\beta\alpha'\beta'}^+ - \frac{1}{2}g_{\alpha\beta} G_{\gamma\alpha'\beta'}^{+\gamma} \right) [x, z_u(\tau')] u^{\alpha'} u^{\beta'} d\tau', \tag{71}$$

³⁶ The evolution of an orbit is lately getting the necessary concern. Pound and Poisson [162] apply osculating orbits to EMRI, but unfortunately their method is not applicable to plunge, for two reasons: the semi-latus rectum of the orbit, which decreases for radiation reaction, is smaller than a given quantity, considered as limit in their study case; the velocities and fields in the plunge are highly relativistic and their post-Newtonian expansion of the perturbing force becomes inaccurate. Non-applicability to plunge stands also for the work by Hinderer and Flanagan [104].

³⁷ Indeed, it affirms that it is preferable to apply successively a first order expansion at x_0 and then at x_1, x_2, \dots, x_m , rather than a second or higher order expansion at solely x_0 . It is evident, though, that self-consistency and perturbation order are decoupled concepts and that the former may be conceptually applicable to higher orders and more specifically, when, and if, a second order formalism will be available. In the same line of reasoning it would be preferable to apply successively a second order expansion at x_0 and then at x_1, x_2, \dots, x_m , rather than a third or higher order expansion at solely x_0 .

the first order in λ deviation of the motion from γ is expressed by (see [92] for the additional spin term):

$$u^\gamma \nabla_\gamma (u^\beta \nabla_\beta Z^\alpha) = -R_{\beta\gamma\delta}{}^\alpha u^\beta Z^\gamma u^\delta - (g^{\alpha\beta} + u^\alpha u^\beta) (\nabla_\delta h_{\beta\gamma}^{\text{tail}} - \frac{1}{2} \nabla_\beta h_{\gamma\delta}^{\text{tail}}) u^\gamma u^\delta. \quad (72)$$

Self-consistency prescribes that rather than using Eqs. 70–72, it is instead preferable to apply the self-force coherently all along the trajectory:

$$\nabla^\gamma \nabla_\gamma \tilde{h}_{\alpha\beta} - 2R^\gamma{}_{\alpha\beta}{}^\delta \tilde{h}_{\gamma\delta} = -16\pi m u_\alpha(t) u_\beta(t) \frac{\delta^{(3)}[x^\mu - z_p^\mu(t)]}{\sqrt{-g}} \frac{d\tau}{dt}, \quad (73)$$

$$u^\beta \nabla_\beta u^\alpha = -(g^{\alpha\beta} + u^\alpha u^\beta) (\nabla_\delta h_{\beta\gamma}^{\text{tail}} - \frac{1}{2} \nabla_\beta h_{\gamma\delta}^{\text{tail}}) u^\gamma u^\delta, \quad (74)$$

$$h_{\alpha\beta}^{\text{tail}}(x) = m \int_{-\infty}^{\tau_{\text{ret}}^-} \left(G_{\alpha\beta\alpha'\beta'}^+ - \frac{1}{2} g_{\alpha\beta} G_{\gamma\alpha'\beta'}^{+\gamma} \right) [x, z_p(\tau')] u^{\alpha'} u^{\beta'} d\tau', \quad (75)$$

where this time $u^\alpha(\tau)$ in Eqs. 74 and 75, normalised in the background metric, refers to the self-consistent motion $z_p(\tau)$, rather than to a background geodesic as in Eqs. 71 and 72; $G_{\alpha\beta\alpha'\beta'}^+$ is the retarded Green function, normalised with a factor of -16π [165]; the symbol τ_{ret}^- indicates the range of the integral being extended just short of the retarded time τ_{ret} , so that only the interior part of the light-cone is used.

The geodesic deviations vanish in Eq. 74, since self-consistency is imposed. Nevertheless, there might be situations where, for a whatever reason, the numerical implementation of the self-consistent prescription may be cumbersome. In this case, the addition of geodesic deviation terms might as in Eq. 72 be necessary.

If furthermore, motion is wished to be expressed in coordinate time, like in the pragmatic approach, it is useful to find a correspondence between the first order perturbation equation, containing geodesic deviations, Eq. 72, and the terms of Eq. 59. It is going to be shown that the α_6 term of Eqs. 59 and 62 corresponds to the self-force term of Eq. 72, when the latter is transferred to coordinate time. Referring to the time and radial components of the self-force, it is obtained from Eqs. 46 and 47:

$$\frac{d^2 t}{d\tau^2} = F_{self}^t - m \Gamma_{\beta\gamma}^t u^\beta u^\gamma = -m g^{t\beta} \left(h_{\beta\gamma;\delta}^* - \frac{1}{2} h_{\gamma\delta;\beta}^* \right) u^\gamma u^\delta - m k^t, \quad (76)$$

$$\frac{d^2 r}{d\tau^2} = F_{self}^r = -m \Gamma_{\beta\gamma}^r u^\beta u^\gamma - m g^{r\beta} \left(h_{\beta\gamma;\delta}^* - \frac{1}{2} h_{\gamma\delta;\beta}^* \right) u^\gamma u^\delta - m k^r, \quad (77)$$

where

$$k^\alpha = \left(h_{\beta\gamma;\delta}^* - \frac{1}{2} h_{\delta\gamma;\beta}^* \right) u^\alpha u^\beta u^\gamma u^\delta. \quad (78)$$

Since

$$\frac{d}{d\tau} = \frac{dt}{d\tau} \frac{d}{dt} \quad \frac{d^2r}{d\tau^2} = \frac{dr}{dt} \frac{d^2t}{d\tau^2} + \frac{d^2r}{dt^2} \left(\frac{dt}{d\tau} \right)^2,$$

after some computation, the term k^α disappears when the self-force is expressed in coordinate time:

$$m \frac{d^2r}{dt^2} = m \left[\Gamma_{\beta\gamma}^t v^\beta v^\gamma + g^{t\beta} \left(h_{\beta\gamma}^*{}_{;\delta} - \frac{1}{2} h_{\gamma\delta}^*{}_{;\beta} \right) v^\gamma v^\delta \right] \dot{z}_u(t) - m \Gamma_{\beta\gamma}^r v^\beta v^\gamma - m g^{r\beta} \left(h_{\beta\gamma}^*{}_{;\delta} - \frac{1}{2} h_{\gamma\delta}^*{}_{;\beta} \right) v^\gamma v^\delta. \quad (79)$$

Furthermore, a tedious computation shows that Eq. 79 is nothing else than the α_6 term of Eq. 59 apart from the regularisations by mode-sum or Riemann–Hurwitz ζ function:

$$\alpha_6 \leftrightarrow g^{t\beta} \left(h_{\beta\gamma}^*{}_{;\delta} - \frac{1}{2} h_{\gamma\delta}^*{}_{;\beta} \right) v^\gamma v^\delta \dot{z}_u(t) - g^{r\beta} \left(h_{\beta\gamma}^*{}_{;\delta} - \frac{1}{2} h_{\gamma\delta}^*{}_{;\beta} \right) v^\gamma v^\delta. \quad (80)$$

The recasting of the Riemann tensor term and of the left-hand side of Eq. 72 into coordinate time determines the relation to the $\alpha_1 \Delta z$ and $\alpha_2 \Delta \dot{z}$ terms in Eqs. 59 and 62. Finally, the geodesic deviation equation³⁸ Eq. 72 is dealt with by Aoudia and Spallicci [7] as difference between background and perturbed motions.

10 Conclusions

It has been shown that free fall is still the arena for a deeper comprehension of gravitation. Furthermore, it still generates acute observations like relativistic gliding for which the asymmetric oscillations of a quasirigid body slow down or accelerate its fall in a gravitational background [96].

But is the problem of radial fall solved? Do we know the laws of motion of a star falling into a black hole, the relativistic modern version of the falling stone? A fair and objective answer leads to a moderate optimism. The general relativistic problem has had undeniable progress from 1997, but a careful analysis of the literature shows that some issues are either still partly open or simply not fully at hand, in terms of clear procedures, by means of which clearly cut answers are obtained.

The remaining steps to be fulfilled for a satisfactory level of comprehension for radial fall of a small particle into a large mass (represented by an SD black hole) within the first perturbative order in m/M are divided in three groups.

Investigations to be pursued before recurring to the self-consistent prescription:

1. Compute in proper time the first order perturbation equation with deviation terms, i.e. all terms of Eq. 72.
2. Identify and compare expressions for renormalised, semi-renormalised, and unrenormalised accelerations in the SD perturbed geometry; evaluate repulsive

³⁸ See Levi-Civita [122, 123], Ciufolini and Wheeler [38], Ciufolini [37].

effects. The last century debate on repulsion has not addressed the effects of mass finitude ($l = 0, 1$) and radiation reaction on the motion ($l \geq 2$). The inclusion of perturbations blurs the issue further and renders it more complex.

3. Compare the waveform corrections obtained with energy-balance by Aoudia [5] with those to be obtained by the integration in the equation of motion of the approximate analytic expression of the self-force of Barack and Lousto [18]. The comparison is of interest only in case of concurring outcomes, as both procedures have a large degree of approximation. An other daring comparison [191] could be made with the numerical results for a 100 : 1 mass ratio, as larger mass ratios imply the solution of the numerical two-scale problem, not yet at hand.
4. Solve the initial conditions for particles with non-null or even relativistic initial velocities with and without large eccentricities. If the outcome of the previous item is successful, coarsely identify possible signatures of radiation reaction.³⁹ It is likely that a non-null initial velocity would act as an amplifier of the waveform shifts due to radiation reaction. In [133], Lousto and Price show that the energy radiated by a non-null initial velocity may rise up to almost two orders of magnitude (see also Fig. 3 in [128]). Incidentally, this line of work is confluent with the interests shown recently in the particle physics community [8, 185, 192, 203, 214] for head-on collisions and the associated radiation reaction [89, 90]. Hopefully, progress in numerical relativity may lead to analysis of larger mass ratios and thereby test future results of perturbation theory⁴⁰ in the range beyond 10^{-3} .

Investigations to be performed solely by the implementation of the self-consistent prescription:

5. Evaluate the trajectory by means of the Gralla and Wald self-consistent method [92]. At each integration step, for a given number of modes: evaluation of the perturbation functions at the position of the particle; regularisation by mode-sum or ζ methods; correction of the geodesic and identification of the cell crossed by the particle; computation of the source term; reiteration of the above. Perturbations and self-force analysis in de Donder's gauge are on-going [19, 23].
6. Repeat steps [i–iv] on the basis of the acquired self-consistent trajectory. An iterative scheme may be envisaged also in coordinate time.

Investigations to be performed independently from the self-consistent prescription:

7. Identify whether in a thought radial fall experiment, there is a physical observable independent of the gauge choice or else manifesting a recognisable effects in a given gauge, following Detweiler [57].

³⁹ For Mino and Brink [149], the energy and momentum radiated are computed on the assumption that the small body falls in a dynamical time scale, with respect to proper time, well short of the radiation reaction time scale and therefore the gravitational radiation back-action on the orbit is considered negligible. The particle plunges on a geodesic trajectory, incidentally starting from a circular orbit, thus at zero initial radial velocity.

⁴⁰ For head-on collisions, Price and Pullin have surprisingly shown the applicability of perturbation theory for the computation of the radiated energy and of the waveform for two equal black holes, starting at very small separation distances [164], the so-called close-limit approximation.

8. Identify the domain of applicability of the self-force. Although self-consistency represents a closer description to perturbed motion, it is limited to cases where local deviations are small, since after all, it remains a perturbative approach. Thus, a quantitative identification of the domain of applicability of such prescription would be of interest to the community developing LISA templates. The adiabatic approximation cannot be evoked to establish the limits of any self-force based analysis. Indeed, the hypothesis on the feeble magnitude of local deviations may be less constraining than the adiabatic hypothesis, for the absence of requirements on averaging. An explicit condition, referring to the geometry of the orbit and to the mass ratio and defining the domain of applicability of the self-force versus fully numerical approaches, is not yet available.
9. Investigate other gauges than Regge–Wheeler and other methods than the self-force for radial fall. A confirmation of the results by Barack, Lousto in [18, 130] by an independent group is missing. It would be also beneficial to carry out an analysis by the EOB method [25, 45].⁴¹

The self-force community and even more the EMRI community are animated by different interests ranging from fundamental physics and theory, to numerical applications, data analysis and astrophysics. To these variegated communities and generally to physicists and astrophysicists, the capture of stars by supermassive black holes will bring a development comparable to the ancient markings made by the leaning Pisa tower, the Woolsthorpe apple tree and the Einstein lift.

Acknowledgements Discussions throughout the years with S. Aoudia, L. Barack, S. Chandrasekhar, S. Detweiler, C. Lousto, J. Martin-Garcia, S. Gralla, E. Poisson, R. Price, R. Wald, B. Whiting are acknowledged. I would like to thank E. Vergès for the support to CNRS School on Mass and the 11th Capra conference held in June 2008 in Orléans. It was my sincere hope that both events could put together different communities working on mass and motion: part of the contributors to this book and their colleagues (Blanchet, Detweiler, Le Tiec and Whiting) have already made concrete steps towards such cooperation [27, 28].

References

1. C. Allègre, *Un peu de science pour tout le monde* (Fayard, Paris, 2003)
2. S. Antoci, in *Meteorological and Geophysical Fluid Dynamics (a book to commemorate the centenary of the birth of Hans Ertel)*, ed. by W. Schröder (Science Edition, Bremen, 2004), p. 343
3. S. Antoci, D.-E. Liebscher, *Astron. Nachr.* **322**, 137 (2001)
4. Apollo 15, <http://nssdc.gsfc.nasa.gov/planetary/lunar/apollo.html>
5. S. Aoudia, *Capture des étoiles par les trous noirs et ondes gravitationnelles*, Doctorate thesis, Dir. A.D.A.M. Spallicci, Université de Nice-Sophia Antipolis, 2008
6. S. Aoudia, A.D.A.M. Spallicci, in *Proc. 12th Marcel Grossmann Mtg*, Paris 12–18 July 2009, ed. by T. Damour, R.T. Jantzen, R. Ruffini (World Scientific, Singapore, to appear), arXiv:1003.3107v3 [gr-qc] (2010); see also S. Aoudia, A.D.A.M. Spallicci, arXiv:1008.2507v1 [gr-qc] (2010)

⁴¹ A non-recent analysis by Simone, Poisson and Will [187], between pN and perturbation methods in head-on collisions, was limited to the computation of the gravitational wave energy flux.

7. S. Aoudia, A.D.A.M. Spallicci, to appear
8. R.F. Aranha, H.P. de Oliveira, I. Damião Soares, E.V. Tonini, *Int. J. Mod. Phys. D* **17**, 2049 (2008)
9. L.I. Arifov, *Problemy Teor. Gravitatsii i Èlement. Chastits* **11**, 96 (1980)
10. L.I. Arifov, *Izv. Vyssh. Uchebn. Zaved. Fiz.* **4**, 61 (1981). English translation: *Rus. Phys. J.* **24**, 346 (1981)
11. Aristotélēs, *Corpus Aristotelicum: Physica* (184a), *De caelo* (268a), *Methaphysica* (980a), in Latin (in parenthesis the Bekker index). English translation: *The Complete Works of Aristotle*, ed. by J. Barnes (Princeton Univ. Press, Princeton, 1984)
12. R. Baierlein, *Phys. Rev. D* **8**, 4639 (1973)
13. L. Barack, *Phys. Rev. D* **62**, 084027 (2000)
14. L. Barack, *Phys. Rev. D* **64**, 084021 (2001)
15. L. Barack, *Class. Q. Grav.* **26**, 213001 (2009)
16. L. Barack, L. Burko, *Phys. Rev. D* **62**, 084040 (2000)
17. L. Barack, C. Cutler, *Phys. Rev. D* **70**, 122002 (2004)
18. L. Barack, C.O. Lousto, *Phys. Rev. D* **66**, 061502 (2002)
19. L. Barack, C.O. Lousto, *Phys. Rev. D* **72**, 104026 (2005)
20. L. Barack, A. Ori, *Phys. Rev. D* **61**, 061502 (2000)
21. L. Barack, A. Ori, *Phys. Rev. D* **64**, 124003 (2001)
22. L. Barack, Y. Mino, H. Nakano, A. Ori, M. Sasaki, *Phys. Rev. Lett.* **88**, 091101 (2002)
23. L. Barack, N. Sago, *Phys. Rev. D* **81**, 084021 (2010)
24. H. Bauer, *Mathematische einföhrung in die gravitationstheorie Einsteins, nebst einer exakten darstellung ihrer wichtigsten ergebnisse* (F. Deuticke, Leipzig and Wien, 1922)
25. S. Bernuzzi, A. Nagar, *Phys. Rev. D* **81**, 084056 (2010)
26. M. Biezunski, *Einstein à Paris: le temps n'est plus* (Presse Univ. de Vincennes, St. Denis, 1991)
27. L. Blanchet, S. Detweiler, A. Le Tiec, B.F. Whiting, *Phys. Rev. D* **81**, 064004 (2010)
28. L. Blanchet, S. Detweiler, A. Le Tiec, B.F. Whiting, *Phys. Rev. D* **81**, 084033 (2010)
29. L. Burko, *Class. Q. Grav.* **22**, S847 (2005)
30. A. Čadež, M. Calvani, A. Gomboc, U. Kostić, in *Albert Einstein Century International Conference*, ed. by J.-M. Alimi, A. Füzfa, *AIP Conf. Proc.* **861** (AIP, Melville, 2006) pp. 566–571
31. G. Cavalleri, G. Spinelli, *Lett. Nuovo Cimento* **6**, 5 (1973)
32. G. Cavalleri, G. Spinelli, *Phys. Rev. D* **15**, 3065 (1977)
33. S. Chandrasekhar, *Proc. R. Soc. Lond. A* **343**, 289 (1975)
34. S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Oxford University Press, Oxford, 1983)
35. S. Chandrasekhar, S. Detweiler, *Proc. R. Soc. Lond. A* **344**, 441 (1975)
36. K.P. Chung, *Nuovo Cimento B* **14**, 293 (1973)
37. I. Ciufolini, *Phys. Rev. D* **34**, 1014 (1986)
38. I. Ciufolini, J.A. Wheeler, *Gravitation and Inertia* (Princeton University Press, Princeton, 1995)
39. F.I. Copperstock, *Phys. Rev. D* **10**, 3171 (1974)
40. P. Crawford, I. Tereno, *Gen. Rel. Grav.* **34**, 2075 (2002)
41. G. Cruciani, *Nuovo Cimento B* **115**, 693 (2000)
42. G. Cruciani, *Nuovo Cimento B* **120**, 1045 (2005)
43. C. Cutler, D. Kennefick, E. Poisson, *Phys. Rev. D* **50**, 3816 (1994)
44. M. Dafermos, I. Rodnianski, in *Clay Summer School*, Zurich 23 June - 18 July 2008, arXiv:0811.0354v1 [gr-qc] (2008)
45. T. Damour, arXiv:0910.5533v1 [gr-qc] (2009)
46. T. Damour, *Sp. Sc. R.* **148**, 191 (2009)
47. M. Davis, R. Ruffini, *Lett. Nuovo Cimento* **2**, 1165 (1972)
48. M. Davis, R. Ruffini, W.H. Press, R.H. Price, *Phys. Rev. Lett.* **27**, 1466 (1971)
49. M. Davis, R. Ruffini, J. Tiomno, *Phys. Rev. D* **5**, 2932 (1972)
50. T. de Donder, *La gravifique Einsteinienne* (Gauthier-Villars, Paris, 1921)
51. T. de Donder, *The Mathematical Theory of Relativity* (M.I.T. Press, Cambridge, 1927)

52. C. de Jans, Mem. Acad. R. Belgique Cl. Sci. **7**, 1 (1923)
53. C. de Jans, Mem. Acad. R. Belgique Cl. Sci. **7**, 1 (1924)
54. C. de Jans, Mem. Acad. R. Belgique Cl. Sci. **7**, 96 (1924)
55. S. Detweiler, Phys. Rev. Lett. **86**, 1931 (2001)
56. S. Detweiler, Class. Q. Grav. **22**, S681 (2005)
57. S. Detweiler, Phys. Rev. D **77**, 124026 (2008)
58. S. Detweiler, E. Poisson, Phys. Rev. D **69**, 084019 (2004)
59. S. Detweiler, B.F. Whiting, Phys. Rev. D **67**, 024025 (2003)
60. B.S. DeWitt, R.W. Brehme, Ann. Phys. (N.Y.) **9**, 220 (1960)
61. P.A.M. Dirac, Proc. R. Soc. Lond. A **167**, 148 (1938)
62. N.A. Doughty, Am. J. Phys. **49**, 412 (1981)
63. S. Drake, *Two New Sciences/A History of Free Fall, Aristotle to Galileo* (Wall and Emerson, Toronto, 2000)
64. J. Droste, Kon. Ak. Wetensch. Amsterdam **23**, 968 (1915). English translation: Proc. Acad. Sci. Amsterdam **17**, 998 (1915)
65. J. Droste, *Het zwaartekrachtsveld van een of meer lichamen volgens de theorie van Einstein*, Doctorate thesis, Dir. H.A. Lorentz, Rijksuniversiteit van Leiden, 1916
66. J. Droste, Kon. Ak. Wetensch. Amsterdam **25**, 163 (1916). English translation: Proc. Acad. Sci. Amsterdam **19**, 197 (1917)
67. P. Drumaux, Ann. Soc. Sci. Bruxelles **56**, 5 (1936)
68. I.G. Dymnikova, Yad. Fiz. **31**, 679 (1980). English translation: Sov. J. Nucl. Phys. **31**, 353 (1980)
69. J. Earman, J. Eisenstaedt, Stud. Hist. Phil. Mod. Phys. **30**, 185 (1999)
70. A.S. Eddington, Nature **105**, 37 (1920)
71. A.S. Eddington, Nature **113**, 192 (1924)
72. A. Einstein, Ann. Phys. (Leipzig) **49**, 769 (1916). English translations: *The Principle of Relativity* (Methuen and Company, London, 1923; reprinted by Dover, New York, 1952); *The Collected Papers of Albert Einstein: The Berlin Years: Writings: 1914–1917*, ed. by M.J. Klein, A.J. Kox, J. Renn, R. Schulmann, vol. 6 (Princeton University Press, Princeton, 1996), p. 146
73. A. Einstein, M. Grossmann, *Entwurf einer verallgemeinerten relativitätstheorie und einer theorie der gravitation* (Teubner, Leipzig, 1913); reprinted by Zeitschrift Math. Phys. **62**, 225 (1914). English translation: *The Collected Papers of Albert Einstein: The Swiss Years: Writings: 1914–1917*, ed. by M.J. Klein, A.J. Kox, J. Renn, R. Schulmann, vol. 4 (Princeton University Press, Princeton, 1996), p. 151
74. A. Einstein, M. Grossmann, Zeitschrift Math. Phys. **63**, 215 (1914). English translation: *The Collected Papers of Albert Einstein: The Berlin Years: Writings: 1914–1917*, ed. by M.J. Klein, A.J. Kox, J. Renn, R. Schulmann, vol. 6 (Princeton University Press, Princeton, 1996), p. 6
75. L.P. Eisenhart, *Riemannian Geometry* (Princeton University Press, Princeton, 1926)
76. J. Eisenstaedt, Arch. Hist. Exact Sci. **27**, 157 (1982)
77. J. Eisenstaedt, Arch. Hist. Exact Sci. **37**, 275 (1987)
78. E.D. Fackerell, Astrophys. J. **166**, 197 (1971)
79. P. Fayet, arXiv:hep-ph/0111282v1 (2001)
80. M. Ferraris, M. Francaviglia, C. Reina, Gen. Rel. Grav. **14**, 243 (1982)
81. D. Finkelstein, Phys. Rev. **110**, 965 (1958)
82. P. Fiziev, Class. Q. Grav. **26**, 2447 (2006)
83. M. Freitag, Astrophys. J. **583**, L21 (2003)
84. J. Friedman, Proc. R. Soc. Lond. A **335**, 163 (1973)
85. V. Frolov, I. Novikov, *Black Hole Physics* (Kluwer, Dordrecht, 1998)
86. G. Galilei, Letter to P. Sarpi (1604)
87. G. Galilei, *Dialogo sopra i due massimi sistemi del mondo, Tolemaico e Copernicano* (G.B. Landini, Firenze, 1632). English translations: *Dialogue Concerning the Two Chief World Systems, Ptolemaic and Copernican* (McMillan, New York, 1914; California University Press, Berkeley, 1967)

88. G. Galilei, *Discorsi e dimostrazioni matematiche intorno à due nuove scienze attenenti alla meccanica e i movimenti locali* (Elsevier, Leiden, 1638). English translation: see reference [63]
89. D.V. Gal'tsov, G. Kofinas, P. Spirin, T.N. Tomaras, arXiv:1003.2982v1 [hep-th] (2010)
90. D.V. Gal'tsov, G. Kofinas, P. Spirin, T.N. Tomaras, Phys. Lett. B. **683**, 331 (2010)
91. D.V. Gal'tsov, P. Spirin, S. Staub, in *Gravitation and Astrophysics*, Proc. VII Asia-Pacific Int. Conf., Chungli 23–26 November 2005, ed. by J.M. Nester, C.-M. Chen, J.-P. Hsu (World Scientific, Singapore, 2006), p. 346
92. S.E. Gralla, R.M. Wald, Class. Q. Grav. **25**, 205009 (2008)
93. G. Green, J. Reine Ang. Math. J. Crelle **39**, 73 (1850)
94. G. Green, J. Reine Ang. Math. J. Crelle **44**, 356 (1852)
95. G. Green, J. Reine Ang. Math. J. Crelle **47**, 161 (1854)
96. E. Guéron, R.A. Mosna, Phys. Rev. D **75**, 081501 (2007)
97. J. Hadamard, *Lectures on Cauchy's Problem in Linear Partial Differential Equations* (Yale University Press, New Haven, 1923)
98. Y. Hagihara, Jap. J. Astron. Geophys. **8**, 67 (1931)
99. M.P. Haugan, Ann. Phys. (N.Y.) **118**, 156 (1979)
100. M.P. Haugan, S.L. Shapiro, I. Wasserman, Astrophys. J. **257**, 283 (1982)
101. D. Hilbert, Nachr. König. Ges. Wiss. Göttingen, Math. Phys. Kl. 395 (1915). English translation: in *The Genesis of General Relativity*, Boston Stud. Phil. Sc., ed. by J. Renn, vol. 4 (Springer, Dordrecht, 2007)
102. D. Hilbert, Nachr. König. Ges. Wiss. Göttingen, Math. Phys. Kl. 53 (1917). English translation: in *The Genesis of General Relativity*, Boston Studies Phil. Sc., ed. by J. Renn, vol. 4 (Springer, Dordrecht, 2007)
103. D. Hilbert, Math. Ann. **92**, 1 (1924) (reproduction with extensive changes of [102])
104. T. Hinderer, É.É. Flanagan, Phys. Rev. D **78**, 064028 (2008)
105. J.M. Hobbs, Ann. Phys. (N.Y.) **47**, 141 (1968)
106. A. Hurwitz, Z. Math. Phys. **27**, 86 (1882)
107. J. Jaffe, I.I. Shapiro, Phys. Rev. D **6**, 405 (1972)
108. J. Jaffe, I.I. Shapiro, Phys. Rev. D **8**, 4642 (1973)
109. A.I. Janis, Phys. Rev. D **8**, 2360 (1973)
110. A.I. Janis, Phys. Rev. D **15**, 3068 (1977)
111. D. Kennefick, *Traveling at the Speed of Thought: Einstein and the Quest for Gravitational Waves* (Princeton University Press, Princeton, 2007)
112. M. Kutschera, W. Zajiczek, arXiv:0906.5088v1 [astro-ph.EP] (2009)
113. C. Lämmerzahl, Eur. Phys. J. Sp. Top. **163**, 255 (2008)
114. L.D. Landau, E.M. Lifshits, *Teoriya polya* (MIR, Moskva, 1941). English translation: *The Classical Theory of Fields* (Pergamon, Oxford, 1951)
115. P.S. Laplace, *Exposition du système du monde*, vol. 2 (Imprimerie du Cercle Social, Paris, 1796). English translation: *The System of the World* (Richard Phillips, London, 1809)
116. P.S. Laplace, *Allgemeine geographische Ephemeriden*, vol. 4 (Im verlage des industrie-comptoirs, Weimar, 1799). English translation: H. Stephani, arXiv:gr-qc/0304087v1 (2003)
117. E.W. Leaver, Proc. R. Soc. Lond. A **402**, 285 (1985)
118. E.W. Leaver, J. Math. Phys. **27**, 1238 (1986)
119. E.W. Leaver, Phys. Rev. D **34**, 384 (1986)
120. *Le Canard Enchaîné*, February 24th, March 3rd, March 10th, March 17th (1999)
121. Leonardo da Vinci, *Code A 22v*, Institut de France. Original manuscript and text online, <http://www.leonardodigitale.com>
122. T. Levi-Civita, *Lezioni di calcolo differenziale assoluto*, compiled by E. Persico (Stock, Roma, 1925). English translation: *The Absolute Differential Calculus* (Blackie, Glasgow, 1927)
123. T. Levi-Civita, Math. Ann. **97**, 291 (1926)
124. LISA, <http://www.esa.int/science/lisa>, <http://lisa.jpl.nasa.gov>
125. A. Loinger, T. Marsico, arXiv:0904.1578v1 [physics.gen-ph] (2009)
126. H.A. Lorentz, Kon. Ak. Wetensch. Amsterdam **23**, 1073 (1915). English translation: Proc. Acad. Sci. Amsterdam **19**, 751 (1915)
127. L. Lorenz, Philos. Mag. **34**, 287 (1867)

128. C.O. Lousto, *Mod. Phys. Lett. A* **12**, 1879 (1997)
129. C.O. Lousto, *Phys. Rev. Lett.* **84**, 5251 (2000)
130. C.O. Lousto, *Class. Q. Grav.* **18**, 3989 (2001)
131. C.O. Lousto, *Class. Q. Grav.* **22**, S543 (2005)
132. C.O. Lousto, H. Nakano, *Class. Q. Grav.* **26**, 015007 (2009)
133. C.O. Lousto, R.H. Price, *Phys. Rev. D* **55**, 2124 (1997)
134. C.O. Lousto, R.H. Price, *Phys. Rev. D* **56**, 6439 (1997)
135. C.O. Lousto, R.H. Price, *Phys. Rev. D* **57**, 1073 (1998)
136. E. Mach, *Die mechanik in ihrer entwicklung historisch-kritisch dargestellt* (Brockhaus, Leipzig, 1883). English translation: *The Science of Mechanics. A Critical and Historical Exposition of Its Principals* (The Open Court Publishing Co., Chicago, 1893)
137. S. Mano, H. Suzuki, E. Takasugi, *Progr. Theor. Phys.* **96**, 549 (1996)
138. F. Markley, *Am. J. Phys.* **41**, 45 (1973)
139. K. Martel, E. Poisson, *Phys. Rev. D* **66**, 084001 (2002)
140. K. Martel, E. Poisson, *Phys. Rev. D* **71**, 104003 (2005)
141. J. Mathews, *J. Soc. Ind. Appl. Math.* **10**, 768 (1962)
142. C.H. McGruder III, *Phys. Rev. D* **25**, 3191 (1982)
143. G.C. McVittie, *General Relativity and Cosmology* (Chapman and Hall, London, 1956)
144. J. Michell, *Phil. Trans. R. Soc. Lond.* **74**, 35 (1784)
145. Y. Mino, *Class. Q. Grav.* **22**, S375 (2005)
146. Y. Mino, *Class. Q. Grav.* **22**, S717 (2005)
147. Y. Mino, *Progr. Theor. Phys.* **113**, 733 (2005)
148. Y. Mino, *Progr. Theor. Phys.* **115**, 43 (2006)
149. Y. Mino, J. Brink, *Phys. Rev. D* **78**, 124015 (2008)
150. Y. Mino, M. Sasaki, T. Tanaka, *Phys. Rev. D* **55**, 3457 (1997)
151. A. Mitra, *Found. Phys. Lett.* **13**, 543 (2000)
152. V. Moncrief, *Ann. Phys. (N.Y.)* **88**, 323 (1974)
153. T. Müller, *Gen. Rel. Grav.* **40**, 2185 (2008)
154. A. Nagar, L. Rezzolla, *Class. Q. Grav.* **22**, R167 (2005); Corrigendum, *ibidem* **23**, 4297 (2006)
155. T. Nakamura, K. Oohara, Y. Kojima, *Progr. Theor. Phys. Suppl.* **90**, 110 (1987)
156. I. Newton, *Philosophiæ naturalis principia mathematica* (S. Pepys Reg. Soc. Præses, London, 1687). English translation: *The Principia, Mathematical Principles of Natural Philosophy – A New Translation* (California University Press, Berkeley, 1999)
157. L. Page, *Nature* **104**, 692 (1920)
158. L.I. Petrich, S.L. Shapiro, I. Wasserman, *Astrophys. J. Suppl. S.* **58**, 297 (1985)
159. E. Poisson, *Living Rev. Rel.* **7**, URL (cited on 4 May 2010): <http://www.livingreviews.org/lrr-2004-6>
160. H. Poli de Souza, Report for the Master thesis, Dir. A.D.A.M. Spallicci, Observatoire de Paris (2008)
161. A. Pound, *Phys. Rev. D* **81**, 024023 (2010)
162. A. Pound, E. Poisson, *Phys. Rev. D* **77**, 044013 (2008)
163. G. Preti, *Found. Phys.* **39**, 1046 (2009)
164. R.H. Price, J. Pullin, *Phys. Rev. Lett.* **72**, 329 (1994)
165. T.C. Quinn, R.M. Wald, *Phys. Rev. D* **56**, 3381 (1997)
166. T.C. Quinn, R.M. Wald, *Phys. Rev. D* **60**, 064009 (1999)
167. T. Regge, *Nuovo Cimento* **5**, 325 (1957); Additional remarks and errata corrige, *ibidem* **6**, 1233 (1957)
168. T. Regge, J.A. Wheeler, *Phys. Rev.* **108**, 1063 (1957)
169. G.F.B. Riemann, *Monatsber. Königl. Preuss. Akad. Wiss. Berlin*, 671 (1859). English translation by H.M. Edwards, *Riemann's Zeta Function* (Academic, New York, 1974)
170. W. Rindler, *Essential Relativity*, 2nd revised edn. (Springer, New York, 1979)
171. W. Rindler, *Relativity* (Oxford University Press, New York, 2006)
172. P. Ritter, Report for the Master thesis, Dir. A.D.A.M. Spallicci, Université de Toulouse III Paul Sabatier, 2010
173. F. Rohrlich, *Found. Phys.* **30**, 621 (2000)

174. T. Rothman, *Gen. Rel. Grav.* **34**, 1541 (2002)
175. R. Ruffini, in *Black Holes*, Les Houches 30 July–31 August 1972, ed. by C. de Witt, B. de Witt (Gordon and Breach Science, New York, 1973), p. 451
176. R. Ruffini, *Phys. Rev. D* **7**, 972 (1973)
177. R. Ruffini, in *Physics and Astrophysics of Neutron stars and Black Holes*, Proc. of the Int. School of Physics E. Fermi Course LXV, Varenna 14–26 July 1975, ed. by R. Giacconi, R. Ruffini (North-Holland, Amsterdam and Soc. It. Fisica, Bologna, 1978), p. 287
178. R. Ruffini, J.A. Wheeler, in *Significance of Space Research for Fundamental Physics*, ESRO colloquium, Interlaken 4 September 1969, ESRO-52, ed. by A.F. Moore, V. Hardy (European Space Research Organisation, Paris, 1971), p. 45
179. R. Ruffini, J.A. Wheeler, in *The Astrophysical Aspects of the Weak Interactions*, Cortona 10–12 June 1970, ANL Quaderno n. 157 (Accademia Nazionale dei Lincei, Roma, 1971), p. 165
180. N. Sago, H. Nakano, M. Sasaki, *Phys. Rev. D* **67**, 104017 (2003)
181. K. Schwarzschild, *Sitzungsber. Preuss. Akad. Wiss., Phys. Math. Kl.* **189** (1916). English translation with foreword by S. Antoci and A. Loinger, arXiv:physics/9905030v1 [physics.hist-ph] (1999)
182. S.L. Shapiro, S.A. Teukolsky, *Black Holes, White Dwarfs, and Neutron Stars: the Physics of Compact Objects* (Wiley, New York, 1983)
183. S.L. Shapiro, I. Wasserman, *Astrophys. J.* **260**, 838 (1982)
184. M. Shibata, T. Nakamura, *Progr. Theor. Phys.* **87**, 1139 (1992)
185. M. Shibata, H. Okawa, T. Yamamoto, *Phys. Rev. D* **78**, 101501 (2008)
186. Simplicios, *De caelo commentaria*. English translation: *On the Heavens* (Duckworth, London and Cornell University Press, Ithaca, 1987)
187. L.E. Simone, E. Poisson, C.M. Will, *Phys. Rev. D* **52**, 4481 (1995)
188. C.F. Sopuerta, P. Laguna, *Phys. Rev. D* **73**, 044028 (2006)
189. A.D.A.M. Spallicci, in *Proc. 8th Marcel Grossmann Mtg*, Jerusalem 22–28 June 1997, ed. by T. Piran, R. Ruffini (World Scientific, Singapore, 1998), p. 1107
190. A.D.A.M. Spallicci, S. Aoudia, *Class. Q. Grav.* **21**, S563 (2004)
191. U. Sperhake, Private communication on 18 September 2009
192. U. Sperhake, V. Cardoso, F. Pretorius, E. Berti, J.A. González, *Phys. Rev. D* **101**, 161101 (2008)
193. G. Spinelli, in *Proc. 5th Marcel Grossmann Mtg*, Western Australia 8–13 August 1988, ed. by D.G. Blair, M.J. Buckingham, R. Ruffini (World Scientific, Singapore, 1989), p. 373
194. K.N. Srinivasa Rao, *Ann. Inst. H. Poincaré (A) Phys. Théor.* **5**, 227 (1966)
195. J. Stachel, *Nature* **220**, 779 (1968)
196. J.L. Synge, *Relativity: the General Theory* (North-Holland Publishing Co., Amsterdam, 1960)
197. Y. Tashiro, H. Ezawa, *Prog. Theor. Phys.* **66**, 1612 (1981)
198. E.F. Taylor, J.A. Wheeler, *Exploring Black Holes* (Addison Wesley Longman, San Francisco, 2000)
199. W. Thirring, *Ann. Phys. (N.Y.)* **16**, 96 (1961)
200. K.S. Thorne, *Rev. Mod. Phys.* **52**, 299 (1980)
201. H.J. Treder, *Die relativität de trägheit* (Akademie Verlag, Berlin, 1972)
202. H.J. Treder, K. Fritze, *Astron. Nachr.* **296**, 109 (1975)
203. G. Veneziano, Seminar at the Institut d’Astrophysique de Paris on 25 January 2010
204. C.V. Vishveshwara, *Phys. Rev. D* **1**, 2870 (1970)
205. M. von Laue, *Die relativitätstheorie. Die allgemeine relativitätstheorie*, vol. 2, 1st edn. (Vieweg und Sohn, Braunschweig, 1921). French translation: *La théorie de la relativité. La relativité générale et la théorie de la gravitation d’Einstein*, vol. 2 (Gauthier-Villars et C^{ie}, Paris, 1926) translation in French of the revised and integrated German 4th edn. (1924)
206. E. von Rabe, *Astron. Nachr.* **275**, 251 (1947)
207. H. Weyl, *Ann. Phys. (Leipzig)* **54**, 117 (1917)
208. J.A. Wheeler, Public lecture held at the Goddard Institute of Space Studies on 29 December 1967; *Am. Sch.* **37**, 248 (1968); *Am. Sci.* **56**, 1 (1968)
209. B.F. Whiting, S. Detweiler, *Int. J. Mod. Phys. D* **9**, 1709 (2003)

210. E.T. Whittaker, *A History of the Theories of Aether and Electricity*, vol. 2 (Nelson, London, 1953)
211. C.M. Will, *Theory and Experiments in Gravitational Physics*, revised edition (Cambridge University Press, Cambridge, 1993)
212. C.M. Will, *Living Rev. Rel.* **9**, URL (cited on 4 May 2010): <http://www.livingreviews.org/lrr-2006-3>
213. M. White, *Isaac Newton: the Last Sorcerer* (Fourth Estate Limited, London, 1997)
214. H. Yoshino, M. Shibata, *Phys. Rev. D* **9**, 084025 (2009)
215. Y.B. Zel'dovich, I.D. Novikov, *Dokl. Akad. Nauk* **155**, 1033 (1964). English translation: *Sov. Phys. Doklady* **9**, 246 (1964)
216. Y.B. Zel'dovich, I.D. Novikov, *Relyativistskaya astrofizika* (Izdatel'svo Nauka Moskva, 1967). English translation (revised and enlarged): *Relativistic Astrophysics* (Chicago University Press, Chicago, 1971)
217. A. Zenginoglu, *Class. Q. Grav.* **27**, 045015 (2010)
218. F.J. Zerilli, *J. Math. Phys.* **11**, 2203 (1970)
219. F.J. Zerilli, *Phys. Rev. Lett.* **24**, 737 (1970)
220. F.J. Zerilli, *Phys. Rev. D* **2**, 2141 (1970); Erratum, in *Black holes*, Les Houches 30 July - 31 August 1972, ed. by C. DeWitt, B. DeWitt (Gordon and Breach Science, New York, 1973)
221. V.I. Zhdanov, *Izv. Vyssh. Uchebn. Zaved. Fiz.* **9**, 34 (1979). English translation: *Sov. Phys. J.* **22**, 951 (1979)

Index

A

- Abraham, 49, 270, 324, 504
- Abraham-Lorentz
 - equation, 49
 - vector, 504
- Abraham-Lorentz-Dirac equation, 270
- acceleration
 - Pioneer, 48, 54
 - scale, 48, 54, 289, 480
 - self-acceleration, 50–52, 410, 566
 - small, *see* small acceleration
 - tidal, 289–290
 - uniqueness of, *see* uniqueness of acceleration
- action
 - Einstein-Hilbert, 90, 98, 128, 462, 468, 480, 533, 537, 543
 - Einstein-Infeld-Hoffmann (EIH), 467
 - Fokker, 461, 465–467, 469, 486
 - Holst, 537
 - matter, 90, 463–467, 485–486
 - Palatini, 537
 - variables, 217, 219
- adiabatic, 368, *see* non-adiabatic, post-adiabatic
 - approximation, 590, 597
 - inspiral, 120
 - kludge waveforms, 453–455
 - last stable orbit, 222
 - limit, 232, 235, 398
 - orbital evolution, 328
 - quasi-circular inspiral, *see* inspiral
 - self-force, 412
 - signals, 228
 - templates, 408
 - waveforms, *see* waveforms
- adiabaticity, 562, 584, 590
- ADM, *see* Arnowitt, Deser and Misner
 - 4-momentum, 102–103
 - and Bondi masses, 115
 - and Bondi-Sachs energy-momentum, 122
 - and Komar Mass, 103
 - angular momentum, 104–105
 - approach, 184
 - canonical formalism, 173
 - coordinates, 159, 184–185
 - energy, 100–101, 103, 105, 107, 112, 182
 - Hamiltonian, 162, 174
 - mass, 102–104, 107–108, 110, 112–113, 119–120, 180, 428, 430
 - mass-energy, 180
 - momentum, 102
- alternative theories of gravity, 461–462, 468
 - aquadratic kinetic terms, 481
 - Aquadratic Lagrangians, 481
 - diagrammatic representation, 465
 - extended bodies, 149, 477, 485
 - mass-dependent models, 481
 - metric coupling, 462–464
 - modified Newtonian dynamics, *see* MOND
 - nonminimal couplings, 486
 - scalar-tensor theories, *see* scalar-tensor
- ampere, 67–68, 70, 75, 77, 84, 351
- analytic
 - case, 257
 - continuation, 133, 138, 153, 155, 214, 235, 420, 423, 426
- analytical waveforms, 212, 243
- angle average, 259
- angular average, 152, 156, 267
- angular momentum, 35, 40, 87–90, 95, 97, 104, 108, 114–115, 121–122, 149, 231–233, 242, 369, 450, 511
 - ADM, 104–105
 - and energy, 33, 91–92, 94, 293, 418
 - Bondi-Sachs, 107
 - conservation, 18, 33
 - defined at spatial infinity, 97, 104–105

- density, 231
- flux, 227, 418
- for the gravitational field, 91–94, 121
- horizon, 116, 118
- in general relativity, 87, 91, 121–122
- intrinsic, 128
- Komar, 95, 105, 114, 120
- orbital, 164, 190, 219, 450
- Quasi-Local, 114
- spin, 148
- anisotropic
 - mass tensor, 30
 - speed of light, 30, 44
- anisotropy (effects), 30, 43, 57
- anti-de Sitter, 99, 534
- anti-hydrogen, 30, 32
- antidamping, 367, 390, 392
- approximation methods, 125–126, 169, 413
- Aristotle (Aristotélēs), 563
- Arnowitt, Deser and Misner (ADM), 97, 102, 158, 169, 216, *see* ADM
- Ashtekar, 115, 533–534, 536–537, 540
 - formulation of gravity, 536
 - variables, 533–534, 537
- Ashtekar-Barbero connection, 537
- Ashtekar-Barbero variable, 537
- Ashtekar-Lewandowski measure, 540
- asphericity of the object, 272
- assumption, unjustified, 264, 325
- astronomical unit, 462
- astronomy, 283, 367
- astrophysical observations, 28, 34, 48, 128
- asymptotic past, 259
- asymptotically flat spacetime, 97, 106, 117, 120, 122, 195, 267
- atomic mass, 73, 77–79, 81
- auxiliary
 - metric, 129, 423
 - worldtube, 357
- average
 - angle, 259
 - angular, 152, 156, 267
- Avogadro constant, 70, 78–79, 81
- axiomatic approach, 264, 311, 322–323, 328, 583
- axioms
 - Detweiler-Whiting, 323
 - Quinn-Wald, 322
- B**
- back reaction, 49, 403
- back-action, 561–569, 590–591, 596
 - Newtonian, *see* Newtonian
- background
 - cosmic microwave, 43, 45, 479
 - field, 387, 561, 574, 588, 589
 - geodesic, 561–562, 584–585, 593–594
 - geometry, *see* geometry
 - metric, 96, 263–264, 287–288, 295, 329, 352, 387–389, 402, 405, 409, 454, 464, 518, 543, 563, 569, 582, 584–585, 587–588, 594
 - motion, 581, 595
 - spacetime, 95, 255, 258, 267–268, 270, 295, 299, 310, 328, 331, 352, 582, 584, 593
- Barbero connection, variable, 537
- Barbero-Immirzi ambiguity, parameter, 537
- barred
 - coordinates, 266
 - metric, 265–266
- Bartnik mass, 113
- base units, new definitions of, 83
- Belinfante-Rosenfeld procedure, 91
- Bianchi identities, 89, 184, 200, 369, 387, 514
- Bianchi identity, 129, 255, 267, 283–284, 296, 387–388, 392
 - linearized, 255, 267
- binary
 - black hole, 26, 37, 120, 125, 175, 177, 182, 196, 211, 214–215, 221–222, 246, 248–249, 395, 416–417, 476
 - coalescing, 211, 213, 248–249
 - neutron stars, 249, 253
 - pulsar, 126–128, 189, 473–475, 483
 - Hulse-Taylor, 126, 473–474
 - system, coalescing, 212
- biscalar, 384, 385
- black hole
 - coalescing binary, 211, 213, 248–249, *see* binary
 - central, 396, 412, 448–453, 455–456, 582
 - charged, 51
 - geometry, 272, 304, 344, 453
 - in three-dimensional Lorentzian anti-de Sitter gravity, 534
 - inspiral, 211–212, 215, 240, 249, 278, 305, 449–450
 - massive, 276, 327, 367, 443–445, 448–449, 561, 564, 582, 597
 - perturbation theory, 226, 343, 417, 433
 - perturbations, 344, 561, 574–575, 577
- body, *see* two-body problem
 - effective one-body, *see* EOB
 - small body, *see* small body
 - sufficiently, 253, 261
 - two-body, *see* two-body problem

- Bondi
 coordinates, 105
 energy, 105, 107
 mass, 107, 115
 Bondi-Metzner-Sachs (BMS), 106
 Bondi-Sachs
 4-momentum, 106–107
 angular momentum, 107
 energy, 111–112
 energy-momentum, 122
 boundedness, 348
 Boyer-Lindquist, 335, 338–339, 341–342, 358, 362, 397
 angular coordinates, 362
 components, 338–339, 342, 358
 coordinates, 338–339, 341, 397
 Schwarzschild coordinates, 335
 Brill-Lindquist, 175–176, 178, 180, 185
 black hole, 175, 185
 solution, 175–176
 Brown-York energy, 109–111, 114
 Brownian motion, 492, 502
 Burke-Thorne
 coordinate condition, 192, 204
 radiation-reaction force, 428
- C**
 caesium, 68, 84
 candela, 68
 candidates for dark matter, 480
 canonical
 conjugate, 173
 energy-momentum tensor, 91
 form, 118, 167, 172–173, 194, 196
 formalism, 167, 172–173, 194
 general relativity, 171, 173, 537
 moments, 145, 147
 quantization, 533
 transformation, 97, 99, 116–118, 188
 variables, 537
 Capra (meeting), 306, 392, 568, 583, 597
 Carter constant, 293, 391, 398
 Casimir
 energy, 491, 494, 505–507, 526
 force, 492, 494, 505–507
 Cassini, 39, 469–470, 474–475, 522, 525
 Cauchy
 hypersurface, 113, 352
 problem, 468, 482, 486, 536
 surface, 352–353
 causality, 323, 412, 482, 496, 499, 501
 centre of mass, 13, 16, 20, 50–52, 149, 157, 160, 177–178, 187, 190, 195–197, 199, 208, 216–217, 220, 267, 270, 272–273, 275, 289, 324, 431, 477, 566–568
 coordinates, 267
 frame, 177–178, 187, 190, 217, 220
 integral, 431
 shift, 566, 568
 challenges, 17, 68, 83, 119, 298, 331, 456–457
 characteristics, 1, 8, 17–18, 294, 332, 345, 449–450, 456, 558
 of the Higgs Boson, 8
 charge conservation, 46, 52–53, 60
 charge-current source, 254, 266
 charged
 black hole, 51
 lepton, 17–18
 particle, 3–4, 36, 49, 76, 277, 322
 Chern-Simons theory, 544
 Christoffel symbol (connection), 33, 57, 259, 285, 574, 582
 circular orbit, 40, 58, 128, 160–161, 178, 190, 206–207, 215, 220–222, 224, 226–227, 231–232, 236, 275, 282, 297–298, 300–301, 303, 349, 354–355, 357, 401, 411–413, 415, 417–418, 421–422, 430–431, 450, 562, 566, 596
 Clebsch-Gordan coefficients, intertwiners, 540
 co-dimension two, 255
 CODATA, 71, 74–75, 77–79, 82–83
 commutation, 329, 511, 512
 compact objects, 103, 125, 128, 147, 150, 157, 396, 418, 444, 448–449, 472, 582
 composition, 27, 30, 42, 50, 253, 565
 Compton frequency, 68, 78, 80, 84
 conformal decomposition, 100, 115
 confusion noise, 452, 454
 conservation
 energy, 498, 562, 564, 584, 592
 equation, 129, 170, 368, 370, 375
 law, 25–27, 49, 89, 141, 170, 369, 387, 398, 468
 probability, 10, 52
 stress-energy tensor, 583
 conservative
 dynamics, 168, 174, 189, 219, 418
 effect, 271, 293, 328, 410, 412–413, 415, 436, 584
 Hamiltonians, 187
 part, 158, 168, 224, 243, 292–293, 298, 411, 418, 427–429

- conserved
 - current, 89–90
 - energy, 159, 163
 - quantity, 88–89, 94–97, 100, 102, 104, 107, 116, 118, 412, 498
 - constants
 - Avogadro, 70, 78–79, 81
 - cosmological, *see* cosmological constant
 - Euler, 160
 - Euler-Mascheroni, 427
 - Fermi, 7
 - fine-structure, *see* fine-structure constant
 - fundamental, *see* fundamental
 - gravitational, *see* gravitational
 - Josephson, 76
 - Planck, *see* Planck
 - Rydberg, 79
 - von Klitzing, 76
 - constraint equation, 174–177, 181, 193, 194, 253
 - conventional values, 77
 - coordinates
 - barred, 266
 - Bondi, 105
 - center of mass, 267
 - components, 265–266, 274
 - Eddington-Finkelstein, 353–354, 578
 - Fermi normal, 258, 267–268, 285, 309, 313
 - fixed, 265
 - harmonic, 129, 131, 139, 140, 142–143, 147, 158–159, 168, 180, 184, 201, 235, 283, 423
 - locally inertial, *see* locally inertial
 - Schwarzschild, *see* Schwarzschild
 - shift, 268
 - Thorne-Hartle-Zhang (THZ), 287, 313
 - transformation, 57, 153, 267, 286, 295, 342, 375, 409, 422, 585
 - unbarred, 266
 - Copernicus, 304
 - correction
 - finite size, 261
 - first-order, 410
 - higher order, 19, 14, 15, 261, 270, 410
 - leading order, 267, 269, 411
 - perturbative, 261, 263, 269, 584
 - second-order, 410
 - self-force, 253, 259, 261
 - correlation coefficient, 75, 78
 - cosmic microwave background, 43, 45, 479
 - cosmological constant, 35, 492–493, 538, 544, 558
 - Coulomb, 30, 217, 271, 288, 305, 336, 348, 368, 378
 - coupled system, 254, 369, 534
 - covariant
 - derivative, 4, 7, 33, 148, 201, 274, 296, 329, 385, 388, 409, 477–478
 - expressions, 313
 - curve, 10, 180, 221, 256, 265, 267, 399, 440, 475
 - cylindrical function, 539–540
- D**
- d'Alembert equation, 423, 425
 - d'Alembertian
 - equation, 129–131, 134, 137
 - operator, 129, 136, 152, 352, 423
 - d-dimensional gravitational constant, 154, 423
 - dark energy, 43, 48, 462, 479, 493
 - dark matter, 20–21, 31, 42–44, 48, 461–462, 476, 479–480, 482, 486, 493
 - around the Earth, 44
 - candidates, 480
 - detection of particles, 42
 - data analysis, 41, 211, 453–457, 597
 - de Broglie-Compton frequency, 80, 84
 - de Donder gauge, 129, 583, 585, 593, 596
 - relaxation, 583, 593
 - de Sitter, 35, 126, 264–265, 534
 - debate, 74–75, 254, 304, 561–562, 564, 569–573, 581, 591, 596
 - decay properties, 254
 - decomposition
 - conformal, 100, 115
 - factorized, 232
 - multiplicative, 216, 227, 232, 241, 244
 - multipole, *see* multipole
 - deformed, 169, 182, 188, 224, 473, 478, 544, 546, 549, 551, 565
 - Delaunay Hamiltonian, 217–218
 - (delta) δ -function, 148, 151, 179, 183, 213–214, 253–256, 316–318, 336, 348–349, 368, 373, 376, 380, 383–385, 396–398, 400–401, 423, *see* Dirac
 - detection of dark matter particles, 42
 - determinant, 4, 128, 201, 257, 352, 423, 462, 505
 - Detweiler-Whiting, 260, 261, 318, 321, 323, 401–403, 419
 - approach, 323, 583
 - Axiom, 323
 - Green's function, 260
 - Observation, 323
 - Reformulation, 260
 - regular two-point function, 318
 - singular Green's function, 318

- deuterium, 79
 - deviation
 - geodesic, *see* geodesic
 - vector field, 268–269
 - second order, 588
 - DeWitt and Brehme, 259, 311, 379, 583
 - DeWitt-Brehme, 305
 - DeWitt-Brehme-Hobbs result, 386
 - diagrammatic representation, 465
 - differential equation, 34, 129, 131, 206, 211, 254, 257–258, 350, 381, 408, 495, 574
 - ordinary, *see* ordinary differential equation
 - partial, 254
 - dimensional
 - analysis, 291, 478
 - regularization, 153–156, 169, 175, 177, 183–184, 196, 214, 415, 420–421, 423, 425–426, 430, 433
 - Dirac
 - coordinate condition, 184
 - delta (δ -distribution) function, 137, 148, 150, 154, 167, 169, 171, 175, 181, 183, 423, 575–576, 581, 589, *see* (delta) δ -function
 - derivation, 368
 - equation, 28, 30, 46
 - gauge condition, 101
 - displacement, 56, 268, 279, 312, 331, 446, 498–499, 503, 506, 517, 587
 - dissipative force, 412, 501
 - distribution of matter, 88, 145
 - distributional
 - solution, 254, 256
 - source, 254
 - divergence, 128, 131–134, 136, 140, 152, 337, 348, 353, 373, 378–379, 381–384, 386–387, 389, 391, 402, 405–406, 425, 496, 499, 502, 509, 515–516, 534, 561, 575, 578, 583, 586, 590
 - free, 89, 107, 112, 114–115, 183–184, 371–372, 388, 390
 - divergence-less, 129
 - domain of dependence, 89, 258
 - Doppler
 - effect, 36, 40, 464, 473
 - measurement, 573
 - ranging, 522
 - residuals, 522
 - shift, 80, 276, 448, 522
 - signal, 523
 - term, 58
 - tracking, 493
 - Drinfeld (quantum) double, 535, 544
 - Droste, 126, 561, 569, 571–572
 - dynamical invariants, 191
- E**
- eccentric orbit, 301, 333, 350–351, 354–355, 396–397
 - Eddington-Finkelstein coordinates, 353–354, 578
 - EEP, 32, 41, 52, 60, *see* Einstein equivalence principle
 - effective one body (EOB), *see* EOB formalism, 211
 - methods, 563, 597
 - effective source, 230, 233, 282, 292, 294, 303–304
 - EIH, *see* Einstein-Infeld-Hoffmann
 - einbein, 389–390
 - Einstein
 - aether, 486
 - equivalence principle (EEP), 462, 464, 485, *see* EEP
 - field equations, 26, 98, 125, 128, 129, 132, 139, 151, 153–154, 168, 173, 181, 184, 200–201, 324, 423, 574
 - linearized, *see* linearized
 - relaxed, 184
 - frame, 469
 - tensor, 90–91, 129, 269, 283, 388
 - second-order, 269
 - Einstein-Hilbert
 - Lagrangian, 514
 - theory, 533
 - Einstein-Infeld-Hoffmann (EIH), 158
 - action, 467
 - surface integral, 158, 169, 214
 - Einstein-Rosen bridge, 179, 180
 - electric
 - and magnetic field, 55
 - charge, 2–4, 49, 51, 121, 279, 309, 413
 - constant, 70
 - current, 76
 - field, 36, 49, 51, 52, 536
 - parity, 269
 - electrodynamics, 70, 253, 322, 368, 377, 391
 - electromagnetic
 - Abraham-Lorentz equation, 49
 - Abraham-Lorentz vector, 504
 - case, 259, 270, 325, 382, 386
 - DeWitt and Brehme, 259, 311, 379, 583
 - DeWitt-Brehme, 305
 - DeWitt-Brehme-Hobbs result, 386
 - Lorentz-Dirac equation, 368, 373, 377, 391

- Lorentz-Dirac force, 380, 382–383
 - radiation reaction, 277, 279
 - self-force, 255, 272, 291, 334, 583
 - electromagnetism, 4–6, 51, 266, 270, 563
 - in Minkowski spacetime, 266
 - electron mass, 71, 79
 - electrostatics, 279
 - electroweak interactions, 5
 - elementary particles, 1–3, 8, 21, 53
 - elements
 - caesium, 68, 84
 - deuterium, 79
 - hydrogen, *see* hydrogen
 - krypton, 69
 - silicon, 79, 81–82, 84
 - elliptic regularity, 256
 - EMRI, 301, 327, 347, 355, 361, 395, 411–413, 416, 443, 449–450, 582, 593, 597,
 - see* extreme mass ratio inspiral
 - and IMRI, 449–451, 456
 - detection of, 327, 396, 451–452, 454
 - observations, 456
 - Signal Characteristics, 452–454
 - waveforms, 327, 408, 455
 - energy flux, 105, 150, 157, 187, 206, 215, 226–230, 233, 236–238, 243–244, 246–248, 591, 597
 - energy in General Relativity, 91
 - absorbed by the black hole, 578, 592
 - Bondi energy, 105, 107
 - Bondi mass, 107, 115
 - energy-momentum conservation, 367, 392, 515
 - energy-momentum tensor, 33, 88–91, 103, 109–110, 352, 369, 371–2, 376, 389, 476, 482, 493–494, 498–499, 505, 508, 510, 512, 514–518, 575
 - ADM and Bondi-Sachs, 122
 - canonical, 91
 - general relativity, 88, 121
 - Landau-Lifshitz, 91, 201
 - entropy conservation law, 170
 - EOB (effective one body)
 - approach, 212–213, 215–216, 223, 238
 - dynamics, 224, 230, 238–242
 - plunge phase, 239
 - effective horizon, 221
 - Hamiltonian, 213, 224, 231
 - late-inspiral dynamics, 248
 - plunge signal, 213
 - method, 212, 216, 219, 224, 249, 597
 - metric, 225, 247
 - EOM, 408, *see* equation of motion
 - ephemeris (ephemerides), 43, 525
 - epicycles, 304
 - equation of motion, 11, 25, 28, 33–34, 46, 49–51, 54–57, 60, 126–128, 150, 153, 156–159, 162–163, 168, 170, 174–175, 177, 181, 184, 186, 192–193, 200–201, 213–214, 216, 224–225, 238–239, 241–242, 254, 259–260, 309, 316, 321–322, 324, 367–368, 370–372, 377, 379, 391–392, 408, 411, 420–421, 425, 431, 451, 453, 465, 510–511, 515–516, 518, 534, 536, 551, 555, 558, 582–583, 596
 - equivalence principle, 41, 49, 76, 389, 462, 464, 471, 475, 479, 485, 491, 493, 514–515, 526–527, 565–566
 - Einstein, *see* Einstein
 - strong, *see* strong
 - weak, 565
 - ESA (European Space Agency), 443
 - Euler-Lagrange equations, 55, 551
 - even-parity, 230–231, 233, 236, 238, 418, 574–575
 - exotic, 46, 427
 - expansion
 - Fourier, 359–360
 - Hadamard, *see* Hadamard
 - higher order, 593
 - Laurent, 133, 151, 155, 380, 420, 497
 - matched asymptotic, 135, 259, 264, 305, 324–325, 328, 332, 583
 - multipole, *see* multipole
 - post-Newtonian (PN), *see* post-Newtonian (PN)
 - Taylor, *see* Taylor
 - extended bodies, 149, 477, 485
 - extensions of the Standard Model, 21
 - extra-dimensions, 378, *see* higher dimensions
 - extreme mass ratio
 - binaries, 416
 - inspiral, 253, 416, 449, *see* EMRI
 - limit, 253, 432
 - extreme situations, 26, 27
 - extrinsic curvature, 93–94, 97–98, 100, 105, 109–110, 112, 114, 173, 184, 254
- F**
- factorized decomposition, 232
 - false alarm rate, 456
 - far zone, 168, 186, 202, 205–206, 259, 267
 - Fermi normal coordinates, 258, 267–268, 285, 309, 313
 - Feynman propagator, 257

- fine-structure constant, 31, 53, 74, 77, 464
 finite difference
 equations, 535–536, 558
 operator, 550, 554
 finite impulse representation, 348
 finite part, 133, 136–139, 151, 156, 344,
 367–368, 379, 387, 419–420, 423,
 425, 427, 583
 finite size
 correction, 261
 effect, 261, 267
 Finsler
 geometry, 57–60
 space-time, 58–59
 first-order, 268, 348, 353, 388, 398, 408,
 410–412, 416, 422, 433
 corrections, 410
 self-force, 411–412
 fixed
 background, 387, 403, 405, 582
 coordinates, 265
 flyby anomaly, 44, *see* Pioneer
 Fock-de Donder approach, 167, 196
 Fokker-Planck equation, 452
 foundations of relativity, 566
 four-momentum, 17, 88, 102–103, 106–107,
 149, 372
 four-vector, 4, 148–149, 477
 four-velocity, 36, 57, 148–149, 171, 255, 263,
 278, 294, 300, 302, 329–330, 337,
 341, 352, 358, 362, 408, 418–419,
 421, 434, 463, 477, 576
 of the worldline, 263
 Fourier
 convolution integral, 254
 expansion, 359–360
 modes, 551
 space, 515
 sum, 350–351, 360
 transform, 254, 299, 454, 496, 546, 548,
 550, 553
 anti-transform, 578
 Fréchet derivative, 169
 frequency domain, 302–303, 327, 334, 343,
 347, 350–351, 395–396, 400,
 496–497, 500–501, 505, 509–510,
 577–578
 function
 cylindrical, 539–540
 delta-function, *see* (delta) δ -function
 Dirac, 137, 148, 150, 154
 puncture, *see* puncture
 Riemann-Hurwitz, *see* Riemann-Hurwitz ζ
 function
 smooth, 132, 256–266, 330, 338, 340, 359,
 544
 step, 241, 317–318, 496
 tail, 382, 386
 window, 281–282, 288, 303
 ζ , *see* (zeta) ζ function
 Functional Representation, 179
 fundamental
 constants, 67–68, 71, 74–78, 532, 541
 interactions, 2–3, 9, 493, 515
 properties, 1, 8, 18
 tests in GR, *see* general relativity
 future light cone, 311, 315, 317, 382
 fuzzy space, 535–536, 546, 550–552, 555, 557
- G**
 gauge
 Burke-Thorne, 192
 confusion, 274–276, 297–298, 568
 de Donder, *see* de Donder gauge
 dependence, 328, 331, 584
 Dirac condition, 101
 harmonic, *see* harmonic gauge
 invariance, 5–6, 421, 467, 577
 Lorenz, *see* Lorenz gauge
 Regge-Wheeler, 296–297, 335, 345, 347,
 403–404, 409, 576–577, 588–589,
 592, 597
 residual freedom, 352, 409
 supplementary conditions, 347
 general relativity, 25, 28, 87–88, 90–92,
 95, 97, 99, 103, 108, 121–122,
 125–128, 153, 156, 158, 167–169,
 171, 173–174, 176–179, 196,
 253–255, 259, 261, 263–265,
 272–276, 282–283, 285, 289, 295,
 305, 324–325, 369, 395, 416, 443,
 449–450, 456–457, 461, 470, 474,
 491–492, 532–533, 535–537, 543,
 562–566, 569, 572
 angular momentum in, 87, 91, 121–122
 approximation in, 125–126, 167
 assumptions of, 259
 binary dynamics in, 178
 black holes in, 179
 canonical, 171, 173, 537
 challenges, 456–457
 constraint equations, 99, 174
 effacing principle, 156
 Einstein theory, 125, 167, 569
 energy, 91
 energy-momentum, 88, 121
 experimental basis, 25

- formulation, 91
 - Hamiltonian, 97, 158
- gauge freedom, 295, 569
- internal consistency, 87
- Isenberg-Wilson-Matthews approach, 179
- isolated systems, 95, 122
- mass, 95
- mathematical, 108
- metric extensions, 491
- motion of bodies, 532–253, 324
- non-linear theory, 169, 325
- numerical, 108
- particle motion, 263
- perturbation theory, 274–275, 282–283
- point particles in, 253–255, 261
 - nonexistence of, 261, 532
- post-Newtonian expansion, 168
- predictions, 125, 416
- quantizing, 533, 536
- quasi-local mass, 122, *see* Quasi-Local
- self-force analysis, 305
- solution space, 92
- symmetries, 153
- tests
 - fundamental, 28, 449–450, 456
 - high precision, 443, 449, 456
 - three-dimensional, 535
 - two body problem, 273, 395, 562
- GEO, 40, 211, 416
- geodesic, 96, 106, 216, 218, 220, 253, 255, 257–258, 261–264, 267, 269–274, 283–285, 287, 289–292, 294–295, 299–300, 305, 309–313, 315, 327–330, 333–335, 338–339, 347, 350, 352, 354–355, 361, 367, 377, 381, 386–389, 398, 401–405, 411, 413, 420–422, 450, 454, 482, 522, 532, 534, 561–562, 565, 573, 575–576, 581–582, 584, 587–588, 593–594, 596
- background, *see* background
- deviation, 261, 263, 269, 591, 593–595
- distance, 257, 311–312, 329–330, 358, 406
- equation, 28, 35–36, 40, 58–59, 260, 292, 294, 297–298, 389, 421, 454, 582, 587
- motion, 181, 253, 255–256, 260–262, 263, 266–267, 269, 271–273, 292–294, 299, 305–306, 386, 403, 515, 566, 593
- non-geodesic, *see* non-geodesic
- post-geodesic, 361
- geometrized units, 329
- geometry, 25–26, 35, 70, 109, 113, 116, 180, 185, 271–274, 283, 327, 337, 418, 449–450, 463, 476, 547, 597
 - background, 285–286, 294, 331, 409, 420, 575, 585
 - black hole, 272, 304, 344, 453
 - Finsler, 57–60
 - Kerr, 344, 397, 450, 563
 - noncommutative, 26, 45, 536, 544
 - perturbed, 276, 331, 595
 - quantum, 533, 535–536, 539, 545–546
 - Riemannian, 28
 - Schwarzschild, 276, 291, 295–298, 300–301, 303, 361, 421, 575
 - spacetime, 25–27, 29, 216, 272, 280, 305, 453
- Geroch energy, 112, 114, 119
- globally hyperbolic spacetimes, 113, 318
- gravitating body, 27, 30, 31, 34, 38, 40–41, 479
- gravitational
 - constant, 44, 50, 59, 101, 154, 168, 423, 469, 471, 491, 514, 517–519, 527, 532
 - time-dependent, 44
 - energy, 41, 87, 90–92, 94, 122
 - fields, propagation of, 515
 - Hamiltonian, 99
 - redshift, 32, 34–36, 464
 - self-force, 255, 259, 261, 263–264, 269, 272–277, 284, 291–292, 294, 297–300, 302–306, 309–311, 324–325, 327, 335, 352, 401, 403–404, 408, 413, 416, 422, 453, 584
 - effect, 263, 298, 306
 - time delay, 38, 40
- gravitational wave, 47, 49, 107, 121–122, 125–128, 157, 161–162, 164, 169, 181, 187–188, 202, 211, 272, 282, 295, 302–303, 305, 327, 396, 416, 443–445, 448–450, 454–456, 464, 471, 476, 483, 529, 569, 575, 582, 584, 597
- amplitude, 129
- astronomy, 283, 367
- detection, 413, 445
- emission, 206, 453
- emitted, 125–126, 162, 169, 188, 574–575
- flux, 128, 159–160, 163
- interferometer, 47, 56
- interferometry, 561
- phase, 164
- tail, 135, 415, 418, 427–428

- templates, *see* templates
- waveform, 207–209, 278, 301, 328, 367, 395, 397, 399, 403, 410, 412–413, 562
- gravitomagnetic, 40–41, *see* Lense-Thirring effect
- graviton, 3, 18, 22, 461–462, 467–469, 471, 479, 482–483, 516
 - propagator, 516
- Green's function, 135, 154, 256–257, 259–260, 273, 305, 309, 332, 339, 373, 378, 380, 382–383, 402, 420, 465, 561, 584, 586, 594
 - advanced, 260, 318
 - Detweiler-Whiting, 260
 - regular, 318
 - singular, 318
 - radiative, 328, 333, 561, 583
 - retarded, 135, 256–258, 264, 316–318, 329, 378, 381–382, 384, 583, 594
 - symmetric, 260
 - tail, 401
- H**
- Haar measure, 540, 545
- Hadamard, 383, 433, 584
 - decomposition, 317
 - disaster, 570
 - expansion, 256–260, 268, 332, 380–381, 420, 583
 - form, 258–259
 - formula, 258
 - partie finie, 151–153, 183, 214
 - regularization, 150–153, 155–156, 158, 175, 196, 213–214, 420–421
 - pure Hadamard-Schwartz, 156
 - regularized 2PN metric, 433
 - series, 257–258
- Hall effect, 76
- Hamilton-Jacobi
 - equation, 109, 218, 224
 - principal function, 109
- Hamiltonian
 - analysis, 107, 109, 116
 - Delaunay, 217–218
 - effective, 224, 231
 - expression, 102, 104
 - formalism, 158, 162, 184, 190
 - formulation, 97, 108
 - of spinning binaries, 167
- harmonic, 164, 184, 327, 332, 344, 346, 348, 351, 353, 399, 450, 546
 - coordinates, *see* coordinates
 - gauge, 129, 140, 184, 268, 369, 421, 467, 583
- Hawking
 - energy, 111–112
 - mass, 119
- Hayward Energy, 112
- Heaviside step (distribution) function, 241, 496, 581, 589, *see* step
- helical
 - Killing vector, 418, 421, 428, 429
 - symmetry, 418–419, 428, 431, 432
- hereditary, integral, 146, 428
- Higgs
 - background events, 16–18
 - boson, 1, 8–13, 15–22
 - coupling, 9, 12, 14–15, 19–20
 - decay, 12, 16–18
 - decay modes, 12, 17
 - detection of, 17
 - mechanism, 1, 5, 7, 18–20
 - particle, *see* particle
 - characteristics, 8
 - production, 14–16, 19
- high-accuracy
 - comparison, 125, 416
 - length measurements, 69
 - waveform, 244
- higher
 - dimensional theories, 26, 48, 52
 - dimensions, 378–380, 385, 391
 - order
 - corrections, 14–15, 19, 261, 270, 410
 - Equation of Motion, 55–56
 - expansion, 593
- Hilbert space, 114, 497, 512, 539–542
- homogeneous
 - equation, 258, 578
 - wave equation, 293, 318, 371, 380, 403, 579, 583
- Hopf algebra, 544, 550
- Hubble constant, 48, 54
- Hulse-Taylor binary pulsar, 126, 473–474
- Huygens principle, 378
- hydrogen, 30, 32, 36, 68, 74, 79, 464, 533
 - atom, 533
 - maser, 32, 36, 464
 - molecule, 74
- hyperfine
 - levels, 84
 - splitting, 32, 68
 - transitions, 53
- hyperplane, 374–375, 538

- hypersurface
 Cauchy, 113, 352
 null, 106, 115–116
 space-like, 88, 109, 111–112, 184, 372, 374, 376
 time-like, 98, 254
- I**
- IH, 115–116, 118, 120, *see* isolated horizon
- Immirzi
 ambiguity, 537
 parameter, 537, 541
- IMRI (intermediate mass ratio inspiral), 449–451, 456
- incoming radiation, 126, 130–131, 135–138, 202, 259, 263, 268, 476
- incompressible fluid, 289
- inertial law, 46, 49, 54
- infinite part, 419–420
- infinite-dimensional, 97, 106, 117, 553
- inhomogeneous
 equation, 382
 wave equation, 202, 403, 578
- initial data, 103, 113, 120–121, 212, 238–239, 241, 253, 353–354, 578
- initial value, 178, 181, 253, 270, 352–353
 formulation, well-posed, 270
- innermost stable circular orbit (ISCO), 190, 413
- inspiral, 49, 127, 158, 212, 215, 239, 241, 243, 245–246, 248, 253, 269, 278–279, 328, 345, 361, 367, 396, 412, 418, 449, 453–454, 456, 582
 adiabatic, 120
 black hole, *see* black hole
 compact binaries, 125–128, 147, 156–157, 416
 EMRI, *see* extreme mass ratio inspiral
 IMRI, 449–451, 456
 late-inspiral dynamics, 248
 phase, 212, 215, 450
 quasi-circular, 120, 214, 225, 239, 298, 418
 templates, 125, 328
 waveforms, 235, 238–239, 328, 243
- integro-differential equation, 131
- internal states of motion, 253
- International System of Units (SI), 67–84
- interpretation, 42–43, 52–53, 60, 79, 89–90, 99, 104, 260, 267, 277, 294–295, 297, 328, 330, 377, 391, 480, 506, 514, 540, 545, 551, 554, 566
 artifact, 480
- irregularity of the field variables, 347
- Isenberg-Wilson-Mathews approach to general relativity, 179
- isolated horizon (IH), 115–116, *see* IH
- J**
- Jacobian, 342
- Josephson
 constant, 76
 effect, 75, 80
- jump discontinuity, 254, 350
- justification, 107, 182, 258, 592
- K**
- k-Essence, 481
- kelvin, 68, 75, 84
- Keplerian orbit, 453
- Kerr geometry, 344, 397, 450, 563
 background, 329, 454
 black hole, 121, 190, 212, 293, 327–328, 398
 metric, 35, 298, 362, 387
 solution, 35
 spacetime, 329, 337, 349, 360–361, 369
- Kerr-de Sitter space-time, 35
- Kerr-Newman, 334
- Kijowski-Liu-Yau
 energy, 111
 expressions, 110
- Killing vector field, 36, 94, 418
- kilogram
 artifact, 67, 69, 71–72, 74–75, 77–79, 83–84
 international prototype, 67, 69, 79
- Kinnersley tetrad, 398
- Kirchhoff formula, 130
- Klein-Gordon, 533
- Komar
 angular momentum, 95, 105, 114, 120
 mass, 94–95, 101, 103, 107, 120
- Kramers-Kronig relations, 496, 499, 505
- krypton, 69
 emission line, 69
- Kubo-Martin-Schwinger (KMS) relations, 497, 500, 506
- L**
- LAGEOS satellites, 40
- Lagrange, 50, 55, 193, 389, 551
- Lagrangian, 4, 6–8, 55, 88, 90, 98, 108, 158, 185, 192, 378, 481–482, 494–495, 498, 508–509, 514–516, 527, 551, 571
 density, 4, 6–7, 98

- Lamb shift, 492
- Landau o symbol, 151, 431
- Landau-Lifshitz (Lifshits), 201, 283, 571, 573
 - pseudo-tensor, 91, 201, 283
- Laplace equation, 134, 256, 257
 - generalized, 257
- lapse and shift, 102, 193
- Large Hadron Collider (LHC), 1, *see* LHC
- Laser Interferometer Space Antenna (LISA), 582, *see* LISA
- last stable orbit (LSO), 216, *see* LSO
- late-inspiral dynamics, 248
- Laurent expansion, *see* expansion
- leading order corrections, 267, 269, 411
- Legendre transformation, 98
- length scale, 48, 133, 146, 274, 278–283, 285, 292, 304, 324, 347, 423, 429, 468, 493, 516–518, 520, 522, 527
- Lense-Thirring effect, 34, 40–41, 49
- Leonardo da Vinci, 564
- Levi-Civita
 - antisymmetric symbol, 429
 - connection, 92–93, 116
 - tensor, 274, 286
- Lewandowski measure, 540
- LHC, 1–2, 8–9, 13–22, 26, 47, *see* Large Hadron Collider
- light deflection, 37, 469, 482–483, 485, 491, 521, 525
- LIGO (Laser Interferometer Gravitational-wave Observatory), 47, 211, 416, 475
- limit
 - extreme mass ratio, 253, 432
 - linearized, 519
 - Newtonian, 101, 218, 300, 463
 - ordinary, 265
 - particle, *see* particle
 - point particle, 253, 255, 261, 264
 - scaled, 265
 - singular, 152
 - test-particle, 191, 221
- limiting worldline, 263
- linear
 - equation, 254, 533
 - wave, 256, 310, 316
 - response, 492–493, 496, 499, 502, 507–508, 510, 516, 526
- linearized
 - approximation, 140, 517
 - Bianchi identity, 255, 267
 - Einstein equation, 255–256, 258, 260, 264, 267, 269–270, 310, 352, 369, 397, 400, 405
 - field equations, 575
 - relaxed, 256, 258, 260
 - gravity, 369, 379, 387
 - limit, 519
 - metric, 142
 - theory, 575
 - vacuum Einstein field equations, 140
- LISA, 327–328, 347, 361, 396, 416, 443–456, 475, *see* Laser Interferometer Space Antenna
 - confusion noise, 452, 454
 - data analysis, 453–457
 - EMRI sources, 396
 - initially estimated candidates, 456
 - frequency bandwidth, 416
 - mission, 443–444, 452
 - Mock Data Challenge, 454
 - templates, 597
- locally inertial coordinates, 274, 283, 285–287, 305, 420, 565
- logarithms
 - SPN, 427, 429–432, 437
 - leading, 230, 232
 - powers of, 427
- loop quantum gravity, 26, 45, 114
- Lorentz, 30, 40, 97, 102, 106, 126, 195, 257, 368, 370, 504, 507, 510–511, 537, 570
 - and Abraham, 324
 - covariant, 28, 167, 200–206
 - force, 49, 278
 - gauge, *see* Lorenz gauge
 - group, 106, 537
 - invariance, 28, 30, 46, 57, 464, 468, 483, 485, 503, 507, 513, 534
- Lorentz-Dirac
 - equation, 368, 373, 377, 391
 - force, 380, 382–383
- Lorentzian, 89, 115–116, 196, 256–258, 260, 534
 - case, 256–258
- Lorenz gauge, 256, 261, 264, 268–269, 296–298, 327, 329–331, 333, 335–336, 339, 343–349, 351–354, 356–357, 397, 400, 404–405, 409, 413, 583
 - condition, 256, 264
 - relaxation, 256, 259, 261, 583
- LSO, 216, 221–222, 224, 226–229, 235–236, 384, 397
- lunar laser ranging, 474

M

- m-mode regularization, 327, 333, 356, 358–360, 395, *see* mode-sum
- Mandelstam's invariant, 220
- manifold, 32, 99, 106, 115–117, 276, 295–296, 299, 512, 515, 532, 535, 537–538, 544, 550, 575
- mass dipole, 147, 267–268
- Mass Metrology, 67–84
- mass-quadrupole tensor, 192, 205
- massless
 - field, 378, 380, 463, 512
 - particle, 29, 224
 - photon, 2–7
- matched asymptotic expansions, *see* expansion
- matter
 - distribution, 131–132, 139, 147, 151, 464
 - source, 126, 151, 168, 202, 254, 424, 428–429, 465–466
- maximal slicing, 176, 184
- Maxwell
 - equation, 28, 30, 46, 52, 56, 254, 277–278, 310, 370
 - field, 36, 368–370, 390
- metric
 - background, *see* background
 - barred, 265–266
 - coefficient, 143, 150–151, 176, 184, 194, 201, 203, 218, 220
 - extensions, 491, 494, 520–525, 527
 - extensions of GR, 494, 521–525, 527
 - linearized, 142
 - scaled, 268
 - Schwarzschild, *see* Schwarzschild
 - Schwarzschild-deSitter, 264
 - unbarred, 266
- Metropolis-Hastings algorithm, 456
- Minkowski spacetime, 88, 96, 111, 259, 266, 369
 - auxiliary metric, 129, 423
- Minkowskian, 57, 89, 102, 129, 130, 336, 515
- MiSaTaQuWa
 - equation, 260–263, 270, 339, 390
 - formula, 327–31, 333–335, 339
 - framework, 327, 329, 333, 343, 583–584
- MiSaTaQuWa-DeWh, 584–585
- Misner-Lindquist solution, 175
- Misner-Sharp energy, 109, 111
- mode-sum, 299, 327, 330, 332–334, 335–343, 344–347, 349–352, 355, 358–359, 361, 401, 403, 561, 586–587, 590, 592, 595–596
 - regularization, 294, 299, 303, 337, 361, 401–402, 404
- modifications of Newtonian dynamics (MOND), 461, *see* MOND
- mole, 70, 75, 78, 84
- molecular clocks, 32
- Moncrief, 347–348, 577
- MOND (modified Newtonian dynamics), 53
 - acceleration scales, 48
 - ansatz, 42, 53–54
 - assumption, 480
 - dynamics, 481, 484–485
 - phenomenology, 482, 486
- Monte Carlo, 434, 452
- Moon, orbit around the Earth, 522
- motion
 - background, 581, 595
 - Brownian, 492, 502
 - circular, 59, 190–191, 428
 - equation of, *see* equation of motion
 - Higher Order, 55–56
 - post-Newtonian, 162
 - Quinn's, 321
 - geodesic, *see* geodesic
 - in general relativity, description of, 532
 - in quantum gravity, 532
 - internal states of, 253
 - non-geodesic, 256, 271, 367, 390–392
 - particle, 263, 389, 567, 578
 - radial, 59, 569
- motivation, 42, 69, 71, 213, 235, 427
- moving puncture, 185
- multiplicative decomposition, 216, 227, 232, 241, 244
- multipolar waveforms, 236–237
- multipole
 - decomposition, 136, 142, 144, 233, 303, 336, 356, 358
 - expansion, 135–137, 139, 167, 205, 332, 336, 338, 343
 - moments, 137, 139–142, 144–145, 147, 156–157, 204, 214, 272, 286–287, 294, 324, 430

N

- NASA (National Aeronautics and Space Administration), 243, 413, 443
- natural units, 71, 73
- navigation around the Earth, 511
- near zone, 126, 131–132, 135–137, 141, 168, 202, 206, 214, 232, 259, 267, 269, 584

- NEH, 115–116, *see* non-expanding horizons
- neighborhood, 106, 130, 133, 266, 271, 273–274, 277–281, 283, 285, 287–288, 291–292, 299, 304–305, 309, 312–313, 315, 322, 403, 419, 420, 517, 582
- normal, 257–258, 260
- of GR, 517, 521, 523, 527
- sufficiently small, 257–258
- neutron stars, 126, 128, 147, 179, 249, 253, 416, 443, 448–449, 471–472, 475–476
- binary, *see* binary
- tidal properties of, *see* tidal
- Newman–Penrose
- gauge-invariant scalars, 344, 346
- scalar, 332, 351
- Newman–Unti energy, 112
- Newtonian
- back-action, 564–569
- constant of gravitation, 71
- Gravity in Canonical Form, 169
- limit, 101, 218, 300, 463
- physics, 289, 484, 561, 585
- misconception, 568–569
- potential, 30, 32, 48, 334–335, 426–427, 464
- self-force, *see* self-force
- system, 275
- tides, 289
- no-hair theorem, 461, 476, 486
- Noether
- current, 88
- theorem, 90, 498
- non-adiabatic, 561–562
- late-inspiral dynamics, 248
- non-expanding horizons (NEH), *see* NEH
- non-geodesic, 255–256, 271, 390–392
- motion, *see* motion
- non-gravitational force, 272, 290, 411
- non-local, 554, 569, 571, 582, 584
- measurement, 571
- Gravitational Energy, 91–92
- non-locality of measurements, 569
- non-radiative modes, 398, 563
- non-vacuum background, 369, 391
- nonanalytic case, 257
- noncommutative
- algebra, 548
- fuzzy, 551, 558
- geometry, 26, 45, 536, 544
- space, 535–536, 543–545, 549–550
- nonlinear
- equation, 254, 533
- radiation-reaction, 139
- regime, 266
- series, 140
- normal neighborhood, 257–258, 260
- null hypersurface, 106, 115–116
- numerical
- relativity, 92, 144, 185, 211, 274, 278, 293, 295, 302, 304, 395–396, 453, 596
- solution, 349–350, 356, 358, 399, 405, 578–581
- waveforms, *see* waveforms
- O**
- odd-parity, 230–231, 233, 236, 238, 418, 574
- one-parameter family, 261, 263–266, 268, 270, 328
- orbit, 32, 37, 41, 44, 53, 59–60, 95, 103, 126–127, 157–164, 182–183, 188–190, 192, 206–208, 211–213, 215–217, 219, 224, 227, 231–232, 235–236, 269, 272–277, 284, 293, 295, 297, 298, 300–301, 304–305, 327–328, 333, 339, 343, 345–352, 354–355, 357–358, 361–362, 367, 395–404, 406, 408–413, 417–419, 422, 427–428, 432, 434, 440, 443–444, 448–450, 452–457, 471–473, 476, 479, 521–522, 562–563, 566, 570, 590–591, 593, 596–597
- adiabatic evolution, 328
- circular, *see* circular orbit
- eccentric, *see* eccentric orbit
- ISCO, *see* innermost stable circular orbit
- Keplerian, 453
- last stable, *see* LSO
- Moon around the Earth, 522
- quasi-circular, *see* quasi-circular orbit
- orbital
- frequency, 157–160, 163–164, 190, 224, 239, 241–242, 245, 276–277, 293, 297, 300, 403, 409, 415, 418, 422, 427–428, 430–431
- phase, 157, 160–162, 164, 208, 242
- ordinary differential equation, 34, 211, 257–258, 350, 408
- ordinary limit, 265
- P**
- Padé
- approximant, 216, 221, 223–224, 227–228
- resummation, 213, 222, 227–229, 237, 241

- Papapetrou equation, 148
parallel propagator, 309, 311–312
parallelly propagated, 269
parametrix, 257
partial
 derivative, 103, 134, 169, 183, 198, 285, 286
 differential equation, 254
 wave, 232, 237
particle
 charged, *see* charged particle
 detection of dark matter, 42
 elementary, 1–3, 8, 21, 53
 Higgs, 2, 8–11, 13–22, 47
 limit, 191, 221, 253, 255, 261, 264, 331, 337–338, 344–345, 578, 586
 point, 253, 255, 261, 264
 test-particle, 191, 221
 masses, generation of, 1, 5, 18, 22
 massless, 29, 224
 motion, 263, 389, 567, 578
 point, *see* point particle
 source, 255, 258, 264
 spinning, 31, 128, 148, 162
PCT symmetry, 30
Penrose
 conjecture, 113, 119
 inequality, 112, 119
periastron
 advance, 189, 473
 precession, 413
 shift, 37
perturbation
 black hole, *see* black hole
 equation, 270, 327, 330, 333, 343–344, 346–347, 349, 352, 356–357, 589, 593–595
 series, 11
 theory, 74, 160, 226, 268, 270, 274–275, 283, 285, 287, 295, 298, 329, 343, 396, 416–417, 433, 453, 582, 596
perturbative
 correction, 261, 263, 269, 584
 equation, 262–263, 270
 self-consistent, 263, 270
 result, 261–263, 270
phase space, 103, 109, 116–117, 224, 239, 454, 538–539
phoney, 264, 270
photon
 field, 4
 mass, 5, 8, 29, 52
 massless, 2–3, 6–7
 speed (velocity), 29, 483
 zero-mass, 4–5
physical property, 68, 72, 90, 103
Pioneer
 acceleration, 48, 54
 anomaly, 43, 44, 48, 54, 60, 462, 485–486, 493, 523–524
 spacecraft, 462, 485
Pisa, leaning tower of, 561, 564, 569, 597
Planck, 11, 20–22, 45, 47–48, 67, 71, 73–74, 79, 83–84, 452, 495, 514, 531–533, 535, 538, 542–545, 552, 557–558
 constant, 67, 71, 74, 79, 83, 495, 514, 531
 energy, 45
 length, 21, 45, 48, 535, 543, 545
 mass, 22, 84, 552
 regime, 533
 scale, 11, 20–21, 47, 531–533, 538, 542, 545, 558
 time, 45, 84
 units, 71, 73, 557
Plebański formulation of gravity, 542
Plebański-Demiański space-times, 35
plunge, 212–213, 215–216, 224, 227, 235, 238–241, 245–246, 248, 328, 354, 455, 591–593, 596
Poincaré, 21, 88, 97, 102–103, 106, 167, 195–196, 198, 504, 510–511, 513, 537, 547
 algebra, 167, 195–196, 198, 511
 Casimirs, 88
 group, 21, 97, 102, 106, 195, 504, 510, 537, 547
 invariant, 103, 513
 symmetry, 97, 537
 transformation, 88
point charge, 254, 278, 302, 315, 369
point particle, 28, 30, 33, 57, 127–128, 148, 150, 178, 181–182, 253, 255, 258–261, 264, 291, 309, 316–317, 319, 321–322, 324–325, 328, 345, 347, 352, 367, 369, 387, 396, 403, 405, 417, 421, 425–426, 462–463, 477–478, 480, 531, 535, 539, 574–576, 582
 in general relativity, 253–255, 261, 263, 532
 limit, 253, 255, 261, 264
 source, 255, 258, 264
Poisson
 bracket, 103, 117, 170–173, 539
 equation, 48, 101, 131–133, 426
 integral, 131–133, 136, 152, 155–156, 425, 430

- polarization, 29, 143, 160–161, 207–208, 447, 522
 - waveforms, *see* waveforms
 - position vector, 173, 195
 - post-adiabatic, 397, 411
 - Self-Force, 408
 - post-Keplerian, 471, 478–479
 - post-Minkowskian (PM), 135, 140, 142, 145, 214, 429, 466
 - series, 142, 168
 - post-Newtonian (PN)
 - and EOB methods, 563
 - ansatz, 132
 - approach, 121
 - approximation, 37, 125–128, 131, 135, 151, 415, 453
 - equations of motion, 162
 - expansion, 131–132, 134–139, 145, 160, 167, 300–301, 453, 593
 - framework, 332
 - perihelion shift, 37
 - series, 126, 134, 168, 185, 202, 216, 233–234, 439
 - formal, 132, 142
 - skeleton Hamiltonian, 177–178
 - source, 135–139
 - templates, 416
 - waveforms, 147, 160
 - pragmatic approach, 587, 594
 - pre-Galilean physics, 564
 - principle of equivalence, *see* equivalence principle
 - probability distribution function, 396
 - propagation
 - of gravitation fields, 515
 - of singularities, 258
 - propagator
 - Feynman, 257
 - graviton, 516
 - parallel, 309, 311–312
 - retarded, 498
 - proper time, 33, 40, 59, 148, 255, 285, 287, 294, 313, 315, 319–320, 329, 331, 370, 374–376, 379, 382, 402–403, 408, 464, 504, 573, 591, 595–596
 - derivative, 319–320
 - parameter, 313, 315
 - pseudo-tensor, 91, 129, 131, 133, 136–139, 142, 574, *see* Landau-Lifshitz
 - energy-momentum, 91, 201
 - stress-energy, 132, 201
 - Ptolemy, 304
 - pulsars, 49, 128, 189, 472–475, 483
 - binary, *see* binary
 - puncture
 - function, 349, 357–358, 360, 407
 - method, 350, 358
- Q**
- QFT, 493–494, 496, 520, 526, 537, 542, *see* quantum field theory
 - QHE, 76, 80, *see* quantized Hall effect
 - quadrupole
 - formula, 126, 569, 572, 574
 - moment, 145–147, 167, 214, 289–290, 391, 428, 430, 573
 - perturbation, 290
 - tensor, 192, 198, 205
 - quantization, 45, 219, 532–539, 544
 - quantized Hall effect, 76, *see* QHE
 - quantum
 - correlations, 497
 - dynamics, 550
 - field theory, 21, 512, 543, *see* QFT
 - fluctuations, 9, 11–12, 14–15, 491, 493–494, 498, 508–509, 514–517, 526–527
 - geometry, *see* geometry
 - gravity, 26, 45–47, 114, 532–536, 540, 542–544, 557–558
 - mechanics, 27, 217, 533
 - metrology triangle, 77
 - quasi-circular orbit, 103, 157–158, 160, 297–298, 408, 418
 - quasi-equilibrium, 103, 115–116
 - Quasi-Local
 - Angular Momentum, 114
 - energy, 109, 111–112
 - Brown-York, 109–111, 114
 - Geroch, 112
 - Hawking, 111
 - Hayward, 112
 - Kijowski-Liu-Yau, 111
 - Misner-Sharp, 109, 111
 - Newman-Unti, 112
 - mass, 109, 111, 113–114, 118–122
 - Bartnik, 113
 - Hawking, 119
 - quasi-normal modes, 241, 282
 - Quinn's equation, 323
 - of motion, 321
 - Quinn-Wald axiom, 322
 - quintessence, 31

R

- radial fall, 404, 562–563, 566, 570, 576, 582, 587, 589–590, 592, 595–597
 - controversy, 569, 573
 - direction (impressed), 591
 - entire history, 562
 - Historical Heritage, 563–564
 - local measurements, 569, 571
 - pragmatic approach, 587, 594
 - repulsion, 561, 569–573, 591, 596
 - self-consistent prescription, 592–596
 - radial infall, 562
 - radial motion, 59, 569
 - radiated energy, 107, 212, 596
 - radiation reaction, 126–127, 134–137, 139, 157–159, 161, 168, 204, 213–216, 223–225, 227, 236, 238, 240–241, 243–246, 253, 271–274, 277–279, 293, 305, 322, 328, 367, 369, 378, 387, 389–391, 398, 412, 418, 427–429, 431, 501–504, 511, 562–563, 569, 582–584, 590, 592–593, 596
 - damping, 427
 - dynamics, 168
 - effect, 139, 157–158, 398, 418, 428, 431
 - force, 126–127, 139, 213, 216, 224, 227, 236, 240–241, 243, 253, 272, 277–278, 390–391, 412, 418, 428, 501–504, 511
 - functions, 135–137
 - timescale, 562, 596
 - radiative
 - field, 583–584
 - Green's function, 328, 333, 561, 583
 - moments, 143, 145–147
 - recurrent relation, 378, 382, 385
 - recursion relation, 257, 258
 - redshift observable, 416, 419, 421, 432, 433, 434, 440
 - Regge-Wheeler
 - equation, 298, 574
 - gauge, *see* gauge
 - self-force, 346
 - Regge-Wheeler-Zerilli, 231, 284, 575
 - regular
 - field, 280, 293, 299, 302, 319–321, 323, 403, 405–406, 583
 - piece, 309, 403
 - remainder, 281, 319, 323
 - regularization, 582, 586, 590, 592, 595–596
 - dimensional, *see* dimensional
 - field, 299
 - Hadamard, *see* Hadamard
 - m-mode, *see* m-mode regularization
 - mode-sum, *see* mode-sum
 - parameters, 586, 592
 - point-splitting, 373, 383, 385
 - pure Hadamard-Schwartz, 156
 - Riemann-Hurwitz ζ -function, *see* Riemann-Hurwitz ζ -function
 - self-field, 128, 150–151, 158
 - relativity principle, 28, 29, 44
 - relaxation, *see* de Donder, Lorenz gauges
 - relaxed
 - (Einstein) field equations, 139, 184, 201, 214, 423
 - linearized Einstein equation, 256, 258, 260
 - repulsion, *see* radial fall
 - residual degree of freedom, 8, 99
 - residual field, 349–350, 357, 360, 407
 - rest frame, 45, 149, 315, 321, 368, 377
 - resummation method, 215–216, 227, 248
 - resummed (expression), 212, 222, 224, 227–230, 232, 236–237, 240, 243–244, 248
 - retarded
 - Green's function, 135, 583, 594
 - integral, 131, 135–136, 138, 185
 - propagator, 498
 - Riemann
 - curvature, 514, 516, 565
 - tensor, 149, 271–272, 284, 286–287, 313, 320, 331, 352, 384, 391, 477, 565, 573, 585, 595
 - Riemann-Hurwitz ζ -function, 561, 587, 589, 590, 595
 - Riemannian
 - background metric, 96
 - case, 119, 256
 - geometry, 28
 - metric, 28, 32, 256
 - space-time, 28, 58
 - ring-down, 125, 212, 241, 248
 - ringing tail, 213
 - Rohrlich analysis, 368
 - Rosenfeld procedure, 91
 - rotational deformation, 196, 198
 - Routh functional, 174, 185
 - Rydberg constant, 79
- S**
- scalar wave equation, 310
 - scalar-tensor
 - gravity, 468, 476, 478–479
 - theories, 461–462, 467–469, 471, 473, 476–479, 481, 483, 485–486

- scaled
 - limit, 265
 - metric, 268
- scattering matrix, 495, 500–503, 505–506, 508–509
- Schiff effect, 34, 41
- Schott term, 248, 367–369, 373–374, 376–377, 379, 387, 391–392
- Schwarzschild
 - black hole, 181, 265, 284, 290, 297, 327, 352, 357, 361, 405, 408, 422
 - coordinates, 147, 179, 190–191, 335, 408, 420–421, 519, 521
 - deformed, 188
 - geometry, *see* geometry
 - metric, 34–35, 179, 184, 188, 190, 220, 265, 288, 290, 296, 298, 303, 386, 419–420, 422, 483
 - singularity, 179–180, 185
 - solution, 485, 570
 - spacetime, 111, 119, 178, 190, 236, 290, 328, 333
- Schwarzschild-de Sitter
 - metric, 264
 - solution, 35
- Schwarzschild-Droste (SD), 569–571, 573–576, 588, 595, *see* Schwarzschild
 - black hole, 561
 - metric, 569
- second order, 49, 55–56, 129, 268–269, 283–285, 292, 298, 348, 388, 395, 397, 408, 410–412, 433, 551, 574, 581, 588–589, 593
 - corrections, 410
 - deviation, 588
 - Einstein tensor, 269
- secular, 44, 126, 270, 412, 524, 562, 584
- self-acceleration, 50–52, 410, 566
- self-adjoint, 258, 260, 512
 - wave equation, 260
- self-consistent, 227, 254, 261, 263, 270, 296, 396, 499, 502, 509, 561–562, 592–596
 - approach, 561–562, 593
 - equations, 269
 - method, 296, 596
 - perturbation, 263, 270
 - perturbative equation, 263, 270
 - radial fall prescription, 592–596
- self-coupling, 7, 10–11, 19–20
- self-field, 127–128, 150–151, 158, 269, 309, 415, 417, 426
 - divergent, 415, 417, 426
 - regularization, 128, 150–151, 158
- self-force (SF), 259, 264, 270, 273–275, 277, 280–282, 285, 291–295, 297–305, 309–313, 316, 321–324, 327, 344, 356, 378, 395, 397–398, 401–404, 407–408, 410–413, 415–416, 440, 453, 532, 561–563, 582–586, 590–597
 - adiabatic, 412
 - post-adiabatic, 408
 - corrections, 253, 259, 261
 - effect, 253, 255, 261, 263, 269, 275–277, 294–295, 298, 300, 306, 412–413
 - gravitational, 263, 298, 306
 - electromagnetic, *see* electromagnetic
 - first-order, 411–412
 - gravitational, *see* gravitational
 - Newtonian, 273, 275, 297, 564, 584
 - problem, 255, 273, 282, 285, 291, 300, 310, 325
 - Regge-Wheeler, 346
 - term, 261, 594
- self-interaction, 6, 9, 151, 383, 468, 552
- semi-latus rectum, 352, 355, 593
- series
 - Hadamard, 257–258
 - nonlinear, 140
 - perturbation, 11
 - post-Minkowskian, 142, 168
 - post-Newtonian, *see* post-Newtonian
 - formal, 132, 142
 - Taylor, 227, 266
 - finite, 266
- SF (self-force), *see* self-force
- shell, 174–175, 254, 321–322, 334, 350–351, 372, 536
 - matter, 254
 - radius, 321
- short distance, 256, 534, 558
- SI (*Système International*), 67–71, 74, 77–81, 83–84
- silicon, 79, 81–82, 84
 - XRCD, *see* XRCD
- singular
 - behavior, 155, 256–257, 259, 349, 583, 586
 - coefficient, 151, 155
 - field, 271, 280, 282, 299, 302, 305, 318–321, 323–324, 331, 339, 358, 403, 419
 - limit, 152
 - piece, 323, 333, 405
- skeleton Hamiltonian, 177–178
- small
 - acceleration, 26, 48, 53–54, 272, 289
 - body, 253, 261, 270, 408, 593, 596

- mass, 3, 13, 20, 37, 160, 271–272, 276, 283, 292–293, 300, 302, 328, 347, 418, 432, 443, 563, 582, 590
- object, 273–275, 279, 282–283, 289
- smooth
 - function, 132, 256–266, 330, 338, 340, 359, 544
 - vector field, 256
- solution
 - Brill-Lindquist, 175–176
 - distributional, 254, 256
 - Kerr geometry, 35
 - Misner-Lindquist, 175
 - numerical, *see* numerical
 - Schwarzschild, 485, 570
 - Schwarzschild-de Sitter, 35
 - unique, 131, 220, 257
 - vacuum, *see* vacuum
 - wave equation, *see* wave equation
- source
 - charge-current, 254, 266
 - distributional, 254
 - effective, *see* effective source
 - EMRI, 396
 - matter, *see* matter
 - moments, 142–143, 145, 147–148
 - point particle, 255, 258, 264
 - post-Newtonian, 135–139
 - term, 129, 133, 138, 140, 154, 181, 198, 233, 253, 348–349, 351, 353, 397–398, 400, 408, 423, 575–576, 579, 596
- space-like
 - hypersurface, *see* hypersurface
 - slice, 184, 253
- spacetime
 - asymptotically flat, *see* asymptotically flat spacetime
 - background, *see* background
 - curvature, 31, 310, 562
 - domain, 87, 92, 108–109
 - Finsler, 58–59
 - geometry, *see* geometry
 - globally hyperbolic, 113, 318
 - Kerr, *see* Kerr geometry
 - Minkowski, *see* Minkowski spacetime
 - Riemannian, 28, 58
 - Schwarzschild, *see* Schwarzschild
 - stationary, 94, 369
- special relativity, 28, 44, 87, 103, 297, 463–464, 467
- speed of light, 2, 28–30, 44, 69, 126, 167, 416, 570, *see* photon speed
- spherical harmonics, 144–145, 161, 286, 294, 336–338, 350–351, 353, 355–358, 405, 576, 586
- spin
 - coupling
 - Papapetrou equation, 148
 - spin-orbit, 162, 196–197
 - spin-spin, 127
 - to curvature, 148, 162
 - force, 261, 263, 269
 - precession, 163, 193, 199
 - spin-2, 18, 128, 294, 462, 469, 471, 479
 - spinning particle, 31, 128, 148, 162
 - supplementary condition, 149
- spin-foam model, 537, 541–543
- spin-networks, 536, 542, 558
- spontaneous emission, 492
- Standard Model, 1–5, 7–11, 15–17, 20–22, 464
 - extensions, 21
- state of the art, 128, 333, 361, 561, 565, 590, 592
 - beyond, 593
- stationary spacetime, 94, 369
- step (Heaviside step function)
 - discontinuity, 360
 - distribution, 581, 589
 - function, 241, 317–318, 496
- STF, 134, 139, 141, 144–145, 204, 286–287, 361, *see* symmetric-trace-free tensor, 145, 286
- Stone-von Neumann theorem, 539
- stress-energy tensor, 93, 129, 131, 147–148, 150–151, 171, 181, 183–184, 200, 253–255, 277, 283–285, 287–288, 291, 398, 405, 423–424, 576, 583, *see* pseudo-tensor
 - conservation, 583
 - stress-energy-momentum, 109
- string theory, 26, 45, 462, 463
- strong
 - equivalence principle, 41, 49, 76, 471, 475, 479, 565
 - interaction, 2, 4, 12, 15–16, 22
 - nuclear interactions, 1
- Strong-Field Predictions, 471
- sufficiently small
 - body, 253, 261
 - neighborhood, 257, 258
- superstring theories, 21
- supersymmetry, 21, 22, 47, 462, 480
- supplementary phase, 230
- symmetric mass ratio, 157, 213, 217, 223, 431, 432

- symmetric-trace-free (STF), 134, *see* STF
 symmetry group, 3, 4–6, 20–21, 106, 195, 537–539, 544
- T**
- T-parity, 371, 386
 tail, 138–139, 146, 209, 259, 330–332, 387, 390, 428, 436, 561, 584–586
- component, 586
 contribution, 329, 332, 378, 428
 effect, 161, 168, 214, 230, 232, 428
 factor, 230, 232–233, 237
 field, 330
 function, 382, 386
 Green's, 401
 GW, *see* gravitational wave
 integral, 146, 161, 429
 part, 378, 401, 403, 584
 ringing, 213
 term, 203, 205, 240, 317, 334, 369, 379, 382–384, 386, 388, 392, 428
- Taylor
 approximants, 227, 236
 T4, 215
 coefficients, 228
 expand, 138, 267, 285
 expansion, 126, 221, 225–227, 237, 286, 312, 317, 373, 581, 587
 series, 227, 266
 finite, 266
- Taylor-expanded, 212–213, 225, 228, 233, 235
 templates, 127, 162, 396, 416, 453–455
- adiabatic, 408
 gravitational wave (GW), 125, 128, 156, 162, 164, 188, 211, 249, 396
 inspiral, 125, 328
 LISA, 597
 post-Newtonian (PN), 416
 theoretical, 127, 328, 396
 waveforms, 211
- tensor
 anisotropic mass, 30
 Einstein, *see* Einstein
 second order, 269
 energy-momentum, *see* energy-momentum tensor
 canonical, 91
 stress, 109
 Levi-Civita, 274, 286
 quadrupole, 192, 198, 205
 mass-quadrupole, 192, 205
 Riemann, *see* Riemann
 STF, 145, 286
- stress-energy, *see* stress-energy tensor
 Weyl, 398, 477, 478
- test-particle limit, 191, 221
 Testing Basic Laws of Gravitation, 25–60
 New Effects, 45
- tetrad, 91, 313–314, 339, 398, 537
 Teukolsky equation, 346, 348, 396–400, 403
 TeVeS (Tensor-Vector-Scalar) model, 483
 theoretical templates, 127, 328, 396
 Thorne-Hartle-Zhang (THZ) coordinates, 313, *see* THZ
- three-dimensional gravity, 534–535, 544
 transplanckian region, 563
 Trautman-Bondi-Sachs, 105
 THZ, 283, 287–288, 299, 313, *see* Thorne-Hartle-Zhang
- tidal
 acceleration, 289–290
 deformation, 367, 391
 distortion, 271–272, 288, 305, 420
 effects, 274, 289, 449, 463
 forces, 449, 478, 565
 properties of neutron stars, 249
- tides, 289
- time
 dependence, 26, 39, 53, 103, 290, 413, 473
 domain, 87, 92, 108–109, 302–303, 327, 333–334, 343, 347–350, 356, 375, 395, 400, 496, 498, 509, 517, 578
 interval, 270
 scale, 51, 126, 274, 283, 287, 289–290, 395–396, 398, 453, 510, 526, 562, 590, 596
- time-dependent gravitational constant, 44
 time-like
 hypersurface, 98, 254
 surface, 255
- two-body problem, 125–164, 211–249, 273, 275, 395, 562, 578, 582
- U**
- UFF, 30–31, 41–42, 46, 50, 57, *see* universality of free fall
- unbarred
 coordinates, 266
 metric, 266
- uncertainty principle, 9
 uniformity requirement, 265, 266
 unique solution, 131, 220, 257
 uniqueness of acceleration, 565–566, 568–569, 585
- Units, 67–84
 ampere, *see* ampere

astronomical unit, 462
 atomic units, 71, 73
 base units, new definitions of, 83
 candela, 68
 geometrized units, 329
 kelvin, 68, 75, 84
 kilogram, *see* kilogram
 meter, 68–69, 74
 natural, 71, 73
 Planck, 71, 73, 557
 second, 67–69, 72, 75, 80, 84
 universality of free fall (UFF), 30, 463, 464, 514, *see* UFF

V

vacuum
 energy, 31, 491–492, 494, 506
 field fluctuations, 491–492, 503, 508, 526
 fluctuation, 492–494, 504, 510
 solution, 35, 175, 265, 273, 278, 283, 286, 287, 291, 292, 346
 state, 492, 497, 501–502, 507, 511
 wave equation, 574
 van der Waals force, 492
 van Vleck-Morette determinant, 257
 Virgo, 211, 416, 475
 von Klitzing constant, 76

W

watt balance, 78–80, 82–84
 wave equation
 homogeneous, *see* homogeneous
 inhomogeneous, 202, 403
 linear, 256, 310, 316
 scalar, 310
 self-adjoint, 260
 solution, 136, 138, 140, 317–319, 383
 vacuum, 574
 wave operator, 264, 283–284, 316, 353, 401, 579
 waveform of compact binaries, 145
 waveforms, 232, 237, 245–246, 399, 411–413, 433, 452–454, 578, 592
 adiabatic, 395, 397–398, 412–413, 453–454
 analytical, 212, 243
 approximate, 453
 EMRI, 327, 408, 455

EOB, 213, 223, 238–239, 243
 gravitational (GW), *see* gravitational wave
 inspiral, *see* inspiral
 kludge, 453–455
 multipolar, 236–237
 numerical (NR), 213, 223, 235, 243, 396
 polarization, 143–145, 160–161, 207
 post-Newtonian, 147, 160
 template, 211
 wavefront set, 254
 weak interaction, 2, 4–8, 11
 weakly isolated horizon, 116, *see* WIH
 well-posed initial value formulation, 270
 Weyl
 curvature, 517, 519
 scalars, 298
 tensor, 398, 477, 478
 white dwarfs, 396, 449
 WIH, 116–118, *see* weakly isolated horizon
 symmetry, 117–118
 wild oscillation, 264
 Wilson loop, 533
 window function, 281–282, 288, 303
 Woolsthorpe apple tree, 561, 597
 worldtube, 115, 264, 357, 477
 auxiliary, 357

X

X-ray crystal density, 79, 81, *see* XRCD
 XRCD, 79, 81–84
 silicon, 79, 81–82, 84

Y

Yang-Mills
 gauge field, 533, 536
 theory, 533, 536, 538

Z

Zerilli, 240, 574, 577, *see* Regge-Wheeler-Zerilli
 equation, 298, 574–575, 581
 potential, 591
 zeroth order, 283
 (zeta) ζ function, *see* Riemann-Hurwitz
 ζ -function
 zitterbewegung, 56