

*Carlo Rovelli (Ed.)*

# GENERAL RELATIVITY: THE MOST BEAUTIFUL OF THEORIES

APPLICATIONS AND TRENDS AFTER 100 YEARS

STUDIES IN MATHEMATICAL PHYSICS 28



# **De Gruyter Studies in Mathematical Physics**

Edited by

Michael Efroimsky, Bethesda, Maryland, USA  
Leonard Gamberg, Reading, Pennsylvania, USA  
Dmitry Gitman, São Paulo, Brazil  
Alexander Lazarian, Madison, Wisconsin, USA  
Boris Smirnov, Moscow, Russia

**Volume 28**

# **General Relativity: The Most Beautiful of Theories**

---

Applications and Trends after 100 Years

Edited by  
Carlo Rovelli

**DE GRUYTER**

**Physics and Astronomy Classification Scheme 2010**  
04.20.-q, 04.60.-m

**Editor**

Prof. Dr. Carlo Rovelli  
Centre de Physique Théorique de Luminy Aix-Marseille  
University  
Case 907, Luminy

**13288 Marseille, France**

[rovelli@cpt.univ-mrs.fr](mailto:rovelli@cpt.univ-mrs.fr)

ISBN 978-3-11034042-6  
e-ISBN (EPUB) 9783110383645

e-ISBN (PDF) 978-3-11-034330-4  
e-ISBN (EPUB) 978-3-11038364-5  
Set-ISBN 978-3-11-034331-1  
ISSN 2194-3532

**Library of Congress Cataloging-in-Publication Data** A CIP catalog record for this book has been applied for at the Library of Congress.

**Bibliographic information published by the Deutsche Nationalbibliothek** The Deutsche Nationalbibliothek lists this

publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at  
<http://dnb.dnb.de>.

© 2015 Walter de Gruyter GmbH, Berlin/Munich/Boston  
Typesetting: le-tex publishing services GmbH, Leipzig Printing  
and binding: CPI books GmbH, Leck

[www.degruyter.com](http://www.degruyter.com)

Epub-production: Jouve, [www.jouve.com](http://www.jouve.com)

# Table of Contents

## **De Gruyter Studies in Mathematical Physics**

Title Page

Copyright Page

## **The most beautiful physical theory**

### **Astrophysical black holes**

#### **1 Introduction**

#### **2 A brief history of astrophysical black holes**

##### **2.1 Early history**

##### **2.2 The Schwarzschild metric**

#### **3 Relativistic astrophysics emerges**

##### **3.1 Rotating black holes**

##### **3.2 Black holes as energy sources**

##### **3.3 Motion in the Schwarzschild metric**

##### **3.4 Circular orbits**

##### **3.5 Stability of circular orbits**

#### **4 Evidence from X-rays, quasars and AGN**

#### **5 The black hole at the centre of the Milky Way**

##### **5.1 Sgr A\***

##### **5.2 GR effects on orbits**

###### **5.2.1 Orbital precession**

#### **6 Galaxies and black holes**

##### **6.1 AGN feedback**

##### **6.2 Jets, Gamma-Ray Bursts and the birth of black holes**

#### **7 Current observations of accreting black holes**

##### **7.1 Particle motion in the Kerr metric**

## 7.2 Current observations of accreting black holes continued

## 7.3 Velocities and frequencies

## 7.4 Further AGN properties

## 8 Measurements of the masses of black holes

## 9 Measurements of black hole spin

### 9.1 Equations for photon motion and redshift

### 9.2 Light bending around a Kerr black hole

### 9.3 Iron line emission

## 10 Future observations of astrophysical black holes

## 11 More general spherically symmetric black holes

## 12 Primordial black holes

### 12.1 Hawking radiation

### 12.2 Link with surface gravity

### 12.3 Astrophysical aspects of black hole evaporation

### 12.4 Black hole entropy

### 12.5 Laws of black hole thermodynamics and the Penrose process

### 12.6 Adiabatic (reversible) changes

### 12.7 Other processes for extracting energy from a spinning black hole

## 13 Conclusions

## Bibliography

## Observations of General Relativity at strong and weak limits

## 1 Introduction

## 2 Tests in the solar system and binary systems

### 2.1 Orbit precession

### 2.2 Gravitational waves

### 2.3 Lense-Thirring effect and relativistic spin-orbit coupling

### 2.4 Bending of light rays and gravitational redshift

### 2.5 Massive spinning black hole test

### 3 Observational discovery of a non-zero cosmological constant (dark energy)

#### 4 Weak limit test near zero gravity surface

##### 4.1 Dark energy antigravity as a test of General Relativity

##### 4.2 Local dark energy test via outflow

##### 4.3 Dynamical structure of a gravitating system within dark energy

### 5 Estimating cosmologically nearby dark energy: the Local Group

### 6 Mass, dark energy density and the lost gravity effect

### 7 Dark energy in the Coma Cluster of galaxies

### 8 Testing the constancy of $\Lambda$

### 9 Strong limit: Spinning black holes and no-hair theorem

### 10 OJ287 binary system

#### 10.1 The binary model

#### 10.2 OJ287 flares and jet

#### 10.3 OJ287 orbit parameters (without using outburst times)

### 11 Modeling binaries with Post Newtonian methods (with outburst times)

### 12 OJ287 results at the strong field limit

### 13 Conclusions

### 14 Appended section; mass, dark energy density and the “lost gravity” effect

### 15 Appended section: dark energy in the Coma cluster of galaxies

#### 15.1 Three mass estimates of the cluster

#### 15.2 Matter mass profile

#### 15.3 Upper limits and beyond

### 16 Appended section: modeling binaries with Post Newtonian methods

### Bibliography

## General Relativity and dragging of inertial frames

### 1 Frame-dragging: the theory

#### 1.1 Dragging of inertial frames and the origin of inertia

#### 1.2 Dragging of inertial frames and the gravitomagnetic analogy

#### 1.3 The gravitomagnetic formal analogy of General Relativity with electrodynamics

#### 1.4 Dragging of inertial frames inside a hollow sphere

#### 1.5 Frame-dragging phenomena on clocks and photons

#### 1.6 Frame-dragging, time-delay and gravitational lensing

##### 1.6.1 Time-delay inside a slowly rotating massive shell

#### 1.7 [An invariant characterisation of frame-dragging]

### 2 The need to further test General Relativity

#### 2.1 The universe and the triumph General Relativity

#### 2.2 The riddle of dark energy and dark matter

#### 2.3 Unified theories, alternative gravitational theories and some limits of General Relativity

#### 2.4 [Frame-dragging, Chern-Simons gravity and String theory]

### 3 The holy grail of experimental General Relativity and its observation with the LAGEOS satellites and Gravity Probe-B

### 4 The LARES space experiment

#### 4.1 The LARES satellite, its structure and its orbit

#### 4.2 The LARES satellite, General Relativity and geodesic motion

#### 4.3 The LARES satellite and its preliminary orbital

[results](#)

[4.4 \[LARES error analyses\]](#)

[5 Conclusions](#)

[Bibliography](#)

[GNSS and other applications of General Relativity](#)

[1 Introduction](#)

[2 Relativity principles](#)

[3 Astronomical and geocentric time scales](#)

[4 Earth-based time scales TT, TAI, UTC](#)

[5 Gravitational frequency shifts](#)

[6 Sagnac effect; realizing coordinate time](#)

[7 Relativistic effects on orbiting clocks](#)

[8 The eccentricity effect](#)

[9 Navigation on the rotating earth](#)

[10 Emission coordinates](#)

[11 JUNO and other missions](#)

[12 Summary](#)

[Bibliography](#)

[The strange world of quantum spacetime](#)

[1 A world with no space](#)

[2 A world without time](#)

[3 Loop gravity](#)

[4 Quantum spacetime](#)

[5 Empirical evidence](#)

[Bibliography](#)

[Index](#)

[List of contributors](#)

[De Gruyter Studies in Mathematical Physics](#)

Carlo Rovelli

## The most beautiful physical theory

*"The most beautiful of all existing theories."*

Landau and Lifshitz

*"Probably the greatest scientific discovery that was ever made."*

Dirac

*"Scarcely anyone who fully understands this theory can escape  
from its magic"*

Einstein

(All quoted in S. Chandrasekhar, J. Astrophys. Astr **5** (1984), 3–  
11.)

There are masterpieces that move us deeply: Mozart's Requiem, the Odyssey, the Sistine Chapel, King Lear.... They give us new eyes to see the world from a novel and deeper perspective. General relativity, the jewel of Albert Einstein, is one of these.

Conceived almost in isolation, without a single element of new empirical data<sup>1</sup>, based on an acute reflection on the scientific knowledge available at the time and on the courage of challenging common sense frontally, general relativity is just a simple idea and a few lines of equations. But an astonishing wealth of incredible predictions follow from these: time runs faster on the mountains than at seaside; light does not travel in straight lines; spacetime bends, and can tremble over the vast interstellar lands like the surface of a lake under a soft wind; it can bend over the weight of a star to the point of collapsing into a hole where all things and even light can enter but not come out; the entire universe cannot stay put, but must contract or expand, and in fact is born from a fiercely hot concentrated core....

All this sounds like a tale told by an idiot, full of sound and fury, more than sober results of physicists's calculation. And

much of this was still seen as such when I studied at the university. Steven Weinberg's 1972 general relativity textbook still calls the black holes's horizons "hypothetical" and says that  $r = 2 m$  Schwarzschild radius "does not seem to have much relevance for the real world" (page 207)!

Instead, it has all come true. The universe has turned out to be truly so kaleidoscopic: one after the other, years after year, all the astonishing predictions derived from the theory have been confirmed by observations.

Today we measure the difference of the rate at which time flows at a few decimetres of elevation difference; the evidence for black holes is overwhelming; the evidence for gravitational waves is indirect, but very strong; nobody questions anymore the idea that the universe expands and has evolved from a hot and dense core... Physics is not new to spectacular predictions: Newton's theory has been used to deduce the existence of Neptune before seeing it, Dirac has predicted antimatter, Maxwell the electromagnetic waves, the standard model of particle physics has predicted the Higgs particle, and so on. But no theory has predicted a sequence of previously unconceivable aspects of reality like general relativity.

**Carlo Rovelli:** CPT-CNRS, Case 907, Luminy, 13288  
Marseille, France

I remember my emotion when I began understanding something about the theory. It was summer. I was on a beach in Calabria, immersed in the Mediterranean glare, perfuming of ancient Greece, at the time of my last university year. Holidays is when one studies better, without the school's distractions. I studied on a book gnawed by mice; I had used it to block the holes of these little creatures, at night, in the hippy house on the Umbrian hill where I used to take refuge from the boredom of University lectures. Every once in a while, I raised my eyes from the worn out book, to stare the sparkle of the sea: I could almost see the bending of space and time. It was like magic, as

a soft friend's voice whispering an extraordinary secret truth at my ear, a veil being moved from reality, revealing a new glimpse of its hidden beauty.

Since we learned that the Earth is round and spinning, we have realised that reality is not as it appears: every time we understand a new piece of it, is an emotion. Another screen that falls. But among all the many leaps forward in our knowledge, one after the other in the course of history, the one made by Einstein from 1907 to 1915 is probably a leap without equals.

For one thing, the theory is of breathtaking simplicity. Newton tried to explain why things fall and planets revolve. He imagined a "force" that pulls bodies toward one another. How could this force pull things far away from each other, without anything acting in between, was quite unclear, and the father of science cautiously didn't feign hypotheses. He also imagined that bodies move in space. The nature of such "space", a sort of "container" of the world, which Newton did feign, was not clear either. But a few years before Albert's birth, Faraday and Maxwell had added an ingredient to Newton's cold universe: the electromagnetic field, a real entity filling space and carrying the electric force. Einstein, fascinated from childhood by the electromagnetic field, that pushed the rotors of the power plants built by his father, soon realised that gravity, like electricity, had to be carried by a field. And here came the extraordinary idea: the gravitational field is not widespread in space: the gravitational field *is* space.

The strange entity that Newton had introduced into physics, "space",<sup>2</sup> is understood to be nothing else than a configuration of the gravitational field. And the entity that "carries along" the force of gravity, like the electromagnetic field carries the electric force, is space itself, which is to say, the gravitational field. The two conceptual puzzles of Newtonianism solve one another. This is the simple idea that grounds the theory of general relativity.

Newton's "space", where things move, and the "gravitational field", which carries the force of gravity, are one and the same thing. It is an impressive simplification of the world: space is no longer something different from matter. It is one of the

components of the world, like any other field: an entity that sways, bends, twists. We are not contained into an invisible rigid shelf: we are immersed in a giant flexible jellyfish.<sup>3</sup>

Einstein had the intuition of a bending spacetime rather early, but had to search for the right mathematics to translate his vision into a theory. Gauss, “prince of mathematicians”, had written the mathematics describing two-dimensional curved surfaces and had then asked a good student to generalize this to higher dimensions. The student, Bernhard Riemann, had produced a ponderous dissertation, the kind that seem completely unnecessary. The result was that the properties of a curved space are captured by a mathematical object, which we now call the Riemann curvature. Einstein wrote an equation stating that this curvature is related to the energy of matter. This is it. The equation is half a line. A vision – spacetime is a field and bends – and an equation. Inside this equation, slowly emerging over the decades, a kaleidoscopic reality where universes can explode, space sink, time slow, and the endless expanses of interstellar space ripple.

All this was emerging slowly in Calabria from my book gnawed by mice. And it was not the effect of the scorching Mediterranean sun, an hallucination on the glimmering sea. It was reality. Or rather, a glimpse into reality, a little less veiled than the blurred one of our everyday’s look. A reality that seemed made of stuff of which dreams are made, but was nevertheless more real than our clouded everyday’s dream.

All of this the result of an elementary intuition: space and the field are the same thing. And a simple equation

$$R_{ab} - \frac{1}{2}(R - 2\Lambda)g_{ab} = T_{ab} .$$

(1)

Sure, it takes a path of apprenticeship to digest the mathematics of Riemann and master the technique to read this equation. A little commitment and effort. But less than those needed to get the rare beauty of one of Beethoven’s late quartets. In either case, the reward is beauty, and new eyes to

see the world.

\*\*\*\*

General relativity is not just an extraordinarily beautiful physical theory providing the best description of the gravitational interaction we have so far. It is more.

In science, there are discoveries that stay. Anaximander discovered that the sky we see above our head continues under our feet, Copernicus that the Earth is not still at the center of the universe, Galileo that velocity is a relative quantity. Faraday and Maxwell discovered that there are fields, Dirac that there are antiparticles, Einstein, in 1905, that there is no absolute simultaneity, Darwin that living beings belonging to different species have common ancestors. These are discoveries, not volatile theories: they will stay with us. Maybe we'll get more insights, but we are not going to unlearn them. We are not going to find, in the future, that Earth is flat, there are no antiparticles, or species do not evolve.<sup>4</sup>

Einstein's general relativity belongs to this class. It reflects a new insight we have reached about Nature. This is the reason of its extraordinary empirical success: we have understood, a bit better, how the world works. What have we actually discovered?

The central discovery underpinning general relativity is that the idea that the world is formed by things located in space and changing as time passes is only an approximation to a reality which has a different structure. The approximation holds only as long as we are in regimes where the dynamics of the gravitational field can be neglected or is sufficiently soft. Beyond such regimes, the world is made differently.

Einstein struggled at length with this deep change of perspective. He called this effort his "struggle with the meaning of the coordinates". He got to conclusions such as his celebrated:

"Space and time [...] lose the last vestige of physical reality"<sup>5</sup>.

The spatiotemporal structure of the world does not come before its dynamics: it comes after. Space and time are aspects of something else, not something that supports the existence of the rest.<sup>6</sup> As far as we have understood so far, the world can be described as made by entities (se still cal them “fields”) which are dynamical, interact with one another and obeys the laws of quantum theory. Space and time come later.

I believe this discovery is still underestimated and not yet digested by the scientific community. The difficulty of the change of perspective implied by this discovery should not be underestimated. Einstein himself got wrong a surprising number of times. He did not understand the meaning of the Schwarzschild singularity.<sup>7</sup> We was long convinced that gravitational waves are a gauge artefact. He rejected the possibility of the expansion of the Universe. He believed that his equations had no solution without matter. The number and the extent of his mistakes are impressive. The later scientists did not score much better. It took many decades to understand that gravitational waves carry energy, that the Schwarzschild singularity can be traversed safely, that the Lemaître-Friedmann solutions were realistic, and so on. For years the accurate measurements of the Earth-Moon distance were presented in the literature giving the (meaningless) coordinate distance rather than the physical distance. It took Dirac to begin unraveling the hamiltonian structure of the theory, apparently so much different from that of the conventional physical theories. When the first GPS satellite system was mounted the persons responsible of the project did not believe that time run faster up there and the system had to be built with a double system just in case... The theory is far more subtle than what it looks at first reading.

Today the theory is familiar to many, but I am convinced that its deep lesson has not yet been absorbed. The majority of the papers on Physical Review are still written as if the author had not digested the insight into Nature obtained thanks to this theory, insight supported by the astonishing empirical success

of the theory. Most papers still assume that there is a universal time in the universe, that the universe is globally Poincaré invariant, that coordinates have necessarily a metrical meaning, that localisation can be defined neglecting the dynamics of the reference, and so on. Today's large majority of papers sound like the works of the intellectuals that after Copernicus were still assuming that the Earth is still at the center of the universe. There is nothing wrong in this as long as they do not pretend to be valid beyond the regime where this a decent approximation. But too many of them forget this.

The world is not located in space and does not evolve in time. At the fundamental level, the dynamical theory cannot be given by a Hamiltonian evolving the system in an external time. In the fundamental quantum theory, there cannot be unitary evolution time, because there is no such an external time in the universe. There is no global Poincaré invariance in the universe. Observables cannot depend on a non-dynamical coordinate system. Locating observables at infinity does not saves from these difficulties, it only establishes an approximation regime, where the theory cannot be fundamental.

Extremely slowly, the realisation that the world is such has been percolating into different domains of physics, from cosmology to astrophysics, from solar system dynamics to high energy physics. But the process is still painfully slow. After all, it took nearly a century between Copernicus' book and the time when Galileo and Kepler begun succeeding convincing people that the Earth does indeed move. A century has passed from Einstein's paper. Everything predicted using the equation in it has turned out to be true. It is time time to begin taking its still astonishing insights more seriously.

\*\*\*\*

The current research in general relativity is far too vast to be summarised in a book. This book focuses on some of the most beautiful and characteristic aspects of general relativistic physics and summarises the state of the art in these domains.

The most unexpected and striking natural phenomenon

revealed by general relativity is the black hole. Black holes, whose actual existence in Nature was still doubted not long ago, are today a common topic of investigation in astrophysics. The current knowledge about black holes in astrophysics is reviewed by Andy Fabian and Anthony Lasenby.

Gene Byrd, Arthur Chernin, Pekka Teerikorpi, and Mauri Valtonen discuss extensively the status of the general relativistic phenomenology in the strong and weak limits. Ignazio Ciufolini covers the intriguing phenomenon of frame dragging. Neil Ashby discusses the applications of the theory and in particular the navigation systems.

In a short final chapter, Carlo Rovelli offers a perspective on the main theoretical question open in gravitational physics: the description of the quantum properties of the gravitational field.

This is not an exhaustive overview of the research in general relativity, but it offers an definite insight on some of its most astonishing and revolutionary aspects.

Andrew C. Fabian and Anthony N. Lasenby

# Astrophysical black holes

**Abstract:** Black holes are exotic relativistic objects which are common in the Universe. It has now been realised that they play a major role in the evolution of galaxies. Accretion of matter into them provides the power source for millions of high-energy sources spanning the entire electromagnetic spectrum. Observations of stars orbiting close to the centre of our Galaxy provide detailed clear evidence for the presence of a 4 million Solar mass black hole. Gas accreting onto distant supermassive black holes produces the most luminous persistent sources of radiation observed, outshining galaxies as quasars. The energy generated by such displays may even profoundly affect the fate of a galaxy. We briefly review the history of black holes and relativistic astrophysics before exploring the observational evidence for black holes and reviewing current observations including black hole mass and spin. In parallel (and in *italic*) we outline the general relativistic derivation of the physical properties of black holes relevant to observation. Finally we speculate on future observations and touch on black hole thermodynamics and the extraction of energy from rotating black holes.

## 1 Introduction

Black holes are exotic relativistic objects which are common in the Universe. It has now been realised that they play a major role in the evolution of galaxies, and accretion around them, and jets launched from them, provide the power source for millions of high-energy sources spanning the entire electromagnetic spectrum. In this chapter we consider black

holes from an astrophysical point of view, and highlight their astrophysical roles as well as providing details of the General Relativistic phenomena which are vital for their understanding.

To aid the reader in appreciating both aspects, we have provided two tracks through the material of this Chapter. Track 1 provides an overview of their astrophysical role and of their history within 20th and 21st century astrophysics. Track 2 (in italic text) provides the mathematical and physical details of what black holes are, and provides derivations of their properties within General Relativity. These two tracks are tied together in a way which we hope readers with a variety of astrophysical interests and persuasions will find useful.

**Andrew C. Fabian:** Institute of Astronomy,  
Madingley Road, Cambridge, CB3 0HA, UK

**Anthony N. Lasenby:** Kavli Institute for  
Cosmology, Madingley Road, Cambridge, CB3 0HA  
and Cavendish Laboratory, J.J. Thomson Avenue,  
Cambridge, CB3 0HE, UK

## 2 A brief history of astrophysical black holes

### 2.1 Early history

Although the term “black hole” was coined by J. A. Wheeler in 1967, the concept of a black hole is over two hundred years old. In 1783, John Michell [41] was considering how to measure the mass of a star by the effect of its gravity on the speed of the light it emitted. Newton had earlier theorized that light consists of small particles. Michell realized that if a star had the same density as the Sun yet was 500 times larger in size, then light could not escape from it. The star would thus be invisible. He noted, however, that if it was orbited by a luminous star, the measurable motion of that star would betray the presence of the invisible one.

This prescient, but largely forgotten paper, embodies two important concepts. The first is that Newtonian light and gravity predicts a minimum radius  $R = 2GM/c^2$  for a body of mass  $M$  from within which the body would not be visible. The second is that it can still be detected by its gravitational influence on neighbouring stars. The radius is now known as the Schwarzschild radius of General Relativity and is the radius of the event horizon of a non-spinning black hole. Black holes are now known to be common due to their gravitational effect on nearby stars and gas. Pursuing Newtonian black holes further leads to logical inconsistencies and also the problem that relativity requires the speed of light to be constant.

The concept re-emerged after the publication of Einstein's General Theory of Relativity in 1915 when Karl Schwarzschild found a solution for a point mass. Einstein himself "had not expected that the exact solution to the problem could be formulated". It was not realised at the time that the solution represented an object which would turn out to be common in the Universe. Chandrasekhar in 1931 [9] discovered an upper limit to the mass of a degenerate star and which implied the formation of black holes (although this was not spelled out). Eddington, who wrote the first book of General Relativity to appear in English, considered the inevitability of complete gravitational collapse to be a *reductio ad absurdum* of Chandrasekhar's formula. The concept was again ignored for a further two decades, apart from work by Oppenheimer and Snyder [50] who considered the collapse of a homogenous sphere of pressureless gas in GR, and found that the sphere becomes cut off from communication with the rest of the Universe. In fact, what they had discovered was the inevitability of the formation of a black hole when there is no pressure support.

With this brief introduction to the early history, we now give some details of what is now understood by a Schwarzschild black hole.

## 2.2 The Schwarzschild metric

The simplest type of black hole is described by the Schwarzschild metric. This is a vacuum solution of the Einstein field equations in the static, spherically symmetric case, and takes the form

$$ds^2 = \left(1 - \frac{2GM}{c^2r}\right)dt^2 - \frac{1}{c^2} \left(1 - \frac{2GM}{c^2r}\right)^{-1}dr^2 - \frac{r^2}{c^2}d\theta^2 - \frac{r^2 \sin^2 \theta}{c^2}d\phi^2$$

(1)

where  $M$  is the mass of the central object. Most of the observational properties of black holes that we need follow from this metric, rather than needing the deeper level of the field equations themselves to understand them, but to give an indication of the achievement of Schwarzschild, and since we will wish to consider briefly more general spherically-symmetric black holes as well, we give a brief sketch of how the metric is derived. For more details, and a description of the sign conventions we are using, see Chapters 8 and 9 of [30].

The field equations themselves are

$$G_{\mu\nu} = -8\pi T_{\mu\nu}$$

(2)

where the Einstein tensor  $G_{\mu\nu}$  is a trace-reversed version of the Ricci tensor  $R_{\mu\nu}$  and  $T_{\mu\nu}$  is the matter stress-energy tensor (SET). If the matter SET is traceless, in particular if we are in a vacuum case, then clearly  $G_{\mu\nu}$  must be traceless as well, and so the vacuum equations are that all components of the Ricci tensor are zero:

$$R_{\mu\nu} = 0$$

(3)

The Ricci tensor is in turn a contraction,  $R_{\mu\nu} = R^\lambda{}_{\mu\nu\lambda}$  of the Riemann tensor, which is defined by

$$R^\mu_{\alpha\beta\gamma} = \frac{\partial \Gamma^\mu_{\alpha\gamma}}{\partial x^\beta} - \frac{\partial \Gamma^\mu_{\alpha\beta}}{\partial x^\gamma} + \Gamma^\mu_{\alpha\beta}\Gamma^\sigma_{\gamma\alpha} - \Gamma^\mu_{\alpha\gamma}\Gamma^\sigma_{\beta\alpha}$$

(4)

where  $\Gamma^\mu_{\alpha\beta}$  is the connection. For a general metric  $g_{\mu\nu}$ , defined by

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu$$

(5)

the connection is given in terms of the metric by

$$\Gamma_{\mu\alpha\beta} = \frac{1}{2} \left( \frac{\partial g_{\mu\beta}}{\partial x^\alpha} + \frac{\partial g_{\alpha\beta}}{\partial x^\mu} - \frac{\partial g_{\alpha\beta}}{\partial x^\mu} \right)$$

(6)

We can thus see that solving the apparently simple [equation \(3\)](#) involves potentially formidably complicated second order partial differential equations, and it is not surprising that Einstein considered it unlikely that an exact solution for the metric around a spherical body would be found. In the current case, where we have assumed everything is static, we can attempt an ansatz for the metric which simplifies things considerably. We let

$$ds^2 = A dt^2 - B dr^2 - r^2 (d\theta^2 + \sin^2 \theta d\phi^2)$$

(7)

where  $A$  and  $B$  are functions of  $r$  alone. This means that our PDEs become ODEs, and although the working is still complicated, we eventually find that we require

$$R_{00} = -\frac{A''}{2B} + \frac{A'}{4B} \left( \frac{A'}{A} + \frac{B'}{B} \right) - \frac{A'}{rB} = 0,$$

(8)

$$R_{11} = \frac{A''}{2A} - \frac{A'}{4A} \left( \frac{A'}{A} + \frac{B'}{B} \right) - \frac{B'}{rB} = 0,$$

(9)

$$R_{22} = \frac{1}{B} - 1 + \frac{r}{2B} \left( \frac{A'}{A} - \frac{B'}{B} \right) = 0 , \quad (10)$$

$$R_{33} = R_{22} \sin^2 \theta = 0 . \quad (11)$$

We can eliminate A and find a simple equation for B by forming the combination

$$\frac{R_{00}}{A} + \frac{R_{11}}{B} + \frac{2R_{22}}{r^2} = 0 \quad (12)$$

which yields the ODE

$$\frac{dB}{dr} = \frac{B(1-B)}{r} \quad (13)$$

with solution

$$B = \frac{r}{r+C} \quad (14)$$

where C is a constant. Inserting this in (10) then yields a first order equation for A of which the solution is

$$A = \frac{a(r+C)}{r} \quad (15)$$

where a is a further constant. This latter constant effectively just changes the units of time, and it is sensible to choose this so that the speed of light c is 1, which we temporarily employ. We have thereby recovered the Schwarzschild metric, (1), and can identify the constant C as  $-2GM$ .

One notices straightaway that some kind of singularity exists at  $r = -C = 2 GM$ , but it is worth noting that this is not a singularity of the Riemann tensor (the only non-zero quantity

transforming as a full tensor we have available for investigation). The entries of this all behave with  $r$  like  $1/r^3$ , corresponding to tidal forces which make neighbouring particles move apart or together in their motion, and none have a singularity at  $r = 2GM$ . Indeed, the form in which Schwarzschild first discovered his metric also had no singularity at this radius (see e.g. [63] and the comments in [1]), and this coupled with lack of singularity of the Riemann tensor except at the origin, perhaps contributed to the uncertainty stretching over many years as to the physical status of the distance  $2 GM$ .

Nowadays, we of course recognise this as the “horizon”, the point where light becomes trapped. We discuss this in more detail below, particularly in connection with the Kerr solution, but we can understand this in simple terms by looking at the ‘coordinate speed’ of a light pulse. For such a pulse, the interval  $ds = 0$ , and for outward radial motion this means

$$Adt^2 - Bdr^2 = 0, \quad \text{i.e.} \quad \frac{dr}{dt} = \sqrt{\frac{A}{B}}$$

(16)

The ‘horizon’ is therefore where the metric coefficient  $B$  goes to infinity, since this marks the point where light is no longer able to move outwards. This happens in the Schwarzschild case at  $r = r_S = 2GM/c^2$ , which is the location of the event horizon.

Before proceeding further, we mention a few physical facts about Schwarzschild black holes. The first is that the radius of the event horizon,  $r_S$ , corresponds to 3 km per Solar mass. This means that the mean density of a black hole is given by

$$\rho = 2 \times 10^{16} \left( \frac{M}{M_\odot} \right)^{-2} \text{ gm cc}^{-1}.$$

(17)

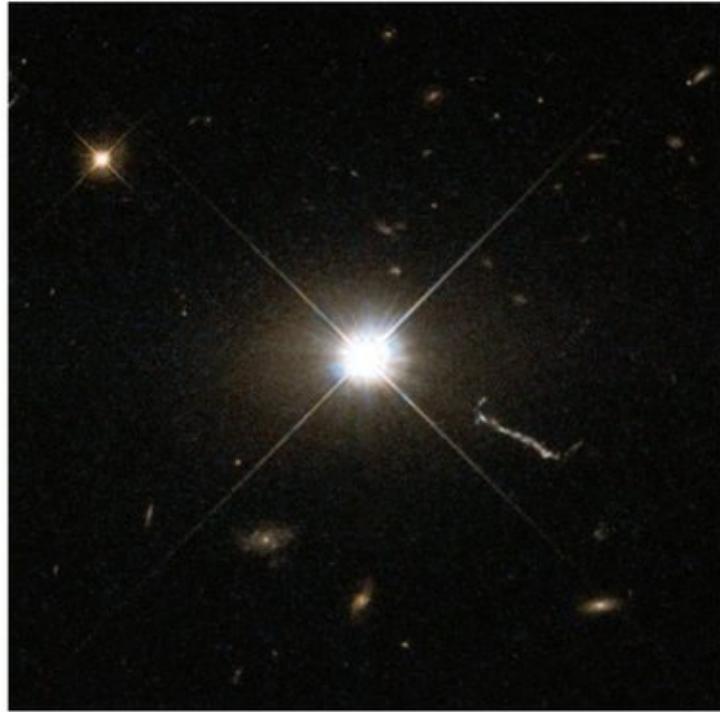
Thus black holes with masses above  $10^8 M_\odot$ , have average densities below that of water, or the Sun. Those above a few billion  $M_\odot$  have densities below that of air. So from a mean

*density point of view, supermassive black holes are not particularly dense. The light-crossing time of the event horizon (i.e. a length equivalent to its diameter) is 0.2 ms for a  $10M_{\odot}$  black hole, 20 s for  $10^6 M_{\odot}$  and about one day for  $5 \times 10^9 M_{\odot}$ .*

### 3 Relativistic astrophysics emerges

The previous neglect of black holes changed in 1963 when two important events occurred, the discovery of quasars by Maarten Schmidt [61] and the discovery of the solution for a spinning black hole by Roy Kerr [31]. These combined to lead to Relativistic Astrophysics as the combination of Astrophysics and GR, embodied in the first Texas Symposium of Relativistic Astrophysics in December 1963. Kerr's brief talk has in retrospect been called the most important announcement at the Symposium but was not mentioned by any of the three summarizers of the meeting [62].

The discovery of quasars resulted from the accurate position for the cosmic radio source 3C273 obtained by Cyril Hazard and collaborators [29] using the lunar occultation technique. Schmidt used the 200" Hale Telescope to take an optical spectrum of the starlike object at that position in the Sky. Some weeks later he identified emission lines in the spectrum with Balmer lines of hydrogen, redshifted by 0.16. Interpreting the redshift as due to cosmic expansion means that it is at a (luminosity) distance of 0.75 Gpc. From its apparent brightness, the object can be inferred to be more than  $10^{12}$  times more luminous than the Sun. 3C273 can be found on optical plates taken since the late 1800s and was soon shown to be variable, at times changing significantly within a week. Ignoring relativistic considerations, causality requires that significant variations cannot occur on timescales shorter than the light crossing time of the object. It therefore appears that the prodigious power output of 3C273 emerges from a region of size similar to the Solar System! [8](#)



**Fig. 1.** The first quasar identified, 3C273, as imaged by the Hubble Space Telescope. It lies over two billion light years away and is the brightest quasar in the visible Sky. The structure to the SW (lower right) is the outer parts of its relativistic jet, which is seen from radio to X-ray wavelengths. The projected length of the jet is over 200 thousand light years, so its true length is nearly an order of magnitude larger than our galaxy. The mass of its black hole has been measured by the optical reverberation technique at almost a billion Solar masses. The X-shape is an instrumental artefact created by support structures in the telescope; it implies that the source is unresolved. 3C273 varies on week-long timescales indicating that the emission region is less than a light week across.

Kerr's solution to the Einstein field equations was not immediately recognised as the exact solution for a spinning black hole. We now know the geometry described is the unique and complete description of the external gravitational field of an uncharged stationary black hole. It is now thought that the power of 3C273 and other quasars is due to accretion onto a Kerr black hole, which we now give some preliminary details of.

### **3.1 Rotating black holes**

*It is now known that the totality of black holes can be described using just three parameters. Whatever makes them up, and however they were formed, in the end only three parameters*

matter - their mass  $M$ , charge  $Q$ , and angular momentum  $J$ . Due to the high conductivity of interstellar matter, black holes are not likely to have a net charge for long, so the only relevant parameter in addition to mass is spin.<sup>9</sup> A metric applicable to a black hole with angular momentum was found by Kerr in 1962 [31]. This metric, in the form later developed by Boyer & Lindquist [6], is

$$ds^2 = \left(1 - \frac{2GMr}{pc^2}\right)dt^2 - \frac{1}{c^2} \left[ \frac{-4GMra\sin^2\theta}{\rho c} dtd\phi + \frac{\rho}{\Delta} dr^2 + \rho d\theta^2 + \left(r^2 + a^2 + \frac{2GMra^2\sin^2\theta}{\rho c^2}\right)\sin^2\theta d\phi^2 \right], \quad (18)$$

where  $a = (J/Mc)$  is the angular momentum of the black hole per unit mass (and has the dimensions of distance),  $\Delta = r^9 - (2GMr/c^9) + a^9$  and  $\rho = r^9 + a^9 \cos^9\theta$ . Substituting this metric into [equation \(6\)](#) for the connection coefficients and calculating the Riemann (4) from these, the Ricci tensor  $R_{\mu\nu} = R^\lambda{}_{\mu\nu\lambda}$  will be found to vanish, and therefore satisfy the vacuum field [equations \(3\)](#). This is a formidable calculation, however, and techniques which derive the metric in the context of more general azimuthally symmetric spacetimes (e.g. [14]) are in fact easier to carry out.

Note if the black hole is non-rotating, then  $J = a = 0$  and the Kerr metric reduces to the standard Schwarzschild metric (1). Almost certainly, all real black holes in the universe will be of the Kerr type. The idea is that just as infalling matter will have angular momentum, so will the material from which the black hole formed, leaving it both with a mass and net angular momentum.

### 3.2 Black holes as energy sources

The enormous luminosity of 3C273 and other quasars was soon linked to black holes by Edwin Salpeter [60] and (separately) Yakov Zeldovich [78]. The energy is due to the deep

gravitational potential well of the black hole which can be liberated by collisions outside the event horizon as gas falls in. If two gas clouds with small amounts of equal but opposite angular momenta fall into a Schwarzschild black hole colliding at  $4r_g$  (note  $r_g$  is defined as  $GM/c^2$ ) and momentarily coming to rest with their kinetic energy emitted as radiation, then about 29% of the rest mass energy of the clouds is released. More generally the angular momentum will be much larger and the radiative efficiency of accretion smaller. A specific model involving an accretion disc was proposed by Donald Lynden-Bell [36] in 1969. Due to the small size of a black hole, it is most unlikely that gas falling into a black hole does so radially, but will have some angular momentum which will cause it to orbit the hole. Viscosity in the swirling gas will cause matter at smaller radii, which is moving faster, to transfer its angular momentum outward to material at larger radii, which is moving slower. The gas then spreads in radius forming an accretion disc in which angular momentum is transferred outward as matter flows inward. The accretion disc model of Lynden-Bell was then studied in detail in the early 1970s by Pringle & Rees [58], Shakura & Sunyaev [65] and, for the Kerr metric, by Novikov & Thorne [49]. The gravitational energy liberated by the inflow heats the disc which radiates locally as a quasi-blackbody. The disc is thin but may extend outward for a considerable distance. This basic picture probably accounts for much of the energy liberated by accreting black holes. There are important modifications due to the magnetic nature of the ionized infalling plasma which will be discussed later. Accretion onto a black hole is the most mass-to-energy efficient process known, apart from direct matter-antimatter annihilation which, due to the rarity of antimatter in our universe, is highly uncommon. Such accretion may account for 20–30% of the energy released in the Universe since the recombination era. To understand the details of how this can happen, we now examine particle motion in GR, first generally, and then applied specifically to the question of efficiency of gravitational energy release around a Schwarzschild black hole. (The more complicated issue of

energy release around a Kerr black hole is treated in [Section 7.1.](#))

### **3.3 Motion in the Schwarzschild metric**

*The key to most astrophysical applications of the Schwarzschild metric is how point particles and photons move in it. In General Relativity, particles move on geodesics of the metric, i.e. the paths with an extremal lapse of proper time (for a massive particle) or ‘affine parameter’ (for a massless particle), along the worldline. If we let  $ds$  be the differential interval along a path, and  $s_{AB}$  the total interval between two points A and B on a given path, then with  $\dot{x}_\mu \equiv dx_\mu/ds$  we have*

$$\begin{aligned} s_{AB} &= \int_A^B ds = \int_A^B [g_{\mu\nu} dx^\mu dx^\nu]^{\frac{1}{2}} = \int_A^B \left[ g_{\mu\nu} \frac{dx^\mu}{ds} \frac{dx^\nu}{ds} \right]^{\frac{1}{2}} ds \\ &= \int_A^B G(x^\mu, \dot{x}^\mu) ds \quad \text{where} \quad G(x^\mu, \dot{x}^\mu) = [g_{\mu\nu} \dot{x}^\mu \dot{x}^\nu]^{\frac{1}{2}} \end{aligned}$$

(19)

and finding the path which extremises  $s_{AB}$  then leads to the Euler–Lagrange equations (one for each  $\mu$ )

$$\frac{d}{ds} \left( \frac{\partial G}{\partial \dot{x}^\mu} \right) - \frac{\partial G}{\partial x^\mu} = 0$$

(20)

Note using the result (6) is is easy to demonstrate the equivalence of (20) with the alternative, and perhaps more familiar form

$$\frac{dx^\mu}{ds^2} + \Gamma^\mu_{\nu\sigma} \frac{dx^\nu}{ds} \frac{dx^\sigma}{ds} = 0$$

(21)

The advantage of the former, (20), comes from the fact that it does not need knowledge of the connection coefficients to

compute it, and also conservation laws for variables of which the metric is not an explicit function, are easy to read off. Note from its definition (compare the second and fifth quantities in the chain of equalities (19)), that as a numerical quantity  $G(x^\mu, \dot{x}^\mu)$  evaluates to 1 (at least for a massive particle), which we can use to simplify formulae once its functional dependence has already been used.

So we now apply these results to the metric in the general static form given by (7). (The advantage of carrying this out for the general form, rather than just Schwarzschild, is that it enables us to consider results for other types of black hole, such as Reissner- Nordstrom, and Schwarzschild-de Sitter – see below.) We can assume w.l.o.g. that the particle motion is confined to the  $\theta = \pi/2$  plane, and then the equations we find are

$$A\dot{t}^2 - B\dot{r}^2 - r^2\dot{\phi}^2 = 1 \quad (22)$$

coming from  $G(x^\mu, \dot{x}^\mu) = 1$ , and

$$A\dot{t} = k, \quad \text{and} \quad r^2\dot{\phi} = h \quad (23)$$

coming from the Euler-Lagrange equations in  $t$  and  $\phi$  respectively, and where  $k$  and  $h$  are constants. These last two equations result from the constancy of the metric coefficients in  $t$  and  $\phi$  and correspond to the conservation of energy and angular momentum. For a particle of mass  $m$ ,  $kmc^2$  can be identified as the particle energy, and  $h$  is the specific angular momentum per unit mass.

There is no point forming the Euler-Lagrange equation in  $r$ , since we already have a simple expression for  $\dot{r}$  available from the combination of (22) and (23), yielding

$$\dot{r}^2 = \frac{1}{B} \left( \frac{k^2}{A} - \frac{h^2}{r^2} - 1 \right) \quad (24)$$

We return to this general form later, but now wish to obtain results specific to the Schwarzschild case. With  $A = B^{-1} = (1 - 2GM/r)$ , and reinstating  $c$  for the remainder of this section, we obtain

$$\dot{r}^2 = c^2(k^2 - 1) - \frac{h^2}{r^2} \left(1 - \frac{2GM}{c^2 r}\right) + \frac{2GM}{r}. \quad (25)$$

Using the usual Newtonian substitution  $u \equiv 1/r$ , and changing the independent variable to azimuthal angle  $\phi$  rather than the interval  $s$ , so that it is easier to discuss the shape of the orbit, we differentiate (25) to obtain

$$\frac{d^2 u}{d\phi^2} + u = \frac{GM}{h^2} + \frac{3GM}{c^2} u^2. \quad (26)$$

In Newtonian gravity, the equation for planetary orbits, in the same notation, is

$$\frac{d^2 u}{d\phi^2} + u = \frac{GM}{h^2}, \quad (27)$$

so we have almost got the Newtonian answer, except for the extra term  $3GM u^2 / c^2$ . This is what gives all the relativistic effects, and we note it correctly goes to zero as  $c \rightarrow \infty$ .

### 3.4 Circular orbits

An important application of the orbit formula in the context of high energy astrophysics, is what it tells us about circular orbits in Schwarzschild geometry. These will be approximately the orbits of material accreting onto black holes, since infalling material nearly always has angular momentum, and we would not generally expect direct radial infall.

If  $r$  is constant, then our equation for  $u$  (26) yields

$$\hbar^2 = \frac{GMr^2}{r - 3GM/c^2} . \quad (28)$$

Putting  $\dot{r} = 0$  in [equation \(25\)](#) gives us

$$\frac{\hbar^2}{c^2 r^2} = \frac{k^2 r}{r - \frac{2GM}{c^2}} - 1 \quad (29)$$

and then putting both these last two results together yields an equation for  $k$  in terms of  $r$  alone. We derive:

$$k = \frac{1 - \frac{2GM}{rc^2}}{\sqrt{1 - \frac{3GM}{rc^2}}} . \quad (30)$$

Remembering  $k = E/mc^2$ , where  $E$  is the particle energy, we have found that the energy of a particle in a circular orbit is

$$E_{\text{circ}} = mc^2 \frac{1 - \frac{2GM}{rc^2}}{\sqrt{1 - \frac{3GM}{rc^2}}} . \quad (31)$$

An obvious check on this equation, is whether it can reproduce the Newtonian expression for the total energy of a circular orbit in the limit of large  $r$ . Using the binomial theorem we see that indeed the first two terms in an asymptotic expansion in  $r$  are

$$E_{\text{circ}} \sim mc^2 - \frac{GMm}{2r} + \dots , \quad (32)$$

Thus we get agreement at large  $r$  with the usual Newtonian expression (derived via

$$E_{\text{tot}} = K.E. + P.E. = \frac{1}{2}mv^2 - \frac{GMm}{r} = \frac{-GMm}{2r} \quad \text{if} \quad \frac{mv^2}{r} = \frac{GMm}{r^2})$$

(33)

*provided we realise that it enters as a correction to the rest mass energy  $mc^2$ , which is the dominant term.*

*The equation we have just found for the energy of a circular orbit, provides us with a remarkable amount of information about the nature of such orbits.*

*First we see that in the limit  $m \rightarrow 0$ , the orbit  $r \rightarrow 3GM/c^2$  is of interest, since the singularity in the denominator can cancel the zero at the top. In fact this is the circular photon orbit at  $r = 3GM/c^2$ , which we will derive below in a treatment of photon motion.*

*Secondly, we can see which orbits (for particles of non-zero rest mass) are bound. This will occur if  $E_{\text{circ}} < mc^2$ , since then we have less energy than the value for a stationary particle at infinity. The condition for  $E_{\text{circ}} = mc^2$  is that*

$$\left(1 - \frac{2GM}{rc^2}\right)^2 = 1 - \frac{3GM}{rc^2}$$

(34)

*This happens for  $r = 4GM/c^2$  or  $r = \infty$ . Thus over the range  $4 < r < \infty$  the circular orbits are bound.*

*This appears to show we can get as close as 2 Schwarzschild radii to a black hole for particles in a circular orbit, suggesting that this is where the inner edge of any accretion disc would be. But is such an orbit stable? We discuss this first in the specific context of the Schwarzschild metric, and then later look at stability in the context of more general metrics of the form (7).*

### **3.5 Stability of circular orbits**

*In Newtonian dynamics the equation of motion of a particle in a central potential is*

$$\frac{1}{2} \left( \frac{dr}{dt} \right)^2 + V(r) = E, \\ (35)$$

where  $V(r)$  is an “effective potential”. For an orbit around a point mass, the effective potential is

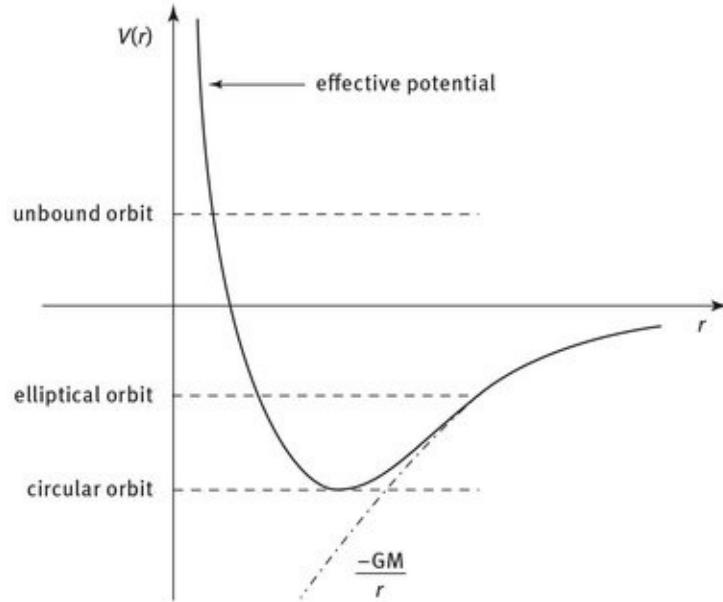
$$V(r) = \frac{h^2}{2r^2} - \frac{GM}{r}, \\ (36)$$

where  $h$  is the specific angular momentum of the particle. Since the  $1/r^2$  term will eventually always exceed the  $1/r$  term as  $r \rightarrow 0$ , we can see that in Newtonian dynamics a non-zero angular momentum provides an angular momentum barrier preventing a particle reaching  $r = 0$  — see [Figure 2](#). In this effective potential, bound orbits have two turning points and a circular orbit corresponds to the special case where the particle sits at the minimum of the effective potential. However, as we have already partially seen in [equation \(48\)](#), the same is not true in General Relativity.

Starting with [equation \(25\)](#) we can rewrite this as

$$\frac{1}{2}\dot{r}^2 + \frac{h^2}{2r^2} \left( 1 - \frac{2GM}{c^2 r} \right) - \frac{GM}{r} = \frac{1}{2}c^2(k^2 - 1) \\ (37)$$

where we recall  $k = E_{\text{part}}/mc^2$  and  $r^2 \dot{\phi} = h$ .



**Fig. 2.** The Newtonian effective potential showing how an angular momentum barrier prevents particles reaching  $r = 0$ .

Thus, although the r.h.s. is not the particle energy here, the fact that it is constant tells us that

$$U(r) = \frac{h^2}{2r^2} \left( 1 - \frac{2GM}{c^2 r} \right) - \frac{GM}{r} \quad (38)$$

is an “effective potential” for the problem, which we can use to study stability in the same way as in the Newtonian case. Note that the relativistic term  $(1 - 2GM/c^2 r)$  weakens the centrifugal effect of angular momentum at small  $r$ .

Differentiating this expression,

$$\frac{dU}{dr} = -\frac{h^2}{r^3} + \frac{3h^2 GM}{c^2 r^4} + \frac{GM}{r^2}, \quad (39)$$

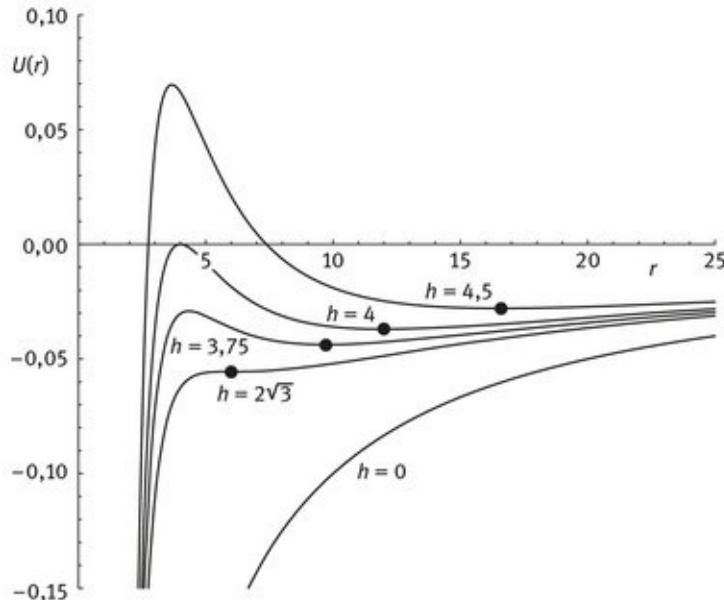
and so the extrema of the effective potential are located at the solutions of the quadratic equation

$$GMr^2 - h^2 r + \frac{3h^2 GM}{c^2} = 0, \quad (40)$$

i.e. at

$$r = \frac{h^2}{2GM} \left\{ 1 \pm \sqrt{1 - 12 \left( \frac{GM}{hc} \right)^2} \right\}. \quad (41)$$

If  $h = \sqrt{12} \frac{GM}{c}$  there is only one extremum, and there are no turning points in the orbit for lower values of  $h$ . At this point  $r = 6GM/c^2 = 3R_s$ . [Figure 3 shows](#) the effective potential for several values of  $h$ . The dots show the locations of stable circular orbits. The maxima in the potential are the locations of unstable circular orbits.



**Fig. 3.** The effective potential  $U(r)$  plotted for several values of the angular momentum parameter  $h$  (units here have  $GM/c^2 = 1$ ).

What is the physical significance of this result? The smallest stable circular orbit has

$$r_{\min} = \frac{6GM}{c^2}. \quad (42)$$

Gas in an accretion disc settles into circular orbits around the

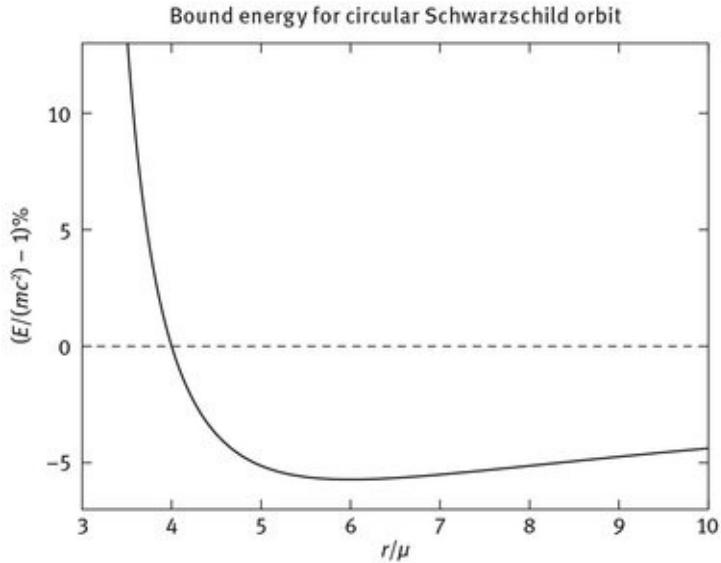
compact object. However, the gas slowly loses angular momentum because of turbulent viscosity (the turbulence is thought to be generated by magnetohydrodynamic instabilities). As the gas loses angular momentum it moves slowly towards the black hole, gaining gravitational potential energy and heating up. Eventually it loses enough angular momentum that it can no longer follow a stable circular orbit and so it falls into the black hole. On this basis, we can estimate the efficiency of energy radiation in an accretion disc via looking at a plot of the ‘fractional binding energy’  $E/(mc^2) - 1 = k - 1$  versus  $r$  (see [Figure 4](#)).

The maximum efficiency is of order the gravitational binding energy at the smallest stable circular orbit divided by the rest mass energy of the gas. From the plot we can see that this will be about 6%. More accurately, from [equation \(31\)](#) we see that at  $r = 6GM/c^2$   $k = E/(mc^2)$  is  $2\sqrt{2}/3$ , hence we obtain for this efficiency

$$\epsilon_{\text{acc}} \approx 1 - 2\sqrt{2}/3 \approx 5.7\% \quad (43)$$

The equivalent Newtonian value, is not far away at

$$\epsilon_{\text{acc}} \approx \frac{1}{2} \frac{GMm}{r_{\min}} \frac{1}{mc^2} \simeq \frac{1}{12} \sim 8\% . \quad (44)$$



**Fig. 4.** A plot of  $E/(mc^2) - 1$  in percent versus  $r$  (the latter measured in units of  $\mu = GM/c^2$ ) where  $E$  is the energy of a particle of mass  $m$  in a circular orbit at radius  $r$  about a Schwarzschild black hole.

*As we will see, this value can be even larger for a black hole with spin, and an accretion disc can convert 5-20 percent of the rest mass energy of the gas into radiation, depending on spin. This can be compared with the efficiency of nuclear burning of hydrogen to helium (26 MeV per He nucleus),*

$$\epsilon_{\text{nuclear}} \sim 0.7\%$$

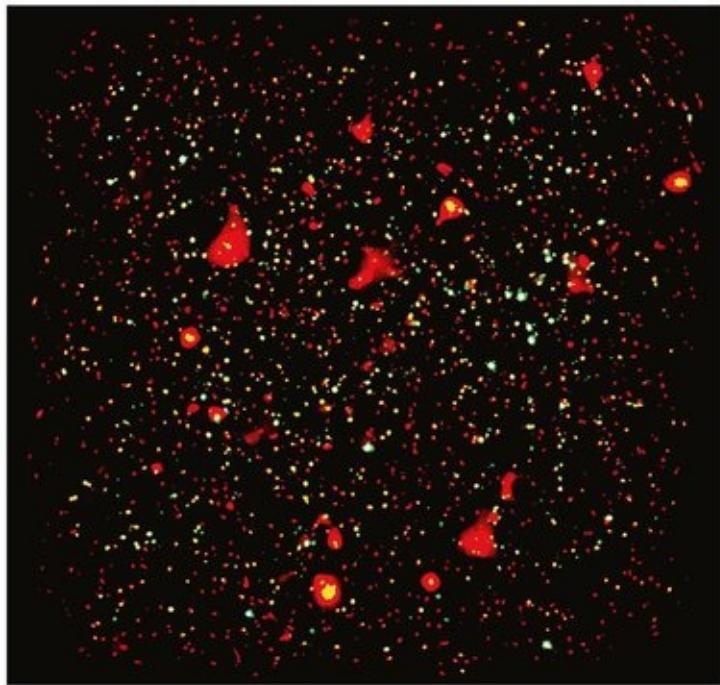
(45)

*Accretion discs are capable of converting rest mass energy into radiation with an efficiency that is about 10 times greater than the efficiency of nuclear burning of hydrogen. The ‘accretion power’ of black holes causes the most energetic phenomena known in the Universe.*

## 4 Evidence from X-rays, quasars and AGN

The early 1960s also saw the birth of X-ray astronomy, when in June 1962 Riccardo Giacconi and colleagues [25] launched a sounding rocket with Geiger counters sensitive to X-rays. As the

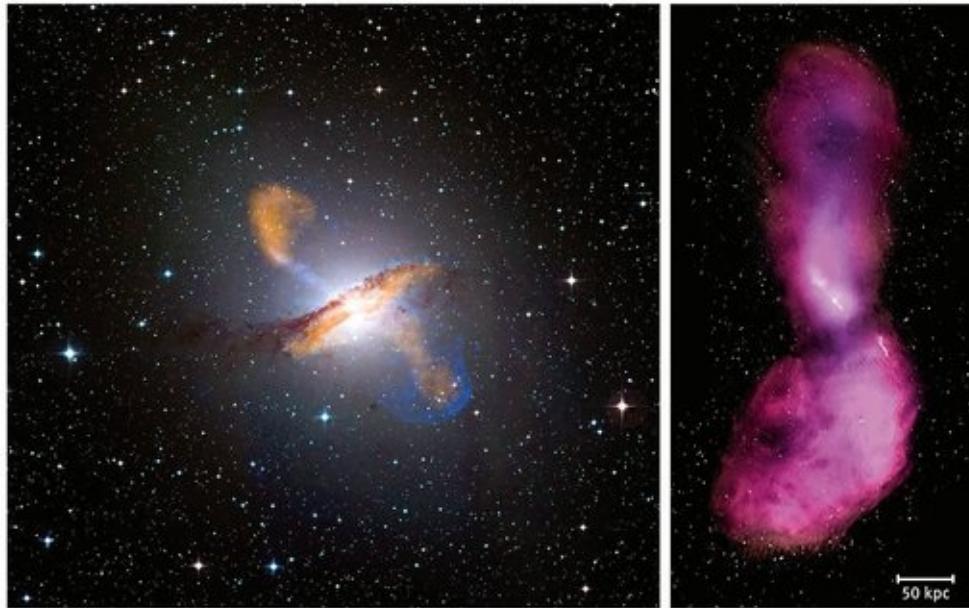
detectors scanned across the Sky during their 10 minutes above the Earth's atmosphere, a broad peak of emission was discovered from the general direction of the Scorpius constellation and a steady background of cosmic X-rays were seen. Sco X-1 and the cosmic X-ray Background had been discovered. The rapidly variable X-ray source Sco X-1 is a neutron star accreting matter from an orbiting companion star and the X-ray Background is now known to be the integrated emission from accreting supermassive black holes at the centres of distant galaxies (see [Figure 5](#)).



**Fig. 5.** The XMM COSMOS field is about 30 arcmin across and shows the X-ray Background resolved into pointlike and extended X-ray sources. The pointlike sources are mostly accreting black holes (distant AGN) and the extended sources are due to hot gas pooled in the gravitational potential wells of clusters and groups of galaxies.

The first clear association of X-ray astronomy and black holes was made after Giacconi and colleagues studied the brightest X-ray source in the constellation of Cygnus, Cyg X-1, using the first X-ray astronomy satellite, Uhuru, which continuously scanned the Sky. Cyg X-1 showed chaotic variability on timescales down to a fraction of a second. It has two relatively long-lived states, one when the spectrum is soft and the other

hard. A radio source was found to appear when it switched into the hard state and accurate measurement of the position of that source enabled it to be identified with a 7th magnitude B star, HDE 226868. Separate teams of optical astronomers, Webster & Murdin [76] and Bolton [5], then found that this massive star was being swung around at  $70 \text{ km s}^{-1}$  by an unseen object on a 5.6 day period. This enabled mass estimates to be made (which relied on understanding the mass of the companion B star) yielding at least  $3.5M_{\odot}$  which is above the likely upper mass limit for a neutron star of about  $3M_{\odot}$ . The small size implied by the rapid variability combined with the lower mass limit pointed to a black hole in Cyg X-1.



**Fig. 6.** Centaurus A is one the nearest radio-loud AGN. The image on the left is a combination of infrared through X-ray images, its jets are seen going from the NE to SW. On the right is a larger scale image of low frequency radio emission showing the large diffuse lobes of radio-emitting plasma ejected by the accreting black hole out to intergalactic space.

Many other black hole X-ray binary systems are now known, some of which have much clearer mass estimates. M33 X-1, for example, in the nearby galaxy M33 has a precise distance and stellar mass estimate and also shows eclipses which enable the orbital inclination to be deduced, leading to a mass for the X-ray

source of  $15.64 \pm 1.45 M_{\odot}$  [51].

By the 1980s, many quasars and X-ray binary systems were known, showing that black holes are common. Quasars were seen to be an extreme part of the more general phenomenon of Active Galactic Nuclei (AGN). The centres of a few percent of all galaxies appear to have some non-stellar activity in the form of bright broad emission lines, a nonthermal radio source or an X-ray source. Quasars occur when the AGN outshines the host galaxy in the optical band. AGN of lesser power are known as Seyfert galaxies or just low luminosity AGN as the power drops. Jets of highly-collimated relativistically outflowing plasma are found in about 10% of AGN. An example outflow in one of the nearest AGN is shown in [Figure 6.](#)

Unusual emission spectra at the centres of some galaxies had been known as a phenomenon since about 1908, and the first jet was reported in M87 by Heber Curtis in 1918 [12]. Little follow-up work was done before the 1960s, apart from Carl Seyfert's PhD thesis in the 1940s [64]. By the end of the 1980s most astronomers considered that AGN were powered by accretion onto black holes, but a minority still favoured some other explanation, such as multiple supernova outbursts. The arguments for black holes were strong but circumstantial. A major observational problem was that black holes are by definition difficult to observe directly.

Quasars appear to be associated with a past phase in the evolution of the Universe, when it was 1–5 billion years old (mostly at redshifts of 2–4). They are now relatively uncommon compared with that era. (3C273 is a rare quasar at low redshift: low is relative here as it was moderately high when discovered in the early 1960s.) The enormous powers involved mean that the accretion rates were high and the mass doubling time of the black holes could have been a few 100 million years. Radiation pressure would have restricted the inflow to below a limit first deduced for stars by Eddington, known as the Eddington limit. The limit is obtained by comparing the force of radiation acting on an electron (through Thomson scattering of cross-section  $\sigma_T$ ) at the surface of the object, with the force of gravity on a proton

there. The electron and proton, although assumed free, are bound together by electrostatic attraction.

$$F_{\text{grav}} = \frac{GMm_p}{R^2} = F_{\text{rad}} = \frac{L}{4\pi R^2 h\nu} \sigma_T \frac{h\nu}{c},$$

(46)

where the radiative force term is the flux of photons (of typical energy  $h\nu$ ) times the cross-section times the momentum of a photon. This gives

$$L_{\text{Edd}} = \frac{4\pi GMm_p c}{\sigma_T},$$

(47)

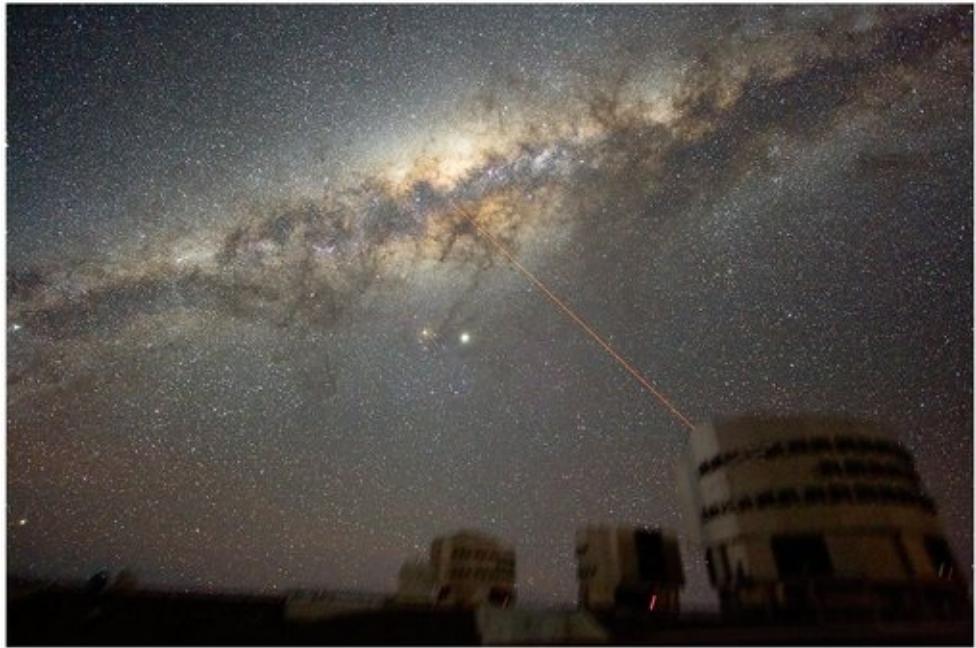
which is about  $10^{31}$  W per Solar mass.

Since the Eddington limit is proportional to the mass of the black hole the black hole mass can grow exponentially, if sufficient fuel is available. If growth is at the Eddington limit, the e-folding time of the black hole mass is about  $4 \times 10^7$  yr. This timescale is sufficient for the typical massive black hole to have grown from stellar mass (say  $30M_\odot$ ) but it does become challenging for the most distant billion Solar mass quasars found above redshift 6.

## 5 The black hole at the centre of the Milky Way

### 5.1 Sgr A\*

The evidence for the existence black holes changed though the 1990s owing to careful observations of the motion of stars moving around the centre of our own Milky Way galaxy, first starting in 1991 by Reinhard Genzel and colleagues [23] using ESO telescopes and later joined by Andre Ghez and colleagues [24] using the Keck telescopes. An interesting picture of the technique used to achieve the required image stability for this work is shown in [Figure 7](#).

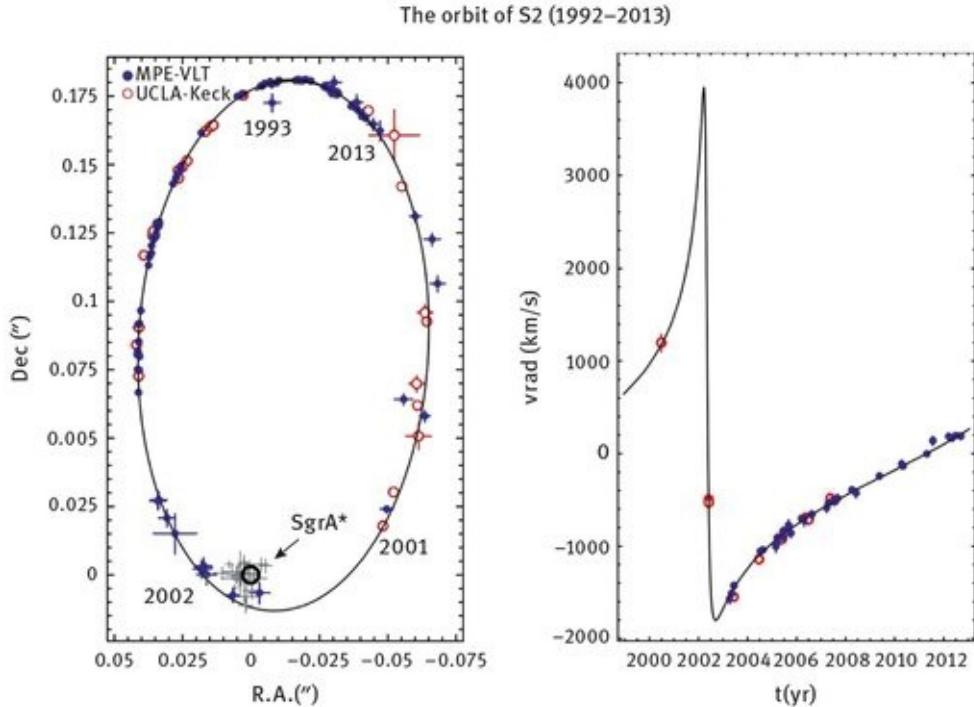


**Fig. 7.** The central stellar bulge of our Galaxy. A laser beam from one of the ESO Very Large Telescopes points at the very centre where Sgr A\* resides. This creates an artificial star in the Earth's upper atmosphere which is used to stabilise images made of the stars orbiting the central black hole.

Most stars, like the Sun, orbit the centre of the Galaxy at about  $230 \text{ km s}^{-1}$ , but within the innermost light year they move faster until at a distance of  $2000r_g$  from the dynamical centre a star orbits at up to  $3000 \text{ km s}^{-1}$ . The inferred mass within that orbit is 4 million  $M_\odot$  and the dynamical centre coincides with a strange radio source in the constellation of Sagittarius long known as Sgr A\*. It is also a flickering infrared and X-ray source. No stellar emission which can be attributed to say a massive cluster of normal stars is seen. The stellar orbits of dozens of stars there are known in 3D, both from their apparent motions mapped in the plane of the Sky and from line-of-sight Doppler shifts. The orbit of one of the closest stars, S2, is shown in [Figure 8](#) [26]. The mass density within the innermost orbits exceeds  $10^{18} M_\odot \text{ pc}^{-3}$  and there is nothing known to Physics which can lie there other than a black hole. Clusters of stars, even neutron stars, would interact and collide. The above lower limit on the mass density of the central object is increased by a further 5 orders of magnitude when constraints on the

maximum size of Sgr A\* obtained from millimetre band Very Long Baseline Interferometry (VLBI) are included.

The origin of the young, massive stars which are bright enough to be tracked around the Galactic Centre is not yet fully understood. The underlying stellar density there is so high that it should inhibit further star formation.



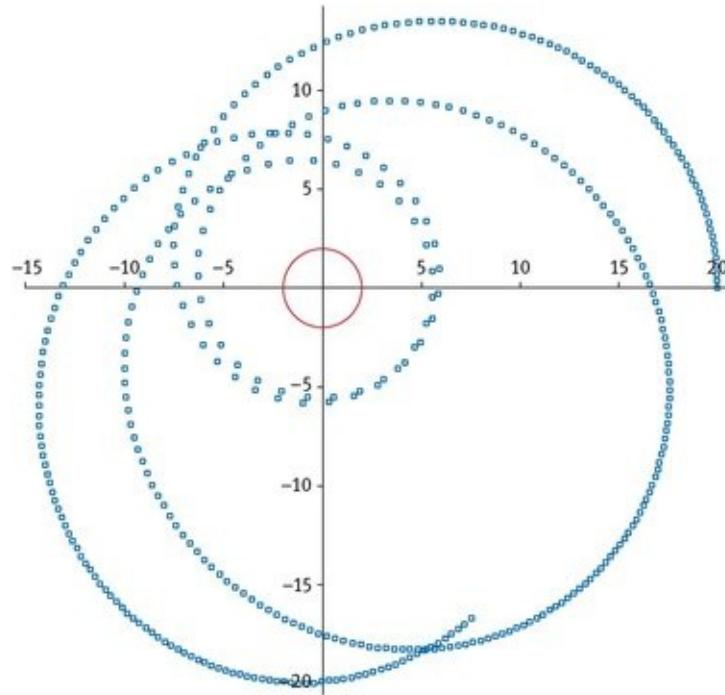
**Fig. 8.** The path of star S2 about Sgr A\*. The elliptical orbit has a period of 15.2 years, a major axis of 5.5 light days and is inclined at 46 degree to the plane of the Sky. It requires that a mass of  $4 \times 10^6 M_\odot$  lies within a radius of 17 light hours. The only long-lived object which is physically consistent is a massive black hole.

As well as the issues of mass within a given radius, there is also hope that general relativistic effects on the orbits of stars and possibly gas clouds near Sgr A\* will be observed in the near future, so we discuss now the details of these effects, including the important aspect of ‘capture’ by a black hole.

## 5.2 GR effects on orbits

We derived above the ‘shape’ equation for orbits around a Schwarzschild black hole, (26). For the Newtonian case, then

due to the harmonic kernel at the left, and the constant at the right, it is obvious that solutions of the form  $u = a + b \cos \varphi$  will work, i.e. ellipses , so if the extra term in the relativistic version (26) is small we expect the orbits to be modified ellipses. In the Solar System, in which a Schwarzschild metric due to the Sun applies, these modifications are very small (e.g. the largest effect is the 42 seconds per century precession of Mercury's orbit), but in the Schwarzschild metric around a black hole, we can expect much larger effects. For example, in [Figure 9](#) we show the motion for an object starting about 10 Schwarzschild radii out in an almost circular orbit. The gaps between the markers on the orbit are laid down at equal intervals of time, and so indicate how fast the object is moving. We can see that strongly precessing ellipses are obtained, but with a motion that looks as though it will continue indefinitely (which it does).



**Fig. 9.** Advance of perihelion in a Schwarzschild metric. The units of distance are  $G M/c^2$  , where  $M$  is the mass of the central body. The particle is launched tangentially and given a specific angular momentum  $h$  of 3.75 where the circular orbit  $h$  would be 4.85. The circle marks the Schwarzschild radius of the central black hole.

However, in [Figure 10](#), we show what happens when starting from the same point, and also moving tangentially, but now with

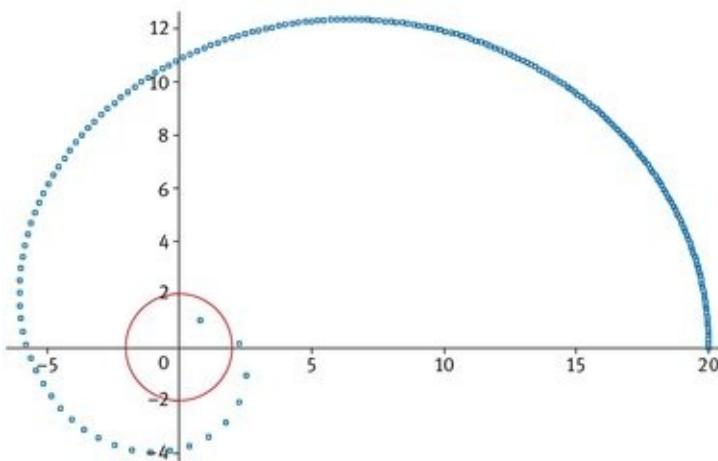
a smaller specific angular momentum. This time, the object is clearly ‘captured’ by the black hole, and ends up crossing the black hole horizon.

We can gain some insight into how the ‘capturing’ comes about, by considering the orbit equation in a different form. For a massive particle, the interval  $s$  we have been using is just the proper time of the particle, which we label  $\tau$ . If we start with [equation \(25\)](#), differentiate with respect to interval  $\tau$  and then remove first derivatives,  $d r/d \tau$  using the original equation again, one arrives at the following:

$$\frac{d^2r}{d\tau^2} = -\frac{GM}{r^2} + \frac{h^2}{r^3} - \frac{3h^2GM}{c^2r^4}.$$

[\(48\)](#)

The first two terms are very Newtonian-like, and correspond to an inward gravitational force and a repulsive term, proportional to angular momentum squared, which is basically the ‘centrifugal force’. What is new is the third term, also proportional to angular momentum squared, but this time acting inwards. This shows that close to the hole, specifically within the radius  $r = 3GM/c^2$ , centrifugal force ‘changes sign’, and is directed inwards, thus hastening the demise of any particle that strays too close to the hole, as in the example of [Figure 10](#).



**Fig. 10.** As for preceding figure but now for a particle projected with  $h=3.5$  (in units with  $G M/c^2=1$ ).

### 5.2.1 Orbital precession

The equation we wish to solve is

$$\frac{d^2u}{d\phi^2} + u = \frac{GM}{h^2} + \frac{3GM}{c^2}u^2 \quad \left( u \equiv \frac{1}{r} \right) \quad (49)$$

in the limit that the departure from Newtonian motion is small. This would apply to the motion of the planets in our solar system for example.

The Newtonian solution to this equation is

$$u = \frac{GM}{h^2}(1 + e \cos \phi), \quad (50)$$

and so we can use this as a first approximation, and then iterate to get a better one. Substituting into the r.h.s. of (49), we obtain the new equation

$$\frac{d^2u}{d\phi^2} + u = \frac{GM}{h^2} + \frac{3(GM)^3}{c^2 h^4} (1 + e \cos \phi)^2. \quad (51)$$

This will be solved by  $u = \frac{GM}{h^2}(1 + e \cos \phi) +$  the particular integral (P.I.) of the equation:

$$\frac{d^2u}{d\phi^2} + u = A(1 + 2e \cos \phi + e^2 \cos^2 \phi) \quad (52)$$

where  $A = 3(GM)^3/c^2 h^4$  is very small. The P.I. can be found to be:

$$A \left( 1 + e \phi \sin \phi + e^2 \left( \frac{1}{2} - \frac{1}{6} \cos 2\phi \right) \right). \quad (53)$$

Now, in this expression, the first and third terms are tiny, since

*A is. However, the second term,  $Ae\phi \sin \phi$ , might be tiny at first, but will gradually grow with time, since the  $\phi$  part (without a cos or sin enclosing it) means it is cumulative. We must therefore retain it, and our second approximation is*

$$u = \frac{GM}{h^2} (1 + e \cos \phi + \delta e \phi \sin \phi) \quad (54)$$

*where*

$$\delta = \frac{3(GM)^2}{h^2 c^2} \ll 1. \quad (55)$$

*Using*

$$\cos(\phi(1 - \delta)) = \cos \phi \cos \delta \phi + \sin \phi \sin \delta \phi \approx \cos \phi + \sin \phi \delta \phi \quad \text{for } \delta \ll 1, \quad (56)$$

*we can therefore write*

$$u \approx \frac{GM}{h^2} (1 + e \cos[\phi(1 - \delta)]) . \quad (57)$$

*u is therefore periodic, but with period  $\frac{2\pi}{1-\delta}$ . The r values thus repeat on a cycle which is slightly larger than  $2\pi$ , and we find*

$$\Delta\phi = \frac{2\pi}{1-\delta} - 2\pi \approx 2\pi\delta = \frac{6\pi(GM)^2}{h^2 c^2} . \quad (58)$$

*But from the geometry of the ellipse, and the Newtonian solution, where we know  $l = h^2/GM$  and  $l = a(1 - e^2)$ , we can get the final result:*

$$\Delta\phi = \frac{6\pi(GM)^2}{c^2 l(GM)} = \frac{6\pi GM}{a(1 - e^2)c^2} . \quad (59)$$

For example, Mercury's orbit has  $a = 5.8 \times 10^{10}$  m, eccentricity  $e = 0.2$  and we know  $M_{\odot} = 2 \times 10^{30}$  kg. Therefore our prediction for the precession is

$$\begin{aligned}\Delta\phi &= 5 \times 10^{-7} \text{ radians per orbit} \\ &= 0''.1 \text{ per orbit.}\end{aligned}$$

Since Mercury's orbital period is 88 days, we would thus expect to accumulate a precession of 43'' per century. This is what is observed after correction for the perturbations due to the other planets, which cause a total precession of more like 5000'' per century.

Turning now to the Galactic Centre, the star S2, whose orbit was shown in [Figure 8](#), has an eccentricity of 0.876 and semi-major axis of 980 AU [16]. Taking the mass of the Galactic Centre black hole as  $4 \times 10^6 M_{\odot}$ , then our formula yields a precession of 11 arcmin per orbit, which takes about 15 years. This sounds large, but projected on the sky at perihelion amounts to only about 0.5 milliarcsec. By the use of near-infrared interferometry, this may be observable in the relatively near future (see Section 10), but as for the solar system it is competing effects from other nearby masses, and the general matter distribution near the centre, that will be key to determining whether it is the GR effect itself which is being seen [44].

## 6 Galaxies and black holes

### 6.1 AGN feedback

It was earlier unclear whether galaxies with central black holes were special in some way or not. Quasars could be long lived in just a few galaxies or short lived and occur in every galaxy. Then careful imaging of the centres of nearby galaxies with HST and other telescopes revealed through the 1990s that most galaxies host a massive black hole. The mass of the black hole

is correlated with the mass of the galaxy bulge<sup>10</sup> that it is embedded in (the Magorrian relation [37]). Current results indicate that the host bulge has a mass 500 times that of the black hole. Some argue that the correlation is better if the velocity dispersion of the surroundings stars (beyond the radius where the black hole's mass dominates) is used (the  $M_{BH} - \sigma$  relation – an example of this is shown in [Figure 11](#)).

Both correlations show considerable scatter but hold over several orders of magnitude.

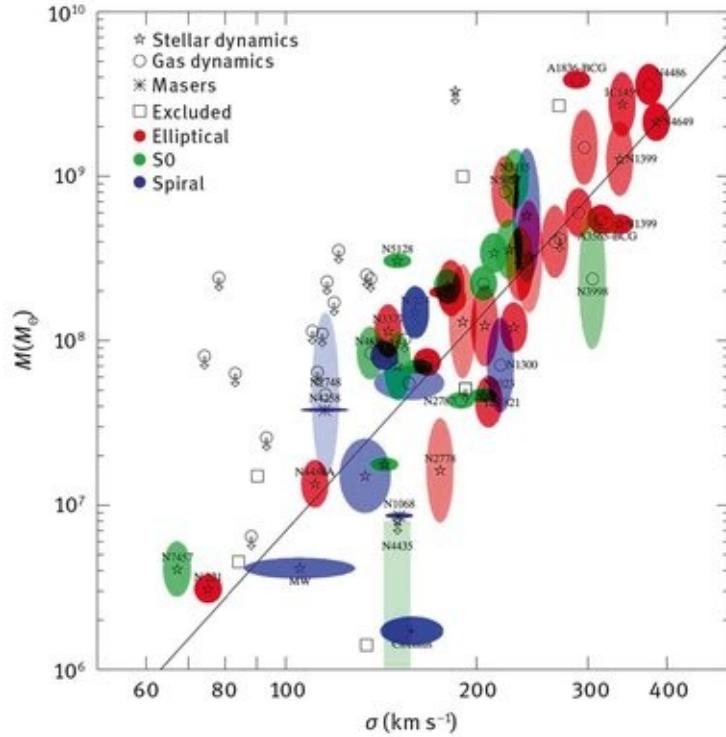
The  $M-\sigma$  relation poses the question of how the black hole at the centre of a galaxy knows the total mass of the host galaxy? The ratio of the physical size of the black hole event horizon to the galaxy can be 100 million or more (a similar scaled comparison is the Earth and a football). Some astronomers argue that larger galaxies just have larger black holes, but the correlations look better than that. An exciting possibility is that the question should be turned round and it is the black hole which controls the mass of the host galaxy. In 1998 Silk & Rees [66] proposed that it was the phenomenal energy output in the quasar phase, as the black hole grew in mass by accretion, which was responsible. Both grow from gas in the galaxy – the galaxy forming stars from the gas, the black hole accreting the gas. Eventually the black hole is so big and powerful that it blasts all the gas from the galaxy so both stop growing and we are left with the observed relation.

There is clear evidence that massive black holes have grown by accretion from the Soltan relation [67]:  $\eta$

$$\mathcal{E}_{\text{acc}}(1+z) = \frac{\eta}{(1-\eta)} \rho_{\text{BH}} c^2 , \quad (60)$$

where  $z$  is the mean redshift at which the accretion occurs. It is a recasting of the famous equation,  $E = mc^2$ , in terms of densities. The factor of  $1 + z$  is due to the redshift-ing of the energy of the radiation; there is no such factor for mass.  $\eta$  is the radiative efficiency of the accretion process  $L = \eta \dot{M}c^2$ . The

energy density of radiation from accreting black holes,  $\mathcal{E}_{\text{acc}}$  can be measured from the summed spectra of AGN (and the X-ray Background) and  $\rho_{\text{BH}}$  can be estimated from the mass function of galaxies together with the  $M_{\text{BH}} - M_{\text{gal}}$  relation. Results show agreement if  $\eta$  is about 0.1 which, as discussed later, is typical for luminous accretion.



**Fig. 11.** The black holes mass – stellar velocity dispersion relation  $M_{\text{bh}} - \sigma$  [27].

It is straightforward to show that the growth of the central black hole by accretion can have a profound effect on its host galaxy. If the velocity dispersion of the galaxy is  $\sigma$  then the binding energy of the galaxy bulge, mass  $M_{\text{gal}}$ , is  $E_{\text{gal}} \approx M_{\text{gal}} \sigma^2$ . The mass of the black hole is on average observed to be  $M_{\text{BH}} \approx 2 \times 10^{-3} M_{\text{gal}}$ . If the radiative efficiency of the accretion process of 10%, then the energy released by the growth of the black hole is given by  $E_{\text{BH}} = 0.1 M_{\text{BH}} c^2$ . Therefore  $E_{\text{BH}}/E_{\text{gal}} \approx 2 \times 10^{-4} (c/\sigma)^2$ . Very few galaxies have  $\sigma < 350 \text{ km s}^{-1}$ , so  $E_{\text{BH}}/E_{\text{gal}} > 100$ . The energy produced by the growth of the black hole therefore

exceeds the binding energy of the host galaxy by about two orders of magnitude!

If even a small fraction of the accretion energy can be transferred to the gas, then an AGN can have a profound effect on the evolution of its host galaxy. In practice, radiation from the accreting black hole will not have any significant influence on existing stars in the galaxy. It can however strongly influence gas clouds from which new stars can potentially form. By ejecting those clouds from the galaxy, AGN feedback can effectively stop further stellar growth of a galaxy.

The original mechanism of Silk & Rees is actually a little too effective and although it predicts a slope to the relation which is acceptable, the normalization is too low. Gas is blasted out too easily. However, energy is likely to be radiated away in the process which is likely to be controlled by conservation of momentum [18, 32, 45]. (Rocket science would be much easier if only energy conservation were at stake.) It has since been shown that this leads to the correct normalization and a slightly flatter slope. The details now centre [19] on the precise mechanisms responsible for the ejection (e.g. radiation pressure or accretion disc winds). Energetic winds are seen from some AGN and UV radiation interacts strongly with dusty gas in all quasars exerting an outward force through radiation pressure.

It is interesting to note that the effective Eddington limit for dusty gas exposed to the ultraviolet radiation of a quasar is about 500 times that of ionized gas such as expected close to the quasar. The  $M - \sigma$  relation means that when the quasar is at its Eddington limit locally, the host galaxy is also at its effective Eddington limit globally.

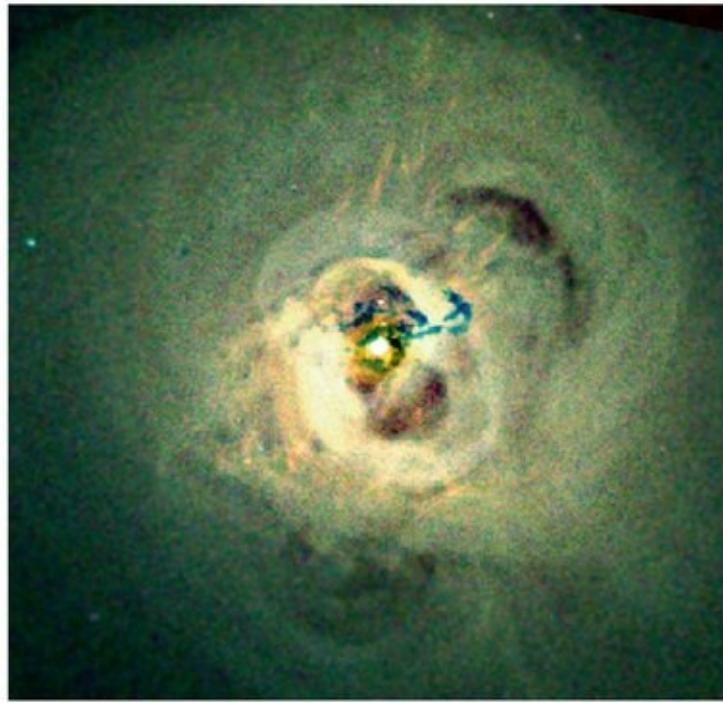
Dramatic evidence of AGN feedback occurs in the cores of many clusters of galaxies. The intergalactic medium in clusters has been squeezed and heated by gravity to create an intracluster medium which is 10 million K or hotter. The baryons in the hot gas exceed all baryons in the galaxy members of the cluster by a factor of more than ten. The most massive galaxies known lie at the dense centres of clusters where accretion onto their massive black holes feeds energy into the intracluster

medium, preventing it from cooling down due to the emission of X-rays. The total mass of gas in a cluster and the depth of the gravitational well is so large that the gas cannot be expelled. Energy instead flows from the black hole accretion process into the gas to maintain a thermodynamic status quo.

The energy flows from the near vicinity of the black hole in the form of relativistic jets which push the hot gas back, forming giant galaxy-sized bubbles of relativistic plasma (cosmic rays and magnetic field) – an example is shown in [Figure 12](#).

The bubbles are buoyant and when large detach and float upward in the cluster potential, while new bubbles form and grow. The power in the bubbling process can be estimated from the energy content of a bubble ( $4PV$  where  $P$  is the pressure of the surrounding gas and  $V$  is the volume of the bubble, the factor 4 being appropriate for relativistic plasma) divided by the bubble growth time. This power matches the radiative energy loss of the inner cluster core, indicating that close feedback has been maintained for billions of years [11].

AGN feedback implies that black holes play a central role in the final growth and evolution of all massive galaxies. If black holes did not exist then massive galaxies would be much larger and brighter than we now see.



**Fig. 12.** X-ray image of the core of the Perseus cluster centred on the giant galaxy NGC1275. The image is about 100 kpc or 300 thousand light years across. The AGN at the centre has blown 10 kpc diameter bubbles of relativistic plasma into the surrounding cluster gas. Detached buoyant bubbles are seen to the NW and S. Concentric ripples are seen in the hot gas due to the bubbling activity created by the AGN. It is clear that a black hole can have dramatic effects on gas structures both in and beyond the host galaxy.

## 6.2 Jets, Gamma-Ray Bursts and the birth of black holes

Relativistic jets occur in about 10% of AGN and in low-state BHB. They are thus a marker of the presence of a black hole. The details of how the material is accelerated and collimated are still unclear, with possibilities discussed later in this Chapter. Apparent superluminal motion indicates that the bulk Lorentz factor in many jets  $\Gamma \sim 10$ . Polarization indicates that the radio and much other emission are due to synchrotron radiation. The jets therefore contain electrons and a component with positive charge is expected, unless the jets represent enormous currents. Whether the positive particles are positrons or protons is not known. If they are protons (or even heavier nuclei), then the power in the jets can be huge, exceeding the radiative power in some objects.

Jets are presumably accelerated magnetically by the accretion disc. Fields emerging from the disc are continuously wound up around and along the central axis. The details are not yet fully understood. Jets are seen from all classes of object exhibiting accretion discs, including young stars and accreting white dwarfs. A long discussed possibility from quasar jets is that in some cases the spin energy of the black hole is being tapped by the Blandford-Znajek mechanism, which is a variant of the Penrose process (see [Sections 12.5](#) and [12.7](#) below). There has been recent progress in successfully modelling jets using numerical simulations of this mechanism.

The most relativistic jets occur in Gamma-Ray Bursts (GRB). These were discovered in the late 1960s by the Vela spacecraft which had a gamma-ray detector to monitor the Cold War Test Ban Treaty (but not made public by Ray Klebesadel and colleagues from Los Alamos until 1973 [33]). They are intense flashes of gamma-rays which have a flux which would make them readily visible to the naked eye if in the optical band. Many theories of their origin were proposed, mostly based on neutron stars. It was not until the late 1990s that a GRB was localized and identified with a distant galaxy, by the BeppoSax satellite, that made it clear that GRB are the most luminous events in the Universe, representing about a Solar rest mass of energy radiated within a few 10s of seconds.

Luminosity and causality point to enormous concentrations of gamma-rays in a very small region which would instantly lead to electron-positron pair creation creating a electron scattering dense fog, preventing any radiation emerging. It is difficult to see how gamma-rays can be seen unless strong relativistic beaming,  $\Gamma \sim 100$ , is involved. Many aspects of GRB, including the radio through X-ray afterglow all support this highly relativistic interpretation [57]. The source of GRB must also be clean of any other matter, which could prevent the radiation emerging. It may then be surprising then that the source of GRB is now believed to be the very centre of dying massive stars as a black hole is formed. The swirling matter powers intense jets which burrow through and out of the star, to produce the GRB

display. The birth of a black hole is therefore marked by one of the most amazing displays in the Universe.

Stars which give birth to black holes are estimated to start life at masses exceeding  $25M_{\odot}$ . Such stars have luminosities which are a million times greater than the Sun and live for just a few million years.

## 7 Current observations of accreting black holes

Apart from the nearest examples, we know of the presence of most black holes from the power emitted as radiation and jets produced as a result of accretion. Matter falling into a black hole is accelerated by gravity up to the speed of light and if collisions occur then the matter can become very hot and radiate. The efficiency of the process depends on the rate of collisions and on whether the matter has time to radiate before falling into the event horizon. If the flow is a thin dense accretion disc then collisions are plentiful and the efficiency depends on the ISCO (innermost stable circular orbit). Defining the radiative efficiency  $\eta$  through  $L = \eta \dot{M}c^2$ , then  $\eta$  ranges from 5.7% to 42% as the spin  $a/M$  increases from 0 to 1. In practice it is impossible to spin a black hole up to exactly 1, and high spins mean  $\eta \sim 20\text{-}30\%$ . Note that  $\eta$  for complete nuclear fusion (hydrogen to iron) is 0.7%, so black hole accretion is about an order of magnitude or more higher in efficiency. We now give some details of where these numbers come from, by looking at the orbits of massive particles around a Kerr black hole.

### 7.1 Particle motion in the Kerr metric

*Before discussing the details of particle motion, we need to consider the general structure of the Kerr solution, since this determines the character of the regions in which the particles are able to move.*

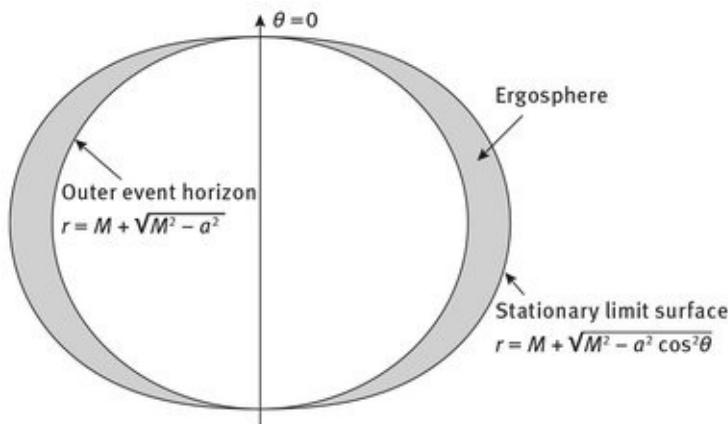
*The most important regions for astrophysical purposes are shown in [Figure 13](#). We have drawn an outer surface, called the*

'Stationary limit surface', and the one lying inside this is called the 'Outer event horizon'. Both of these are complemented by further surfaces (not drawn) lying closer to the black hole centre (the 'Inner Stationary limit' and 'Inner horizon' surfaces), but these are shielded by the outer horizon, and so do not have an immediate astrophysical role.

The stationary limit surface marks the point where it is no longer possible for an observer to remain stationary in the  $(r, \theta, \phi)$  coordinate system. Try as he or she might, it is inevitable that they will be swept around by the rotation of the hole. Mathematically, this corresponds to the point where the  $g_t$  Killing vector corresponding to invariance of the metric under time translations, changes from timelike to spacelike, i.e. to where the  $g_{tt}$  component of the metric (18) passes through zero. This requires (setting  $c = G = 1$  for clarity in the expressions)

$$r^2 - 2Mr + a^2 \cos^2 \theta = 0, \quad \text{i.e.} \quad r = M \pm \sqrt{M^2 - a^2 \cos^2 \theta} \quad (61)$$

and the outer of these surfaces is shown.



**Fig. 13.** The stationary limit and outer horizon surfaces for the Kerr solution.

The other surface drawn corresponds to the point where the  $g_{rr}$  component of the metric (18) becomes infinite, and marks the position of the event horizon. This is the position from

beyond which a particle can never escape, whatever its mass or state of motion. It is different from the stationary limit, since particles in the inbetween region, although they cannot remain stationary, can in principle escape to infinity. This inbetween region is known as the ‘ergosphere’, since it is possible to arrange particle motions so that work can be extracted from it (see [Section 12.5](#) below). The position of the horizon itself can be found by requiring

$$\Delta = r^2 - 2Mr + a^2 = 0, \quad \text{i.e.} \quad r = M \pm \sqrt{M^2 - a^2} \quad (62)$$

and again the outer of these two surfaces is shown. Given that this radius, unlike the stationary limit surface, does not depend on  $\theta$ , it might be wondered why the surface is drawn oblate in the diagram. This is because that Boyer-Lindquist coordinates do not correspond to spherical polars even in the case where the mass vanishes, and an embedding of the 2d surfaces we are working with into Euclidean space (for the purposes of visualisation) results in ellipsoids even for  $r = \text{const.}$  – see Section 13.6 of [30] for details.

The Lagrangian methods used above in the Schwarzschild case also work here, and one finds quickly the following results for the  $t$  and  $\phi$  of a particle moving in the equatorial plane in terms of its conserved specific energy and specific angular momentum,  $k$  and  $h$ :

$$\begin{aligned} \dot{t} &= \frac{k((r+2M)a^2 + r^3) - 2haM}{r\Delta} \\ \dot{\phi} &= \frac{h(r-2M) + 2kaM}{r\Delta} \end{aligned} \quad (63)$$

As before, for the  $r$  equation it is simpler to use the fact that evaluated numerically the interval function  $G(x^\mu, \dot{x}^\mu)$  is 1 for a massive particle, and 0 for a photon. We start with discussing a massive particle, since our initial emphasis is on the energy liberated from accretion disc orbits. Here we find

$$\dot{r}^2 = k^2 - 1 + \frac{2M}{r} - \frac{1}{r^2} (h^2 - a^2(k^2 - 1)) + \frac{2M}{r^3}(h - ak)^2$$

(64)

We now have three equations we can integrate to find the motion. A key quantity for the energy and stability is the effective potential in  $r$ . We define this similarly to before via

$$\frac{1}{2}\dot{r}^2 + U_{\text{kerr}}(r) = \text{const.}$$

(65)

resulting in

$$U_{\text{kerr}}(r) = -\frac{M}{r} + \frac{1}{2r^2} (h^2 - a^2(k^2 - 1)) - \frac{M}{r^3}(h - ak)^2 ,$$

(66)

where an irrelevant overall constant has been ignored.

If we compare with the equivalent in the Schwarzschild case, [equation \(38\)](#), we can see that despite the increased complexity of the Kerr metric, the effective potential has terms in just  $1/r$ ,  $1/r^2$  and  $1/r^3$  as before, resulting in overall similar behaviour (at least for small  $a$ ). A difference, however, is that now the coefficients of the second and third terms depend on the particle energy as well as the angular momentum.

For a circular orbit, we can use the fact that  $\dot{r}$  and  $r \ddot{r}$  are zero to find expressions for the specific energy and angular momentum. For the prograde orbits we are currently considering, these turn out to be

$$k = \frac{1 - \frac{2M}{r} + a\sqrt{\frac{M}{r^3}}}{\sqrt{1 - \frac{3M}{r} + 2a\sqrt{\frac{M}{r^3}}}} , \quad h = \frac{\sqrt{Mr} - \frac{2aM}{r} + a^2\sqrt{\frac{M}{r^3}}}{\sqrt{1 - \frac{3M}{r} + 2a\sqrt{\frac{M}{r^3}}}}$$

(67)

An important quantity for us to find is the radius of the innermost stable circular orbit, expected to be the inner edge of an accretion disc around a Kerr black hole, and the value the

energy function takes there.

As in the Schwarzschild case, we can use the effective potential to do this, asking that its second derivative is positive where the first derivative vanishes. Calculating this for the  $U(r)$  in (66) leads to the relatively simple criterion for stability

$$r^2 - 6Mr + 8a\sqrt{Mr} - 3a^2 > 0$$

(68)

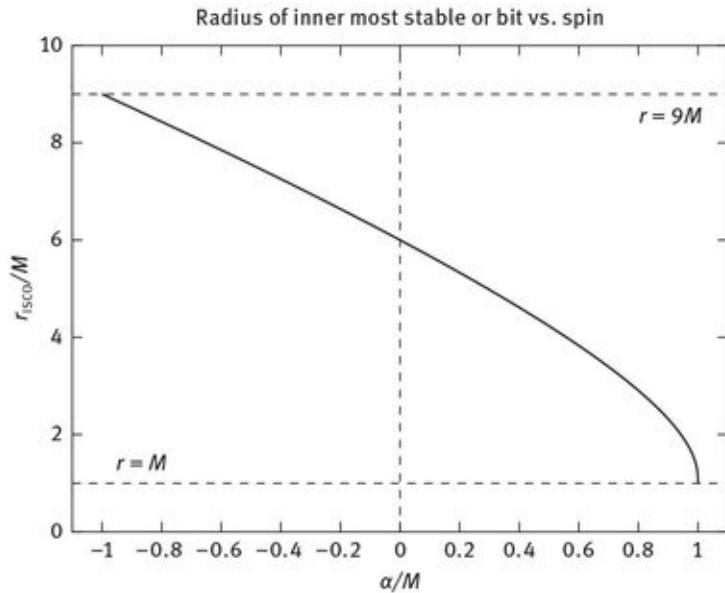
This is soluble analytically for  $r$ , but perhaps more illuminating to is to look at a plot of  $r_{\text{ISCO}}$  (the  $r$  corresponding to the innermost stable circular orbit) versus spin, as shown in [Figure 14](#). If  $a = M$ , an ‘extremal’ black hole, then the solutions are  $r = M$  and  $r = 9M$ . If  $a = 0$ , we get the single solution  $r = 6M$  corresponding to the Schwarzschild case. The former of these shows that stable circular orbits (in the prograde direction) persist right up to the event horizon for a black hole spinning at the maximum mathematically allowed rate, i.e.  $a = M$ . We note, however, that various effects are likely to intervene before this maximum spin rate can be achieved (mainly the counteracting torque felt by the hole in absorbing radiation from the accretion disc [70]), and  $a = 0.998M$  is considered the most likely maximum attainable value, leading to a minimum attainable  $r_{\text{ISCO}}$  of  $1.24M$ .

More generally, by eliminating  $r$  between (68) and the expression for  $k$  in (67) we can get a useful relation between the black hole spin and the efficiency of energy release for a particle which has reached the ISCO. For prograde orbits this relation is

$$\frac{a}{M} = \frac{2(2\sqrt{2}\sqrt{1-k^2} - k)}{3\sqrt{3}(1-k^2)}$$

(69)

This provides the plot of efficiency versus spin shown in [Figure 15](#). Again taking the maximum attainable value of  $a/M$  as 0.998, this leads to a maximum attainable efficiency of 32%.



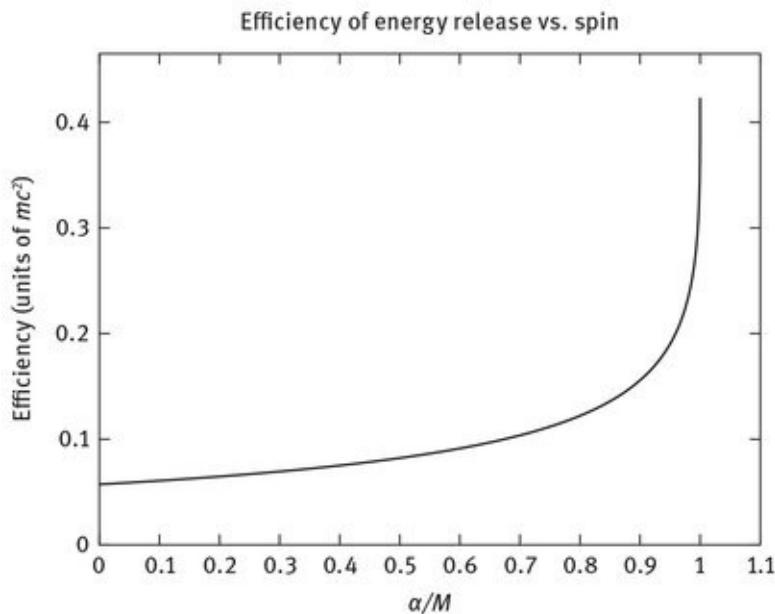
**Fig. 14.** Plot of radius of innermost stable orbit vs. spin.

## 7.2 Current observations of accreting black holes continued

The accreted matter can be supplied by a companion star, in the case of a BHB, or from surrounding gas in AGN. How much is captured by the black hole in the AGN case will depend on the density, temperature, magnetic field and angular momentum of the gas. The case when there is no angular momentum and magnetic fields can be ignored is known as Bondi accretion, which assumes that the gas can be treated as a fluid. Basically gas within the Bondi radius, defined as the radius beyond which the gas particles have the escape velocity, will be captured and this can be many orders of magnitude larger than would occur if the particles were free particles with no interactions or collisions. Zero angular momentum is a rather extreme assumption and may not apply to any real system, but it presents an easily calculated order of magnitude estimate of the accretion rate. Angular momentum is likely to reduce any inflow rate and could even choke off the flow. Indeed accretion flows can be unstable when heating by radiation from the central accreting mass is considered. Magnetic fields and turbulence also contribute to complexity and accretion flows are

commonly observed to show chaotic variability.

The accretion disc will be thin, with the ratio of thickness to radius,  $h/r < 0.1$ , when the luminosity is above about 1%  $L_{\text{Edd}}$ . At lower accretion rates the disc thickens as the collision rates drop which can then cause the efficiency to drop with further decreases in  $\dot{M}$ . (The black hole swallows both the matter and energy released.) The properties of such radiatively inefficient flows (RIAFS) depend on whether the electrons (which radiate) and the protons (which have much of the energy) remain coupled, *i.e.* have the same temperature. There are still theoretical issues remaining as to whether this happens in such a highly magnetized plasma, but observations do indicate changes in behaviour in objects with low accretion rates. Not least is a tendency for jets to become common. The low luminosity of Sgr A\* suggests it is in an extreme RIAF state.



**Fig. 15.** Plot of efficiency of energy release vs. spin (prograde orbits).

The luminous accretion flows, however, are usually above 1%  $L_{\text{Edd}}$  and so are thin and dense and optically thick resulting in much radiation being emitted as a black body. The spectrum of the disc is then produced by the sum of blackbodies from different radii which gives a  $\nu^{-1/3}$  spectrum at low energies up

to the highest energies, which come from the smallest radii, where it turns over into a Wien tail. The peak temperature scales as  $\dot{M}r_g^{-1/4}$ . This means that the most massive luminous black hole systems have spectral peaks in the UV, while stellar mass black holes peak in the X-ray band.

Significant power can also emerge as a result of magnetic fields wound up in the accretion flow. They can emerge from the disc, become twisted and create a mess of reconnecting field lines powering a hot corona, with a temperature which can easily be 100 times hotter than the disc itself. The million degree Solar corona with associated flares and outflows is a nearby example but even low luminosity AGN have a power  $10^{12}$  times larger than the Sun, so the analogy should not be taken too literally. The corona is the source of most of the X-rays seen from AGN. Evidence discussed later indicates that the corona in bright objects is located above and close to the black hole ( $\sim 10r_g$ ). The magnetic field structure also appears to power jets in some objects, and outflows in others. The degree to which the spin of the black hole is important (and maybe also a small ISCO) are still debated.

BHB have stellar-mass black holes and accrete from their companion star. Disc formation is implicit due to the orbital angular momentum. A few sources such as Cyg X-1 are persistent, with their long-term luminosity varying by less than an order magnitude, but sometimes switching states from hard (coronal dominated) to soft (thermal disc dominated). Many of the BHB are however transient, having outbursts every few years or decades, during which the luminosity can rise from  $10^{33}$  to  $10^{39}$  erg s $^{-1}$ . Outbursts follow a similar pattern in the luminosity-colour (hardness) plane, starting hard as the luminosity builds up towards  $L_{\text{Edd}}$ , then switching to being soft as the thermal disc emission dominates over the corona, before subsiding back to the hard state as they decline to quiescence [59]. At the peak of an outburst the nearest of these sources (eg A0620-00 at 2.5 kpc) can become the brightest in the sky, outshining Sco X-1 and even the Sun (except when it is flaring)

by a large factor.

There are also many similar X-ray binaries in our Galaxy which host a neutron star instead of a black hole. We know about the neutron star by observation of rapid pulsations due to the spin of the highly magnetic field, or from X-ray bursts which are thermonuclear flashes taking place very near the neutron star surface. Many of these are also transient and behave in a similar way to BHB. The depth of quiescence has been proposed as evidence of an event horizon in BHB by Ramesh Narayan and colleagues [47]. Black holes have no hard surface and can accrete matter and radiation whereas neutron stars must emit the gravitational energy released. BHB generally have lower quiescent luminosities than neutron star binaries which fits in with this picture. Whether these systems have similar enough accretion rates in quiescence for this to be a proof has yet to be determined.

Both BHB and some neutron star systems also show quasi-periodic oscillations, discovered by Michiel van der Klis in 1986 [74]. This phenomenon is complex and may be a rich source of information about the accretion flow and geometry, but it has yet to be fully understood and interpreted. To understand at least one of the frequencies we would expect to be seen, we now calculate the expected orbital frequency seen by a distant observer for material in the innermost orbit.

### **7.3 Velocities and frequencies**

*For astrophysical black holes, further important quantities for circular orbits are the velocity in the orbit, and the orbital period. For the former, the velocity determines (essentially via special relativistic effects) an important component of the shift in frequency which occurs for line radiation coming from the accretion disc, and for the latter, we can in some cases see directly the modulation of radiation associated with an orbital period. These are thought to provide one of the frequencies within the ‘quasi-periodic oscillations’ which are occasionally observed from accreting stellar mass black holes and can have*

frequencies in the several hundred Hz region [74].

Working with a Schwarzschild black hole, then if we let  $v$  be the ‘ordinary’ velocity in a circular orbit, and  $\alpha$  be the corresponding ‘rapidity’, then (temporarily dropping factors of  $c$  since this makes the formulae clearer), we have

$$v = \tanh \alpha \quad \text{and} \quad r\dot{\phi} = yv = \sinh \alpha$$

(70)

We can thus use (28) for  $h$  and the definition  $r^2\dot{\phi} = h$ , to calculate  $v$  as a function of  $r$ , obtaining the simple result

$$v^2 = \frac{GM}{r - 2GM}$$

(71)

For an object rotating in the innermost stable circular orbit (ISCO), at  $r = 6GM$ , we therefore find that its velocity is half the speed of light. (Note that, perhaps surprisingly, exactly the same velocity is obtained for the ISCO about a spinning black hole with any magnitude for its spin.)

As regards orbital frequency, then since the  $t$  coordinate is the time as measured by a stationary observer at spatial infinity, if we can form  $d\phi/dt$  then setting this equal to  $2\pi\nu_{\text{orb}}$  will immediately give us the orbital frequency, in Hz, as measured by an external observer.

Thus using

$$\frac{d\phi}{dt} = \frac{d\phi}{d\tau} \frac{d\tau}{dt} = \frac{\dot{\phi}}{\dot{t}}, \quad \dot{\phi} = \frac{h}{r^2} = \sqrt{\frac{GM}{r^3} \frac{1}{(1 - \frac{3GM}{r})}}$$

(72)

and

$$\dot{t} = \frac{k}{1 - \frac{2GM}{r}} = \frac{1}{\sqrt{1 - \frac{3GM}{r}}},$$

(73)

we find

$$\frac{d\phi}{dt} = 2\pi v_{\text{orb}} = \frac{\dot{\phi}}{t} = \sqrt{\frac{GM}{r^3}},$$

which despite being derived in full GR is exactly the Newtonian (Keplerian) result.

As an example, evaluating at the innermost stable circular orbit,  $r = 6GM/c^2$ , for a 10 solar mass black hole, yields the impressively high result of  $v_{\text{orb}} = 218$  Hz. Thus an emitting blob orbiting at this radius could potentially produce this frequency in the frequency spectrum of the source.

## 7.4 Further AGN properties

AGN range in luminosity from below  $10^{10}$  to  $10^{14}$   $L_\odot$  and are situated at the centre of the host galaxy. They appear to respect the Eddington limit and show strong evolution in luminosity with redshift (and thus cosmic time). The quasar and therefore mass-building phase peaked at about redshift 2.5 with quasars being observed back to redshift 7.1. An important issue with AGN is that a plentiful supply of fuel leads to obscuration of the source itself. If the absorbing column density is Thomson thick (i.e. most photons are scattered by electrons on passing out of the source), then only heavily reprocessed emission is all that remains, emerging predominantly in the infrared.

Type I AGN are largely unobscured and Type II obscured. The difference may be geometrical and due to our line of sight. In a broad class of source the obscuring medium is in the form of a thick torus at a radius of about a parsec. If viewed in the plane of the torus we see a Type II obscured object and if we view it face-on then it is Type I. If at the same time there is a jet, then face-on down the jet the object appears highly beamed as a blazar [72].

## 8 Measurements of the masses of black holes

The mass of a compact object can be measured from the motions of stars or gas in its vicinity using Newton's Law of Gravitation. If a star makes a circular orbit around a massive black hole with velocity  $v$  at radius  $r$  then the mass of the black hole  $M = v^2 r / G$ . Multiplying factors emerge for different shaped orbits but the essence of measuring velocity and radius remains the same. For the black hole at the centre of our Galaxy, Sgr A\*, the 3D motion of many nearby stars has been mapped to give the best estimate of  $4 \times 10^6 M_\odot$ .

With stellar mass black holes orbiting a normal star, then constraints on the mass are obtained by measuring periodic doppler shifts of spectral lines of that star to yield its radial velocity curve of semi-amplitude  $K$  and period  $P$ . These are combined to give the mass function

$$f = \frac{(M_{\text{bh}} \sin i)^3}{(M_s + M_{\text{bh}})^2} = \frac{PK^3}{2\pi G}$$

(74)

Further progress requires estimates of the mass of the normal star, perhaps from its spectral type, and of the inclination of the orbit  $i$ . If the system is eclipsing then  $i$  must be high, but if not then other means must be used such as observation of periodic ellipsoidal variations in the lightcurve of the star due to tidal distortions. These can be modelled to yield the inclination. There are now several dozen black holes with masses estimated to about 10% in this manner.

The mass of supermassive black holes (SMBH) is obtained from the motions of nearby stars and gas if they lie in relatively nearby galaxies where the region of influence of the black hole can be resolved. This is roughly the radius within which the black hole-induced motions are faster than the general motions in the galaxy. The masses of more distant active SMBH can be determined using the reverberation technique. Delayed variations in the ionizing UV radiation from the nucleus are seen in orbiting broad-line clouds - the whole system appears to reverberate with delays ranging from days to weeks. (The

clouds show doppler-broadened emission lines due to orbiting close to the SMBH.) The light travel-time delay then gives the radius and the line width of the clouds gives the velocity required to determine the central mass. Assumptions need to be made about the geometry but careful work of the past two decades has established the success of the method [55]. It has enabled correlations to be built between the width of selected emission lines, optical luminosity and black hole mass.

Current results indicate a maximum mass for observed neutron stars at just over  $2M_{\odot}$ , while stellar mass black holes have masses from  $\sim 4\text{--}20M_{\odot}$ , with an average of about  $7M_{\odot}$ . Observed supermassive black holes have masses ranging from just below  $10^6$  to  $2 \times 10^{10} M_{\odot}$ . It is possible that some UltraLuminous X-ray sources (ULXs), which have luminosities greater than the Eddington limit for a  $10M_{\odot}$  black hole, are Intermediate Mass Black Holes (IMBHs) with masses of  $10^2\text{--}10^5 M_{\odot}$ .

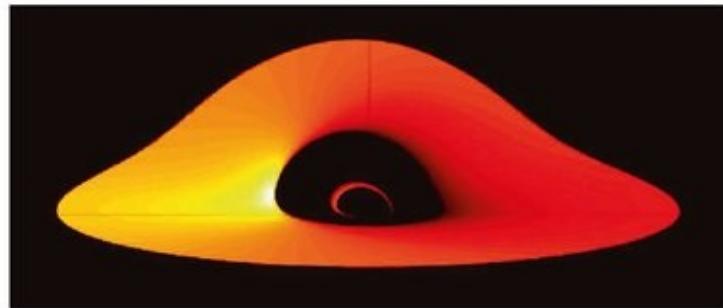
## 9 Measurements of black hole spin

Whilst mass measurements can be made at a large distance from an object, spin requires a probe which is close in, within  $10r_g$ <sup>11</sup>. The standard approach is to identify the inner edge of the accretion disc with the ISCO and then convert that radius to spin. This means that the inner disc must be luminous and detectable.

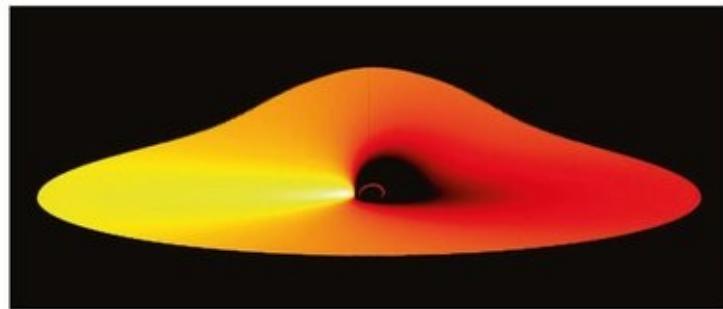
Bright Galactic X-ray binaries in the soft state generally have a quasi-blackbody X-ray spectrum which can be used to measure the emitting area using the Stefan-Boltzmann law in an analogous way to that by which stellar radii are determined (blackbody luminosity  $L = 4\pi R^2\sigma T^{11}$ ) [39]. Unlike the surface of a star, an accretion disc is not all at the same temperature, so the emission profile of the disc must be assumed and the black hole mass and distance determined in order to measure the ISCO and thus spin.

Light bending effects are very significant for the accretion disc near a black hole [43], so we now give some details of light bending effects in GR, first in the Schwarzschild metric, and then for Kerr black holes. An impression of what we would actually see for the accretion disc around a Schwarzschild black hole, if sufficient resolution was available, is shown in [Figure 16](#).

This can be compared with the more distorted and asymmetric image we would see from a black hole rotating at the expected maximum rate of  $a = 0.998M$ , in [Figure 17](#).



**Fig. 16.** Pseudoimage of a luminous thin accretion disc around a black hole with doppler and gravitational redshift effects included. Light bending causes the distant side of the disc to appear as an arch above the hole. The smaller loop near the centre is due to scattering of photons from the (unstable) photon orbit at  $r = 3G M/c^2$ . The colour scale indicates flux received.



**Fig. 17.** As for [Figure 16](#), but for a Kerr black hole rotating at the maximum attainable value of  $a = 0.998M$  (see [Section 7.1](#)). The unstable photon orbit is now near  $r = GM/c^2$ .

## 9.1 Equations for photon motion and redshift

The other results we need for understanding astrophysical aspect of black holes, are the equations of motion for photons,

and results for how their energy changes during propagation.

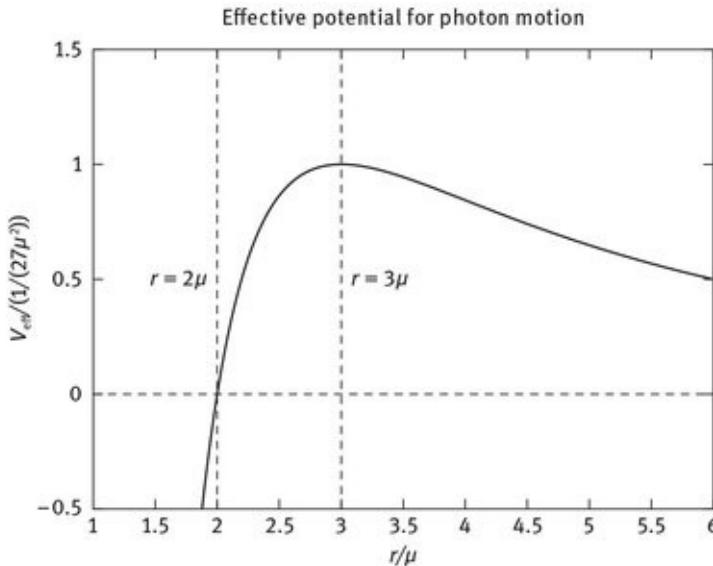
For the equations of motion, we can use the above analysis, but with the interval function  $G(x^\mu, \dot{x}^\mu)$  set equal to 0 rather than 1, once its functional variation has been taken.

Tracking through the changes this causes, we find [equation \(25\)](#) is replaced by

$$\dot{r}^2 = k^2 c^2 - \frac{\hbar^2}{r^2} \left( 1 - \frac{2GM}{rc^2} \right) \quad (75)$$

and for the ‘shape’ equation, (26), one finds that it is now the ‘Newtonian’ term that disappears on the r.h.s., and we just get

$$\frac{d^2 u}{d\phi^2} + u = \frac{3GM}{c^2} u^2. \quad (76)$$



**Fig. 18.** Plot of the effective potential for photon motion in the Schwarzschild geometry. (Units of  $V_{\text{eff}}$  are  $1/(27\mu^2)$  where  $\mu = GM/c^2$ .)

One can immediately confirm from this equation that there is a circular photon orbit at  $r = 3 GM/c^2$ , as mentioned above, but is it stable?

To do the stability analysis, we rewrite the energy equation as

$$\frac{\dot{r}^2}{h^2} + V_{\text{eff}}(r) = \frac{1}{b^2},$$

(77)

where  $b = h/c k$ ,  $\mu = GM/c^2$  and the effective potential

$$V_{\text{eff}}(r) = \frac{1}{r^2} \left( 1 - \frac{2\mu}{r} \right).$$

(78)

Let us look at a plot of this function, [Figure 18](#). We can see  $V_{\text{eff}}(r)$  has a single maximum at  $r = 3\mu$  where the value of the potential is  $1/(27\mu^2)$ . This shows that the circular orbit at  $r = 3\mu$  is unstable and in fact no stable circular photon orbit is possible in the Schwarzschild geometry.

Another immediate use of [equation \(76\)](#) is in connection with light bending. Referring to [Figure 19](#), we can see that a suitable first solution in which the term  $3GMu^2/c^2$  is completely ignored, is

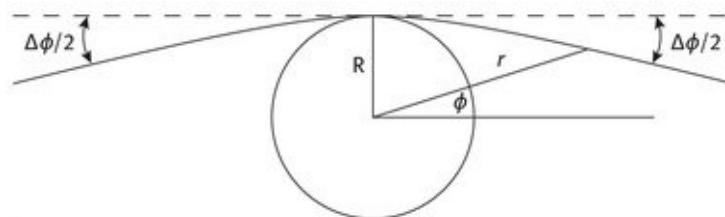
$$u = \frac{\sin \phi}{R},$$

(79)

where  $R$  is the radius of the body the gravitational deflection due to which we wish to work out. We iterate this equation by putting  $\sin^2 \phi/R^2$  for  $u^2$  on the r.h.s. of (76), i.e.

$$\frac{d^2 u}{d\phi^2} + u = \frac{3GM}{c^2 R^2} \sin^2 \phi.$$

(80)



[Fig. 19.](#) Gravitational deflection of light

*This is satisfied by the particular integral*

$$u_1 = \frac{3GM}{2c^2R^2} \left( 1 + \frac{1}{3} \cos 2\phi \right), \quad (81)$$

*and adding this into the original solution yields*

$$u = \frac{\sin \phi}{R} + \frac{3GM}{2c^2R^2} \left( 1 + \frac{1}{3} \cos 2\phi \right). \quad (82)$$

*Now consider the limit  $r \rightarrow \infty$ , i.e.  $u \rightarrow 0$ . Clearly we can take  $\sin \phi \approx \phi$ ,  $\cos 2\phi \approx 1$  there, and we obtain  $\phi = -2GM/(c^2 R)$  so that the total deflection (see figure) is*

$$\Delta\phi = \frac{4GM}{c^2R}. \quad (83)$$

*This is the famous gravitational deflection formula. For the Sun it yields 1.77 seconds of arc, and was first verified in the 1919 eclipse expedition. More recent high precision tests use radio sources, since these can be observed near the Sun, even when there is no eclipse, and there is now no doubt that the GR prediction (which incidentally is twice what had previously been worked out using a Newtonian approach) is accurate to a fraction of a percent.*

*In the black hole context, much stronger bending can occur, for which this first approximation is insufficient, and the full equations of motion need to be integrated. This can lead to some interesting effects, illustrated below in the context of a Kerr black hole.*

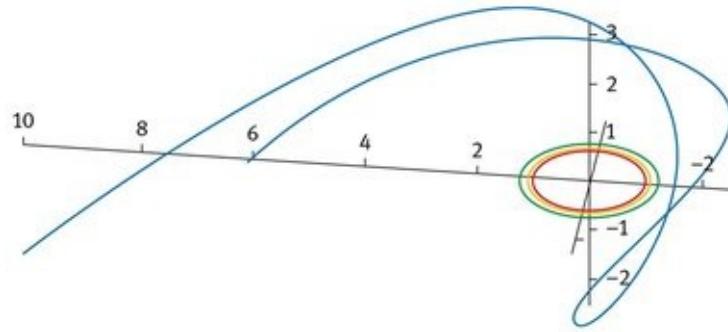
*The other important aspect to discuss is gravitational redshift. Quite generally, for a metric with a Killing vector field  $g_t$ , the redshift due to gravitational time dilation will be given by the ratio of  $p \cdot g_t$  at emission and reception, where  $p$  is the photon 4-momentum.  $g_t$  becoming null, i.e.  $g_{tt}$  going to zero, marks the*

*point where infinite redshift is possible since  $p$  can then be in the direction of  $g_t$ , e.g. for counter-rotating photons at the Kerr Stationary limit. This redshift from  $p \cdot g_t$  acts multiplicatively on any further redshifts which are occurring at a point due to special relativistic effects. E.g. for a point in an accretion disc, there will be a doppler redshift due to its motion relative to the initial path of an emitted photon, and then this will be multiplied by the gravitational redshift the photon experiences between that point and infinity. The combined effects of the redshifts can be seen in pattern of received flux seen in [Figures 16 and 17](#), which are for the case where the accretion disc itself has constant emissivity.*

## **9.2 Light bending around a Kerr black hole**

*The treatment of photon orbits for the Kerr case can proceed in the same way as in [Section 7.1](#), which dealt with massive particles, except that now the interval function  $G(x^\mu, \dot{x}^\mu)$  evaluates to 0 instead of 1. Several textbooks, e.g. [30] and [22], deal with the details of the resulting orbits, and in particular show that as in the Schwarzschild case, there are still no stable circular photon orbits.*

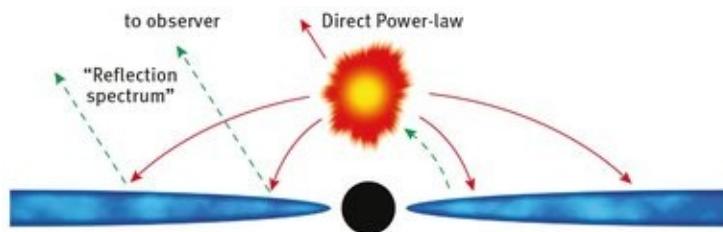
*The possible orbits for non-equatorial photons can be quite exotic. As an example, we show in [Figure 20](#) the path of a photon emitted from an accretion disc in the equatorial plane, moving towards the black hole but with an initial upward component to its velocity. This moves round behind the black hole and is eventually able to escape to infinity, moving in a roughly opposite direction to its initial motion. This example is for a Kerr black hole with the maximum expected value for  $a$  of  $0.998M$  — the ISCO and horizon radii are also shown.*



**Fig. 20.** Path of a photon emitted from an accretion disk that eventually escapes from the black hole in a direction about  $180^\circ$  from its initial direction.

### 9.3 Iron line emission

The method discussed above for Galactic X-ray binaries does not work for AGN since the much higher black hole mass means that the blackbody disc emission emerges in the far UV band which is unobservable because of absorption by Galactic hydrogen clouds. Observations are made instead of broad iron-line emission produced by X-ray reflection. This is the fluorescent and back-scattered emission produced by irradiation of the disc by the coronal X-ray continuum – see [Figure 21](#). X-rays absorbed in the disc lead to line emission, particularly iron K emission at 6.4 to 6.95 keV, depending on the ionization state of the emitter. At higher energies the X-rays are more likely to be Compton-scattered back out of the disc rather than absorbed, producing what is called the Compton hump in the reflection spectrum.



**Fig. 21.** Schematic cross-section of an accretion disc around a black hole with a central corona providing irradiation. The production of the reflection spectrum is illustrated.

The spectral features are produced in the innermost parts of

the accretion disc where gravitational effects are strong and lead to large doppler shifts and gravitational redshifts which skew and broaden emission lines [20, 21] - an example is shown in [Figure 22](#). Detailed modelling of the line shape, and generally the shape of the whole observed reflection spectrum, leads to an estimate of the inner radius of the disc and thus the spin. Radii are determined in this method directly in gravitational units so the distance and mass of the black hole are not required.

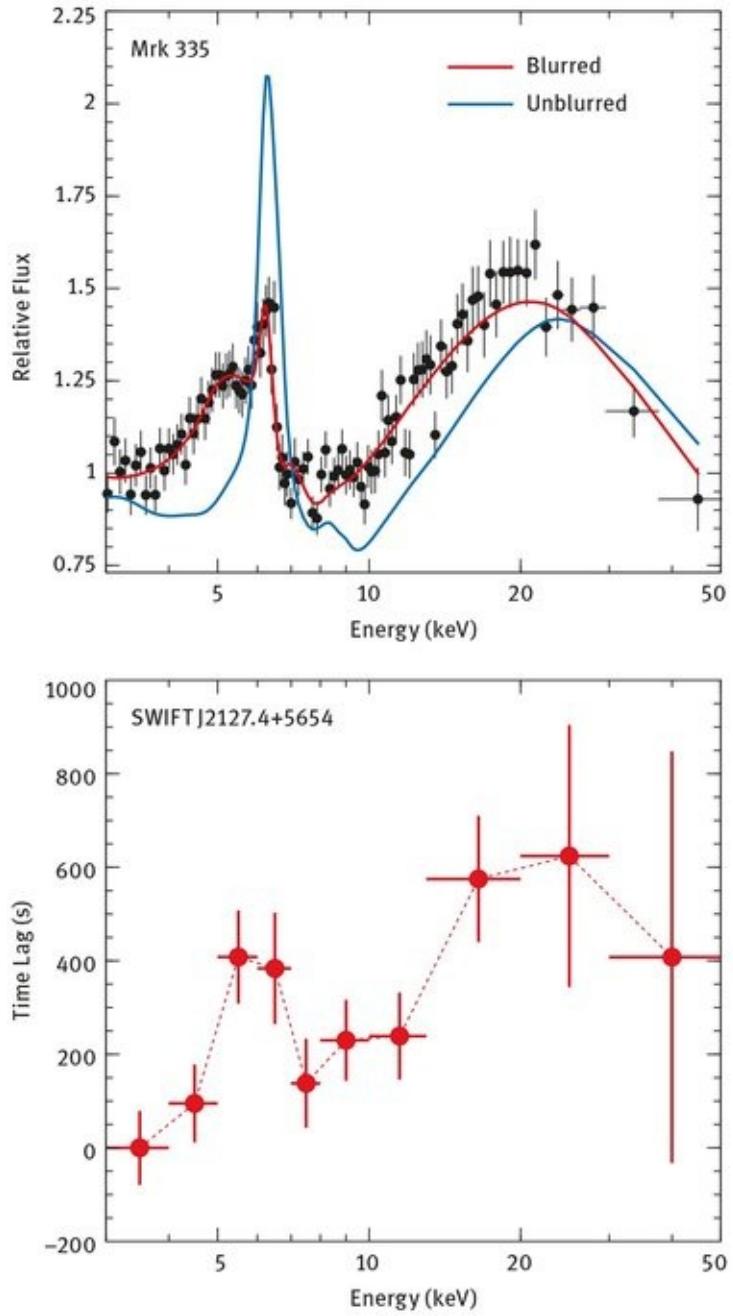
The broad iron line method can also be used in certain states of BHB [42] and has been found to give consistent results with the blackbody method.

Confirmation that the reflection modelling is correct has recently been obtained from X-ray reverberation [73]. This involves the small (tens of seconds to tens of minutes) time delays or lags whch occur between the detection of an intrinsic change in luminosity of the corona and the corresponding change in the reflection. The energy spectrum of the high frequency time lags shows a broad iron line indicative of reflection (lower panel of [Figure 22](#)).

The results for a number of objects imply that high spin is common for the brightest observed sources and the corona in those objects is compact and at a height of  $5-10r_g$ . In this regime, strong light bending occurs such that much of the coronal emission is bent down to the disc making the reflection strong. Month-long temporary episodes have been seen in a few sources in which the continuum drops to being almost unobservable, with only slight changes to the reflected spectrum. It appears that the corona has dropped below  $5r_g$  in these cases and that light bending is preventing much of the continuum being observed at all. Most of the reflection spectrum is then emerging from within  $2-3r_g$ , which is the closest emission easily seen from around any black hole [53].

Further evidence of a compact corona is obtained from microlensing studies of distant quasars which are lensed by intervening galaxies. This produces several images of the quasar which follow different light paths through the galaxy.

Stars in the galaxy crossing one of those paths amplify the image by lensing to a degree depending on how the angular sizes of the emission region of the quasar and the Einstein ring of the star match one another. For typical values this means that higher amplitude variability is expected from smaller emission regions. Both optical and X-ray observations have been made of several lensed quasars which show that the X-ray emission region is much smaller than the optical one, and in several cases is less than  $10r_g$  in size [3].



**Fig. 22.** Upper: The unblurred reflection spectrum including the iron-K emission line and Compton hump (solid line) appears relativistically blurred to the observer as shown in the line fitted to the data points. The obvious dominant redshift is gravitational due to the proximity to the black hole. The source is the Seyfert galaxy Mkn335, observed by NuSTAR during a low state. Modelling of the data shows that the black hole in Mkn335 has a spin  $a/M > 0.97$ . Most of the emission seen in the spectrum emerges from within  $2r_g$  of the event horizon [53]. Lower: Time lags measured by NuSTAR from the AGN SWIFTJ2127 showing both iron K and Compton peaks.

## 10 Future observations of astrophysical black holes

Many exciting developments are expected over the next decade or so in observational studies of black holes. One important step will be the networking together of several millimetre-wavelength telescopes to form the Event Horizon Telescope which will resolve the “shadow” of the black hole in Sgr A\* [13]. Extreme light bending of the emission from the accretion flow onto the black hole should reveal a dark patch, the shadow, at the position of the black hole. The diameter of the shadow is determined by the paths of light rays emitted from just above the event horizon to the observer and corresponds to about  $10r_g$  almost independent of the black hole spin.

### ***Calculation of the shadow radius:***

We use the results above for the photon effective potential in the Schwarzschild case. Putting together [equations \(77\)](#) and [\(78\)](#), we find

$$\frac{\dot{r}^2}{\hbar^2} + \frac{1}{r^2} \left( 1 - \frac{2\mu}{r} \right) = \frac{1}{b^2}$$

(84)

where  $\mu = GM/c^2$  and  $b$  has been defined as  $\hbar/kc$ . Note the dot here is not derivative w.r.t. proper time, which is zero for a photon, but w.r.t. an affine parameter along the path. We do not need to identify this parameter explicitly, since we are going to work with the shape of the orbit. Specifically, noting  $r^2\dot{\phi} = \hbar$ , we have

$$\left( \frac{\dot{r}^2}{r^4\dot{\phi}^2} \right)^{1/2} = \frac{1}{r^2} \frac{dr}{d\phi} = \left( \frac{1}{b^2} - \frac{1}{r^2} \left( 1 - \frac{2\mu}{r} \right) \right)^{1/2}$$

(85)

Solving (85) for  $\mu = 0$ , which is the limit where the only role of the central mass is to define the origin of the  $(r, \phi)$  system), we get  $r = b/\sin \phi$ . This is the equation of a straight line, which if it continued all the way past the central object, would pass it with perpendicular distance  $b$ , so we can identify this  $b$  as the impact parameter. This generalises what we did above, where we considered the case of a light ray just grazing the surface of an object of radius  $R$ , for which the unperturbed solution was given by (79), i.e.  $r = R/\sin \phi$ .

Now we consider the exact equation again, and relate to a black hole. We know from the plot of photon effective potential in [Figure 18](#), that photons straying inside  $r = 3GM/c^2$  are not free to escape to infinity. We therefore want to use this distance as the radius of an ‘object’ which the incoming photon just grazes. Using  $R$  for this object radius, we have

$$\left. \frac{dr}{d\phi} \right|_{r=R} = 0$$

This tells us that

$$\frac{1}{b^2} = \frac{1}{R^2} \left( 1 - \frac{2\mu}{R} \right) \quad \text{i.e.} \quad b = R \left( \frac{R}{R - 2\mu} \right)^{1/2}$$

This expression is valid generally, and e.g. tells us that the Sun’s effective shadow is about 3 km bigger than its coordinate radius, but applied with  $R = 3GM/c^2$  for a black hole, we find  $b = 3\sqrt{3}GM/c^2$ , which is therefore the radius of the geometric shadow cast. Interestingly, the main effect of including black hole spin in the calculation is to cause a shift in the centroid of the shadow rather than its size – see e.g. [22] for details.

For Sgr A\* this distance we have just worked out corresponds to an angular diameter of 52 microarcsec. This is  $2.5 \times 10^{-10}$  radians, and is very small indeed. Nevertheless, the angular resolution, approximately equal to the ratio of observing wavelength to telescope diameter, is about 20 microarcsec for intercontinental mm VLBI where the effective telescope

diameter is 10 000 km.

Interestingly, the 6 billion  $M_\odot$  black hole in M87 is 1500 times more massive than Sgr A\* and 2000 times further away which makes its event horizon appear only slightly smaller, at 40 microarcsec. Plans are also underway to resolve the M87 black hole shadow, together with the base of its jet which should help us understand how it is accelerated. All other black holes subtend smaller angular sizes. 3C273, for example, is 34 times more distant than M87, so its shadow will be about 1 microarcsec in size.

Near infrared interferometry at ESO's Very Large Telescope in Chile will use wavelengths around 2 microns and effective diameters of a 100 m to give a resolution of 4 milliarcsec. This will not resolve the black hole shadow but will resolve the motion of the orbiting stars much more precisely, particularly as the positional accuracy can be more than an order of magnitude better than the resolution. General relativistic effects on the orbits will be seen and fainter, possibly closer, stars resolved.

Of great interest would be a pulsar orbiting Sgr A\*, which if a millisecond pulsar could prove to be a very stable and reliable orbiting clock. None have been found yet and the ionized plasma that bathes the inner region is a problem for sensitive radio searches. A type of pulsar called a magnetar was found in 2013 within 3 arsec of Sgr A\*, but this magnetar unfortunately appears to be rather unreliable as a clock [10].

A gas cloud, G2, was discovered on a highly eccentric orbit close to Sgr A\* in 2012 [56]. At pericentre in 2014, about 3 Earth masses of material passes within about  $2000r_S$ . Ram pressure from the accretion flow should cause it to spiral inward over the following few years leading to an increased accretion rate and associated display. It is in a similar path to an earlier discovered cloud G1 and other clouds. Their apocentre and orbital plane lie close to the orbits of some of the surrounding luminous stars. The winds of the stars may be the origin of the inspiralling gas clouds.

Further past activity of Sgr A\* has been revealed by X-ray fluorescent emission lines seen from molecular clouds many

tens to hundreds of light years out from the Galactic Centre. One interpretation suggests that the luminosity of Sgr A\* may have been almost a million times higher about 100 yr ago [7].

Sufficient numbers of black hole spin measurements are expected from X-ray studies of AGN that the spin history can begin to be understood. Present work indicates many high spin objects in the mass range from  $10^6$ – $5 \times 10^7 M_\odot$ . The accretion flow within  $10r_g$  and particularly within  $3r_g$  will be mapped and the corona understood. Indeed the whole power output of accreting black holes in terms of radiation, winds and jets will be explored.

Tidal disruption events due to stars straying close to a massive black hole are beginning to be observed. Black holes with masses above  $10^8 M_\odot$  can swallow stars whole in the sense that they are not tidally disrupted until within the event horizon. The tidal forces of the more common lower-mass black holes can however destroy stars at greater distances leading to the production of a short lived, very luminous, accretion disc which decays with a timescale of about a year.

Studies of AGN feedback will accelerate with use of telescopes such as the Atacama Large Millimetre Array ALMA, the James Webb Space Telescope JWST, the Large Synoptic Survey Telescope LSST and the Advanced Telescope for High Energy Astronomy Athena. The growth of massive black holes through accretion and mergers will become understood.

Actually testing GR with black holes is difficult. The astrophysical phenomena observed so far are explained within the Kerr metric. Some precision tests may occur if a pulsar is serendipitously found in a close orbit around a stellar mass black hole in our Galaxy or around Sgr A\*. The discovery of gravitational waves from merging black holes will of course transform this aspect of the study of black holes.

More generally, the detection of gravitational waves emitted in events involving black holes has great potential for tying down aspects of astrophysical black holes that are poorly understood at present. These include the numbers of

intermediate mass holes in the range between stellar mass black holes and SMBHs in galactic centres, and the evolution of the number density of black holes with cosmic epoch. Both these aspects are key targets for a future space-borne gravitational wave detector mission, such as eLISA or a future variant, and form a key component of its science mission [17]. There are many further aspects to the importance of gravitational waves in black hole astrophysics, but we defer coverage of these to other chapters of this book.

## **11 More general spherically symmetric black holes**

*Real black holes are embedded in a universe in which we know there are effects matching those of a cosmological constant, accounting for about 0.7 of the total energy density. Thus it is of interest to embed a black hole in a de Sitter universe. Additionally, black holes in anti-de Sitter universes are very interesting objects theoretically, since via the AdS/CFT correspondence (see e.g. [38]), the strong gravity effects of black holes can be replaced by weak perturbative effects in a conformal field theory living on the boundary of an anti-de Sitter space (though we note that the dimensionality involved does not necessarily match that of 4d spacetime).*

*Returning to the derivations of [Section 2.2](#), we find that to satisfy the new field equation*

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = -8\pi T_{\mu\nu}$$

(86)

*arising from introducing a cosmological constant  $\Lambda$ , the new equations the Ricci tensor has to satisfy are*

$$R_{\mu\nu} = \Lambda g_{\mu\nu}$$

(87)

*Using the components already given for the Ricci in (10) and*

above, one finds that this leads to a simple modification to the Schwarzschild values of  $A$  and  $B$ . We find that  $B$  is still equal to the inverse of  $A$ , and that now

$$A = 1 - \frac{2GM}{rc^2} - \frac{\Lambda}{3}r^2 \quad (88)$$

These are therefore the metric coefficients for a black hole of mass  $M$  embedded in a de Sitter universe, if  $\Lambda$  is positive, and anti-de Sitter, if  $\Lambda$  is negative.

Horizons occur (for our current form of the metric) when  $B \rightarrow \infty$ , hence here where  $A = 0$ . If  $\Lambda > 0$ , this now leads to an additional horizon being present, beyond the Schwarzschild one near the black hole centre, at a position close to  $r = \sqrt{3/\Lambda}$  (we are adopting a convention here in which  $\Lambda$  has dimensions length $^{-2}$ ). This is a version of the 'de Sitter horizon' – its exact position is modified slightly by the presence of the mass.

A further generalisation we can make, is to the case where the black hole is charged. As mentioned earlier, this is less compelling physically, since we do not expect significant charge separation to have occurred in the formation of a black hole, but the solution for this type of black hole, called the Reissner-Nordstrom solution (see e.g. [75] for references), is still of great interest theoretically. This also falls into our general scheme, and in fact the most general spherically symmetric black hole, which we can dub RNdS, for Reissner-Nordstrom-de Sitter, is given by

$$B = A^{-1}, \quad A = 1 - \frac{2GM}{rc^2} - \frac{\Lambda}{3}r^2 + \frac{q^2}{r^2} \quad (89)$$

where  $q$  is the black hole charge. Note there are now (in general) three horizons. The Reissner-Nordstrom solution on its own introduces two horizons – one like the normal Schwarzschild one, and another usually much smaller one associated with the charge – and then for  $\Lambda > 0$  there is an

*additional de Sitter horizon.*

*An advantage of dealing with all these cases via the A, B form of the metric, is that we can give a unified treatment of energies and angular momenta in orbits. Repeating the analysis of Section 3 for this new case, we already have  $\dot{r}$  from (25) and can get  $\ddot{r}$  from*

$$\frac{d}{dr}(\dot{r}^2) = \frac{d}{d\tau}(\dot{r}^2) \frac{d\tau}{dr} = \frac{2\ddot{r}\dot{r}}{\dot{r}} = 2\ddot{r}$$

*For a circular orbit, both  $\dot{r}$  and  $\ddot{r}$  have to vanish and this combination of expressions therefore says that*

$$-\frac{k^2}{A^2}A' + \frac{2h^2}{r^3} = 0$$

*where  $A' \equiv dA/dr$ . Combining again with the expression for  $\dot{r} = 0$  we can solve for  $k^2$  and  $h^2$ , obtaining*

$$k^2 = \frac{2A^2}{2A - rA'} , \quad \text{and} \quad h^2 = \frac{r^3 A'}{2A - rA'}$$

*So we now have the energy and angular momentum for a particle in a circular orbit for quite a wide range of metrics. To calculate criteria for stability, one can show that it is useful to work in terms of the quantity (see Lasenby (2014) in preparation)*

$$T(r) = \left(\frac{v}{c}\right)^2 = \tanh^2 u \quad (90)$$

*where  $v$  is the velocity in a circular orbit at radius  $r$  and  $u$  is the corresponding rapidity parameter. From what we have already derived, one can show*

$$T = \frac{rA'}{2A} \quad \text{and} \quad r\dot{\phi} = yv = \sinh u \quad (91)$$

*and then the criterion for stability turns out to be that*

$$2T^2 - rT' - 2T$$

(92)

*should be negative. This simple criterion applies across all the spherically symmetric black hole cases. For example, inserting the  $A$  appropriate for a Schwarzschild-de Sitter metric, (88), we find the stability criterion*

$$18M^2 - 15M\Lambda r^3 - 3Mr + 4\Lambda r^4 < 0$$

(93)

*for stable circular orbits (in units with  $c = G = 1$ ). For appropriate astrophysical values of  $M$  and  $\Lambda$ , and restoring units, the outer solution to this is well approximated by*

$$r_{\text{stab}} \approx \left( \frac{3GM}{4\Lambda c^2} \right)^{1/3}$$

(94)

*We can apply this formula for any central mass, not just a black hole, and one can show that e.g. for a cluster of galaxies with  $M = 10^{15} M_\odot$ , and for the measured value of the cosmological constant, then  $r_{\text{stab}} \approx 7.2$  Mpc. This is interestingly close to the maximum cluster sizes observed, though of course circular orbit stability is not likely to be the direct criterion one would apply in this case, where effects on velocity dispersion would be more appropriate (see [46] for further details).*

## 12 Primordial black holes

Primordial black holes could have formed as a result of the high densities present soon after the Big Bang [8]. Their masses would be comparable to the particle horizon mass at their formation so it can range from the Planck mass ( $10^{-8}$  kg) at the Planck time ( $10^{-43}$  s), to  $10^5 M_\odot$  at a time of 1 s. Primordial black holes of mass  $10^{12}$  kg would have formed at  $10^{-23}$  s and would

be evaporating now due to the emission of Hawking radiation (see below), which would peak at 100 s of MeV. Observations provide a limit on the background intensity of 100 MeV gamma-rays such that evaporating primordial black holes cannot account for more than  $10^{-8}$  of the critical density of the Universe [52]. The frequency, energies and timescales of the final explosions do not match typical observed gamma-ray bursts. It is therefore unlikely that they are an important constituent of the Universe.

## **12.1 Hawking radiation**

*An initially very surprising result about black holes was proved by Hawking in 1974 [28]. This is that despite their name, the horizon radiates energy, as though it was a blackbody at the temperature of*

$$T = \frac{\hbar c^3}{8\pi k_B GM} \quad (95)$$

*where  $k_B$  is the Boltzmann constant and  $M$  is the black hole's mass. The types of black hole we have mainly considered have a minimum mass of several  $M_\odot$ , and so their temperatures are less than  $\sim 6 \times 10^{-8}$  K, and therefore negligibly tiny. However, a  $10^{12}$  kg primordial black hole of the type mentioned above, would have a temperature of  $\sim 1.2 \times 10^{11}$  K, and would therefore produce highly energetic radiation.*

*The emission process as calculated originally by Hawking was explained in terms of particle production in a second quantised treatment of a massless scalar field surrounding the black hole, but the principles extend to all kinds of field, massive or massless, which will be radiated with 'statistics' (i.e. Bose-Einstein or Fermi-Dirac) as appropriate to their spin, and with a thermal distribution with the temperature (95).*

*A heuristic way of demonstrating this effect is to consider vacuum fluctuations in the space just outside the horizon, which*

*will lead to the production of virtual particle /antiparticle pairs there. Sufficiently close to the horizon, one of the pair will be able to ‘tunnel’ through the horizon and inside can become a real particle with negative energy. (The possibility of negative energy arises since as we saw above ([Section 7.1](#)), the timelike vector in the t-direction tips over to become spacelike inside the horizon, which coincides with the Stationary Limit surface in the Schwarzschild case, and so the projection of the particle momentum onto the local t-vector may be negative.) The particle remaining outside, which has positive energy, can therefore be emitted off to infinity, whilst the absorption of a negative energy particle by the black hole, decreases the mass M, thus providing the overall energy for the emission.*

*This type of derivation can be used to calculate the energy of the emitted particle, which turns out to be independent of exactly where (though assumed to be close to the horizon) the virtual pair is produced, and yields an energy  $E = k_B T$  which is only a factor of  $2\pi$  higher than the correct result from (95) – see Chapter 11 of [30] for details.*

*A further possibility for a calculation avoiding the full rigours of quantum field theory (QFT), is to work with a first-quantised field surrounding the black hole. For example, a detailed derivation for a massive Dirac field is given in Section 8 of [35]. This yields precisely the result (95) together with the correct Fermi-Dirac statistics for a spin-1/2 particle, and the calculation can be extended to electromagnetic and scalar particles, with the same results that would be obtained from QFT in each case (Lasenby et al unpublished). However, it is only in a QFT context that one can be properly certain of the physics involved, and of how to go about making unique choices for the branches of analytic functions that are involved, and so a QFT approach is still necessary to be fully confident of the results.*

## **12.2 Link with surface gravity**

*What a full QFT approach leads to is a very general result that singles out the ‘surface gravity’ of a black hole (of any type) as*

*the important quantity. The surface gravity would appear to be infinite at the horizon, since no particle can escape the black hole's pull there. However, the definition of surface gravity intended here is that appropriate to the force felt at the horizon, but which would be evaluated by an observer at infinity. In e.g. [75], Exercise 6.4, this force is shown to differ from that at the horizon by the redshift factor from the horizon to infinity, which is also infinite. Specifically, one finds for a Schwarzschild black hole that the local force necessary to keep an observer of mass  $m$  stationary at radial distance  $r$  is (see e.g. Section 7 of [35])*

$$F = \frac{GMm}{r^2} \left(1 - \frac{2GM}{rc^2}\right)^{-1/2} \quad (96)$$

*The gravitational redshift factor between the horizon and infinity is  $(1 - \frac{2GM}{rc^2})^{1/2}$ , hence multiplying by this, we obtain the simple expression  $GMm/r^2$  as the force that the observer at infinity would ascribe as being necessary to keep the observer at radius  $r$  at rest. (In the exercise in Wald ([75]), this difference is explained in terms of the different tensions necessary at the ends of a (massless) rope that extends from the observer at infinity to the one at radius  $r$ .)*

*This may seem a mundane and wholly Newtonian result, and exactly what we would expect 'at infinity'. However, we are now entitled to put  $r = 2GM/c^2$ , i.e. the value of  $r$  at the horizon, into this expression, in order to evaluate the 'surface gravity' of the black hole. This is defined as acceleration per unit mass, and customarily denoted  $\kappa$ , so we have found*

$$\kappa = \frac{c^4}{4GM} \quad (97)$$

*The contribution of QFT at this point, is to tell us that a horizon with surface gravity  $\kappa$  radiates particles with a thermal temperature*

$$T = \frac{\hbar\kappa}{2\pi k_B c}$$

(98)

*Unruh in 1976 [71] showed that this is the temperature of a thermal heat bath that develops around an observer accelerating at rate  $\kappa$  in Minkowski space. We can now understand the link with surface gravity in black holes, at least heuristically. By the equivalence principle and working close to the horizon so that global curvature effects are not important, we can expect that the heat bath seen by a stationary observer feeling the effects of the black hole surface gravity  $\kappa$ , should be the same as the heat bath seen by an observer in Minkowski space accelerating at the same rate, and we can see indeed that [equations \(97\)](#) and (98) agree with (95) for this Schwarzschild case. The advantage of this route, however, is that if we work out the surface gravity in more complicated cases, such as for Kerr or Kerr-Newman or even de Sitter spacetimes (which as we have seen, also possess a horizon), then we can make the same transition to temperature as via (98). For example, for our general spherically symmetric metric (7), it is easy to show that the surface gravity is*

$$\kappa = \left. \frac{A'}{2\sqrt{AB}} \right|_{B^{-1}=0}$$

(99)

*(The point where  $1/B = 0$  is picked out as being the position of the event horizon.) Evaluating this for the Reissner-Nordstrom metric discussed above, and for clarity temporarily putting  $G = c = 1$  etc., so  $1/B = A = 1 - 2M/r + q^2/r^2$ , we find that the surface gravity at the outer horizon is*

$$\kappa_{RN} = \frac{\sqrt{M^2 - q^2}}{(M + \sqrt{M^2 - q^2})^2}$$

(100)

*and the Hawking temperature of the radiation emitted in this*

case is given by inserting this  $\kappa$  into (98).

### **12.3 Astrophysical aspects of black hole evaporation**

We can compute the rate at which a Schwarzschild black hole loses mass via the formula for the luminosity of a blackbody at temperature  $T$ , i.e.  $\sigma T^4$  where  $\sigma = \pi^2 k_B^4 / (60 \hbar^3 c^2)$  is the Stefan-Boltzmann constant, together with the area of the horizon, which is  $4\pi(2GM/c^2)^2$ . This yields

$$\dot{M}c^2 = -\text{power radiated} = -\frac{c^6 \hbar}{15 360 \pi G^2 M^2}$$

(101)

Solving this differential equation, we find that the cube of the mass declines linearly with time: 4

$$M^3(t) = M_0^3 - \frac{c^6 \hbar}{5120 \pi G^2} t$$

(102)

where  $M_0$  is the initial mass and  $t$  is measured from formation. This gives an evaporation time in terms of  $M_0$  of

$$t_{\text{evap}} = 2.66 \times 10^{-24} M_0^3 \text{ yrs}$$

(103)

If the current age of the universe is 13.8 billion years, this means only black holes with a mass less than  $1.7 \times 10^{11}$  kg will have had a chance to decay by now. Due to the  $M^{-2}$  dependence in the expression for radiated power, (101), nearly all the energy emitted is confined to the very last moments of the black hole's life.

### **12.4 Black hole entropy**

The fact that black holes radiate like black bodies, suggest that they should have an entropy. If we identify  $Mc^2$  as the hole's

*energy, then the thermodynamic relation*

$$dU = TdS \quad (104)$$

*along with our identification of the BH temperature in (95), yields (first working in units with  $G = \hbar = c = k_B = 1$  for clarity)*

$$dM \times 8\pi M = dS, \quad \text{so that} \quad S = 4\pi M^2 = \frac{\mathcal{A}}{4} \quad (105)$$

*where  $A = 16\pi M^2$  is the black hole area. Putting back the units, we find that the black hole entropy in units of  $k_B$ , is*

$$\frac{S}{k_B} = \frac{\mathcal{A}}{4\ell_p^2} \quad (106)$$

*where  $\ell_p = \sqrt{\hbar G/c^3}$  is the Planck length and  $\mathcal{A} = 4\pi(2GM/c^2)^2$  is the BH area. For an astrophysical black hole, this entropy is large, and as we have seen scales like  $M^2$ . The result of this is that it is thought that supermassive black holes in the centres of galaxies dominate the entropy budget of the observable universe. Estimates of this are given in [15], with the result that SMBH may contribute up to 7 orders of magnitude more entropy than stellar mass BHs, and approximately 15 orders of magnitude more than the next largest component, due to photons.*

*This identification of entropy with 1/4 of the event horizon area expressed in units of  $\ell_p^2$ , is the starting point for a great deal of work connected with whether a microphysics approach to black hole entropy is possible. In this approach we would seek to obtain the same result by counting states, and then assigning  $S = k_B \ln N$ , where  $N$  is the number of available microstates. A related question is what happens to the information about the matter and radiation from which the hole was originally formed. This information is screened beyond the horizon, and our lack of knowledge of the internal arrangements*

*and compositions of this original material provides a satisfactory understanding of why the black hole should have an entropy at all, as already proposed by Beckenstein [2], before Hawking's discovery of black hole radiation. However, at the end stage of evaporation of the black hole, we no longer have a screen, and unless the information is somehow encoded in the phases of the emitted particles and fields composing the Hawking radiation, so that this is not random after all, then the information has been irretrievably lost. This would seem to violate the basic principles of quantum mechanics, which demand unitarity of transformations between beginning and end states, and would not allow the disappearance of information in this way, given that there is no longer a causal horizon beyond which the information is hidden.*

*These are deep and fundamental questions, which we do not venture further with, but are discussed elsewhere in this volume. We note that our discussion so far has been just in terms of a Schwarzschild black hole, but equivalences of the same nature work for all types of black hole, and are encoded in the 'Laws of black hole thermodynamics'. We now briefly discuss these concentrating on their astrophysical implications, and in particular the possibility of extracting energy from rotating black holes.*

## **12.5 Laws of black hole thermodynamics and the Penrose process**

*The generalisation of (104) to an uncharged rotating black hole reads (in natural units again, for clarity):*

$$dU = TdS + \Omega_H dJ$$

i.e.  $dM = \frac{\kappa}{8\pi} dA + \Omega_H dJ$

(107)

*where  $\Omega_H$  is the black hole angular velocity at the horizon, and we have again made the identification of temperature with surface gravity/ $2\pi$  and of entropy with one quarter of the*

horizon area. The new feature here is that work can be done to change the angular momentum  $J$  of the black hole. (Recall that this is related to the mass and spin parameters via  $J = aM$ .)

[Equation \(107\)](#) is known as the ‘First law of black hole thermodynamics’. There are equivalent versions for black holes of the other three laws of thermodynamics as well – see e.g. Section 9.9 of [22] for a discussion of these. For astrophysical processes, the other law of immediate interest is the equivalent of the Second law, which states that in any classical process, the area of a black hole horizon, which we know measures the black hole’s entropy, does not decrease. The restriction to classical processes is necessary, since as we have seen, Hawking radiation succeeds in reducing both mass and surface area.

The most interesting immediate application of these laws is to (classical) processes in which we attempt to remove energy from the hole. This cannot be done for a non-rotating hole, but the  $\Omega_H dJ$  term in the First law provides a route through to this for rotating holes.

The first process of this kind to be discussed, was the Penrose process [54]. In this process, an incoming particle enters the ergosphere of the black hole. It then decays into two particles. Particles can escape from the ergosphere (which lies outside the horizon), and we arrange the initial trajectory and decay such that one of the decay products escapes, and the other falls into the black hole. The key observation is now that inside the ergosphere the  $g_t$  Killing vector corresponding to invariance of the metric under time displacements, becomes spacelike, (remember the definition of the ergosphere is where  $g_t^2$  changes sign), and it is possible for a particle to have negative energy when its 4-momentum is projected onto it. If this happens, then from conservation of energy, the energy of the emitted particle will be greater than that of the original particle, and we will have succeeded in extracting energy from the black hole. This energy has come via the BH’s absorption of the infalling negative energy particle.

One can also analyse what happens to angular momentum in

this process, by considering the projections onto the other Killing vector (corresponding to  $\varphi$  invariance of the metric)  $g_\varphi$ . This reveals (see [77] or [22] for details) that the hole absorbs negative angular momentum as well as negative energy (meaning the emitted particle enjoys a boost to its angular momentum as well as energy, relative to the incoming one) and that the black hole parameter changes obey the inequality

$$\delta M \geq \Omega_H \delta J$$

(108)

Now, the expressions in the Kerr case for the quantities appearing in the First law (107) are

$$\begin{aligned} \text{Horizon area} \quad & \mathcal{A} = 8\pi \left( M^2 + \sqrt{M^4 - J^2} \right) \\ \text{Horizon angular velocity} \quad & \Omega_H = \frac{J}{2M \left( M^2 + \sqrt{M^4 - J^2} \right)} \\ \text{Surface gravity} \quad & \kappa = \frac{\sqrt{M^4 - J^2}}{2M \left( M^2 + \sqrt{M^4 - J^2} \right)} \end{aligned}$$

(109)

(again see [77] or [22] for details). We thus have that the response of the horizon area to a general change in  $M$  and  $J$  is

$$\delta \mathcal{A} = \frac{\partial \mathcal{A}}{\partial M} \delta M + \frac{\partial \mathcal{A}}{\partial J} \delta J = \frac{8\pi J}{\Omega_H \sqrt{M^4 - J^2}} (\delta M - \Omega_H \delta J)$$

(110)

From (108), we see that despite the fact that both  $M$  and  $J$  decrease, the r.h.s. here is always positive in the Penrose process. This is an example of the Second law of black hole thermodynamics in action - for a classical process, the BH entropy, as measured by the horizon area, must always increase.

## 12.6 Adiabatic (reversible) changes

An interesting case to consider astrophysically, is the limit of gradual changes in mass and angular momentum, which satisfy the thermodynamic notion of adiabaticity, in particular the changes are reversible. This is possible, for example, for particular versions of the Penrose process. In the black hole context, since horizon area equates to entropy, an adiabatic process will have  $\delta A = 0$ .

Solving (110) for  $\delta A = 0$ , immediately yields the following differential equation for  $M(J)$ :

$$\frac{dM}{dJ} = \Omega_H = \frac{J}{2M(M^2 + \sqrt{M^4 - J^2})}$$

(111)

for which the following (implicit) solution works:

$$2M_0^2 = M^2(J) + \sqrt{M^4(J) - J^2}$$

(112)

where  $M_0$  is a constant. As an example, suppose we start with a Schwarzschild black hole with mass  $M_0$ , and gradually feed in material with positive angular momentum. We will be able to do this until an extremal black hole with  $J = M^2$  is reached, at which point the mass is  $\sqrt{2}M_0$ . An interesting point is what happens to the horizon radius  $r_{\text{outer}}$  in this process. We know the general expression for it (in terms of  $M$  and  $J = aM$ ) is

$$r_{\text{outer}} = M + \frac{1}{\sqrt{M}} \left( M^2 + \sqrt{M^4 - J^2} \right)$$

(113)

and naively we might expect this to increase as the hole gains mass. But in fact we can see for the adiabatic process just discussed,

$$r_{\text{outer}} = \frac{\mathcal{A}}{8\pi M} = \frac{16\pi M_0^2}{8\pi M} = \frac{2M_0^2}{M}$$

(114)

The horizon radius therefore goes down, reaching a minimum of  $\sqrt{2}M_0$  at the end of the process (corresponding to the  $r_{\text{outer}} = M$  value for an extremal black hole), having started at  $2M_0$ .

## 12.7 Other processes for extracting energy from a spinning black hole

As well as the Penrose process, there are other ways of extracting energy from a rotating black hole. One of these is ‘superradiance’, which is the analogue for waves of what happens for particles in the Penrose process. The possibility of such a process was first drawn attention to specifically in the context of the Kerr solution by Starobinski in 1973 [68]. It involves an incoming radiation field of the form

$$\phi \sim \phi_0(r, \theta) e^{-i\omega t} e^{im\phi}$$

(115)

for which part of the wave (the ‘transmitted’ wave) is absorbed by the black hole, and the other part (the ‘reflected’ wave) reaches infinity again. In the same way as for the Penrose process, due to the spacelike nature of the  $g_t$  Killing vector inside the ergosphere, the transmitted wave can have negative energy, so that the reflected wave carries increased energy to infinity (specifically it will have greater amplitude at the same frequency as compared to the incident wave). We can derive the condition for this to occur (at least in a heuristic way), as follows (see e.g. Section 8.8 of [22]).

We can think of the wave as composed of quanta with energy  $\hbar\omega$  and angular momentum  $\hbar m$ . When these are absorbed by the black hole, the change in the black hole’s parameters will satisfy

$$\frac{\delta J}{\delta M} = \frac{\hbar m}{\hbar\omega}$$

(116)

But we also know that  $\delta M - \Omega_H \delta J \geq 0$  from the fact the horizon area must not decrease. Putting these together leads to

$$\frac{\delta M}{\omega} (\omega - \Omega_H m) \geq 0$$

(117)

and so the condition for superradiance ( $\delta M < 0$ ) is

$$0 < \omega < m\Omega_H$$

(118)

Further details of this process, which only works for bosonic fields, can be found in [Section 12.4](#) of [75].

Our final example, concerns what may be an important way in which spinning black holes return energy to the environment, and which was referred to briefly above in [Section 6.2](#), namely the Blandford-Znajek effect [4].

Here, one considers a Kerr black hole immersed in an ambient magnetic field, typically associated with an accretion disc around the black hole. A possible picture of the process (see e.g. [22], Section 8.9) represents the black hole horizon as a moving conductor within the magnetic field. Indeed the Kerr black hole horizon, in common with all stationary event horizons, can be modelled as having a resistance of  $4\pi$  in geometrical units (377 Ohms in ordinary units). This rotating conductor coupled to the magnetic field, generates a current which flows between the poles and equator – it is in this sense like a dynamo. The main analytic field configuration chosen by Blandford & Znajek was a ‘split monopole’ – i.e. a different sign magnetic monopole solution in each hemisphere, with magnetic field lines pointing radially outwards or inwards. This needs a current sheet on the join between them in the equitorial plane, which is presumed to be supplied at least outside the hole by currents in the accretion disc.

The radiated power can be calculated from the Poynting vector, and gives similar results as for a rotating magnetic dipole pulsar model evaluated on the light cylinder. Specifically

we find (see e.g. [40], Section 14.3)

$$\begin{aligned} \mathcal{P}_{\text{EM}} &= \frac{B_E^2 r_E^4 \Omega_f^2}{c} = \frac{B_H^2 r_H^4 \Omega_f^2}{c} \\ &\sim 4.1 \times 10^{47} \text{ erg s}^{-1} \left( \frac{B_H}{10^5 \text{ G}} \right)^2 \left( \frac{a}{M} \right)^2 \left( \frac{M}{10^9 M_\odot} \right)^2 \end{aligned} \quad (119)$$

Here  $B_E$  and  $B_H$  are the magnetic field values at the ergosphere and horizon respectively,  $r_E$  and  $r_H$  the corresponding  $r$  values, and  $\Omega_f$  the magnetic field rotation rate, which is typically  $\sim \frac{1}{2}\Omega_H$ . This EM power will presumably be manifested in a magnetised particle wind and jet, removing angular momentum and energy from the black hole in the process. Subsequent work (see e.g. Komissarov [34]) has shown that the Blandford- Znajek (BZ) mechanism is stable, and clarified that important components of it take place within the ergosphere itself, emphasising its links with the Penrose process and superradiance. The BZ mechanism is now thought to be an important component in generating at least some of the high energy jets we see in astrophysics [69].

## 13 Conclusions

Black holes are now an integral features of our cosmic landscape. Many millions of stellar mass black holes reside in our Galaxy. A supermassive black hole lies at the centre of all massive galaxies. We know little about most of them, unless they are massive and nearby or accreting gas from a stellar companion or the surrounding interstellar medium. The inner part of the accretion flow can be extremely luminous. making the immediate surroundings of the blackest parts of the Universe into the brightest.

The behaviour of observed black holes can so far be explained well by General Relativity and the Kerr metric. The physics of accretion onto black holes, and of the outflows and jets that

often accompany inflows, is complex and leads to complicated observational phenomena which lie at the forefront of astrophysics. Understanding how quasars work remains a significant astrophysical challenge. The consequences for a galaxy hosting a supermassive black hole are profound. It is likely that the energy released by the growth of the black hole plays a decisive factor in its final stellar mass and possibly its physical size.

Black holes are intrinsically relativistic objects which have stimulated and tested physical understanding to the extreme. The internal structure of the black hole, within the event horizon, is beyond direct observation and we have not discussed it here, although it is clearly of great theoretical interest. In this Chapter we have looked at black holes as observable physical objects and considered how they work as engines of gravitational energy release.

The future of research into astrophysical black holes is very bright, as more telescopes probe ever deeper into more wavebands to uncover new objects, features and phenomena. Serendipitous discovery plays a key role in the history of astronomy and we look forward to the discoveries of a millisecond pulsar orbiting close to a massive black hole, a rogues gallery of black hole shadows, a range of quasars above redshift 10, a complete X-ray spectral-timing deconstruction of the innermost accretion flow around a rapidly spinning black hole and, most of all, some new phenomena of black holes that we have not yet anticipated.

**Acknowledgment:** We thank Dan Wilkins for provision of [Figures 16 and 17](#), Michael Parker and Erin Kara for [Fig. 22 and](#) Reinhard Genzel for [Fig. 8](#). [Fig. 7 is](#) due to ESO/Yuri Beletsky. ANL thanks Michael Hobson and George Efstathiou as co-authors of the book [30], referred to at several points in the text above.

## Bibliography

- [1] S. Antoci and D.-E. Liebscher, Editor's Note: On the Gravitational Field of a Mass Point According to Einstein's Theory by K. Schwarzschild, *General Relativity and Gravitation* **35** (2003), 945–950.
- [2] J. D. Bekenstein, Black holes and the second law, *Lett. Nuovo Cim.* **4** (1972), 737–740.
- [3] J. A. Blackburne, C. S. Kochanek, B. Chen, X. Dai and G. Chartas, The Optical, Ultraviolet, and X-Ray Structure of the Quasar HE 0435-1223, *ApJ* **789** (2014), 125.
- [4] R. D. Blandford and R. L. Znajek, Electromagnetic extraction of energy from Kerr black holes, *Mon. Not. Roy. Astron. Soc.* **179** (1977), 433–456.
- [5] C. T. Bolton, Identification of Cygnus X-1 with HDE 226868, *Nature* **235** (1972), 271–273.
- [6] R. H. Boyer and R. W. Lindquist, Maximal analytic extension of the Kerr metric, *J. Math. Phys.* **8** (1967), 265.
- [7] R. Capelli, R. S. Warwick, D. Porquet, S. Gillessen and P. Predehl, The X-ray lightcurve of Sagittarius A\* over the past 150 years inferred from Fe-K $\alpha$  line reverberation in Galactic centre molecular clouds, *A&A* **545** (2012), A35.
- [8] B. J. Carr, K. Kohri, Y. Sendouda and J. Yokoyama, New cosmological constraints on primordial black holes, *Phys. Rev. D* **81** (2010), 104019.
- [9] S. Chandrasekhar, The Maximum Mass of Ideal White Dwarfs, *ApJ* **74** (1931), 81.
- [10] J. Chennamangalam and D. R. Lorimer, The Galactic Centre pulsar population, *MNRAS* **440** (2014), L86–L90.
- [11] E. Churazov, R. Sunyaev, W. Forman and H. Böhringer, Cooling flows as a calorimeter of active galactic nucleus mechanical power, *MNRAS* **332** (2002), 729–734.
- [12] H. D. Curtis, Descriptions of 762 Nebulae and Clusters Photographed with the Crossley Reflector, *Publications of Lick Observatory* **13** (1918), 9–42.

- [13] S. Doeleman, E. Agol, D. Backer, F. Baganoff, G. C. Bower, A. Broderick, A. C. Fabian, V. Fish, C. Gammie, P. Ho, M. Honman, T. Krichbaum, A. Loeb, D. Marrone, M. Reid, A. Rogers, I. Shapiro, P. Strittmatter, R. Tilanus, J. Weintroub, A. Whitney, M. Wright and L. Ziurys, Imaging an Event Horizon: submm-VLBI of a Super Massive Black Hole, in: *astro2010: The Astronomy and Astrophysics Decadal Survey*, Astronomy 2010, p. 68, 2009.
- [14] C. Doran and A. N. Lasenby, New techniques for analyzing axisymmetric gravitational systems. 1. Vacuum fields, *Class. Quant. Grav.* **20** (2003), 1077–1102.
- [15] C. A. Egan and C. H. Lineweaver, A Larger Estimate of the Entropy of the Universe, *ApJ* **710** (2010), 1825–1834.
- [16] F. Eisenhauer, R. Genzel, T. Alexander, R. Abuter, T. Paumard, T. Ott, A. Gilbert, S. Gillessen, M. Horrobin, S. Trippe, H. Bonnet, C. Dumas, N. Hubin, A. Kaufer, M. Kissler-Patig, G. Monnet, S. Ströbele, T. Szeifert, A. Eckart, R. Schödel and S. Zucker, SINFONI in the Galactic Center: Young Stars and Infrared Flares in the Central Light-Month, *ApJ* **628** (2005), 246–259.
- [17] The eLISA Consortium, The Gravitational Universe, *ArXiv e-prints* 1305.5720 (2013).
- [18] A. C. Fabian, The obscured growth of massive black holes, *MNRAS* **308** (1999), L39–L43.
- [19] A. C. Fabian, Observational Evidence of Active Galactic Nuclei Feedback, *ARA&A* **50** (2012), 455–489.
- [20] A. C. Fabian, E. Kara and M. L. Parker, Relativistic Disc lines, *ArXiv e-prints* (2014).
- [21] A. C. Fabian, M. J. Rees, L. Stella and N. E. White, X-ray fluorescence from the inner disc in Cygnus X-1, *MNRAS* **238** (1989), 729–736. [22] V. P. Frolov and A. Zelnikov, *Introduction to black hole physics*, Oxford University Press, 2011.
- [23] R. Genzel, F. Eisenhauer and S. Gillessen, The

Galactic Center massive black hole and nuclear star cluster, *Reviews of Modern Physics* **82** (2010), 3121-3195.

- [24] A. M. Ghez, M. Morris, E. E. Becklin, A. Tanner and T. Kremenek, The accelerations of stars orbiting the Milky Way's central black hole, *Nature* **407** (2000), 349-351.
- [25] R. Giacconi, H. Gursky, F. R. Paolini and B. B. Rossi, Evidence for x Rays From Sources Outside the Solar System, *Physical Review Letters* **9** (1962), 439-443.
- [26] S. Gillessen, F. Eisenhauer, T. K. Fritz, H. Bartko, K. Dodds-Eden, O. Pfuhl, T. Ott and R. Genzel, The Orbit of the Star S2 Around SGR A\* from Very Large Telescope and Keck Data, *ApJ* **707** (2009), L114-L117.
- [27] K. Güttekin, D. O. Richstone, K. Gebhardt, T. R. Lauer, S. Tremaine, M. C. Aller, R. Bender, A. Dressler, S. M. Faber, A. V. Filippenko, R. Green, L. C. Ho, J. Kormendy, J. Magorrian, J. Pinkney and C. Siopis, The M- $\sigma$  and M-L Relations in Galactic Bulges, and Determinations of Their Intrinsic Scatter, *ApJ* **698** (2009), 198-221.
- [28] S. W. Hawking, Black hole explosions, *Nature* **248** (1974), 30-31.
- [29] C. Hazard, M. B. Mackey and A. J. Shimmmins, Investigation of the Radio Source 3C 273 By The Method of Lunar Occultations, *Nature* **197** (1963), 1037-1039.
- [30] M. P. Hobson, G. P. Efstathiou and A. N. Lasenby, *General Relativity: An Introduction for Physicists*, Cambridge University Press, 2006.
- [31] R. P. Kerr, Gravitational field of a spinning mass as an example of algebraically special metrics, *Phys. Rev. Lett.* **11** (1963), 237-238.
- [32] A. King, The AGN-Starburst Connection, Galactic Superwinds, and  $M_{BH} - \sigma$ , *ApJ* **635** (2005), L121-L123.
- [33] R. W. Klebesadel, I. B. Strong and R. A. Olson, Observations of Gamma-Ray Bursts of Cosmic Origin, *ApJ* **182** (1973), L85.

- [34] S. S. Komissarov, Electrodynamics of black hole magnetospheres, *MNRAS* **350** (2004), 427–448.
- [35] A. Lasenby, C. Doran and S. Gull, Gravity, gauge theories and geometric algebra, *Royal Society of London Philosophical Transactions Series A* **356** (1998), 487.
- [36] D. Lynden-Bell, Galactic Nuclei as Collapsed Old Quasars, *Nature* **223** (1969), 690–694.
- [37] J. Magorrian, S. Tremaine, D. Richstone, R. Bender, G. Bower, A. Dressler, S. M. Faber, K. Gebhardt, R. Green, C. Grillmair, J. Kormendy and T. Lauer, The Demography of Massive Dark Objects in Galaxy Centers, *AJ* **115** (1998), 2285–2305.
- [38] J. M. Maldacena, The Large N limit of superconformal field theories and supergravity, *Int. J. Theor. Phys.* **38** (1999), 1113–1133.
- [39] J. E. McClintock, R. Narayan, S. W. Davis, L. Gou, A. Kulkarni, J. A. Orosz, R. F. Penna, R. A. Remillard and J. F. Steiner, Measuring the spins of accreting black holes, *Classical and Quantum Gravity* **28** (2011), 114009.
- [40] D. L. Meier, *Black Hole Astrophysics: The Engine Paradigm*, Springer, 2009.
- [41] J. Michell, On the Means of Discovering the Distance, Magnitude, &c. of the Fixed Stars, in Consequence of the Diminution of the Velocity of Their Light, in Case Such a Diminution Should be Found to Take Place in any of Them, and Such Other Data Should be Procured from Observations, as Would be Farther Necessary for That Purpose. By the Rev. John Michell, B. D. F. R. S. In a Letter to Henry Cavendish, Esq. F. R. S. and A. S., *Royal Society of London Philosophical Transactions Series* **74** (1784), 35–57.
- [42] J. M. Miller, Relativistic X-Ray Lines from the Inner Accretion Disks Around Black Holes, *ARA&A* **45** (2007), 441–479.
- [43] G. Miniutti, A. C. Fabian, R. Goyder and A. N.

Lasenby, The lack of variability of the iron line in MCG-6-30-15: general relativistic effects, *MNRAS* **344** (2003), L22–L26.

- [44] M. R. Morris, L. Meyer and A. M. Ghez, Galactic center research: manifestations of the central black hole, *Research in Astronomy and Astrophysics* **12** (2012), 995–1020.
- [45] N. Murray, E. Quataert and T. A. Thompson, On the Maximum Luminosity of Galaxies and Their Central Black Holes: Feedback from Momentum-driven Winds, *ApJ* **618** (2005), 569–585.
- [46] R. Nandra, A. N. Lasenby and M. P. Hobson, The effect of an expanding universe on massive objects, *MNRAS* **422** (2012), 2945–2959.
- [47] R. Narayan and J. E. McClintock, Advection-dominated accretion and the black hole event horizon, *New. Astron. Rev.* **51** (2008), 733–751.
- [48] E T. Newman, R. Couch, K. Chinnapared, A. Exton, A. Prakash and R. Torrence, Metric of a Rotating, Charged Mass, *J. Math. Phys.* **6** (1965), 918–919.
- [49] I. D. Novikov and K. S. Thorne, Astrophysics of black holes., in: *Black Holes (Les Astres Occlus)* (C. DeWitt and B. S. DeWitt, eds.), pp. 343–450, 1973.
- [50] J. R. Oppenheimer and H. Snyder, On Continued Gravitational Contraction, *Physical Review* **56** (1939), 455–459.
- [51] J. A. Orosz, J. E. McClintock, R. Narayan, C. D. Bailyn, J. D. Hartman, L. Macri, J. Liu, W. Pietsch, R. A. Remillard, A. Shporer and T. Mazeh, A 15.65-solar-mass black hole in an eclipsing binary in the nearby spiral galaxy M 33, *Nature* **449** (2007), 872–875.
- [52] D. N. Page and S. W. Hawking, Gamma rays from primordial black holes, *ApJ* **206** (1976), 1–7.
- [53] M. L. Parker, D. R. Wilkins, A. C. Fabian, D. Grupe, T. Dauser, G. Matt, F. A. Harrison, L. Brenneman, S. E. Boggs, F. E. Christensen, W. W. Craig, L. C. Gallo, C. J. Hailey, E. Kara, S. Komossa, A. Marinucci, J. M. Miller, G.

- Risaliti, D. Stern, D. J. Walton and W. W. Zhang, The NuSTAR spectrum of Mrk 335: extreme relativistic effects within two gravitational radii of the event horizon?, *MNRAS* **443** (2014), 1723–1732.
- [54] R. Penrose, Gravitational Collapse: the Role of General Relativity, *Nuovo Cimento Rivista Serie* **1** (1969), 252.
- [55] B. M. Peterson, Measuring the Masses of Supermassive Black Holes, *Space Sci. Rev.* (2013).
- [56] O. Pfuhl, S. Gillessen, F. Eisenhauer, R. Genzel, P. M. Plewa, T. Ott, A. Ballone, M. Schartmann, A. Burkert, T. K. Fritz, R. Sari, E. Steinberg and A.-M. Madigan, The Galactic Center cloud G2 and its gas streamer, *ArXiv e-prints* 1407.4354 (2014).
- [57] T. Piran, The physics of gamma-ray bursts, *Reviews of Modern Physics* **76** (2004), 1143–1210.
- [58] J. E. Pringle, M. J. Rees and A. G. Pacholczyk, Accretion onto Massive Black Holes, *A&A* **29** (1973), 179.
- [59] R. A. Remillard and J. E. McClintock, X-Ray Properties of Black-Hole Binaries, *ARA&A* **44** (2006), 49–92.
- [60] E. E. Salpeter, Accretion of Interstellar Matter by Massive Objects., *ApJ* **140** (1964), 796–800.
- [61] M. Schmidt, 3C 273: A StarLike Object with Large RedShift, *Nature* **197** (1963), 1040.
- [62] E. L. Schucking, The first Texas symposium on relativistic astrophysics, *Physics Today* **42** (1989), 46–52.
- [63] K. Schwarzschild, On the Gravitational Field of a Mass Point According to Einstein's Theory, *General Relativity and Gravitation* **35** (2003), 951–959.
- [64] C. K. Seyfert, Nuclear Emission in Spiral Nebulae., *ApJ* **97** (1943), 28.
- [65] N. I. Shakura and R. A. Sunyaev, Black holes in binary systems. Observational appearance., *A&A* **24** (1973), 337–355.
- [66] J. Silk and M. J. Rees, Quasars and galaxy formation,

*A&A* **331** (1998), L1-L4.

- [67] A. Soltan, Masses of quasars, *MNRAS* **200** (1982), 115–122.
- [68] A. A. Starobinskii, Amplification of waves during reflection from a rotating “black hole”, *Soviet Journal of Experimental and Theoretical Physics* **37** (1973), 28.
- [69] A. Tchekhovskoy, R. Narayan and J. C. McKinney, Efficient generation of jets from magnetically arrested accretion on a rapidly spinning black hole, *MNRAS* **418** (2011), L79–L83.
- [70] K. S. Thorne, Disk-Accretion onto a Black Hole. II. Evolution of the Hole, *ApJ* **191** (1974), 507–520.
- [71] W. G. Unruh, Notes on black hole evaporation, *Phys. Rev. D* **14** (1976), 870.
- [72] C. M. Urry and P. Padovani, Unified Schemes for Radio-Loud Active Galactic Nuclei, *PASP* **107** (1995), 803.
- [73] P. Uttley, E. M. Cackett, A. C. Fabian, E. Kara and D. R. Wilkins, X-ray reverberation around accreting black holes, *ArXiv e-prints* (2014).
- [74] M. van der Klis, Millisecond Oscillations in X-ray Binaries, *ARA&A* **38** (2000), 717–760.
- [75] R. M. Wald, *General Relativity*, University of Chicago Press, 1984.
- [76] B. L. Webster and P. Murdin, Cygnus X-1-a Spectroscopic Binary with a Heavy Companion?, *Nature* **235** (1972), 37–38.
- [77] A. Zee, *Einstein gravity in a nutshell*, Princeton University Press, 2013.
- [78] Y. B. Zel'dovich, The Fate of a Star and the Evolution of Gravitational Energy Upon Accretion, *Soviet Physics Doklady* **9** (1964), 195.

Gene G. Byrd, Arthur Chernin, Pekka Teerikorpi, and Mauri Valtonen

## **Observations of General Relativity at strong and weak limits**

**Abstract:** Einstein's General Relativity theory has been tested in many ways during the last hundred years as reviewed in this chapter. Two tests are discussed in detail in this article: the concept of a zero gravity surface, the roots of which go back to Järnefelt, Einstein and Straus, and the no-hair theorem of black holes, first proposed by Israel, Carter and Hawking. The former tests the necessity of the cosmological constant  $\Lambda$ , the latter the concept of a spinning black hole. The zero gravity surface is manifested most prominently in the motions of dwarf galaxies around the Local Group of galaxies. The no-hair theorem is testable for the first time in the binary black hole system OJ287. These represent stringent tests at the limit of weak and strong gravitational fields, respectively. In this article we discuss the current observational situation and future possibilities.

### **1 Introduction**

In his theory of General Relativity, Einstein concluded that matter causes curvature in the surrounding spacetime, and bodies react to this curvature in such a way that there appears to be a gravitational attraction which causes acceleration [1]. From the geometry of spacetime, it is possible to calculate the orbits of bodies which are influenced by gravity. In flat spacetime the force-free motion happens on a straight line, but in a spacetime curved by mass/energy the force free motion can create practically closed orbits as seen in the elliptical motion of a planet around the Sun.

The above relationship is described mathematically by the Einstein equations [2]:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = -\frac{8\pi G}{c^4}T_{\mu\nu} .$$

[\(1\)](#)

The Ricci tensor  $R_{\mu\nu}$  and the Ricci scalar  $R$  are functions of the metric tensor  $g_{\mu\nu}$ . The metric tensor describes the geometry of the spacetime while the Ricci tensor and the Ricci scalar measure its curvature. Two dimensional examples are the above-mentioned flat space surface with infinite radius of curvature corresponding to  $R$ , the surface of a sphere with finite positive curvature and a hyperboloid “saddle” surface with finite negative curvature. The energy momentum tensor  $T_{\mu\nu}$  has the dominant component  $T_{00} = \rho_M c^2$  where  $\rho_M$  is the matter density.  $G$  is the gravitational constant and  $c$  is the speed of light. The Einstein equations contain  $\Lambda$ , the so-called cosmological constant which we will discuss in more detail later. Here it suffices to say that the cosmological constant acts opposite matter’s gravitation. See e.g. [3, 4] for background details.

**Gene G. Byrd:** University of Alabama, Tuscaloosa, AL, 35487-0324, USA

**Arthur Chernin:** Sternberg Astronomical Institute, Moscow University, 119899, Moscow, Russia

**Pekka Teerikorpi:** Tuorla Observatory, Department of Physics and Astronomy, University of Turku, FI-21500, Piikkiö, Finland

**Mauri Valtonen:** FINCA, University of Turku, FI-21500 Piikkiö, Finland

In the presence of the cosmological constant  $\Lambda$  the metric of a spherically symmetric object of mass  $M$  was calculated by Lemaître and McVittie:

$$ds^2 = - \left( 1 - \frac{2GM}{rc^2} - \frac{\Lambda r^2}{3} \right) dt^2 + \left( 1 - \frac{2GM}{rc^2} - \frac{\Lambda r^2}{3} \right)^{-1} dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)$$

(2)

where  $ds$  is the line element in the 4-dimensional spacetime, and  $t, r, \theta, \phi$  are the spherical polar coordinates, centered on the body ([5, 6] and [7]). For a simple unchanging flat space in spherical coordinates with no interior matter or no cosmological constant,

$$ds^2 = dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2).$$

(3)

For the metric with the cosmological constant above, the deviation from the Minkowski “flat” metric is minimal when

$$r = R_V = \left( \frac{3GM}{\Lambda c^2} \right)^{1/3}.$$

[\(4\)](#)

The surface defined by this radial distance from the center is called the zero gravity surface. Inside this radius, matter gravitation dominates while outside the cosmological constant’s antigravity repulsion dominates. It is the weak field limit of General Relativity in the presence of the cosmological constant. It becomes significant in the smallest scales of cosmological expansion, such as in the Hubble flow around the Local Group of galaxies. Järnefelt was the first to derive [Equation \(4\)](#) [8]. The zero gravity surface is now understood to appear within the “vacuole” introduced by Einstein and Straus [9, 10] to describe the environments of bound mass concentrations in expanding space. This weak field limit is the domain of our first tests of General Relativity. We will later describe this weak field limit’s application to the Local Group and other nearby systems in detail as the closest study in space and time of the cosmological constant (or dark energy).

The weak field is often also defined by [11]

$$v/c \ll 1, \quad r_s/r \ll 1$$

(5)

while in case of a strong field we have

$$v/c \sim 1, \quad r_s/r \sim 1.$$

(6)

Here  $v$  is the orbital speed or other characteristic velocity,  $r$  is the orbital radius or other characteristic length scale in the system and  $r_s$  is the Schwarzschild radius of the primary body of mass  $M$

$$r_s = \frac{2GM}{c^2}.$$

(7)

We will first review briefly a number of tests in the first weak field category. Historically important is the precession of Mercury's orbit which served as a stimulus to Einstein. Also important is the observation of the predicted amount of bending of light near the Sun which served to substantiate General Relativity. More recent weak field tests are the gravitational red shift, the precession of a binary pulsar orbit, gravitational radiation from a binary pulsar, relativistic geodetic precession and the precession due to the relativistic Lense-Thirring effect. In these tests, the cosmological constant (dark energy) plays no role. We review the observational discovery at cosmological distances and times of dark energy's effects on the observed redshift-distance relation. Since they reflect the closest cosmological distance and present-day effects, we review in detail motions of satellite galaxies around the Local Group (and other nearby systems) near the zero gravity surface. In contrast to the weak field tests, for the strong field category we have at present only the binary black hole system OJ287, and hopefully in future other merging black hole binaries. OJ287 is thus also discussed in detail since it represents the best case for testing General Relativity in the stong field case.

## 2 Tests in the solar system and binary systems

### 2.1 Orbit precession

Before the detailed discussions, to provide background, we give a review of tests historically important as an impetus for or in the acceptance of General Relativity. The first test case (which served as an impetus for General Relativity) is the non-Newtonian precession of the major axis of the planet Mercury. In Newtonian theory, a small object orbiting a concentrated spherically symmetric central body should retrace the same elliptical path repeatedly. In case of Mercury, the major axis of its elliptical orbit precesses slowly. The precession is mostly due to gravitation of other planets. However, in 1859, the mathematician Urbain Le Verrier of Paris Observatory concluded, after the influence of other planets have been deducted, that there was a remaining observed shift of  $38''/\text{century}$  in the perihelion of Mercury [12]. He ascribed this to an unknown planet inside Mercury's orbit. This presumed planet was never found despite extensive searches. In 1895, the United States Naval Observatory's Simon Newcomb [13] recalculated the observed perihelion shift and obtained  $41'' \pm 2''/\text{century}$  (also see [14] and [15]).

With General Relativity Einstein [16] derived an additional precession term to be added to the Newtonian precession. Einstein's calculation gave a "post Newtonian" (PN) force law which is almost but not quite the inverse square law proposed by Newton. For the post-Newtonian case, in one orbit, the ellipses traced by the two bodies orbiting each other precess through an angle given by

$$\Delta\phi = \frac{6\pi GM}{c^2 a(1 - e^2)} = 3\pi \frac{r_s}{a(1 - e^2)},$$

(8)

where  $M$  is the total mass of the two bodies,  $a$  is the semimajor axis of the relative motion and  $e$  is the eccentricity. When one of

the bodies is much greater than the other, as in the Sun-Mercury system, the latter equality is useful since it scales the result to the Schwarzschild radius of the primary body.

To calculate a numerical value for the relativistic perihelion shift  $\Delta\phi$  for Mercury, substitute the Schwarzschild radius for the Sun,  $r_s = 2.96$  km, the semimajor axis of Mercury's orbit  $a = 57.91 \times 10^6$  km, and its  $e = 0.2056$  into the Equation above. Multiply the result by 206,265 to go from radians to arcsec and by 415, the number of revolutions per 100 years to obtain

$$\Delta\phi = 43''$$

(9)

per century. The agreement found in this early test of General Relativity gave great confidence in the theory. Today's accuracy of testing General Relativity in the Solar System is at a level of 0.01%. Here the characteristic velocity is  $v/c \sim 10^{-4}$ , and we are dealing with weak fields.

Another more extreme and recent test of General Relativity is the radio pulsar PSR 1913+16 , a smaller member of a binary system of two neutron stars about 1/60 as far from one another as Mercury is from the Sun. Using the semimajor axis of the binary  $a \sim 1.95 \times 10^6$  km, the total mass  $M \sim 2.83M_\odot$  and eccentricity  $e \sim 0.617$ , we find that in one year (1131 orbital revolutions) the periastron advances

$$\Delta\phi \sim 4.2^\circ .$$

(10)

## 2.2 Gravitational waves

Einstein [278, 279] proposed that the elasticity of space can create gravitational waves, small changes in the curvature of space which propagate in space with speed of light. For more description see [3]. Observations show that the binary system does lose energy which cannot be explained in other ways

beside gravitational wave emission. The orbital parameters of PSR 1913+16 result in a prediction for the rate of loss of the orbital energy by gravitational radiation which matches observations with the accuracy of 0.2% ([17-19], and [20]). This coincidence is usually taken as proof that gravitational waves do exist, even though the radiation from PSR 1913+16 is not presently directly measurable by gravitational wave antennas. In binary pulsars, a most interesting object is the double pulsar PRS J0737-3039 with the perihelion advance of  $16.9^\circ$  per year ([21] and [22]), we have  $v/c \sim 2 \times 10^{-3}$ . Here we are also testing General Relativity in weak fields, but including gravitational radiation.

## **2.3 Lense-Thirring effect and relativistic spinorbit coupling**

In 1918, the dragging of space around rotating bodies in General Relativity was proposed by Austrian physicists Joseph Lense and Hans Thirring [23]. By 2004 the Lense-Thirring effect was measured in space surrounding the rotating Earth. Following the motions of two Earth satellites LAGEOS I and II , a team led by Ignazio Ciufolini of University of Lecce, Italy, and Erricos Pavlis of University of Maryland, found that the planes of the orbits of the satellites have shifted by about two meters per year in the direction of the Earth's rotation due to dragging of space around the Earth [24] (see the next chapter in this book). The result is in agreement with the prediction of Lense and Thirring within the 10% accuracy of the experiment. The satellite Gravity Probe B, designed for the measurement of space dragging confirmed these results in 2007 [25]. The Gravity Probe B Lense-Thirring value is  $0.04'' \text{ yr}$  *in contrast to the much larger  $6.6'' \text{ yr}$  for the geodetic precession of its 7,027 km, 97.65 minute orbit around the Earth.*

In the 1970s theoretical calculations for relativistic spinorbit coupling were made. This coupling may cause the PSR 1913+16 pulsar's spin axis to precess ([26-29], and [30]). This precession can result in changes in pulse shape as the pulsar-observer

geometry changes. Under the assumption that relativistic precession is occurring, these changes have been used to model the two-dimensional structure of the pulsar beam.

## 2.4 Bending of light rays and gravitational redshift

An early convincing test of the General Theory of Relativity proposed by Einstein is the bending of light rays which pass close to the Sun. Consider a photon passing by a mass  $M$  with  $b$  as the minimum distance between the photon and the mass  $M$ . The total deflection from a straight line path is [31]

$$\Delta\phi = \frac{4GM}{c^2 b} = 2\frac{r_s}{b}$$

(11)

radians. For a light ray passing the surface of the Sun,  $r_s = 2.96$  km and  $b = 6.96 \times 10^5$  km. General Relativity gives

$$\Delta\phi = 1.75'' ,$$

(12)

in good agreement with observations a few years later by Eddington and Dyson [32, 33]. Today better measurements using radio sources agree with Einstein's theory within 1% accuracy. A classical Newtonian calculation only gives half the relativistic value. Bending of light provided the second early convincing test of the General Relativity theory. Via the deflection, stars close to the edge of the Sun appear to shift radially outward from the center of the Sun during a solar eclipse compared to a photograph of the same area six months earlier.

In a remarkable extension of the classical test of relativity beyond using the Sun, deflection of light from background sources has been seen around massive objects resulting in gravitational lensing, multiple images or strong distortions of the sources. The apparent source flux may also change. Multiple images of the quasar Q0957+561 were detected in 1979 by a

team led by Dennis Walsh of the University of Manchester [34] using the 2.1 meter telescope of Kitt Peak National Observatory. Nowadays gravitational lenses are detected frequently, and are used in astrophysical studies, in particular in estimating the relative amounts of dark matter in clusters of galaxies and matter's importance relative to dark energy in the universe.

Another test proposed by Einstein is gravitational redshift [25] in which the frequency an emitted photon is fractionally decreased by an amount

$$\left(1 - \frac{2GM}{rc^2}\right)^{1/2} \quad (13)$$

where  $r$  is the radius of emission and  $M$  is the mass of the body and the observer is at a large distance. After attempts to observe it in spectra of white dwarf stars, in 1959 this was definitely verified by Pound and Rebka [36] on the Earth using nuclear resonance.

## 2.5 Massive spinning black hole test

A promising case for the study of relativistic spinorbit coupling (as well as instrumental observation of gravitational radiation) is the binary black hole system of quasar OJ287 to be discussed later [37]. Here one of the members is more massive than a star by a factor of  $10^{10}$ . Thus gravitational wave emission from this source should be very much more powerful than from the neutron stars of PSR 1913+16. The next generation of gravitational wave antennas should be able to directly confirm the emission of gravitational waves ([38, 39]).

The curvature of spacetime around a rotating black hole was first calculated in 1963 by the New Zealand mathematician Roy Kerr [40]. By conservation of angular momentum, a black hole arising from a rotating body must also rotate. The rotation of the black hole influences the surrounding spacetime even well beyond the black hole's Schwarzschild radius.

The quadrupole moment of the spinning primary black hole in OJ287 has a measurable effect on the orbit of the secondary. In OJ287,  $v/c \sim 0.25$ . Thus here we are carrying out strong field tests of General Relativity. For example, it has been shown by Valtonen and collaborators that the loss of orbital energy from the system agrees with General Relativity with the accuracy of 2% [41]. More importantly, we may test the no hair theorems of black holes for the first time ([42–46], and also [47]). These theorems relate the spin and the mass of a black hole to its quadrupole moment in a unique way (see Section 9).

### **3 Observational discovery of a nonzero cosmological constant (dark energy)**

We will describe the details of a cosmologically nearby test of General Relativity including the cosmological constant. But first, we will briefly review how the constant's nonzero value was discovered to be important by observations at cosmological distances and times. Based on General Relativity, in 1917, Einstein proposed a model for a curved finite (but still boundless) universe [2]. The cosmological constant,  $\Lambda$ , was specified so as to produce a static universe with no origin in time. The evolving models of the universe, standard today, were derived in the 1920s by the Russian Alexander Friedmann [48, 49]. These papers, a turning point in the study of cosmology, remained almost unnoticed. In 1927 Belgian astronomer Georges Lemaître [50] rediscovered these models, now known as Friedmann universes.

Friedmann found a solution of the General Relativity equations which is more general than Einstein's solution. The Friedmann solution includes the Einstein static solution as a special case, where the cosmological constant is nonzero and related to the matter density so as to attain the gravity-antigravity balance. However, generally, the solution puts no restrictions on the cosmological constant. It might be zero or nonzero, positive or negative, related or unrelated to the matter

density. The solution depends explicitly on time. The universe is not static: it expands or contracts as a whole.

Friedmann preferred expansion over contraction citing observational evidence supporting his choice in Slipher's data of galaxies that are moving away from us. Friedmann died in 1925 at the age of 37 before the Hubble's 1929 discovery of the redshift distance relation [51]. We can thus regard the discovery of the Hubble expansion as a predictive verification of General Relativity.

After the discovery of the expansion of the universe, and other evidence for an origin in time, the so called Big Bang, it became obvious that  $\Lambda$  could not be as big as Einstein had calculated for his static universe model. It became common to assume that  $\Lambda = 0$ . The ideas changed in the late 1990s when it became possible to use extremely luminous standard candles, supernovae of type Ia, to estimate distances of galaxies whose redshifts  $z$  are comparable to unity. In 1998–99 two groups headed by Riess and Perlmutter discovered the nonzero  $\Lambda$ , usually interpreted as an indication of cosmic vacuum or dark energy [52, 53].

## 4 Weak limit test near zero gravity surface

### 4.1 Dark energy antigravity as a test of General Relativity

In recent years it has become customary to move the  $\Lambda$  term from the left hand side of [Equation \(1\)](#) to the right hand side. Then it may be viewed as a contribution to the energy momentum tensor. The corresponding density is called the vacuum density  $\rho_V$ , with the equation of state

$$\rho_V = -P/c^2 , \quad (14)$$

where  $P$  is the vacuum pressure. The value of the vacuum

density is related to  $\Lambda$  by

$$\Lambda = \frac{8\pi G}{c^2} \rho_V . \quad (15)$$

Instead of  $\Lambda$  it is common to state its normalized value

$$\Omega_\Lambda = \frac{c^2 \Lambda}{3H_0^2} \quad (16)$$

which is a dimensionless number of the order of unity.  $H_0$  is the Hubble constant. From the cosmological recession of distant galaxies using Ia supernovae, the analysis of the microwave background radiation (CMB) and by many other ways (see Section 5), we find

$$\rho_V \approx 7 \times 10^{-30} \text{ g cm}^{-3}, \quad \Omega_\Lambda \approx 0.73 . \quad (17)$$

According to General Relativity, gravity depends on pressure as well as density: the effective gravitating density

$$\rho_{\text{eff}} = \rho + 3P/c^2 . \quad (18)$$

It is negative for a vacuum:  $\rho_{\text{eff}} = -2\rho_V$ , and this leads to repulsion (“antigravity”). Hence the study of the antigravity in our neighbourhood, and on short scales in general, is an important test of General Relativity and its concept of Einstein’s  $\Lambda$ -term in weak gravity conditions.

## 4.2 Local dark energy test via outflow

One recently proposed way to study the local dark energy is via an outflow model ([54, 55] and [4]) which describes expansion flows around local masses. It was motivated by the observed

picture of the Local Group with outflowing dwarf galaxies around it ([56] and [57]). The model treats the dwarfs as “test particles” moving in the force field produced by the gravitating mass of the group and the possible dark energy background. A static and spherically symmetric gravitational potential is a rather good approximation despite of the binary structure of the Local Group [58].

Soon after the discovery of the universal acceleration from observations of distant supernovae, researchers returned to the old theoretical question of Järnefelt, Einstein and Straus about what happens to the spacetime around a mass concentration in an expanding universe and asked: at what distance from the Local Group do the gravity of its mass and the antigravity of the dark energy balance each other ([54] and [59])?

Treating a mass concentration as a point mass  $M$  on the background of the antigraviting dark energy, its gravity produces the radial force  $-GM/r^2$ , where  $r$  is the distance from the group barycenter. The antigravity of the vacuum produces the radial force  $G2\rho_V(4\pi/3)r^3/r^2 = (8\pi/3)\rho_V r$ . Here  $-2\rho_V$  is the effective gravitating density of a vacuum. Then the radial component of motion in this gravity/antigravity force field obeys the Newtonian equation

$$\ddot{r} = -GM/r^2 + r/A^2,$$

(19)

where  $r$  is the distance of a particle to the barycenter of the mass concentration. The constant

$$A = [(8\pi G/3)\rho_V]^{-1/2}$$

(20)

is the characteristic vacuum time and has the value  $= 5 \times 10^{17}$  s (or 16 Gyr) for  $\rho_V = 7 \times 10^{-30} \text{ g cm}^{-3}$ .

[Equation \(19\)](#) shows that the gravity force ( $\propto 1/r^2$ ) dominates the antigravity force ( $\propto r$ ) at short distances. At the “zero-gravity distance”

$$R_V = (GMA^2)^{1/3} = [(3/8\pi)(M/\rho_V)]^{1/3}$$

(21)

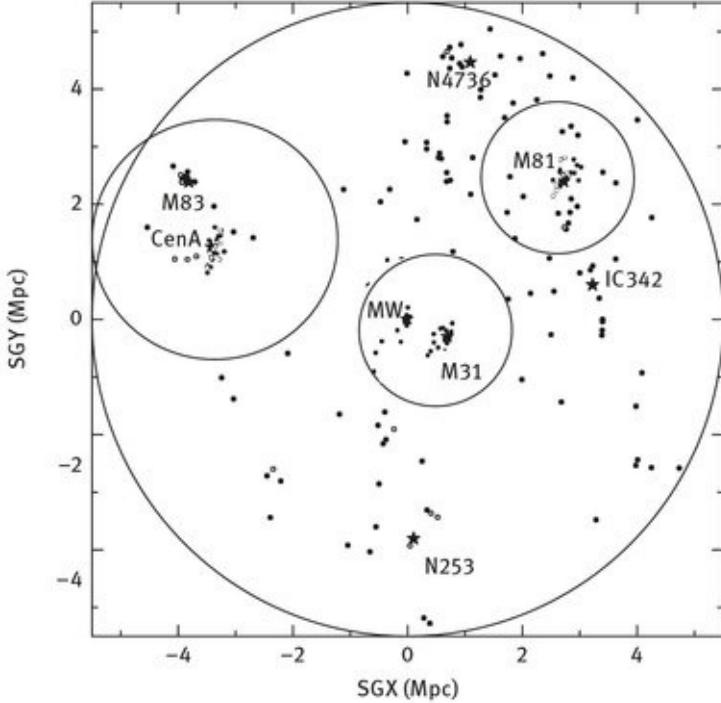
the gravity and antigravity balance each other, and the acceleration is zero. At larger distances,  $r > R_V$ , antigravity dominates, and the acceleration is positive. For the Local Group mass  $M \approx 4 \times 10^{12} M_\odot$  and the global dark energy density, the zero-gravity distance is  $R_V \approx 1.6$  Mpc. [Equation \(4\)](#) shows that we arrive at the same concept also in full General Relativity.

The zero-gravity sphere for a point mass  $M$  has special significance in an expanding universe. A light test particle at  $r > R_V$  experiences an acceleration outwards relative to the point mass. If it has even a small recession velocity away from  $M$ , it participates in an accelerated expansion. For an isolated system of two identical point masses  $M$ , the zero-gravity distance, where the two masses have zero-acceleration relative to their center-of-mass, is about  $1.26R_V$  [60]. If separated by a larger distance, the masses will experience outward acceleration.

These examples illustrate the general result that in vacuum-dominated expanding regions perturbations do not grow. They also lead one to consider the Einstein-Straus [9] solution where any local region may be described as a spherical expanding “vacuole” embedded within the uniform distribution of matter [55]. The metric inside the vacuole is static and the zero-gravity radius and the (present-time) Einstein-Straus vacuole radius are simply related:

$$R_{ES}(t_0) = (2\rho_V/\rho_M)^{1/3} R_V .$$

(22)



**Fig. 1.** Approximate zero-gravity spheres around the Local Group (at the center) and two nearby galaxy groups. The radii have been calculated using the masses  $2 \times 10^{12} M_{\odot}$  for the LG and the M81/M82 Group and  $7 \times 10^{12} M_{\odot}$  for the CenA/M83 Group (the underlying map presents the local environment up to about 5 Mpc as projected onto the supergalactic plane; it is from [59] and <http://arxiv.org/pdf/astroph/0211011v1.pdf>).

For instance, for the ratio  $\rho_V/\rho_M = 0.7/0.3$  one obtains  $R_{ES}(t_0) = 1.67R_V$ . The Einstein-Straus vacuole can be seen as the volume from which gravitation has gathered the matter to form the central mass concentration.

[Figure 1 shows](#) the map of our local extragalactic environment up to about 5 Mpc, together with the approximate zero-gravity spheres drawn around the Local Group and two nearby galaxy groups. The spheres do not intersect. This suggests that the groups are presently receding from each other with acceleration.

Using [Equation \(21\)](#) one may calculate typical zero-gravity radii for different astronomical systems, for the standard value of  $\rho_V$ . For stars, star clusters, galaxies and tight binary galaxies the zero-gravity radius is much larger than the size of the system which is located deep in the gravity-dominated region.

For galaxy groups and clusters,  $R_V$  is near or within the region where the outflow of galaxies begins to be observed (about 1.6 Mpc for the Local Group). It is especially on such scales where the system and its close neighborhood could shed light on the local density of dark energy. We thus will discuss this in detail.

### 4.3 Dynamical structure of a gravitating system within dark energy

The particles move radially as predicted by the Newtonian equation of motion ([Equation \(19\)](#)), but where the forces are the gravity of the central mass and the antigravity of the dark energy. The first integral of this equation expresses the mechanical energy conservation:

$$\frac{1}{2}\dot{r}^2 = E - U(r), \quad (23)$$

where  $E$  is the total mechanical energy of a particle (per unit mass) and  $U(r)$  is the potential energy

$$U(r) = -\frac{GM}{r} - \frac{1}{2}\left(\frac{r}{A}\right)^2. \quad (24)$$

Because of the vacuum, the trajectories with  $E < 0$  are not necessarily finite. Such behavior of the potential has a clear analogue in General Relativity applied to the same problem.

The total energy of a particle that has escaped from the gravity potential well of the system must exceed the maximal value of the potential:

$$E > U_{\max} = -\frac{3}{2}\frac{GM}{R_V}. \quad (25)$$

It is convenient to normalize the equations to the zero-gravity distance  $R_V$  and consider the Hubble diagram with normalized

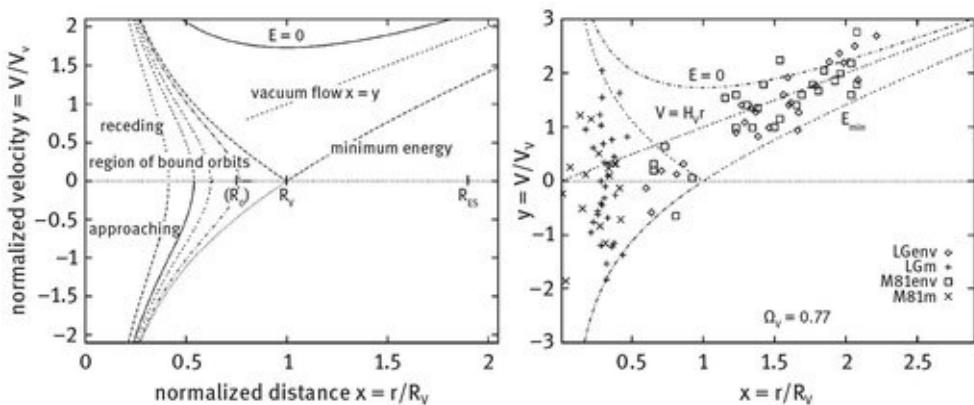
distance and velocity:  $x = r/R_V$  and  $y = (V/H_V)R_V$  where  $V$  is the radial velocity and  $H_V$  is a constant to be discussed below [62]. Then radially moving test particles will move along curves which depend only on the constant total mechanical energy  $E$  of the particle:

$$y = x(1 + 2x^{-3} - 2\alpha x^{-2})^{1/2}. \quad (26)$$

Here  $\alpha$  parameterizes the energy, so that  $E = -\alpha GM/R_V$ . Each curve has a velocity minimum at  $x = 1$ , i.e. at  $r = R_V$ .

The energy with  $\alpha = 3/2$  is the minimum energy which still allows a particle initially below  $x = 1$  to reach this zero-gravity border (and if the energy is slightly larger) to continue to the vacuum-dominated region  $x > 1$ , where it starts accelerating. In the ideal case one does not expect particles with  $x > 1$  below this minimum velocity curve.

[Figure 2](#) (first panel) shows different regions in the normalized Hubble diagram. Below  $r = R_V$ , we have indicated the positive minimum velocity curve and its negative symmetric counterpart. This region defines the bound group: a galaxy will not escape beyond  $R_V$  unless it obtains sufficient energy from an interaction.



**Fig. 2.** First panel: Different regions in the normalized Hubble diagram around a point mass in the vacuum. The quantities  $x$  and  $y$  are defined in the text. In the region of bound orbits a dwarf galaxy cannot move into the vacuum flow region unless it receives extra energy as a result of an interaction with other galaxies. Second panel:

The normalized Hubble diagram for the observed smaller galaxies in the environments of the Local and M 81 groups, for the near standard vacuum density. The velocity-distance relation for the vacuum flow is shown. The curve for the lower limit velocity is given below and above  $x = 1$ ; below  $x = 1$  its negative counterpart is shown. The members of the groups are found within these curves. (Adapted from [62].)

The diagonal line  $x = y$  gives the “vacuum” flow with the Hubble constant  $H_V$ , when dark energy is fully dominating. It is asymptotically approached by the outflying particles beyond  $x = 1$ . This limit is described by de Sitter’s static solution which has the metric of [Equation \(2\)](#) with  $M = 0$ . The spacetime of de Sitter’s solution is determined by the vacuum alone, which is always static. It leads to the linear velocity-distance law,  $V = r/A = H_V r$ , with the constant expansion rate  $H_V = (8\pi G \rho_V / 3)^{1/2}$ :

$$H_V = 61 \times \left( \frac{\rho_V}{7 \times 10^{-30} \text{ g/cm}^3} \right)^{1/2} \text{ km/s/Mpc} . \quad (27)$$

The normalized vacuum energy density depends on  $H_V: \Omega_\Lambda = H_V^2 / H_0^2$ . The vacuum Hubble time  $T_V = 1/H_V$  exceeds the global Hubble time ( $= 1/H_0$ ) by the factor  $(1 + \rho_M/\rho_V)^{1/2} = (\Omega_\Lambda)^{-1/2}$  for a flat universe. In the standard model  $T_V = 16 \times 10^9$  yr and the age of the universe ( $13.7 \times 10^9$  yr) is about  $0.85 T_V$ .

[Figure 2](#) (second panel) shows a combined normalized Hubble diagram for the Local Group, the nearby M81 group, and their environments. The M81 group is at a distance of about 4 Mpc. We have used in this diagram the mass  $2 \times 10^{12} M_\odot$  for both groups. The energy condition to overcome the potential well ( $E > -3/2GM/R_V$ ) is not violated in the relevant range  $x = r/R_V > 1$ .

## 5 Estimating cosmologically nearby dark energy: the Local Group

One expects inward acceleration at distances  $r < R_V$  and

outward acceleration at distances  $r > R_V$  within the region where the point-mass model is adequate. However, such accelerations in the nearby velocity field around the Local Goup are very small (of the order of 0.001 cm/s/yr) and impossible to measure directly. We also cannot follow a dwarf galaxy in its trajectory for millions of years in order to see the location of the minimum velocity which defines the zero-gravity distance.

The objects were likely expelled in the distant past within a rather narrow time interval (e.g. [63]). What we see now is a locus of points at different distances from the center and lying on different energy curves; they make the observed velocity-distance relation.

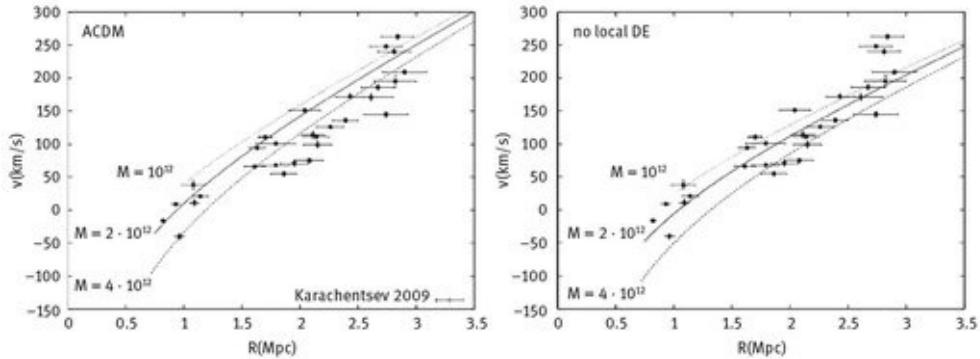
The prediction for the present-day outflow of dwarf galaxies near the Local Group (or other galaxy systems) depends on the mass of the group, the flight time (< the age of the universe), and the local dark energy density. The existence of and the value for the local dark energy density is a local test of this aspect of General Relativity. However, to study dark energy density, one must include known values for the other quantities listed above.

Peirani & de Freitas Pacheco derived the velocity-distance relation using the Lemaître-Tolman model containing the cosmological constant, and compared this with the model for  $\Lambda = 0$ , using the central mass and the Hubble constant as free parameters [64]. Chernin and collaborators considered the total energy for different outflow velocities at a fixed distance, and then calculated the required time for the test particle to fly from near the group's center up to this distance [56]. The locus of the points corresponding to a constant age ( $\approx$  the age of the Universe) gave the expected relation.

Here we show some results from a modification of the above methods which easily allows one to vary the values of the relevant parameters and to generalize the model in various ways [65]. One generates particles close to the center of mass of the group and gives them a distribution of speeds. Then they are allowed to fly along the radial direction for a time  $T$ , and their distances from the center and velocities are noted. The

flight time  $T$ , during which the integration of the equations of motion is performed, is at most the age of the standard Universe, 13.7 Gyr.

We will compare two extreme “universes” local outflows compared to the observations around the Local Group to show how the “local” cosmological constant can be studied. When comparing the standard  $\Lambda$ CDM model, with its constant dark energy density on all scales, it is relevant to consider the “Swiss cheese model”, where the Universe has the same age as the standard model, but where the dark energy density is zero on small scales in systems of galaxies. This model could correspond to the case where dark energy (or analogous effects) operates on large scales only. Karachentsev and collaborators have obtained accurate outflow distances and speeds of galaxies in and outside the Local Group to 3 Mpc [57] which can be compared to the model predictions.



**Fig. 3.** First panel: The location of test particles as injected from the mass centre (curves) after the flight time  $T_0 = 13.7$  Gyr, for different masses ( $1 \times 10^{12} M_\odot$ ,  $2 \times 10^{12} M_\odot$ ,  $4 \times 10^{12} M_\odot$ ). Here the standard model is used (local dark energy = global dark energy). Second panel: The location of test particles after the flight time  $T_0 = 13.7$  Gyr, for different masses. Here the Swiss cheese model is used (no local dark energy). The data points are for the Local Group. (Credit. J. Saarinen)

[Figure 3 shows](#) the predicted distance-velocity curves for different masses in the standard model (which has dark energy in the plot volume) (first panel) and for the Swiss cheese model (which has no dark energy in the plot volume) (second panel). It is seen that the location of the curve depends rather strongly on

the adopted mass. For instance, the Swiss cheese curve with  $M = 2 \times 10^{12} M_{\odot}$  approximates the LG data along the distance range. The  $\Lambda$ CDM model which has nonzero dark energy with  $M \approx 4 \times 10^{12} M_{\odot}$  has a good fit beyond 1.5 Mpc. A general conclusion from the outflow data is that a low-mass Local Group ( $M \leq 2 \times 10^{12} M_{\odot}$ ) is associated with the case of no local dark energy, while a local density near the global dark energy density requires a higher mass,  $M \approx 4 \times 10^{12} M_{\odot}$ .

[Figure 3 shows](#) the cases when the particles were ejected just after the Big Bang. This case would correspond to a classical situation where the outflow around the central group is “primordial”. In practice, the age of the group and the outflowing dwarf galaxies must be less than the age of the Universe, and the origin of the outflow may be due to early interactions within the system, making galaxies escape from it (e.g., [66, 67], and [63]). The dark energy antigravity enhances the escape probability because it makes the particle potential energy barrier lower than in the presence of gravity only.

The calculations show clear differences between the two cases in the sense expected: the  $\Lambda$ CDM curves are above the no-local-dark-energy curves and steeper, making the zero-velocity distance longer in the latter case, as already noted by Peirani & de Freitas Pacheco [64] and Chernin *et al.* [58]. However, in practice, the difference may be difficult to detect because: First, the observed distance-velocity data points are rather scattered. Second, the independently known mass of the Local Group is uncertain. Thirdly, the model of the galaxy group has uncertainty in the exact ejection time.

As is diagramed in [Figure 2 and](#) mentioned in the following section, another way to obtain the local  $\Omega_{\Lambda}$ , is to take the measured values of  $H_V$  from the sloping line beyond the zero gravity radius in the above figures and  $H_0$  from observations of distances and Doppler shifts of other cosmologically local clusters. From the equations in the first part of Section 4,

$$\Omega_\Lambda = \left( \frac{H_V}{H_0} \right)^2 .$$

(28)

Using  $H_V = 59$  km/s/Mpc and  $H_0 = 70$  km/s/Mpc,  $\Omega_\Lambda = 0.71$ , again with  $\pm 10\%$  (the second line of [Table 1](#)).1).

## 6 Mass, dark energy density and the lost gravity effect

A mutual dependence between the system mass and the derived dark energy density is typical for various local dynamical tests. A proper determination of the Local Group mass is thus needed in this test of General Relativity. Mass determination in systems of galaxies (including dark energy) is discussed in detail in the Appended Section, “Mass, dark energy density and the lost gravity effect” placed after the Conclusions section. For those who wish to skip the details, we briefly summarize the Local Group mass estimate below. The virial theorem is the “classical” method of estimating cluster or group mass using the group size and velocity dispersion. Another, powerful method is the Kahn–Woltjer method which uses the Milky Way Galaxy and the Andromeda Galaxy as members of a binary which have moved apart radially since the Big Bang and are now approaching for first time. The pair separation, mutual radial velocity and the age of the universe are used in this second method. Cosmological simulations can also be used to model the Local Group. Dark energy should be included in all these estimates which modifies the estimated mass somewhat, with a range of 3 to  $5 \times 10^{12}$  solar masses. On the basis of the analysis in the detailed discussion, a Local Group mass of  $M = 4 \times 10^{12}$  solar masses, appears to be acceptable for fits to the outflow data, like that in [Figure 3](#). With this value we get  $\Omega_\Lambda = 0.75$ , the value given on the first line of [Table 1](#) (Section 8). The one standard deviation uncertainty may be estimated as  $\pm 10\%$ .

## **7 Dark energy in the Coma Cluster of galaxies**

More details are in the Appended Section “Dark energy in the Coma Cluster of galaxies”, given after the conclusions. Dark energy is less important in virial theorem mass estimates of rich clusters than in groups which will be helpful in local dark energy estimates. More and better distance estimates are needed to extend the process used for the Local Group to the Coma Cluster.

## **8 Testing the constancy of $\Lambda$**

The concept of dark energy (or more specifically Einstein’s cosmological constant) has become a routine factor in global cosmology on Gpc scales. We have shown above that it is also relevant in the local extragalactic universe. In the local weak gravity conditions the antigravity effect of dark energy becomes measurable and has to be included in dynamical studies performed in galaxy group and cluster scales of a few Mpc. In general, extragalactic mass determinations should include a correction term due to the ‘lost gravity’ effect caused by dark energy. Here we briefly mention a few other applications.

Valtonen and Byrd noted how the Hubble outflow is one factor in explaining the well-known redshift anomaly in local galaxy group data via a selection effect ([108, 109] and [110]). This asymmetry between redshifts and blueshifts of group members can be seen as a signature of local dark energy. See [4] for an updated review.

The zero-gravity radius is an important quantity, which has existed (with roughly the same value) since the formation of a mass concentration in the expanding universe. It defines a natural upper limit to the size of a gravitationally bound system, allowing one to give an upper limit to the cosmic gravitating matter density in the form of galaxy systems. For the local universe this was discussed in [101].

The measurement of the local value of the density of dark

energy is naturally very important for our understanding of the nature of dark energy. Is it really constant on all scales and at all times, as suggested by the original concept of Einstein's cosmological constant? After the initial period of rediscovery of the cosmological constant  $\Lambda$  in 1998–1999, there has been a large amount of activity in trying to determine the exact value of the constant, and to study its possible dependence on redshift ([111, 112], and [113]). As we stated before, in General Relativity  $\Lambda$  is an absolute constant.

Tables 1–3 list a sample of  $\Omega_\Lambda$  found in the literature from year 2005 onwards together with the one standard deviation error limits. The values cluster around  $\Omega_\Lambda = 0.73$  with a standard deviation of 0.044. This is somewhat smaller than the typical error in individual measurements. Only the cosmic microwave background (CMB) models give a significantly better accuracy, but in this case one may suspect hidden systematic errors due to foreground corrections (Whitbo [114]). Within the error bounds,  $\Omega_\Lambda$  is constant over the redshift range that can be studied. Note that the CMB model values have been placed at redshift  $z = 3$  since the higher redshift universe does not affect the derived  $\Omega_\Lambda$  significantly. The  $\Omega_\Lambda$  values obtained at the weak limit are the same as the cosmological determinations within errors. Thus there does not appear to be any variation in  $\Omega_\Lambda$  with time or distance. The  $\Lambda$  appears to be a constant just as specified in the Einstein equations.

**Table 1.**  $\Omega_\Lambda$  determinations at different redshifts.

Redshift	$\Omega_A$	Uncertainty	Reference
0	0.75	0.07	this work
0	0.71	0.07	Chernin [115]
0	0.76	0.07	Karachentsev et al. [57]
0.005	0.78	0.08	Mohayaee and Tully [116]
0.01	0.70	0.14	Ma and Scott [117]
0.01	0.76	0.13	Nusser and Davis [118]
0.01	0.71	0.10	Park and Park [119]
0.05	0.71	0.05	Kowalski et al. [120]
0.06	0.68	0.10	Henry et al. [121]
0.07	0.75	0.02	Beutler et al. [122]
0.1	0.71	0.03	Tinker et al. [123]
0.1	0.71	0.03	Simha and Cole [124]
0.1	0.72	0.03	Cacciato et al. [125]
0.1	0.76	0.02	Cole et al. [126]
0.1	0.78	0.02	Sanchez et al. [127]
0.16	0.72	0.08	Hicken et al. [128]
0.2	0.72	0.07	Rozo et al. [129]
0.2	0.78	0.08	Lampeitl et al. [130]
0.2	0.72	0.02	Hajian et al. [131]
0.22	0.72	0.05	Blake et al. [132]
0.23	0.73	0.03	Kessler et al. [133]
0.275	0.71	0.02	Percival et al. [134]
0.3	0.75	0.05	Cabre and Cazañaga [135]
0.3	0.75	0.07	Cao et al. [136]
0.3	0.80	0.15	Rest et al. [137]
0.3	0.75	0.10	Suyu et al. [138]
0.3	0.78	0.06	Mantz et al. [139]
0.3	0.75	0.08	Vikhlinin et al. [140]
0.3	0.76	0.06	Campbell et al. [141]
0.3	0.75	0.06	Ho et al. [142]
0.3	0.73	0.02	Del Popolo et al. [143]
0.35	0.64	0.12	Zhang and Wu [144]
0.35	0.72	0.04	Reid et al. [145]
0.35	0.74	0.02	Sanchez et al. [146]
0.35	0.72	0.02	Mehta et al. [147]
0.35	0.76	0.04	Tegmark et al. [148]
0.35	0.73	0.04	Eisenstein et al. [149]
0.36	0.77	0.05	Okumura et al. [150]
0.4	0.85	0.08	Taylor et al. [151]
0.4	0.72	0.06	Allen et al. [152]
0.4	0.73	0.04	Mandelbaum et al. [153]
0.4	0.82	0.03	Cabre et al. [154]
0.4	0.71	0.09	Blake et al. [155]
0.4	0.76	0.20	Freedman et al. [156]
0.41	0.71	0.07	Blake et al. [155]
0.43	0.73	0.04	Wood-Vasey et al. [157]

**Table 2.**  $\Omega_A$  table continued.

Redshift	$\Omega_A$	Uncertainty	Reference
0.45	0.85	0.15	Sullivan et al. [158]
0.48	0.73	0.15	Wilson et al. [159]
0.5	0.79	0.12	Oguri et al. [160]
0.5	0.72	0.03	Hinshaw et al. [161]
0.5	0.85	0.15	Cao and Zhu [162]
0.5	0.74	0.04	Ferramacho et al. [163]
0.5	0.71	0.06	Clocchiatti et al. [164]
0.5	0.71	0.02	Parkinson et al. [165]
0.5	0.72	0.02	Samushia et al. [166]
0.52	0.75	0.10	Huff et al. [167]
0.55	0.75	0.10	Del Popolo [168]
0.55	0.70	0.15	Ross et al. [169]
0.55	0.73	0.04	Blake et al. [170]
0.5	0.70	0.10	Gaztañaga [171]
0.55	0.73	0.04	Astier et al. [172]
0.57	0.74	0.05	Reid et al. [173]
0.57	0.74	0.04	Chuang et al. [174]
0.57	0.72	0.03	Zhao et al. [175]
0.57	0.71	0.02	Anderson et al. [176]
0.57	0.72	0.04	Chen et al. [177]
0.59	0.70	0.20	Farooq and Ratra [178]
0.6	0.69	0.13	Blake et al. [132]
0.6	0.79	0.10	Guy et al. [179]
0.6	0.61	0.05	Feoli et al. [180]
0.6	0.75	0.10	Conley et al. [181]
0.6	0.70	0.04	Addison et al. [182]
0.6	0.64	0.12	Ganeshalingam et al. [183]
0.6	0.76	0.05	Lin et al. [184]
0.6	0.68	0.05	Carneiro et al. [185]
0.65	0.73	0.04	Davis et al. [186]
0.66	0.71	0.02	Sereno and Paraficz [187]
0.7	0.84	0.10	Heymans et al. [188]
0.7	0.72	0.02	Seljak et al. [189]
0.7	0.75	0.02	Freedman et al. [190]
0.75	0.75	0.07	Benson et al. [191]
0.75	0.73	0.02	March et al. [192]
0.78	0.78	0.09	Blake et al. [132]
0.78	0.76	0.15	Benjamin et al. [193]
0.80	0.65	0.10	Ettori et al. [194]
0.80	0.73	0.10	Giannantonio et al. [195]
0.88	0.72	0.08	Chen and Ratra [196]
0.9	0.68	0.08	Semboloni et al. [197]
0.95	0.75	0.05	Fu et al. [198]
1.0	0.73	0.05	Jassal et al. [199]
1.0	0.73	0.03	Firmani et al. [200]
1.0	0.76	0.06	Basilakos and Plionis [201]

**Table 3.**  $\Omega_A$  table continued.

Redshift	$\Omega_A$	Uncertainty	Reference
1.02	0.73	0.02	Suzuki et al. [202]
1.3	0.75	0.06	Wei et al. [203]
1.3	0.76	0.09	Lopez-Corredoira [204]
1.3	0.73	0.02	Suzuki et al. [205]
1.4	0.65	0.15	da Angela et al. [206]
1.5	0.75	0.08	da Angela et al. [207]
1.6	0.63	0.13	Kodama et al. [208]
1.6	0.64	0.10	Tsutsui et al. [209]
2.3	0.75	0.10	Busca et al. [210]
2.5	0.73	0.06	Busti et al. [211]
3	0.65	0.20	Da Angela et al. [212]
3	0.75	0.03	Dunkley et al. [213]
3	0.69	0.02	The Planck Collaboration: Ade et al. [214]
3	0.72	0.03	Bennett et al. [215]
3	0.76	0.02	Spergel et al. [216]
3	0.74	0.03	Reichardt et al. [217]
3	0.73	0.02	Komatsu et al. [218]
3	0.73	0.02	Komatsu et al. [219]
3	0.75	0.12	Wang [220]
3	0.75	0.05	Liang et al. [221]
3	0.73	0.12	Schaefer [222]
3	0.61	0.14	Sherwin et al. [223]
3	0.65	0.15	Pietrobon et al. [224]
3	0.70	0.02	Spergel et al. [225]
3	0.65	0.10	Sievers et al. [226]
3	0.74	0.02	Larson et al. [227]

## 9 Strong limit: Spinning black holes and no-hair theorem

Up to now we have described weak limit tests of General Relativity. Strong limit tests involve black holes. There is plenty of evidence in support of the existence of black holes having masses in the range from a few  $M_\odot$  to a few  $10^{10} M_\odot$  (see the preceding chapter in this book). However, to be sure that these objects are indeed the singularities predicted by General Relativity, we have to ascertain that at least in one case the spacetime around the suspected black hole satisfies the no-hair theorems. The black hole no-hair theorems state that an electrically neutral rotating black hole in GR is completely described by its mass  $M$  and its angular momentum  $S$ . This implies that the multipole moments, required to specify the external metric of a black hole, are fully expressible in terms of

$M$  and  $S$ . It is important to note that the no-hair theorems apply only in General Relativity, and thus they are a powerful discriminator between General Relativity and various alternative theories of gravitation that have been suggested ([228, 231], and [230]).

A practical test was suggested by Thorne and collaborators ([231] and [232]). In this test the quadrupole moment  $Q$  of the spinning body is measured. If the spin of the body is  $S$  and its mass is  $M$ , we determine the value of  $q$  in

$$Q = -q \frac{S^2}{Mc^2} . \quad (29)$$

For true black holes  $q = 1$ , and for neutron stars and other possible bosonic structures  $q > 2$  ([233] and [234]). In terms of the Kerr parameter  $\chi$  and the dimensionless quadrupole parameter  $q_2$  the same equation reads

$$q_2 = -q\chi^2 , \quad (30)$$

where  $q_2 = c^4 Q/G^2 M^3$  and  $\chi = c S/G M^2$ .

An ideal test of the no-hair theorem is to have a test particle in orbit around a spinning black hole, and to follow its orbit. Fortunately, there exists such a system in nature. The BL Lacertae object OJ287 is a binary black hole of very large mass ratio, and it gives well defined signals during its orbit. These signals can be used to extract detailed information on the nature of the orbit, and in particular, to find the value of the parameter  $q$  in this system.

Valtonen, Sillanpää and collaborators found that the optical light curve of the quasar OJ287 displays periodicities of 11.8 and 55 years ([235] and [236]), as well as a  $\sim 50$ -day period of outbursts [237]. A model which explains these cycles, as well as a wealth of other information on OJ287 is discussed in the next section.

One of the pieces of information that we are able to find out is the spin of the primary black hole,  $\chi_1 = 0.25 \pm 0.04$ . The timing of the next outburst at the beginning of 2016 should help to improve the accuracy of  $\chi$  to about  $\pm 5\%$ . The mass of the primary is already determined with the accuracy of  $\pm 1\%$  which means that at least in principle we could reach the accuracy of  $\pm 10\%$  in measuring  $q$  [238].

There exist a number of other proposals to test the black hole no-hair theorems. The scenarios include the radio timing of eccentric millisecond binary pulsars which orbit an extreme Kerr black hole [233]. Such systems are yet to be discovered. Also one may use several stars orbiting the massive Galactic center (Sgr A\*) black hole at milliarcsec distances, if such stars are discovered and they are followed by infrared telescopes of the future, capable of doing astrometry at  $\sim 10 \mu$  arcseconds level [234]. Observations of gravitational waves from mergers of supermassive black holes, when they become possible some decades from now, may also be used to test the theorems [239].

The imaging of accretion flow around Sgr A\*, when it becomes possible, may allow the testing of no-hair theorems. The test relies on the argument that a bright emission ring characterizing the flow image will be elliptical and asymmetric if the theorems are violated [240]. Finally, quasiperiodic oscillations, relativistically broadened iron lines, continuum spectrum and X-ray polarization in the accretion disk surrounding a spinning black hole may also be used as a probe of the no-hair theorems ([241–250] and [251]).

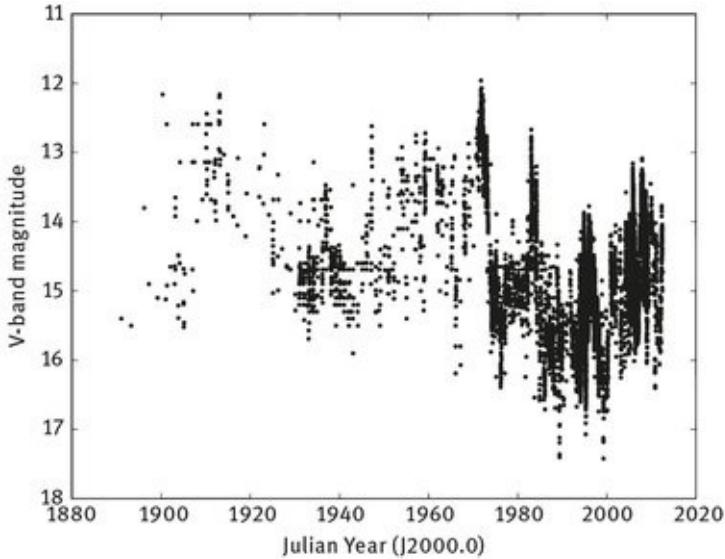
It is expected that some of these other tests should be possible during the next decade. Generally, there may be some difficulty due to the potential degeneracy between accretion physics close to the last stable orbit, which itself is not fully understood, and deviations from General Relativity [238]. In contrast, the test in OJ287 has already been done, and there are good prospects of improving the accuracy to  $\pm 10\%$  later in this decade. It does not depend on the accretion physics so close to the black hole.

# 10 OJ287 binary system

## 10.1 The binary model

In what follows, we describe current knowledge of the one current strong limit test of General Relativity, the OJ287 system, and the present constraints on the value of  $q$ . The identification of the OJ287 system as a likely binary was made already in 1980s, but since the mean period of the system is as long as 12 yr, it has taken a quarter of century to find convincing proof that we are indeed dealing with a binary system ([252, 253], and [254]). The primary evidence for a binary system comes from the optical light curve. By good fortune, the quasar OJ287 was photographed accidentally since 1890s, well before its discovery in 1968 as an extragalactic object. The pre-1968 observations are generally referred to as “historical” light curve points. The light curve of over one hundred years ([Figure 4](#)) shows a pair of outbursts at  $\sim 12$  yr intervals. The two brightness peaks in a pair are separated by 1–2 yrs.

The system is not strictly periodic, but there is a simple mathematical rule which gives all major outbursts of the optical light curve record. To define the rule, take a Keplerian orbit and demand that an outburst is produced at a constant phase angle and also at the opposite phase angle. Due to the nature of Keplerian orbits, this rule cannot be written in a closed mathematical form, but the outburst times are easily calculated from it. According to this rule two outburst peaks arise per period. By choosing an optimal value of eccentricity (which turns out to be  $e \sim 0.7$ ) and by allowing the semimajor axis of the orbital ellipse to precess in forward direction at an optimal rate (which turns out to be  $\Delta\phi \sim 39^\circ$  per period), the whole historical and modern outburst record of OJ287 is well reproduced.



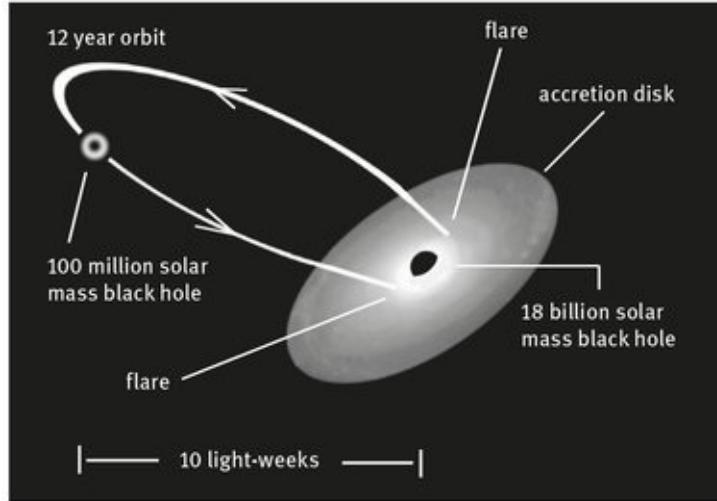
**Fig. 4.** Observations (+) of the brightness of OJ287 from late 1800s until today.  
Model predictions (–). <http://arxiv.org/pdf/1111.3484.pdf>.

The type of model that follows from this rule is immediately obvious. It consists of a small black hole in orbit around a massive black hole. The secondary impacts the accretion disk of the primary twice during each full orbit (Figure 5). The two impacts produce the two flares that are observed 1–2 years apart, and the flares are repeated in every orbital cycle. However, because of relativistic precession, the pattern of flares is not exactly periodic. It is this fact that allows the determination of the precession rate, and from there the mass of the primary in a straightforward way. Note that it is not necessary to know the inclination of the secondary black hole orbit nor the orientation of the system relative to the observer in order to carry out these calculations.

## 10.2 OJ287 flares and jet

The orbit that follows from the observed mathematical rule can be solved as soon as 5 flares are observed. Five flares have four intervals of time as input parameters, and the solution gives four parameters of the orbit uniquely: primary mass  $m_1$ , orbit eccentricity  $e$ , precession rate  $\Delta\phi$ , and the phase angle  $\varphi_0$  at a given moment of time. There does not need to be any solution

at all if the basic model is not correct. However, Lehto and Valtonen found a solution from five flares which already proved the case in the first approximation [255].



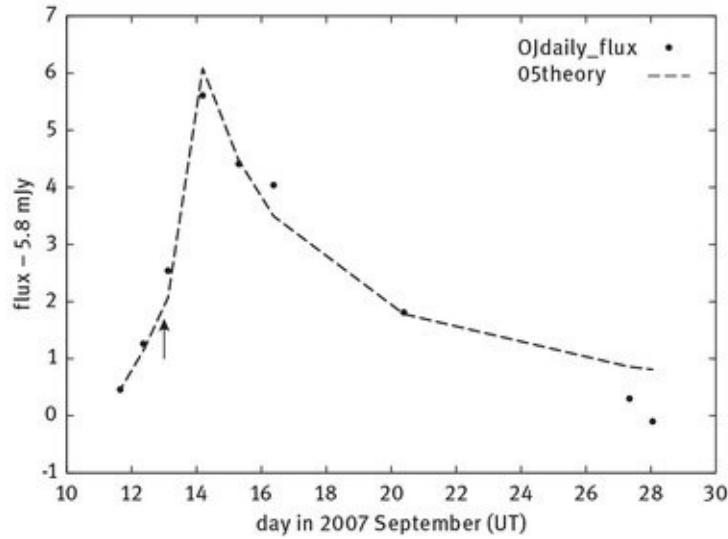
**Fig. 5.** An illustration of the OJ287 binary system. The jets are not shown, but they may be taken to lie along the rotation axis of the accretion disk. The two black holes are not resolved in current observations; the required resolution is  $\sim 10 \mu\text{arcsec}$ . However, the model explains *all* observations from radio to X-rays and the time variability of these data.

At a deeper level one has to consider the astrophysical processes that generate the flares. The flares start very suddenly, with the rise time of only about one day ([Figure 6](#)). This fact alone excludes many possibilities that one might be otherwise tempted to consider: Doppler boosting variation from a turning jet or increased accretion rate due to varying tidal force. The time scales associated with these processes are months to years quite independent of a detailed model. It turns out that the base level of emission of OJ287, which is synchrotron radiation from the jet, is affected by both of these mechanisms. The Doppler boosting variation accounts for the 55 year cycle [256] while the varying tides change the base level in a 11.8 year cycle [257]. Also the  $\sim 50$  day cyclic component is due to tides ([Figure 7](#)), but via a density wave at the innermost stable orbit of the primary accretion disk [237].

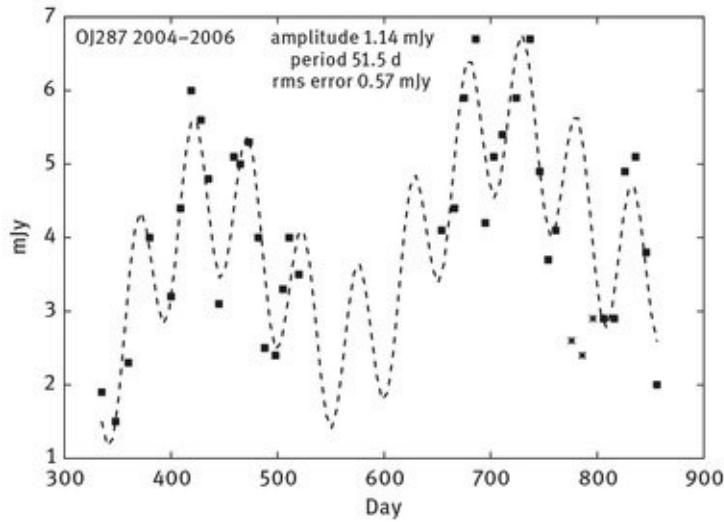
[Figure 8 illustrates](#) the light curve “hump” arising from increased accretion during the 2007 close approach of the two

black holes. This feature is well modeled by simulations, and its time scale is about one year. In the same timescale the impact flare is a sharp “spike” on top the hump.

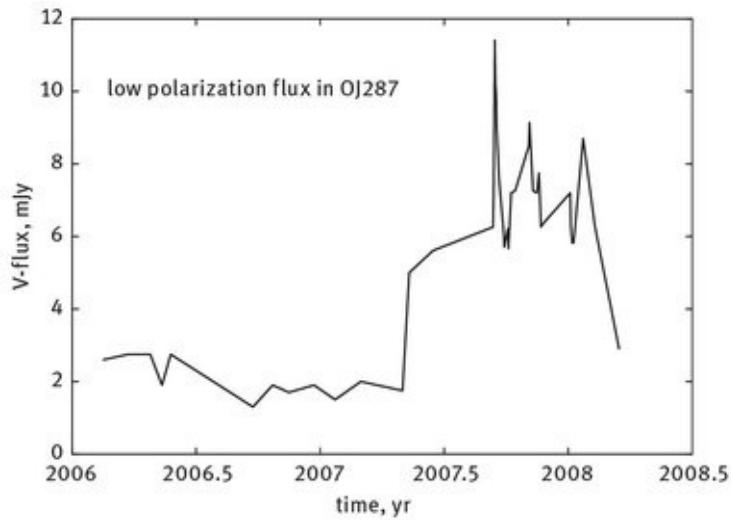
The correct astrophysical explanation of the flares is that the impact on the accretion disk releases hot bubbles of gas from the disk [258]. These bubbles expand until they become optically thin, and the radiation from the whole volume is seen. The calculated light curve profiles agree with this model ([Figure 6](#)). The radiation is thermal bremsstrahlung at the temperature of about  $3 \times 10^5$  K [259]. The addition of an un-polarised component to the emission lowers the degree of polarization [254] which is another piece of evidence pointing to bremsstrahlung radiation. In contrast, the radiation from the “hump” is clearly synchrotron radiation with a raised number density of radiating particles [261]. OJ287 is the only quasar which is known to have bremsstrahlung flares, and as such these flares give a unique set of signals to be used in orbit determination.



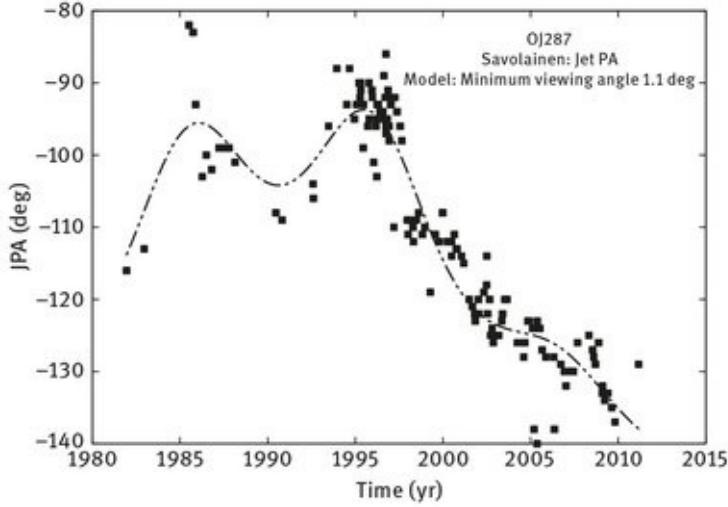
**Fig. 6.** The optical light curve of OJ287 during the 2007 September outburst. Only low polarization (less than 10%) data points are shown. The dashed line is the theoretical fit. The arrow points to September 13.0, the predicted time of origin of the rapid flux rise.



**Fig. 7.** The optical light curve of OJ287 during 2004–2006, using 10 day averages (squares). The 2005 impact flare is excluded. The three stars are points which are not used in the periodicity fit. The best fit for the remaining data is a 51.5 day period. If interpreted as the half-period of the innermost stable orbit, this implies  $\chi_1 = 0.25$ .



**Fig. 8.** The optical light curve of OJ287 during 2006–2008. Only low polarization (less than 10%) data points are shown.



**Fig. 9.** Observations of the jet position angle in OJ287 at cm wavelengths. The line represents a model where the orientation change propagates outward in the jet with speed  $0.85 c$ . In observer's time, it takes over 200 years to propagate the change from the optical core to the cm-wave jet. Due to the small viewing angle of  $\sim 2^\circ$  of the jet, the small changes in its direction are magnified.

What is more important, the model is able to predict future outbursts. The prediction for the latest outburst was 2007 September 13 ([261] and [252]), accurate to one day, leaving little doubt about the capability of the model (see [Figure 6](#); the predicted time refers in this case to the start of the rapid flux rise.) The 2007 September 13 outburst was an observational challenge, as the source was visible only for a short period of time in the morning sky just before the sunrise. Therefore a coordinated effort was made starting with observations in Japan, then moving to China, and finally to central and western Europe. The campaign was a success and finally proved the case for the binary model [254].

We see from [Figure 7](#) that the observed flux rise coincides within 6 hours with the expected time. The accuracy is about the same with which we were able to predict the return of Halley's comet in 1986!

The astrophysical model introduces a new unknown parameter, the thickness of the accretion disk. For a given accretion rate which is determined from the brightness of the quasar (considering the likely Doppler boosting factor), the thickness is a function of viscosity in the standard  $\alpha$  disk theory

of Shakura and Sunyaev [262] and in its extension to magnetic disks by Sakimoto and Corotini [263]. The value of the viscosity coefficient  $\alpha \sim 0.3$  is rather typical for other accreting systems (King *et al.* 2007) [264]. Different values of  $\alpha$  lead to different delay times between the disk impact and the optical flare. The delay can be calculated exactly except for a constant factor; this factor is an extra parameter in the model. The problem remains mathematically well defined. In fact, using only 6 outbursts as fixed points in the orbit, it is possible to solve the four orbital parameters plus the time delay parameter (in effect  $\alpha$ ) in the first approximation [261]. The success of this model in predicting the 2007 outburst was encouraging.

The future optical light curve of OJ287 was predicted from 1996 to 2030 [257] for every two-week interval between 1900 and 2030. During the first fifteen years OJ287 has followed the prediction with amazing accuracy, producing five outbursts at expected times, of expected light curve profile and size. It is extremely unlikely that such a coincidence should have happened by chance: these optical flux variations alone have excluded alternative models such as quasiperiodic oscillations in an accretion disk at the  $5\sigma$  confidence level [265].

OJ287 is unresolved in optical, but in radio and X-ray wavelengths we observe a long jet. The radio jet has been observed since early 1980s and its orientation shows interesting, but in no way simple 12 year cycles. In [Figure 9](#) we show the position angle (PA) observations of the jet as a function of time [266]. The line follows a model where the binary shakes the accretion disk gravitationally. This shaking is transmitted to the jet. With suitably chosen parameters the binary action explains not only the cm-wave observations of [Figure 9](#), but also quite different mm-wave jet wobble, in addition to the changes in the optical polarization angle [256]. All these phenomena occur in the jet, far from the primary sites of action, and thus they cannot be used for high precision determination of the orbit.

### **10.3 OJ287 orbit parameters (without using outburst**

**times)**

The orbital parameters may be determined from the basic binary model without further knowledge of the orbit solution. The precession rate may be estimated by taking the ratio of the two dominant variability frequencies in the optical light curve, averaged over one month to minimize the effects of the impact flares on the analysis. If one of the frequencies (11.8 yr) relates to the orbital period while the other frequency (55 yr) arises from precession [234], then their ratio will tell what fraction of the angle  $\pi$  the major axis of the orbit precesses per period. Due to the symmetry of the accretion disk relative to its midplane, the precessional effects should repeat themselves after rotation by  $180^\circ$ . Therefore the precession rate per orbital period should be  $11.8 \times 180/55 = 38.^{\circ}6$ . Once the precession rate is known, we get a good first estimate for the mass of the primary  $m_1 \sim 1.8 \cdot 10^{10} M_\odot$ .

Lehto and Valtonen determined the secondary mass [255],  $m_2 = 1.44 \cdot 10^8 M_\odot$ , when adjusted to the ‘current’ Hubble constant of 70 km/s/Mpc and to the 5.6 mJy outburst strength, as observed in 2007. The mass value is based on the astrophysics of the impact and the strength of the maximum signal [258] and [259]. A large mass ratio is necessary for the stability of the accretion disk. The primary black hole needs to be at least  $\sim 130$  times more massive than the secondary [259]. The values of  $m_1$  and  $m_2$  satisfy this requirement.

The moderately high eccentricity  $e \sim 0.7$  is expected at a certain stage of inspiral of binary black holes of large mass ratio ([267, 268], and [269]). Dynamical interaction with the stars of the galactic nucleus drives the eccentricity to  $e \sim 0.99$  before the gravitational radiation takes over in the evolution of the binary major axis. When the binary is at the evolutionary stage where OJ287 is now, the eccentricity has dropped to  $e \sim 0.7$ .

The value of  $\chi_1$  may be determined from the period of the innermost stable orbit. The best determination for this period is  $103 \pm 4$  days which corresponds to  $\chi_1 = 0.25 \pm 0.04$ .

The first two columns of [Table 4](#) summarize the parameters determined by astrophysics of the binary black hole system. No information on the outburst timing has been used in the astrophysical model.

[Table 4](#). Solution parameters.

Parameter	Astrophysics	Orbit
$\Delta\phi$	$38.^{\circ}6 \pm 1.^{\circ}0$	$39.^{\circ}1 \pm 0.^{\circ}1$
$m_1$	$(1.8 \pm 0.1) \cdot 10^{10} M_{\odot}$	$(1.84 \pm 0.01) \cdot 10^{10} M_{\odot}$
$m_2$	$(1.4 \pm 0.4) \cdot 10^8 M_{\odot}$	$(1.46 \pm 0.1) \cdot 10^8 M_{\odot}$
$\chi_1$	$0.25 \pm 0.04$	$0.28 \pm 0.03$
$\phi_0$	$56.^{\circ}0 \pm 4.^{\circ}0$	$56.^{\circ}3 \pm 1.^{\circ}0$
$e$	$0.7 \pm 0.03$	$0.70 \pm 0.001$
$\alpha$	$0.3 \pm 0.2$	$0.3 \pm 0.1$
$q$	-	$1.0 \pm 0.3$

## 11 Modeling binaries with Post Newtonian methods (with outburst times)

Post Newtonian methods and results are described in detail in the appended version of this section after the conclusion of this chapter. For readers who may wish to skip the details, here is a brief summary of this material. The OJ287 binary black hole is modeled to contain a spinning primary black hole with an accretion disk and a much lower mass secondary black hole whose spin is less important. Summarizing, the Post Newtonian approximation to General Relativity provides corrections to the Newtonian equations of motion of a compact binary in powers of  $(v/c)^2 \sim GM/(c^2 r)$ , where the variables are the characteristic

orbital velocity, the total mass, and typical orbital separation of the binary, respectively. The correction may be to different orders. The 2PN refers to corrections to Newtonian dynamics in powers of  $(v/c)^4$ . The 2.5 PN and 3 PN include general relativistic, classical spinorbit and radiation reaction effects for describing the evolution of a binary black hole. Differential equations describe the relative acceleration of the binary and the precessional motion for the spin of the primary black hole.

The leading order spinorbit interactions in a binary black hole force the orbital plane to precess. The orbital angular momentum vector, characterizing the orbital plane, precesses around the spin of the primary so that the angle between the orbital plane and the primary's spin vector is nearly constant. The spin-vector precesses tracing a narrow cone. Binary parameter solutions can be uniquely obtained from the first *nine* accurate times of outbursts as listed in [Table 4](#), column 3. Using the tenth outburst as a test gives for the spin  $\chi_1 \sim 0.23$ , only somewhat below that given in [Table 4](#).

## 12 OJ287 results at the strong field limit

As shown in [Table 4](#), using PN-accurate dynamics, we infer that the primary black hole should spin approximately at one quarter of the maximum spin rate allowed in general relativity. In addition, the ‘no-hair theorems’ of black holes [47] are supported by the model, although the testing is possible only to a limited precision. The solutions concentrate around  $q = 1.0$  with one standard deviation of 0.3 units. These results are achieved with the help of new data on historical outbursts as well as using the most recent outburst light curves together with the timing model for OJ287 outbursts [259]. Further, a polar orbit is assumed. Any other high inclination model would give us the same timing results. Sundelius and collaborators carried out simulations with a number of different inclinations between the disk and the orbit, and found that the inclination makes no difference [255]. This is not surprising since the impacts occur

along the line of nodes, and the timing of the impacts is not a function of the impact angle. We have tried varying the spin angle relative to the disk, and found that within reasonable limits (we considered inclinations up to  $10^\circ$ ) the results do not change. Again this is what is expected since the line of nodes circulates slowly in comparison with the orbital angular motion.

The reason why we would expect at least approximate alignment between the black hole spin and the disk spin is the Bardeen-Petterson effect which tends to align the two spins. However, the alignment time scale is much slower (about  $10^7$  yr, [274]) than the spin precession time, and thus the disk does not follow the black hole spin exactly but only on average.

In this scenario, we have a unique solution and also a unique prediction for the next OJ287 outburst, expected in 2016. We should then be able to judge the correctness of the present solution. Note that an outburst is not expected in 2016 in any simple extrapolation from past observations, as it is well before the average 12 yr cycle is due, and thus it is a sensitive test of the general model as well as a test for the spin of the primary black hole. The observations of the 2019 outburst will then improve the accuracy of the ‘no-hair’ theorem test by a factor of two, to the level of 15%.

There are at least 5 additional outbursts in the historical record which have not yet been accurately observed. If new data of these outbursts are found it will open up the possibility of improving the model and improving that accuracy of the no-hair theorem test to the level of 10%. Searches of plate archives are recommended for this purpose.

## 13 Conclusions

The weak field tests of General Relativity correspond to speeds much less than that of light or distances much larger than the Schwarzschild radius of the primary body of an orbiting pair. In the weak field, we briefly reviewed the precession of Mercury’s orbit, bending of light near the Sun, precession of a binary

pulsar orbit, gravitational radiation from the binary pulsar, relativistic geodetic precession and the precession due to the relativistic Lense-Thirring effect. In all these cases General Relativity accounts for the observations but, as expected for these distance scales, the cosmological constant (manifested as dark energy) is not significant.

However, the existence of dark energy in General Relativity can be tested in the weak field limit at large scales. Einstein proposed a cosmological constant to produce a static universe. Evolving models of the universe were derived by Friedmann and Lemaitre. The Einstein static solution is a special case, where the cosmological constant is related to the matter density so as to give gravity-antigravity balance. There are no theoretical restrictions on the cosmological constant. The universe expands or contracts as a whole. Hubble's discovery of the redshift distance relation can be regarded as a predictive verification of General Relativity.

After the discovery of the Big Bang expansion, it seemed that the cosmological constant could not be as big as required for a static universe. It was subsequently commonly assumed to be zero. Surprisingly, when supernovae of type Ia were used to estimate distances of galaxies whose redshifts  $z \sim 1$ , a positive cosmological constant was indicated.

We describe in detail how to test the existence of dark energy at the weak field limit of General Relativity via outflow in groups and clusters of galaxies. The Local Group was found to have outflowing dwarf galaxies around it whose motion indicates not only the gravitational mass of the group but also the effect of dark energy background. Studies of the Virgo and Fornax clusters and even rich clusters like Coma show similar structure.

We review the large amount of effort to determine the value of the cosmological constant, and its possible dependence on redshift or, equivalently, time. In the literature from year 2005 onwards, values cluster around  $\Omega_\Lambda = 0.73$  with a standard deviation of 0.044, the typical error in individual measurements. Only the cosmic microwave background (CMB) models give a significantly better accuracy, but one may suspect hidden

systematic errors due to foreground corrections. In our analysis of the literature including nearby groups and clusters, within the error limits,  $\Lambda$  appears to be a constant over the redshift range from cosmological to Local Group outflows, an impressive validation.

The other extreme, strong field tests, corresponds to speeds comparable to that of light or distances comparable to the Schwarzschild radius of the primary body of a pair. An important strong field test is to be sure that black holes are indeed the singularities predicted in General Relativity. We have to ascertain that at least in one case the spacetime around a suspected black hole satisfies the no-hair theorems which state that an electrically neutral rotating black hole in General Relativity is completely described by its mass,  $M$ , and its angular momentum,  $S$ , implying that the multipole moments, required to specify the external metric are fully expressible in terms of  $M$  and  $S$ . No-hair theorems apply only in General Relativity, and thus are a powerful eliminator of various alternative theories of gravitation.

An ideal test of the no-hair theorem is to have a test particle in orbit around a spinning black hole. We described results for one such system, the BL Lacertae object OJ287, a binary black hole of very large mass ratio. OJ287 gives outbursts during the secondary's periodic orbital disk crossings of an accretion disk around the more massive member. The timing of the repeating outbursts gives the orbit and the primary's quadrupole parameter. Post-Newtonian dynamics, necessary for such a system, implies that the primary black hole should spin at about one quarter of the maximum spin rate allowed in General Relativity. The ‘no-hair theorems’ of black holes are supported by the model but with a limited precision at present. Observations of predicted 2016 and 2019 outbursts and collection of data on more past outbursts should improve the accuracy of the ‘no-hair’ theorem test. This extreme test at the strong field limit supports General Relativity. Calculating the approximate orbit speed for OJ287, we obtain a significant fraction of the speed of light, 0.4 the speed of light at close

approach compared to less than 2/10 000 for the planet Mercury. We find an amazing range of validity of General Relativity in tests from the weak to strong field limits.

## **14 Appended section; mass, dark energy density and the “lost gravity” effect**

We may estimate conservative limits of local dark energy density  $\rho_{loc}$  as follows. If the value of  $R_V$  were known from the velocity-distance diagram and the mass  $M$  of the group is independently measured, the dark energy density may be estimated in the outflow region:

$$\frac{\rho_{loc}}{\rho_V} = \left( \frac{M}{1.3 \times 10^{12} M_\odot} \right) \left( \frac{1.3 \text{ Mpc}}{R_V} \right)^3. \quad (31)$$

In fact, attempts to probe dark energy with nearby outflows were first made (e.g. [55, 69], and [62]) by using the Equation above. Also here the derived dark energy density depends directly on the assumed mass in addition to the strong inverse dependence on the used zero gravity distance.

The size of the group is a strict lower limit to  $R_V$ , giving an upper limit to the local dark energy density  $\rho_{loc}$ , for a fixed group mass. Hence, for the Local Group  $R_V > 1 \text{ Mpc}$  leads to  $\rho_{loc} < 2.2 \rho_V$ , for the mass  $M = 2 \times 10^{12} M_\odot$ .

An upper limit for  $R_V$  would give the interesting lower limit to  $\rho_{loc}$ . One way to study  $R_V$  would be to find the distance  $R_{ES}$ , the Einstein-Straus radius where the local outflow reaches the global Hubble rate, see [Equation \(22\)](#) [70].

If there is no local dark energy, the outflow reaches the global expansion rate only asymptotically in this idealized point mass model. However, assuming that  $\rho_{loc} = \rho_V$ , one may expect the local flow to intersect the global Hubble relation at a distance  $R_{ES} = 1.7 R_V$ . For instance, with the Local Group mass of (2 – 4)

$\times 10^{12} M_{\odot}$  and using the global dark energy density, one calculates  $R_{ES} = 1.7R_V = 2.2 - 2.6$  Mpc. This range is indeed near the distance where the local expansion reaches the global rate ([Figure 3](#); see also [55]).

Starting with the illustrative value  $R_{ES} \approx 2.6$  Mpc, one may estimate that  $R_V \approx 1.5$  Mpc [67] and a lower limit for the dark energy density around the Local Group would be  $\rho_{loc}/\rho_V \geq 0.4$  for the mass of  $2 \times 10^{12} M_{\odot}$ . The limit is directly proportional to the adopted mass value.

In view of the uncertainties in using the outflow kinematics only, one should use it in concert with other independent methods for putting limits on the value of the local dark energy density. Such include the Kahn-Woltjer [68] method and the virial theorem, both of which can be modified to take into account the “lost gravity” effect of dark energy [58].

In 1959 Kahn and Woltjer used a simple one-dimensional two body problem to describe the relative motion of the Milky Way and M31 galaxies [71]. The motion of the galaxies was described as a bound system. The (currently) observed values  $r = 0.7$  Mpc and  $V = -120$  km/s lead to a limit for the estimated binary mass:  $M > 1 \times 10^{12} M_{\odot}$ . The mass is  $M \approx 4.5 \times 10^{12} M_{\odot}$ , if the maximum separation was about 4.4 Gyr ago, corresponding to 13.2 Gyr as the time since the two galaxies started to separate from each other.

With a minimal modification of the original method, the first integral of the equation of motion ([equation \(19\)](#)), in the presence of dark energy background, becomes:

$$\frac{1}{2}V^2 = \frac{GM}{r} + \left(\frac{G4\pi}{3}\right)\rho_V r^2 + E. \quad (32)$$

Now the total energy for a bound system embedded in the dark energy background must be smaller than an upper limit which depends on the mass and the dark energy density:

$$E < -\frac{3}{2} GM^{3/2} \left[ \left( \frac{8\pi}{3} \right) \rho_V \right]^{1/3}. \quad (33)$$

The limiting value corresponds to the case where the distance between the component galaxies could just reach the zero-gravity distance. Now the lower mass limit increases to  $M > 3.2 \times 10^{12} M_\odot$ . Also the timing argument leads to an increased mass,  $M \approx 5.3 \times 10^{12} M_\odot$  ([58] and [72]). On the other hand, dynamical activity during the formation and settling down of the Local Group tends to reduce “timing” mass by a large factor [66]. Though the uncertainty due to the various imperfectly known factors is rather large, the method suggests

$$M \sim 4 \times 10^{12} M_\odot \quad (34)$$

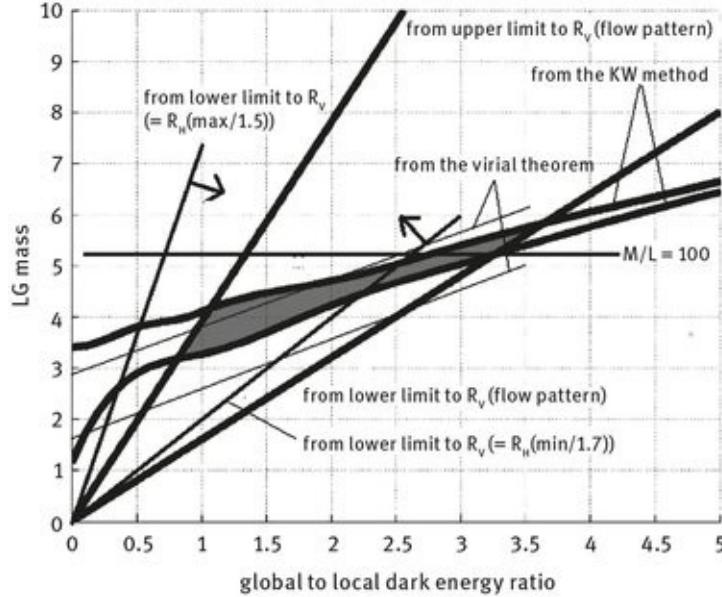
for the Local Group. A local volume cosmological simulation with this mass value agrees nicely with observations [73].

The classical virial theorem is a well-known way to determine the mass of a quasi-stationary gravitationally bound many-particle system. For a system within dark energy, the virial theorem needs an extra term due to the contribution of the particle-dark energy interaction to the total potential energy ([74] and [75]). When positive, the cosmological constant leads to a correction upwards for the mass estimates [58] and one can show that the modified virial theorem should be written, in terms of the total mass  $M$ , a characteristic velocity  $V$  and a characteristic sizes  $R$ , as:

$$M = V^2 \frac{R}{G} + \frac{8\pi}{3} \rho_V R^3, \quad (35)$$

where the second term on the right hand side is the correction due to the dark energy. It is interestingly equal to the value of the effective (anti)gravitating mass of dark energy within the sphere of radius  $R$ . It is a measure of the lost-gravity effect,

which can be significant in galaxy groups. In the Local Group the correction term contributes about 30 percent of the total mass.



**Fig. 10.** Results on the very local density of dark energy from the Local Group and its near environment (adapted from [56]).

Chernin and collaborators [58] used the modified Kahn-Woltjer method together with the outflow data to derive the mass of the Local Group and the local dark energy density, resulting in  $M = (3-6) \times 10^{12} M_{\odot}$  and  $\rho_{loc}/\rho_V = 0.8-3.8$ . The virial estimator which uses the mass and the velocity dispersion within the Local Group [68] gave  $\rho_{loc}/\rho_V = 0.7-2.8$ .

We present various Local Group results in [Figure 10 where](#) the horizontal axis gives the assumed local-to-global dark energy density and the vertical axis is the mass of the Local Group. A simpler version of this diagram appeared in [58], where the admissible range from the modified Kahn-Woltjer-method was shown, together with the limiting straight lines corresponding to the upper and lower limit of  $R_V$  as estimated from the appearance of the outflow pattern. These defined the darkened area as the possible range of the mass and the local dark energy density. The present [Figure 10 includes](#) additional constraints:

1. The range of the virial mass based on van den Bergh's [56] result  $M_{LG} = 2.3 \pm 0.6 M_\odot$  and using the correction term in the appendix to [58].
2. The location of the mass-to-luminosity ratio  $M/L = 100$  (an upper limit for small groups) as the horizontal line, based on the result that  $M_{LG} = 2.3 \times 10^{12} M_\odot$  [54] corresponds to  $M/L = 44$ , and
3. The lines corresponding to upper and lower limits to  $R_V$  as inferred from the distance range 2.2–3 Mpc where the local flow reaches the global Hubble rate and using the calculations in [70]. These limits may be less subjective than the original thick lines which were based on visual inspection of the flow pattern.

We see that the different constraints from the Local Group put the local dark energy density into the range  $\rho_{loc}/\rho_V = 0.5\text{--}2.5$ .

The preceding discussion did not make any use of the prior knowledge of the dwarf galaxy outflow. Dwarf galaxies are thrown out of the Local Group during its early assembly and later evolution. In cosmological N-body simulations where this process is seen, it is sometimes called a “back splash”. The “back splash” follows the normal rules of dynamical ejection in a potential well [76]. The dwarf ejection times have a wide distribution, with a higher rate in the beginning. However, since the Local Group is dynamically young, as the two major galaxies have not yet completed even one full orbit, the dwarf ejection rate has not yet declined significantly.

The distribution of the ejection speeds  $P(V)$  is a steeply declining function of  $V$ . The probability that the escape velocity is in the interval  $V, V + dV$  is

$$PdV \sim V^{-4.5} dV$$

(36)

(from [76], Equation (11.33)). What it means that in practice all dwarfs cross the zero gravity surface with a speed which is very close to zero ( $\sim 40$  km/s). It has two consequences: the flow

speed of dwarfs beyond the surface is very smooth, and the time spent by the dwarfs at the zero gravity sphere is long and consequently we find an accumulation of the flow at this boundary. This makes the identification of the zero gravity radius simple. For the Local group

$$R_V = 1.6 \pm 0.1 \text{ Mpc} . \quad (37)$$

The two main galaxies lose orbital energy from their relative motion every time a galaxy is ejected from the group. Therefore the classic Local Group timing argument [71], when applied to the isolated pair, gives necessarily an overestimate of their combined mass. The ejections shorten the major axis of the relative orbit, and thus less mass is required to close the orbit. The importance of this effect depends on the total mass of the ejected galaxies. In the scenarios calculated in [66] the effect is 5% in the universe of 14 billion yr in age. A greater reduction in the “timing mass”, up to 25%, comes from the possibility that the rotation speed of the Galaxy is greater than the standard 220 km/s. On the other hand, the lost gravity effect would tend to increase the “timing mass” by about 30%.

Therefore, all in all, the full N-body simulation model of Garrison-Kimmel *et al.* [73], with Local Group mass of  $M = 4 \times 10^{12}$  solar mass, appears to be an acceptable model of the Local Group and the surrounding flows. With these values we get  $\Omega_\Lambda = 0.75$ , the value given on the first line of [Table 1](#) (Section 8). The one standard deviation uncertainty may be estimated as  $\pm 10\%$ .

Another way to obtain the local  $\Omega_\Lambda$ , is to take the measured values of  $H_V$  and  $H_0$  from observations, and calculate

$$\Omega_\Lambda = \left( \frac{H_V}{H_0} \right)^2 . \quad (38)$$

Even though both quantities on the right hand side have

associated uncertainties, we get using  $H_V = 59$  km/s/Mpc and  $H_0 = 70$  km/s/Mpc, an estimate  $\Omega_\Lambda = 0.71$  [77], again with the estimated one standard deviation uncertainty of  $\pm 10\%$ . This value is given on the second line of [Table 1](#).

## 15 Appended section: dark energy in the Coma cluster of galaxies

In this Section, we extend our studies of the local dark energy effects from groups of galaxies to clusters of galaxies and address the Coma cluster considering it as the largest regular, nearly spherically-symmetrical, quasi-stationary gravitationally bound aggregation of dark matter and baryons embedded in the uniform background of dark energy. Is antigravity produced by dark energy significant in the volume of the cluster? Does it affect the structure of the cluster? Can antigravity put limits on the major gross parameters of the system? In a search for answers to these questions, we will use and develop the general considerations on the local dark energy given in the sections above.

### 15.1 Three mass estimates of the cluster

The mass of the Coma cluster was first measured by Zwicky in the 1930s ([78] and [79]). Using the virial theorem he found that it was  $3 \times 10^{14} M_\odot$  when normalized to the presently adopted value of the Hubble constant  $H_0 = 70$  km/s/Mpc which is used here. Later The & White [80] found an order of magnitude larger value,  $2 \times 10^{15} M_\odot$ , with a modified version of the virial theorem. Hughes ([81] and [82]) obtained a similar value  $(1-2) \times 10^{15} M_\odot$  with X-ray data under the assumption that the hot intergalactic gas in the cluster is in hydrostatic equilibrium. With a similar assumption,  $4.4 \times 10^{14} M_\odot$  inside the radius of 1.4 Mpc is obtained [83]. A weak-lensing analysis gave

the mass of  $2.6 \times 10^{15} M_{\odot}$  [84] within 4.8 Mpc radius.

Valtonen and Byrd proposed that the action of the central binary in the Coma cluster has the effect of ejecting galaxies from the cluster. It leads to an overestimate of the cluster mass if the virial theorem is used ([85, 86], and [87]).

An examination of the outskirts of the cluster with the use of the caustic technique ([88] and [89]) gave the mass  $2.4 \times 10^{15} M_{\odot}$  within the 14 Mpc radius. Taken at face value, it appears that the mass within 14 Mpc is smaller than the mass within 4.8 Mpc. Most probably, this is due to uncertainties in mass determination. Indeed, the  $2\sigma$  error is  $1.2 \times 10^{15} M_{\odot}$  in Geller's *et al.* ([90] and [91]) data, and within this uncertainty, the result does not contradict the small-radius data.

In each of the works mentioned here, the measured mass is treated as the matter (dark matter and baryons) mass of the cluster at various clustercentric distances. However, the presence of dark energy in the volume of the cluster modifies this treatment, since dark energy makes its specific contribution to the mass of the system. This contribution is naturally measured by effective gravitating mass of dark energy in the volume of the cluster at various clustercentric distances  $R$  ([Equation \(35\)](#)):

$$M_V(R) = \frac{4\pi}{3} \rho_{V\text{eff}} R^3 = -\frac{8\pi}{3} \rho_V R^3 = -0.85 \times 10^{12} \left[ \frac{R}{1 \text{ Mpc}} \right]^3 M_{\odot}. \quad (39)$$

For the largest radius  $R = 14$  Mpc we have:

$$M_V = -2.3 \times 10^{15} M_{\odot}, \quad R = 14 \text{ Mpc}. \quad (40)$$

The total gravitating mass within the radius  $R$  is the sum

$$M_G(R) = M_M(R) + M_V(R), \quad (41)$$

where  $M_M(R)$  is the matter (dark matter and baryons) mass of the cluster inside the same radius  $R$ . It is this mass  $M_G(R)$  that is only available for astronomical measurements via gravity (with virial, lensing, caustic, etc. methods). Because of this, we identify the gravitating mass  $M_G(R)$  with the observational masses quoted above for various clustercentric radii. In particular, the gravitating mass for the largest radius  $R = 14$  Mpc in the Coma cluster is this:

$$M_G(R) = M_V(R) + M_M(R) = 2.4 \times 10^{15} M_\odot . \quad (42)$$

Then the matter mass  $M_M$  at the same radius

$$M_M(R) = M_G(R) - M_V(R) = 4.7 \times 10^{15} M_\odot . \quad (43)$$

We see that the value of the matter mass  $M_M$  at  $R = 14$  Mpc obtained with the presence of dark energy is a factor of (almost) two larger than that in the traditional treatment. This implies that the antigravity effects of dark energy are strong at large radii of the Coma cluster.

## 15.2 Matter mass profile

In the spherically symmetric approximation, the cluster matter mass  $M_M(R)$  may be given in the form:

$$M_M(R) = 4\pi \int \rho(R) R^2 dR , \quad (44)$$

where  $\rho(R)$  is the matter density at the radius  $R$ . According to the widely used NFW density profile [92]

$$\rho = \frac{4\rho_s}{\frac{R}{R_s} \left(1 + \frac{R}{R_s}\right)^2} ,$$

where  $\rho_s = \rho(R_s)$  and  $R_s$  are constant parameters. At small radii,  $R \ll R_s$ , the matter density goes to infinity,  $\rho \propto 1/R$  as  $R$  goes to zero. At large distances,  $R \gg R_s$ , the density slope is  $\rho \propto 1/R^3$ . With this profile, the matter mass profile is

$$M_M(R) = 16\pi\rho_s R_s^3 \left[ \ln(1 + R/R_s) - \frac{R/R_s}{1 + R/R_s} \right]. \quad (46)$$

To find the parameters  $\rho_s$  and  $R_s$ , we may use the small-radii data (as quoted above):  $M_1 = 4.4 \times 10^{14} M_\odot$  at  $R_1 = 1.4$  Mpc,  $M_2 = 2.6 \times 10^{15} M_\odot$  at  $R_2 = 4.8$  Mpc. At these radii, the gravitating masses are practically equal to the matter masses there. The values of  $M_1$ ,  $R_1$  and  $M_2$ ,  $R_2$  lead to two logarithmic equations for the two parameters of the profile, which can easily be solved:  $R_s = 4.7$  Mpc,  $\rho_s = 1.8 \times 10^{-28}$  g/cm<sup>3</sup>. Then we find the matter mass within  $R = 14$  Mpc,

$$M_M \approx 8.7 \times 10^{15} M_\odot, \quad (47)$$

to be considerably larger (over 70%) than given by our estimation above.

The Hernquist density profile [93] is also popular,

$$\rho(R) \propto \frac{1}{R(R + \alpha)^3}. \quad (48)$$

Its small-radius behavior is the same as in the NFW profile:  $\rho \rightarrow \infty$ , as  $R$  goes to zero. The slope at large radii is different:  $\rho \propto 1/R^4$ . The corresponding mass profile is

$$M_M(R) = M_0 \left[ \frac{R}{R + \alpha} \right]^2. \quad (49)$$

The parameters  $M_0$  and  $\alpha$  can be found from the same data as above on  $M_1$ ,  $R_1$  and  $M_2$ ,  $R_2$ :  $M_0 = 1.4 \times 10^{16} M_\odot$ ,  $\alpha = 6.4$  Mpc, giving another value for the mass within 14 Mpc:

$$M_M = 6.6 \times 10^{15} M_\odot, \quad R = 14 \text{ Mpc}. \quad (50)$$

Now the difference from our estimated figure is about 40%.

In a search for a most suitable mass profile for the Coma cluster, we may try the following simple new relation:

$$M_M(R) = M_* \left[ \frac{R}{R + R_*} \right]^3. \quad (51)$$

This mass profile comes from the density profile:

$$\rho(R) = \frac{3}{4\pi} M_* R_* (R + R_*)^{-4}. \quad (52)$$

The density goes to a constant as  $R$  goes to zero; at large radii,  $\rho \propto 1/R^4$ , as in Hernquist's [91] profile.

The parameters  $M_*$  and  $R_*$  are found again from the data for the radii of 1.4 and 4.8 Mpc:  $M_* = 8.7 \times 10^{15} M_\odot$ ,  $R_* = 2.4$  Mpc. The new profile leads to a lower matter mass at 14 Mpc:

$$M_M = 5.4 \times 10^{15} M_\odot, \quad R = 14 \text{ Mpc}, \quad (53)$$

which is equal to our estimate above within 15% accuracy.

### 15.3 Upper limits and beyond

It is obvious that a system of galaxies can be gravitationally bound only if gravity dominates over antigravity in its volume. In terms of the characteristic masses introduced above, this condition may be given in the form:

$$M_M(R) \geq |M_V(R)| .$$

(54)

The condition is naturally met in the interior of the system. But the dark energy mass  $|M_V(R)|$  increases with the radius as  $R^3$ , while the matter mass increases slower in all the three versions of the matter mass profile  $M_M(R)$  discussed above. As a result, an absolute upper limit arises from this condition to the total size and the total matter mass of the cluster. For the new matter mass profile introduced above one has:

$$R_{\max} = 20 \text{ Mpc} , \quad M_{\max} = 6.2 \times 10^{15} M_\odot .$$

(55)

These upper limits are consistent with the theory of large-scale structure formation that claims the range  $2 \times 10^{15} < M < 10^{16} M_\odot$  for the most massive bound objects in the Universe ([94] and [95]).

For comparison, the two traditional matter profiles mentioned above lead to somewhat larger values of the size and mass:

$$R_{\max} = 25 \text{ Mpc} , \quad M_{\max} = 1.5 \times 10^{16} M_\odot \quad (\text{NFW}) ,$$

(56)

$$R_{\max} = 22 \text{ Mpc} , \quad M_{\max} = 9.1 \times 10^{15} M_\odot \quad (\text{Hernquist}) .$$

(57)

Generally, the limit condition  $M_G(R_{\max}) = |M_V(R_{\max})|$  leads to the relation between the upper mass limit and the upper size limit:

$$R_{\max} = \left( \frac{3M_{\max}}{8\pi\rho_V} \right)^{1/3} .$$

(58)

The relation shows that the upper size limit is identical to the

zero-gravity radius introduced in Section 1:  $R_{\max} = R_V$ .

Studies of nearby systems like the Local Group and the Virgo and Fornax clusters ([69, 77, 96–103]) show that their sizes are indeed near the corresponding zero-gravity radii, so that each of the systems occupies practically all the volume of its gravity-domination region ( $R \leq R_V$ ). These examples suggest that the Coma cluster may have the maximal possible size and its total matter mass may be near the maximal possible value. If this is the case, the mean matter density in the system is expressed in terms of the dark energy density and does not depend on the density profile ([10]; [105] and [106]):

$$\langle \rho_M \rangle = \frac{M_M}{\frac{4\pi}{3} R_V^3} = 2\rho_V . \quad (59)$$

Another theoretical prediction may be made about the environment of the Coma cluster at the dark energy domination area  $R > R_V$ . Earlier studies have shown that outflows of galaxies exist around the Local group and some other groups and clusters of galaxies. We may assume that such an outflow may be observed around the Coma cluster as well at the distance  $R > R_V \approx 20$  Mpc from its center. The outflow is expected to have quasi-regular kinematical structure with the nearly linear velocity-distance relation. The galaxies in the outflow have been ejected by dynamical interactions in the cluster center ([107] and [3]).

These theoretical predictions may be tested in current and future astronomical observations of the Coma cluster and its environment. The next generation of ground based and space telescopes should be able to resolve individual stars in the galaxies of the Coma cluster, and allow us to construct a three-dimensional map of the cluster together with the radial velocities of each galaxy. We may then be able to separate the “warm” inflow of galaxies in the outskirts of the cluster from the “cool” outflow, and to determine the position of the zero-gravity radius experimentally.

## 16 Appended section: modeling binaries with Post Newtonian methods

The Post Newtonian approximation to General Relativity provides the equations of motion of a compact binary as corrections to the Newtonian equations of motion in powers of  $(v/c)^2 \sim GM/(c^2 r)$ , where  $v$ ,  $M$ , and  $r$  are the characteristic orbital velocity, the total mass, and the typical orbital separation of the binary, respectively. The approximation may be extended to different orders. The terminology 2PN, for example, refers to corrections to Newtonian dynamics in powers of  $(v/c)^4$ . Valtonen and collaborators use the 2.5 PN [270] and 3PN-accurate [238] orbital dynamics that includes the leading order general relativistic, classical spinorbit and radiation reaction effects for describing the evolution of a binary black hole [271]. The following differential equations describe the relative acceleration of the binary and the precessional motion for the spin of the primary black hole at 2.5PN order:

$$\ddot{\mathbf{x}} \equiv \frac{d^2 \mathbf{x}}{dt^2} = \ddot{\mathbf{x}}_0 + \ddot{\mathbf{x}}_{1PN} + \ddot{\mathbf{x}}_{SO} + \ddot{\mathbf{x}}_Q + \ddot{\mathbf{x}}_{2PN} + \ddot{\mathbf{x}}_{2.5PN}, \quad (60)$$

$$\frac{d\mathbf{s}_1}{dt} = \Omega_{SO} \times \mathbf{s}_1, \quad (61)$$

where  $\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2$  stands for the center-of-mass relative separation vector between the black holes with masses  $m_1$  and  $m_2$  and  $\ddot{\mathbf{x}}_0$  represents the Newtonian acceleration given by

$$\ddot{\mathbf{x}}_0 = -\frac{G m}{r^3} \mathbf{x}, \quad (62)$$

$m = m_1 + m_2$  and  $r = |\mathbf{x}|$ . Kerr parameter  $\chi_1$  and the unit vector  $\mathbf{s}_1$  define the spin of the primary black hole by the relation  $\mathbf{s}_1 = G m_1^2 \chi_1 \mathbf{s}_1 / c$  and  $\chi_1$  is allowed to take values between 0 and 1 in

general relativity.  $\Omega_{SO}$  provides the spin precessional frequency due to spinorbit coupling. The PN contributions occurring at the conservative 1PN, 2PN and the reactive 2.5PN orders, denoted by  $\ddot{\mathbf{x}}_{1PN}$ ,  $\ddot{\mathbf{x}}_{2PN}$  and  $\ddot{\mathbf{x}}_{2.5PN}$ , respectively, are non-spin by nature. The explicit expressions for these contributions, suitable for describing the binary black hole dynamics [272], are

$$\begin{aligned}\ddot{\mathbf{x}}_{1PN} &= -\frac{Gm}{c^2 r^2} \left\{ \hat{\mathbf{n}} \left[ -2(2+\eta) \frac{Gm}{r} + (1+3\eta)v^2 - \frac{3}{2}\eta \dot{r}^2 \right] - 2(2-\eta)\dot{r}\mathbf{v} \right\}, \\ \ddot{\mathbf{x}}_{2PN} &= -\frac{Gm}{c^4 r^2} \left\{ \hat{\mathbf{n}} \left[ \frac{3}{4}(12+29\eta) \left( \frac{Gm}{r} \right)^2 + \eta(3-4\eta)v^4 + \frac{15}{8}\eta(1-3\eta)\dot{r}^4 \right. \right. \\ &\quad \left. \left. - \frac{3}{2}\eta(3-4\eta)v^2 \dot{r}^2 - \frac{1}{2}\eta(13-4\eta) \left( \frac{Gm}{r} \right) v^2 \right. \right. \\ &\quad \left. \left. -(2+25\eta+2\eta^2) \left( \frac{Gm}{r} \right) \dot{r}^2 \right] \right\},\end{aligned}\tag{63}$$

$$-\frac{1}{2}\dot{r}\mathbf{v} \left[ \eta(15+4\eta)v^2 - (4+41\eta+8\eta^2) \left( \frac{Gm}{r} \right) - 3\eta(3+2\eta)\dot{r}^2 \right] \right\},\tag{64}$$

$$\ddot{\mathbf{x}}_{2.5PN} = \frac{8}{15} \frac{G^2 m^2 \eta}{c^5 r^3} \left\{ \left[ 9v^2 + 17 \frac{Gm}{r} \right] \dot{r}\hat{\mathbf{n}} - \left[ 3v^2 + 9 \frac{Gm}{r} \right] \mathbf{v} \right\},\tag{65}$$

where the vectors  $\hat{\mathbf{n}}$  and  $\mathbf{v}$  are defined to be  $\hat{\mathbf{n}} \equiv \mathbf{x}/r$  and  $\mathbf{v} \equiv d\mathbf{x}/dt$ , respectively, while  $\dot{r} \equiv dr/dt = \hat{\mathbf{n}} \cdot \mathbf{v}$ ,  $v \equiv |\mathbf{v}|$  and the symmetric mass ratio  $\eta = m_1 m_2 / m^2$ .

Going to the next level, Barker and O'Connell calculated the leading order spinorbit contributions to  $\mathbf{x}$ , appearing at 1.5PN order [27],

$$\begin{aligned}\ddot{\mathbf{x}}_{SO} &= \frac{Gm}{r^2} \left( \frac{Gm}{c^3 r} \right) \left( \frac{1 + \sqrt{1 - 4\eta}}{4} \right) \\ &\quad \chi_1 \left\{ [12 [\mathbf{s}_1 \cdot (\hat{\mathbf{n}} \times \mathbf{v})]] \hat{\mathbf{n}} + \left[ \left( 9 + 3\sqrt{1 - 4\eta} \right) \dot{r} \right] (\hat{\mathbf{n}} \times \mathbf{s}_1) - \left[ 7 + \sqrt{1 - 4\eta} \right] (\mathbf{v} \times \mathbf{s}_1) \right\},\end{aligned}\tag{66}$$

while

$$\Omega_{SO} = \left( \frac{G m \eta}{2c^2 r^2} \right) \left( \frac{7 + \sqrt{1 - 4\eta}}{1 + \sqrt{1 - 4\eta}} \right) (\hat{\mathbf{n}} \times \hat{\mathbf{v}})$$

(67)

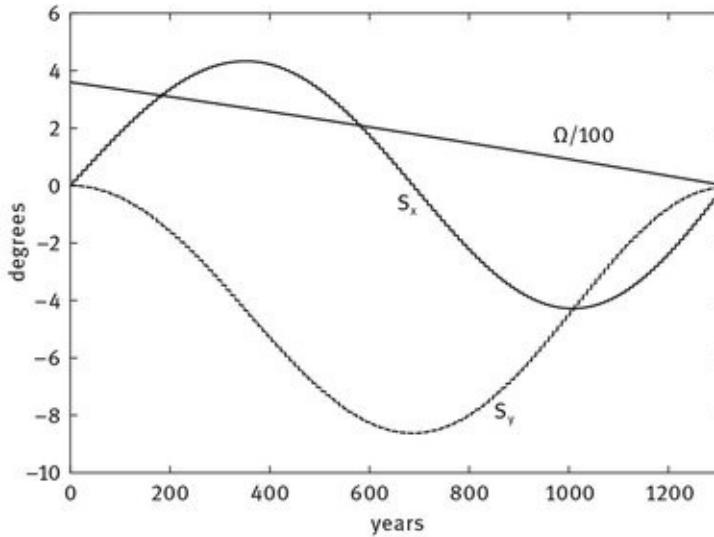
Finally, Barker and O'Connell calculated the quadrupole-monopole interaction term  $\ddot{\mathbf{x}}_Q$ , entering at the 2PN order [27],

$$\ddot{\mathbf{x}}_Q = -q \chi_1^2 \frac{3 G^3 m_1^2 m}{2 c^4 r^4} \left\{ [5(\hat{\mathbf{n}} \cdot \mathbf{s}_1)^2 - 1] \hat{\mathbf{n}} - 2(\hat{\mathbf{n}} \cdot \mathbf{s}_1) \mathbf{s}_1 \right\},$$

(68)

where the parameter  $q$ , whose value is 1 in general relativity, is introduced to test the black hole 'no-hair' theorems [232].

It turns out that adding the 3PN contributions to  $d^2\mathbf{x}/dt^2$ , are not necessary; their influence falls within error limits of the OJ287 problem. It has been important to verify that the PN series converges rapidly enough to ignore the terms of order higher than 2.5PN. Similarly, the terms related to the spin of the secondary turn out to be negligible in the OJ287 problem [238]. The precessional motion for the spin of the primary black hole is 2PN accurate in our calculations.



**Fig. 11.** The circulation of the ascending node ( $\Omega$ ) of the orbit (straight line) in the coordinate system in which the initial spin vector  $\mathbf{s}_1$  points to the direction of the z-axis. The quantities plotted are  $\Omega/100$  in degrees and the symbols  $s_x$  and  $s_y$  correspond to the x and y components of the spin-direction vector  $\mathbf{s}_1$  in degrees.

The main consequence of including the leading order spin-orbit interactions to the dynamics of a binary black hole is that it forces the orbital plane to precess. The orbital angular momentum vector, characterising the orbital plane, precesses around the spin of the primary in such a way that the angle between the orbital plane and the spin vector  $\mathbf{s}_1$  remains almost constant (roughly within  $\pm 0.5^\circ$  in the model). The spin-vector itself precesses drawing a cone with an opening angle of about 8 degrees (see [Figure 11](#)).

The precessional period for both the orbital plane and the spin of the binary, provided by  $|\Omega_{SO}|$ , is about 1300 years for the orbital parameters of OJ287. The precession cone axis coincides with the mean accretion disk axis. It is reasonable to expect such a situation due to the Bardeen-Petterson effect [273]. Because the time scale of the Bardeen-Petterson effect is much longer than the black hole spin precession time scale [274], the disk axis is not able to follow the rapid precession of the primary spin axis.

The orbit solutions currently involve *nine* accurately timed outbursts as listed in [Table 5](#). The orbit solutions provide a unique set of parameters ([Table 4](#), column 3). Recently also the tenth outburst from 1906 has been added [275]. In this case there are only limits for the timing; the main influence of including the 1906 data is to bring  $\chi_1$  to  $\sim 0.23$ , somewhat below the range given in [Table 4](#). Note that the eccentricity  $e$  is defined as

$$\frac{r_a}{r_p} = \frac{1+e}{1-e},$$

(69)

where  $r_a$  and  $r_p$  are the apocenter and pericenter distances in the orbit, respectively. The instantaneous eccentricity varies during a relativistic binary orbit; in this way we get a definite number that resembles the eccentricity of a Keplerian orbit.

**Table 5.** Outburst times with estimated uncertainties. These are starting times of

the outbursts.

<b>Time</b>	<b>uncertainty</b>
1912.970	$\pm 0.010$
1947.282	$\pm 0.0005$
1957.080	$\pm 0.020$
1972.94	$\pm 0.005$
1982.964	$\pm 0.0005$
1984.130	$\pm 0.002$
1995.843	$\pm 0.0005$
2005.745	$\pm 0.005$
2007.692	$\pm 0.0005$

Because we are using 8 time intervals to fit 8 parameters, the fit has no degrees of freedom. The tolerance limits of [Table 4](#) are a consequence of having a tolerance in the fixed points of the outburst times ([Table 5](#)). The no-hair parameter  $q$  is the eighth parameter which of course should be  $q = 1$  if General Relativity is correct.

As part of the solution we obtain the list of all past and future outburst times with their uncertainties. For example, the well recognized outbursts in 1959, 1971 and 1994 are timed at  $1959.213 \pm 0.002$ ,  $1971.1265 \pm 0.002$  and  $1994.6085 \pm 0.005$ , respectively. In all these cases data are missing at the crucial time of expected rapid flux rise, and thus these predictions cannot be verified at present.

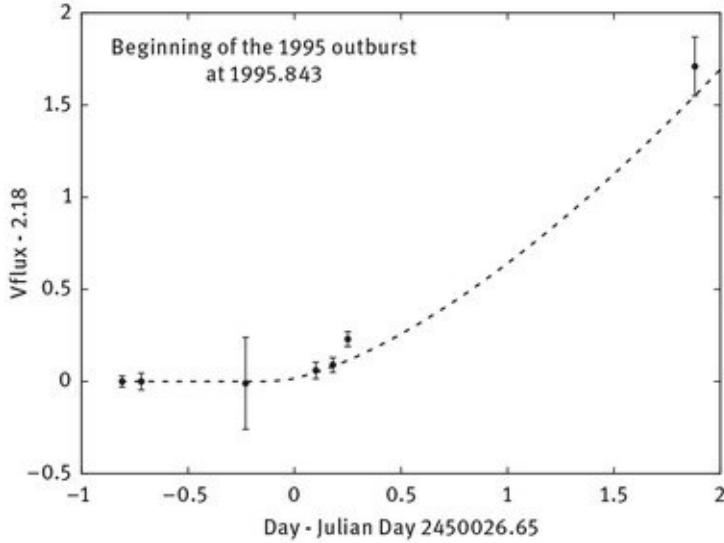
It is remarkable that the two sets of parameters in second and third columns of [Table 4 agree](#) so closely. If the basic model were not correct, there would be no reason for this agreement. Also, the mere fact that an orbit solution exists at all is a strong argument for the model.

Let us now turn our attention to the expected future outbursts

in OJ287. We expect three more outbursts during the next two decades, occurring in 2016, 2019 and 2022. The 2016 outburst should be an easy one to detect, as it comes in January of that year. Its timing is spin-sensitive; the exact date will give us a good spin value from the following formula:

$$\chi_1 = 0.25 - 0.5 \times (t - 2016.0).$$

(70)



**Fig. 12.** The observations of OJ287 during 4–7 November 1995, transformed to optical V-band. Overlaid is the theoretical light curve profile, corresponding to  $q = 1$ . The zero point of the time axis is at 3:36 UT of 1995, November 5. The peak flux of 5.1 mJy above the base level was predicted for November 8 [255], but there were no observations to confirm it. The data on November 9 and later suggest that the outburst maximum was probably missed.

Here  $t$  is the time of the beginning of the outburst in years. The expected accuracy is 0.005 units in  $\chi_1$ . The dependence on  $q$  is secondary, and thus the 2016 outburst timing is of no use by itself for the testing of the no-hair theorem.

The value of  $q$  is currently best determined by the 1995 outburst. There was an intensive monitoring campaign of OJ287 (called OJ94) during this outburst season, but unfortunately there exists a gap in these observations just at the crucial time ([274], [Figure 12](#)). It may still be possible that there are measurements somewhere which are not recorded in the OJ94 campaign light curve, and which would be valuable in narrowing

down  $q$  even from these data. The line in [Figure 12 is](#) drawn using the well observed 2007 outburst as a template to compare with 1995. It is should be noted that even a few more measurements of 1995 could narrow down the range of  $q$ .

The 2019 outburst timing is also sensitive to the  $q$ :

$$q = 1.0 - 1.2 \times (t - t_0),$$

(71)

where  $t$  is time of the beginning of the outburst (in days) and  $t_0$  is 2019 July 14, 11:19 UT. The expected accuracy for the  $q$  determination is 0.15 units, *i.e.* optimally the accuracy of the no-hair theorem test is doubled in comparison with the present uncertainty. Observing OJ287 during the rising flux of the 2019 outburst, from July 14 to July 17, is a challenge for ground based observers. This is because the angular distance between the Sun and OJ287 is only about 19 degrees at the beginning of the event, and goes down to about 16 degrees by the time of the peak flux on July 17. T. Pursimo has measured OJ287 at the Nordic Optical telescope on 2007 July 12 [277] which demonstrates that the July measurements should also be feasible in 2019. Space observations would also be useful in order to study the outburst over a wider spectral range. The 2022 outburst is scheduled practically at the same time of year as the 2019 outburst. Obviously it would also be of interest to observe this event as it would tie down the parameters of the general model more narrowly. However, it will not give any further information on  $q$ .

**Acknowledgment:** This work has been supported by The Finnish Society of Sciences and Letters and The Finnish Academy of Science and Letters.

We note that Bruce Barker (who died in 1988 at the age of 54) was a co-author of two important theoretical papers testing General Relativity in extreme binary systems like those discussed here. He was a friend and colleague of M. V. and G. B. at the University of Alabama.

## Bibliography

- [1] A. Einstein, Die Grundlage der allgemeinen Relativitätstheorie, *Annalen der Physik*, **354** (1916a), 769–822.
- [2] A. Einstein, Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie, *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, Berlin (1917), 142–152.
- [3] G. G. Byrd, A. D. Chernin and M. J. Valtonen, *Cosmology: Foundations and Frontiers*, Moscow, URSS (2007).
- [4] G. Byrd, A. D. Chernin, P. Teerikorpi and M. Valtonen, *Paths to Dark Energy*, De Gruyter, (2012).
- [5] G. Lemaître, The expanding universe, *Monthly Notices of the Royal Astronomical Society*, **91** (1931), 490–501.
- [6] G. C. McVittie, Condensations in an expanding universe, *Monthly Notices of the Royal Astronomical Society*, **92** (1932), 500–518.
- [7] G. C. McVittie, The mass-particle in an expanding universe, *Monthly Notices of the Royal Astronomical Society*, **93** (1933), 325–339.
- [8] G. Järnefelt, Über die Bewegung eines Massenpunktes in einem Raum mit zentralesymmetrischer Massen- und Druckverteilung, *Zeitschrift für Astrophysik*, **7** (1933), 326–327.
- [9] A. Einstein and E. G. Straus, The Influence of the Expansion of Space on the Gravitation Fields Surrounding the Individual Stars, *Reviews of Modern Physics*, **17** (1945), 120–124.
- [10] A. Einstein and E. G. Straus, Corrections and Additional Remarks to our Paper: The Influence of the Expansion of Space on the Gravitation Fields Surrounding the Individual Stars, *Reviews of Modern Physics*, **18** (1946), 148–149.
- [11] D. Psaltis, Probes and Tests of Strong-Field Gravity with Observations in the Electromagnetic Spectrum,

*Living Reviews in Relativity*, **11** (2008), 9:1–9:61.

- [12] U. Le Verrier, Lettre de M. Le Verrier à M. Faye sur la theorie de Mercure et sur le mouvement du perihelie de cette planete *Comptes rendus hebdomadaires des seances de l'Academie des sciences (Paris)*, **49** (1859), 379–383.
- [13] S. Newcomb, *The elements of the four inner planets and the fundamental constants*, (Washington, DC, US Govt. Office, 1895).
- [14] A. Hall, A suggestion in the theory of Mercury, *Astron. J.*, **14** (1894), 49–51.
- [15] I. I. Shapiro, C. C. Counselman III and R. W. King, Verification of the principle of equivalence for massive bodies, *Phys. Rev. Lett.*, **36** (1976), 555–558.
- [16] A. Einstein, Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie *Sitz. Preuss. Akad. Wiss.*, **11** (1915), 831–839.
- [17] R. A. Hulse and J. H. Taylor, Discovery of a pulsar in a binary system, *Astrophys. J.*, **195** (1975), L51–L53.
- [18] J. H. Taylor and J. M. Weisberg, Further experimental tests of relativistic gravity using the binary pulsar PSR 1913 + 16, *Astrophys. J.*, **345** (1989), 434–450.
- [19] T. Damour and J. H. Taylor, On the orbital period change of the binary pulsar PSR1913+16, *Astrophys. J.*, **366** (1991), 501–511.
- [20] R. A. Hulse, The discovery of the binary pulsar, *Rev. Mod. Phys.*, **66** (1994), 699–710.
- [21] A. G. Lyne et al., A Double-Pulsar System: A Rare Laboratory for Relativistic Gravity and Plasma Physics, *Science*, **303** (2004), 1153–1157.
- [22] M. Kramer et al., Tests of General Relativity from Timing the Double Pulsar, *Science*, **314** (2006), 97–102.
- [23] J. Lense and H. Thirring, Über den Einfluss der Eigenrotation der Zentralkörper auf die Bewegung der Planeten und Monde nach der Einsteinschen Gravitationstheorie. *Physikalische Zeitschrift*, **19**

(1918), 156–163.

- [24] I. Ciufolini and E. C. Pavlis, A confirmation of the general relativistic prediction of the Lense-Thirring effect, *Nature*, **431** (2004), 958–960.
- [25] C. W. F. Everitt et al., Gravity Probe B: Final Results of a Space Experiment to Test General Relativity, *Physical Review Letters*, **106** (2011), 221101-1–221101-5
- [26] T. Damour and R. Ruffini, Certain new verifications of general relativity made possible by the discovery of a pulsar belonging to a binary system, *Acad. Sci. Paris Comptes Rendus Ser. Sci. Math.*, **279** (1974), 971–973.
- [27] B. M. Barker and R. F. O’Connell, Gravitational two-body problem with arbitrary masses, spins, and quadrupole moments, *Physical Review D*, **12** (1975a), 329–335.
- [28] B. M. Barker and R. F. O’Connell, Relativistic effects in the binary pulsar PSR 1913+16, *Astrophys. J. Lett.*, **199** (1975b), L25–L26.
- [29] J. M. Weisberg and J. H. Taylor, 2005, The Relativistic Binary Pulsar B1913+16: Thirty Years of Observations and Analysis, *Binary Radio Pulsar, ASP Conference Series*, Eds: F. A. Rasio and I. H. Stairs, **328** (2005), 25–31.
- [30] J. M. Weisberg, D. J. Nice and J. H. Taylor, 2010, “Timing Measurements of the Relativistic Binary Pulsar B1913+16” *Astrophys. J.*, **722** (2010), 1030–1034.
- [31] A. Einstein, Über den Einfluss der Schwerkraft auf die Ausbreitung des Lichtes *Annalen der Physik*, **340** (1911), 898–908.
- [32] A. S. Eddington, The total eclipse of May 29 and the influence of gravitation on light, *Observatory*, **42** (1919), 119–122.
- [33] F. W. Dyson, A. S. Eddington and C. R. Davidson, C., A Determination of the Deflection of Light by the Sun’s Gravitational Field, from Observations Made at the Total Eclipse of May 29, 1919, *Philosophical*

*Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, **220** (1920), 291–333.

- [34] A. Einstein, “Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen” *Jahrb. Radioakt. Elekt.* **4** (1907), 411.
- [35] R. V. Pound and G. A. Rebka, “Gravitational redshift in nuclear resonance”, *Phys. Rev. Lett.* **3** (1959), 439.
- [36] D. Walsh, R. F. Carswell and R. J. Weymann, 0957 + 561 A, B – twin quasistellar objects or gravitational lens, *Nature*, **279** (1979), 381–384.
- [37] M. J. Valtonen and H. J. Lehto, Outbursts in OJ 287: A New Test for the General Theory of Relativity, *Astrophysical Journal Letters*, **481** (1997), L5–L7.
- [38] Y.-T. Sun, J.-Y. Liu, J.-Z. Liu, Y.-L. Ai, M. Zhou and E.-L. Qiao, A Research on the Gravitational Wave Radiation of OJ 287, *Chinese Astronomy and Astrophysics*, **35** (2011), 123–132.
- [39] J. Liu, Y. Zhang, H. Zhang, Y. Sun and N. Wang, Determining gravitational wave radiation from close galaxy pairs using a binary population synthesis approach, *Astronomy and Astrophysics*, **540** (2012), A67:1–A67:6.
- [40] R. P. Kerr, Gravitational Field of a Spinning Mass as an Example of Algebraically Special Metrics, *Physical Review Letters*, **11** (1963), 237–238.
- [41] M. J. Valtonen et al., Measuring black hole spin in OJ287, *Celestial Mech. and Dynamical Astr.*, **106** (2010b), 235–243.
- [42] W. Israel, Event horizons in static vacuum spacetimes, *Phys. Rev.* **164** (1967), 1776–1779.
- [43] W. Israel, Event horizons in static electrovac spacetimes, *Commun. Math. Phys.* **8** (1968), 245–260.
- [44] B. Carter, An axisymmetric black hole has only two degrees of freedom, *Phys. Rev. Lett.* **26** (1970), 331–333.
- [45] S. W. Hawking, Gravitational radiation from colliding

- black holes, *Phys. Rev. Lett.* **26** (1971), 1344–1346.
- [46] S. W. Hawking, Black holes in general relativity, *Commun. Math. Phys.* **25** (1972), 152–166.
- [47] C. W. Misner, K. S. Thorne and J. A. Wheeler, *Gravitation* (W. H. Freeman & Sons, San Francisco, p. 876, 1973).
- [48] A. A. Friedmann, Über die Krümmung des Raumes, *Zeitschrift für Physik*, **10** (1922), 377–386.
- [49] A. A. Friedmann, Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes, *Zeitschrift für Physik*, **21**, 326–332.
- [50] G. Lemaitre, Un Univers homogene de masse constante et de rayon croissant rendant compte de la vitesse radiale des nebuleuses extra-galactiques, *Annales de la Societe Scientifique de Bruxelles*, **A47** (1927), 49–59.
- [51] E. Hubble, A Relation between Distance and Radial Velocity among ExtraGalactic Nebulae, *Proceedings of the National Academy of Sciences of the United States of America*, **15** (1929), 168–173.
- [52] A. G. Riess et al., Observational evidence from supernovae for an accelerating universe and a cosmological constant, *Astron. J.*, **116** (1998) 1009–1038.
- [53] S. Perlmutter et al., Measurements of Omega and Lambda from 42 High-Redshift Supernovae, *Astrophys. J.*, **517** (1999), 565–586.
- [54] A. D. Chernin, Reviews of topical problems: Cosmic vacuum, *Physics Uspekhi*, **44** (2001), 1099–1118.
- [55] A. D. Chernin, P. Teerikorpi and Yu. V. Baryshev, Non-Friedmann cosmology for the Local Universe, significance of the universal Hubble constant, and short-distance indicators of dark energy, *Astronomy and Astrophysics*, **456** (2006), 13–21.
- [56] S. van den Bergh, The local group of galaxies, *Astronomy and Astrophysics Review*, **9** (1999), 273–318.

- [57] I. D. Karachentsev, O. G. Kashibadze, D. I. Makarov and R. B. Tully, The Hubble flow around the Local Group, *Monthly Notices of the Royal Astronomical Society*, **393** (2009), 1265–1274.
- [58] A. D. Chernin et al., Local dark matter and dark energy as estimated on a scale of 1 Mpc in a self-consistent way, *Astronomy and Astrophysics*, **507** (2009), 1271–1276.
- [59] Yu. V. Baryshev, A. D. Chernin and P. Teerikorpi, The cold local Hubble flow as a signature of dark energy, *Astronomy and Astrophysics*, **378** (2001), 729–734.
- [60] P. Teerikorpi, A. D. Chernin and Yu. V. Baryshev, The quiescent Hubble flow, local dark energy tests, and pairwise velocity dispersion in a  $\Omega = 1$  universe, *Astronomy and Astrophysics*, **440** (2005), 791–797.
- [61] I. D. Karachentsev et al., Local galaxy flows within 5 Mpc, *Astronomy and Astrophysics*, **398** (2003), 479–491.
- [62] P. Teerikorpi, A. D. Chernin, I. D. Karachentsev and M. J. Valtonen, Dark energy in the environments of the Local Group, the M 81 group, and the CenA group: the normalized Hubble diagram, *Astronomy and Astrophysics*, **483** (2008), 383–387.
- [63] A. D. Chernin et al., The very local Hubble flow: Computer simulations of dynamical history, *Astronomy and Astrophysics*, **415** (2004), 19–25.
- [64] S. Peirani and J. A. de Freitas Pacheco, Dynamics of nearby groups of galaxies: the role of the cosmological constant, *Astronomy and Astrophysics*, **488** (2008), 845–851.
- [65] J. Saarinen and P. Teerikorpi, The Hubble diagram for a system within dark energy: influence of some relevant quantities. *Astronomy and Astrophysics*, **568** (2014), A33. *arXiv: 1402.0950* (2013).
- [66] M. J. Valtonen, G. G. Byrd, M. L. McCall and K. A. Innanen, A revised history of the Local Group and a generalized method of timing, *Astronomical Journal*,

**105** (1993), 886–893.

- [67] G. Byrd, M. Valtonen, M. McCall and K. Innanen, Orbits of the Magellanic Clouds and Leo I in local group history, *Astronomical Journal*, **107** (1994), 2055–2059.
- [68] A. D. Chernin et al., Local dark energy: HST evidence from the vicinity of the M81/M82 galaxy group, *Astrophysics*, **50** (2007), 405–415.
- [69] P. Teerikorpi and A. D. Chernin, The Hubble diagram for a system within dark energy: the location of the zero-gravity radius and the global Hubble rate, *Astronomy and Astrophysics*, **516** (2010), A93:1–A93:4.
- [70] F. D. Kahn and L. Woltjer, Intergalactic Matter and the Galaxy, *Astrophysical Journal*, **130** (1959), 705–717.
- [71] J. Binney and S. Tremaine, *Galactic Dynamics: Second Edition*, Princeton University Press, Princeton, NJ USA, (2008).
- [72] S. Garrison-Kimmel, M. Boylan-Kolchin, J. Bullock and K. Lee, ELVIS: Exploring the Local Volume in Simulations, *arXiv*: 1310.6746 (2013).
- [73] W. R. Forman, A Reduction of the Mass Deficit in Clusters of Galaxies by Means of a Negative Cosmological Constant, *Astrophysical Journal*, **159** (1970), p.719–722.
- [74] J. C. Jackson, The dynamics of clusters of galaxies in universes with nonzero cosmological constant, and the virial theorem mass discrepancy, *Monthly Notices of the Royal Astronomical Society*, **148** (1970), 249–260.
- [75] A. D. Chernin et al., Virial estimator for dark energy, *Gravitation and Cosmology*, **18** (2012), 1–5.
- [76] M. Valtonen and H. Karttunen, *The Three-Body Problem*, Cambridge University Press, Cambridge (2006).
- [77] A. D. Chernin, Dark energy in the nearby Universe: HST data, nonlinear theory, and computer simulations, *Physics-Uspekhi*, **56** (2013), 704.
- [78] F. Zwicky, Die Rotverschiebung von

- extragalaktischen Nebeln, *Helvetica Physica Acta*, **6** (1933), 110-127.
- [79] F. Zwicky, On the Masses of Nebulae and of Clusters of Nebulae, *Astrophysical Journal*, **86** (1937), p.217-246.
- [80] L. S. The and S. D. M. White, The mass of the Coma cluster, *Astronomical Journal*, **92** (1986), 1248-1253.
- [81] J. P. Hughes, The mass of the Coma Cluster - Combined X-ray and optical results, *Astrophysical Journal*, **337** (1989), 21-33.
- [82] J. P. Hughes, Untangling Coma Berenices: A New Vision of an Old Cluster, *Proceedings of the meeting held in Marseilles, June 17-20, 1997*, Eds.: A. Mazure, F. Casoli, F. Durret, D. Gerbal, Word Scientific, p 175 (1998).
- [83] M. Colless, The 2dF Galaxy Redshift Survey, in *Encyclopedia of Astronomy and Astrophysics*, P. Murdin, ed., Bristol: Institute of Physics Publishing, UK, article 5485 (2002).
- [84] J. M. Kubo et al., The Mass of the Coma Cluster from Weak Lensing in the Sloan Digital Sky Survey, *The Astrophysical Journal*, **671** (2007), 1466-1470.
- [85] M. J. Valtonen and G. G. Byrd, A binary model for the Coma cluster of galaxies, *Astrophysical Journal*, **230** (1979), 655-666.
- [86] M. J. Valtonen, K. A. Innanen, T.-Y. Huang and S. Saarinen, No missing mass in clusters of galaxies?, *Astronomy and Astrophysics*, **143** (1985), 182-187.
- [87] S. Laine, J.-Q. Zheng and M. J. Valtonen, Improved Models for the Evolution of the Coma Cluster of Galaxies, *Astronomical Journal*, **127** (2004), 765-770.
- [88] A. Diaferio and M. J. Geller, Infall Regions of Galaxy Clusters, *Astrophysical Journal*, **481**, (1997) 633-643.
- [89] A. Diaferio, Mass estimation in the outer regions of galaxy clusters, *Monthly Notices of Royal Astronomical Society*, **309** (1999), 610-622.
- [90] M. J. Geller, A. Diaferio and M. J. Kurtz, The Mass

Profile of the Coma Galaxy Cluster, *The Astrophysical Journal*, **517** (1999), L23-L26.

- [91] M. J. Geller, A. Diaferio and M. J. Kurtz, Mapping the Universe: The 2010 Russell Lecture, *The Astronomical Journal*, **142** (2011), 133:1-133:7.
- [92] J. F. Navarro, C. S. Frenk and S. D. M. White, A Universal Density Profile from Hierarchical Clustering, *Astrophysical Journal*, **490** (1997), 493-508.
- [93] L. Hernquist, An analytical model for spherical galaxies and bulges, *Astrophysical Journal*, **356** (1990), 359-364.
- [94] D. E. Holz and S. Perlmutter, The Most Massive Objects in the Universe, *The Astrophysical Journal Letters*, **755** (2012), L36:1-L36:4.
- [95] M. T. Busha, A. E. Evrard, F. C. Adams and R. H. Wechsler, The ultimate halo mass in a  $\Lambda$ CDM universe, *Monthly Notices of the Royal Astronomical Society: Letters*, **363** (2005), L11-L15.
- [96] I. D. Karachentsev, A. D. Chernin and P. Teerikorpi, The Hubble Flow: Why Does the Cosmological Expansion Preserve Its Kinematical Identity from a Few Mpc Distance to the Observation Horizon?, *Astrophysics (English translation of Astrofizika)*, **46** (2003), 399-414.
- [97] A. D. Chernin, Physics of our days: Dark energy and universal antigravitation, *Physics Uspekhi*, **51** (2008), 253-282.
- [98] A. D. Chernin, P. Teerikorpi and Yu. V. Baryshev, Why is the Hubble flow so quiet? *Advances in Space Research*, **31** (2003), 459-467.
- [99] A. D. Chernin et al., Dark energy domination in the Virgocentric flow, *Astronomy and Astrophysics*, **520** (2010), A104:1-A104:7.
- [100] A. D. Chernin et al., Dark energy in six nearby galaxy flows: Synthetic phase diagrams and self-similarity, *Astronomy Reports*, **56** (2012a), 653-663.
- [101] A. D. Chernin et al., Dark energy and extended

dark matter halos, *Astronomy and Astrophysics*, **539** (2012b), A4:1-A4:6.

- [102] A. D. Chernin, G. S. Bisnovatyi-Kogan, P. Teerikorpi, M. J. Valtonen, G. G. Byrd and M. Merafina, Dark energy and the structure of the Coma cluster of galaxies, *Astronomy and Astrophysics*, **553** (2013), A101:1-A101:4.
- [103] F. D. A. Hartwick, The Velocity Field around Groups of Galaxies, *The Astronomical Journal*, **141** (2011), 198:1-198:5.
- [104] M. Merafina, G. S. Bisnovatyi-Kogan and S. O. Tarasov, A brief analysis of self-gravitating polytropic models with a nonzero cosmological constant, *Astronomy and Astrophysics*, **541** (2012), A84:1-A84:5.
- [105] G. S. Bisnovatyi-Kogan and A. D. Chernin, Dark energy and key physical parameters of clusters of galaxies, *Astrophysics and Space Science*, **338** (2012), 337-343.
- [106] G. S. Bisnovatyi-Kogan and M. Merafina, Galactic cluster winds in presence of a dark energy, *Monthly Notices of the Royal Astronomical Society*, **434** (2013), 3628-3632.
- [107] S. Saarinen and M. J. Valtonen, Simulations of clusters of galaxies with massive cores, *Astronomy and Astrophysics*, **153** (1985), 130-138.
- [108] G. G. Byrd and M. J. Valtonen, Origin of redshift differentials in galaxy groups, *Astrophysical Journal*, **289** (1985), 535-539.
- [109] M. J. Valtonen and G. G. Byrd, Redshift asymmetries in systems of galaxies and the missing mass, *Astrophysical Journal*, **303** (1986), 523-534.
- [110] S.-M. Niemi and M. Valtonen, The origin of redshift asymmetries: how  $\Lambda$ CDM explains anomalous redshift, *Astronomy and Astrophysics*, **494** (2009), 857-865.
- [111] J. A. Frieman, M. S. Turner and D. Huterer, Dark Energy and the Accelerating Universe, *Annual Review of Astronomy and Astrophysics*, **46** (2008), 385-432.

- [112] A. Blanchard, Evidence for the fifth element. Astrophysical status of dark energy, *Astronomy and Astrophysics Review*, **18** (2010), 595–645.
- [113] D. H. Weinberg et al., Observational probes of cosmic acceleration, *Physics Reports*, **530** (2013), 87–255.
- [114] J. R. Whitbourn, T. Shanks and U. Sawangwit, Testing WMAP data via Planck radio and SZ catalogues, *Monthly Notices of the Royal Astronomical Society*, **437** (2014), 622–640.
- [115] A. D. Chernin, Dark Energy in Systems of Galaxies, *JETP Letters*, **98** (2013), 394–407.
- [116] R. Mohayaee and R. B. Tully, The Cosmological Mean Density and Its Local Variations Probed by Peculiar Velocities, *Astrophys. J.*, **635** (2005), L113–L116.
- [117] Y.-Z. Ma and D. Scott, Cosmic bulk flows on  $50 h^{-1}$  Mpc scales: a Bayesian hyper-parameter method and multishell likelihood analysis, *Monthly Notices of the Royal Astronomical Society*, **428** (2013), 2017–2028.
- [118] A. Nusser and M. Davis, The Cosmological Bulk Flow: Consistency with  $\Lambda$ CDM and  $z \sim 0$  Constraints on  $\sigma_8$  and  $\gamma$ , *The Astrophysical Journal*, **736** (2011), 93:1–93:9.
- [119] C.-G. Park and C. Park, Power Spectrum of Cosmic Momentum Field Measured from the SFI Galaxy Sample, *Astrophys. J.*, **637** (2006), 1–11.
- [120] M. Kowalski et al., Improved Cosmological Constraints from New, Old, and Combined Supernova Data Sets, *Astrophys. J.*, **686** (2008), 749–778.
- [121] J. P. Henry, A. E. Evrard, H. Hoekstra, A. Babul and A. Mahdavi, The X-Ray Cluster Normalization of the Matter Power Spectrum, *Astrophysica. J.*, **691** (2009), 1307–1321.
- [122] F. Beutler et al., The 6dF Galaxy Survey:  $z \approx 0$  measurements of the growth rate and  $\sigma_8$ , *Monthly Notices of the Royal Astronomical Society*, **423** (2012),

3430–3444.

- [123] J. L. Tinker et al., The Cosmological Constraints from Galaxy Clustering and the Mass-to-number Ratio of Galaxy Clusters, *Astrophys. J.*, **745** (2012), 16:1–16:22.
- [124] V. Simha and S. Cole, Cosmological constraints from applying SHAM to rescaled cosmological simulations, *Monthly Notices of the Royal Astronomical Society*, **436** (2013), 1142–1151.
- [125] M. Cacciato, F. C. van den Bosch, S. More, H. Mo and X. Yang, Cosmological constraints from a combination of galaxy clustering and lensing – III. Application to SDSS data, *Monthly Notices of the Royal Astronomical Society*, **430** (2013), 767–786.
- [126] S. Cole et al., The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications, *Monthly Notices of the Royal Astronomical Society*, **362** (2005), 505–534.
- [127] A. G. Sanchez et al., Cosmological parameters from cosmic microwave background measurements and the final 2dF Galaxy Redshift Survey power spectrum, *Monthly Notices of the Royal Astronomical Society*, **366** (2006), 189–207.
- [128] M. Hicken et al., Improved Dark Energy Constraints from  $\sim 100$  New CfA Supernova Type Ia Light Curves, *The Astrophysical Journal*, **700** (2009), 1097–1140.
- [129] E. Rozo et al., Cosmological Constraints from the Sloan Digital Sky Survey maxBCG Cluster Catalog, *Astrophysica. J.*, **708** (2010), 645–660.
- [130] H. Lampeitl et al., First-year Sloan Digital Sky Survey-II supernova results: consistency and constraints with other intermediate-redshift data sets, *Monthly Notices of the Royal Astronomical Society*, **401** (2010), 2331–2342.
- [131] A. Hajian et al., Measuring the thermal Sunyaev-Zel'dovich effect through the cross correlation of Planck and WMAP maps with ROSAT galaxy cluster

catalogs, *Journal of Cosmology and Astroparticle Physics*, **11** (2013), 064:1–064:30.

- [132] C. Blake et al., The WiggleZ Dark Energy Survey: measuring the cosmic expansion history using the Alcock-Paczynski test and distant supernovae, *Monthly Notices of the Royal Astronomical Society*, **418** (2011a), 1725–1735.
- [133] R. Kessler et al., First-Year Sloan Digital Sky Survey-II Supernova Results: Hubble Diagram and Cosmological Parameters, *Astrophysical J. Supplement*, **185** (2009), 32–84.
- [134] W. L. Percival et al., Baryon acoustic oscillations in the Sloan Digital Sky Survey Data Release 7 galaxy sample, *Monthly Notices of the Royal Astronomical Society*, **401** (2010), 2148–2168.
- [135] A. Cabre and E. Gaztañaga, Clustering of luminous red galaxies – II. Small-scale redshift-space distortions, *Monthly Notices of the Royal Astronomical Society*, **396** (2009), 1119–1131.
- [136] S. Cao, G. Covone, and Z.-H. Zhu, Testing the Dark Energy with Gravitational Lensing Statistics, *Astrophys. J.*, **755** (2012), 31:1–31:13.
- [137] A. Rest et al., Cosmological Constraints from Measurements of Type Ia Supernovae discovered during the first 1.5 years of the Pan-STARRS1 Survey, *arXiv:*, **1310.3828** (2013).
- [138] S. H. Suyu et al., Two Accurate Time-delay Distances from Strong Lensing: Implications for Cosmology, *Astrophys. J.*, **766** (2013), 70:1–70:19.
- [139] A. Mantz, S. W. Allen, D. Rapetti and H. Ebeling, The observed growth of massive galaxy clusters – I. Statistical methods and cosmological constraints, *Monthly Notices of the Royal Astronomical Society*, **406**, (2010), 1759–1772.
- [140] A. Vikhlinin et al., Chandra Cluster Cosmology Project III: Cosmological Parameter Constraints, *Astrophys. J.*, **692** (2009), 1060–1074.

- [141] H. Campbell et al., Cosmology with Photometrically Classified Type Ia Supernovae from the SDSS-II Supernova Survey, *Astrophys. J.*, **763** (2013), 88:1-88:28.
- [142] S. Ho, C. Hirata, N. Padmanabhan, U. Seljak and N. Bahcall, Correlation of CMB with large-scale structure. I. Integrated Sachs-Wolfe tomography and cosmological implications, *Physical Review D*, **78** (2008), 043519:1-043519:35.
- [143] A. Del Popolo, V. Costa and G. Lanzafame, Improvements in the X-ray luminosity function and constraints on the cosmological parameters from X-ray luminous clusters, *Astronomy and Astrophysics*, **514** (2010), A80:1-A80:8.
- [144] Q.-J. Zhang and Y.-L. Wu, Dark Energy and Hubble Constant From the Latest SNe Ia, BAO and SGL, *arXiv*: 0905.1234 (2009).
- [145] B. A. Reid et al., Cosmological constraints from the clustering of the Sloan Digital Sky Survey DR7 luminous red galaxies, *Monthly Notices of the Royal Astronomical Society*, **404** (2010), 60-85.
- [146] A. G. Sanchez, M. Crocce, A. Cabre, C. M. Baugh and E. Gaztañaga, Cosmological parameter constraints from SDSS luminous red galaxies: a new treatment of large-scale clustering, *Monthly Notices of the Royal Astronomical Society*, **400** (2009), 1643-1664.
- [147] K. T. Mehta, A. J. Cuesta, X. Xu, D. J. Eisenstein and N. Padmanabhan, A 2 per cent distance to  $z = 0.35$  by reconstructing baryon acoustic oscillations - III. Cosmological measurements and interpretation, *Monthly Notices of the Royal Astronomical Society*, **427** (2012), 2168-2179.
- [148] M. Tegmark et al., Cosmological constraints from the SDSS luminous red galaxies, *Physical Review D*, **74** (2006), 123507:1-123507:36.
- [149] D. J. Eisenstein et al., Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function

- of SDSS Luminous Red Galaxies, *Astrophysical Journal*, **633** (2005), 560–574.
- [150] T. Okumura et al., Large-Scale Anisotropic Correlation Function of SDSS Luminous Red Galaxies, *Astrophys. J.*, **676** (2008), 889–898.
- [151] J. E. Taylor et al., Measuring the Geometry of the Universe from Weak Gravitational Lensing behind Galaxy Groups in the HST COSMOS Survey, *Astrophys. J.*, **749** (2012), 127:1–127:12.
- [152] S. W. Allen et al., Improved constraints on dark energy from Chandra X-ray observations of the largest relaxed galaxy clusters, *Monthly Notices of the Royal Astronomical Society*, **383** (2008), 879–896.
- [153] R. Mandelbaum et al., Cosmological parameter constraints from galaxy-galaxy lensing and galaxy clustering with the SDSS DR7, *Monthly Notices of the Royal Astronomical Society*, **432** (2013), 1544–1575.
- [154] A. Cabre, E. Gaztañaga, M. Manera, P. Fosalba and F. Castander, Cross-correlation of Wilkinson Microwave Anisotropy Probe third-year data and the Sloan Digital Sky Survey DR4 galaxy survey: new evidence for dark energy, *Monthly Notices of the Royal Astronomical Society: Letters*, **372** (2006), L23–L27.
- [155] C. Blake et al., The WiggleZ Dark Energy Survey: mapping the distance-redshift relation with baryon acoustic oscillations, *Monthly Notices of the Royal Astronomical Society*, **418** (2011b), 1707–1724.
- [156] W. L. Freedman et al., The Carnegie Supernova Project: First Near-Infrared Hubble Diagram to  $z \sim 0.7$ , *Astrophysical Journal*, **704** (2009), 1036–1058.
- [157] W. M. Wood-Vasey et al., Observational Constraints on the Nature of Dark Energy: First Cosmological Results from the ESSENCE Supernova Survey, *Astrophysical Journal*, **666** (2007), 694–715.
- [158] M. Sullivan et al., SNLS3: Constraints on Dark Energy Combining the Supernova Legacy Survey Three-year Data with Other Probes, *Astrophys. J.*, **737**

(2011), 102:1-102:19.

- [159] K. M. Wilson, G. Chen and B. Ratra, Supernova Ia and Galaxy Cluster Gas Mass Fraction Constraints on Dark Energy, *Modern Physics Letters A*, **21** (2006), 219-2204.
- [160] M. Oguri et al., The Sloan Digital Sky Survey Quasar Lens Search. VI. Constraints on Dark Energy and the Evolution of Massive Galaxies, *Astron. J.*, **143** (2012), 120:1-120:14.
- [161] G. Hinshaw et al., Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results, *Astrophys. J. Suppl.*, **208** (2013), 19:1-19:25.
- [162] S. Cao, and Z.-H. Zhu, Constraints on cosmological models from lens redshift data, *Astronomy and Astrophysics*, **538** (2012), A43:1-A43:7.
- [163] L. D. Ferramacho, A. Blanchard and Y. Zolnierowski, Constraints on CDM cosmology from galaxy power spectrum, CMB and SNIa evolution, *Astronomy and Astrophysics*, **499** (2009), 21-29.
- [164] A. Clocchiatti et al., Hubble Space Telescope and Ground-based Observations of Type Ia Supernovae at Redshift 0.5: Cosmological Implications, *Astrophysical Journal*, **642** (2006), 1-21.
- [165] D. Parkinson et al., The WiggleZ Dark Energy Survey: Final data release and cosmological results, *Physical Review D*, **86** (2012), 103518:1-103518:23.
- [166] L. Samushia et al., The clustering of galaxies in the SDSS-III DR9 Baryon Oscillation Spectroscopic Survey: testing deviations from Lambda and general relativity using anisotropic clustering of galaxies, *Monthly Notices of the Royal Astronomical Society*, **429** (2013), 1514-1528.
- [167] E. M. Huff et al., Seeing in the dark - II. Cosmic shear in the Sloan Digital Sky Survey, *arXiv* : 1112.3143 (2011), 1-23.
- [168] A. Del Popolo, Constraints on the cosmological

parameters by means of the clusters mass function, *Memorie della Societa Astronomica Italiana*, **81** (2010), 157–162.

- [169] N. P. Ross et al., The 2dF-SDSS LRG and QSO Survey: the LRG 2-point correlation function and redshift-space distortions, *Monthly Notices of the Royal Astronomical Society*, **381** (2007), 573–588.
- [170] C. Blake et al., The WiggleZ Dark Energy Survey: small-scale clustering of Lyman-break galaxies at  $z < 1$ , *Monthly Notices of the Royal Astronomical Society*, **395** (2009), 240–254.
- [171] E. Gaztañaga, Dark Energy from the millimeter sky, *Revista Mexicana de Astronomia y Astrofisica (Serie de Conferencias)*, **24** (2005), 40–44.
- [172] P. Astier et al., The Supernova Legacy Survey: measurement of  $\Omega_M$ ,  $\Omega_\Lambda$  and  $w$  from the first year data set, *Astronomy and Astrophysics*, **447** (2006), 31–48.
- [173] B. A. Reid et al., The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: measurements of the growth of structure and expansion rate at  $z = 0.57$  from anisotropic clustering, *Monthly Notices of the Royal Astronomical Society*, **426** (2012), 2719–2737.
- [174] C.-H. Chuang et al., The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: single-probe measurements and the strong power of  $f(z)\sigma_8(z)$  on constraining dark energy, *Monthly Notices of the Royal Astronomical Society*, **433** (2013), 3559–3571.
- [175] G.-B. Zhao et al., The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: weighing the neutrino mass using the galaxy power spectrum of the CMASS sample, *Monthly Notices of the Royal Astronomical Society*, **436** (2013), 2038–2053.
- [176] L. Anderson et al., The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Release 9

spectroscopic galaxy sample, *Monthly Notices of the Royal Astronomical Society*, **427** (2012), 3435–3467.

- [177] Y. Chen, C.-Q. Geng, S. Cao, Y.-M. Huang and Z.-H. Zhu, Constraints on the inverse power-law scalar field dark energy model from strong gravitational lensing data and updated Hubble parameter measurements, *arXiv*: 1312.1443 (2013).
- [178] O. Farooq and B. Ratra, Hubble Parameter Measurement Constraints on the Cosmological Deceleration-Acceleration Transition Redshift, *Astrophys. J. Lett.*, **766** (2013), L7:1–L7:4.
- [179] J. Guy et al., The Supernova Legacy Survey 3-year sample: Type Ia supernovae photometric distances and cosmological constraints, *Astronomy and Astrophysics*, **523** (2010), A7:1–A7 :34.
- [180] A. Feoli, L. Mancini, V. Rillo and M. Grasso, Cosmological constraints from supernova data set with corrected redshift, *Journal of Physics: Conference Series*, **354** (2012), 012005:1–012005:10.
- [181] A. Conley et al., Supernova Constraints and Systematic Uncertainties from the First Three Years of the Supernova Legacy Survey, *Astrophys. J. Supplement*, **192** (2011), 1:1–1:29.
- [182] G. E. Addison, G. Hinshaw and M. Halpern, Cosmological constraints from baryon acoustic oscillations and clustering of large-scale structure, *Monthly Notices of the Royal Astronomical Society*, **436** (2013), 1674–1683.
- [183] M. Ganeshalingam, W. Li and A. V. Filippenko, Constraints on dark energy with the LOSS SN Ia sample, *Monthly Notices of the Royal Astronomical Society*, **433** (2013), 2240–2258.
- [184] H. Lin et al., The SDSS Co-add: Cosmic Shear Measurement, *Astrophys. J.*, **761** (2012), 15:1–15:15.
- [185] S. Carneiro, C. Pigozzo, H. A. Borges and J. S. Alcaniz, Supernova constraints on decaying vacuum cosmology, *Physical Review D*, **74** (2006), 023532:1–

023532:7.

- [186] T. M. Davis et al., Scrutinizing Exotic Cosmological Models Using ESSENCE Supernova Data Combined with Other Cosmological Probes, *Astrophys. J.*, **666** (2007), 716–725.
- [187] M. Sereno and D. Paraficz, Hubble constant and dark energy inferred from free-form determined time delay distances, *Monthly Notices of the Royal Astronomical Society*, **437** (2014), 600–605.
- [188] C. Heymans et al., CFHTLenS tomographic weak lensing cosmological parameter constraints: Mitigating the impact of intrinsic galaxy alignments, *Monthly Notices of the Royal Astronomical Society*, **432** (2013), 2433–2453.
- [189] U. Seljak et al., Cosmological parameter analysis including SDSS Ly alpha forest and galaxy bias: Constraints on the primordial spectrum of fluctuations, neutrino mass, and dark energy, *Physical Review D*, **71** (2005), 103515:1–103515:21.
- [190] W. L. Freedman et al., Carnegie Hubble Program: A Mid-infrared Calibration of the Hubble Constant, *Astrophysical Journal*, **758** (2012), 24:1–24:10.
- [191] B. A. Benson et al., Cosmological Constraints from Sunyaev-Zel'dovich-selected Clusters with X-Ray Observations in the First  $178 \text{ deg}^2$  of the South Pole Telescope Survey, *Astrophysica. J.*, **763** (2013), 147:1–147:21.
- [192] M. C. March, R. Trotta, P. Berkes, G. D. Starkman and P. M. Vaudrevange, Improved constraints on cosmological parameters from Type Ia supernova data, *Monthly Notices of the Royal Astronomical Society*, **418** (2011), 2308–2329.
- [193] J. Benjamin et al., Cosmological constraints from the  $100\text{-deg}^2$  weak-lensing survey, *Monthly Notices of the Royal Astronomical Society*, **381** (2007), 702–712.
- [194] S. Ettori et al., The cluster gas mass fraction as a cosmological probe: a revised study, *Astronomy and*

*Astrophysics*, **501** (2009), 61–73.

- [195] T. Giannantonio et al., Combined analysis of the integrated Sachs-Wolfe effect and cosmological implications, *Physical Review D*, **77** (2008), 123520:1–123520:24.
- [196] Y. Chen and B. Ratra, Hubble parameter data constraints on dark energy, *Physics Letters B*, **703** (2011), 406–411.
- [197] E. Semboloni et al., Cosmic shear analysis with CFHTLS deep data, *Astronomy and Astrophysics*, **452** (2006), 51–61.
- [198] L. Fu et al., Very weak lensing in the CFHTLS wide: cosmology from cosmic shear in the linear regime, *Astronomy and Astrophysics*, **479** (2008), 9–25.
- [199] H. K. Jassal, J. S. Bagla and T. Padmanabhan, Understanding the origin of CMB constraints on dark energy, *Monthly Notices of the Royal Astronomical Society*, **405** (2010), 2639–2650.
- [200] C. Firmani, V. Avila-Reese, G. Ghisellini and G. Ghirlanda, The Hubble diagram extended to  $z \gg 1$ : the gamma-ray properties of gamma-ray bursts confirm the Lambda cold dark matter model, *Monthly Notices of the Royal Astronomical Society: Letters*, **372** (2006), L28–L32.
- [201] S. Basilakos and M. Plionis, Breaking the  $\sigma_8 - \Omega_m$  Degeneracy Using the Clustering of High- $z$  X-ray Active Galactic Nuclei, *Astrophys. J. Lett.*, **714** (2010), L185–L189.
- [202] N. Suzuki et al., The Hubble Space Telescope Cluster Supernova Survey. V. Improving the Dark-energy Constraints above  $z > 1$  and Building an Early-type-hosted Supernova Sample, *Astrophys. J.*, **746** (2012), 85:1–85:24.
- [203] J.-J. Wei, X.-F Wu, and F. Melia, The Gamma-Ray Burst Hubble Diagram and Its Implications for Cosmology, *Astrophysica. J.*, **772** (2013), 43:1–43:11.
- [204] M. Lopez-Corredoira, Alcock–Paczynski

cosmological test, *arXiv*: 1312.0003 (2013).

- [205] N. Suzuki et al., The Hubble Space Telescope Cluster Supernova Survey. V. Improving the Dark-energy Constraints above  $z > 1$  and Building an Early-type-hosted Supernova Sample, *Astrophys. J.*, **746** (2012), 85:1–85:24.
- [206] J. da Angela et al., The 2dF QSO Redshift Survey-XV. Correlation analysis of redshift-space distortions, *Monthly Notices of the Royal Astronomical Society*, **360** (2005a), 1040–1054.
- [207] J. da Angela et al., The 2dF-SDSS LRG and QSO survey: QSO clustering and the L-z degeneracy, *Monthly Notices of the Royal Astronomical Society*, **383** (2008), 565–580.
- [208] Y. Kodama et al., Gamma-ray bursts in  $1.8 < z < 5.6$  suggest that the time variation of the dark energy is small, *Monthly Notices of the Royal Astronomical Society: Letters*, **391** (2008), L1–L4.
- [209] R. Tsutsui et al., Constraints on  $w_0$  and  $w_a$  of dark energy from high-redshift gamma-ray bursts, *Monthly Notices of the Royal Astronomical Society: Letters*, **394** (2009), L31–L35.
- [210] N. G. Busca et al., Baryon acoustic oscillations in the Ly $\alpha$  forest of BOSS quasars, *Astronomy and Astrophysics*, **552** (2013), A96:1–A96:18.
- [211] V. C. Busti, R. C. Santos and J.A.S. Lima, Constraining the dark energy and smoothness parameter with type Ia supernovae and gamma-ray bursts, *Physical Review D*, **85** (2012), 103503:1–103503:7.
- [212] J. da Angela, P. J. Outram and T. Shanks, Constraining beta(z) and  $\Omega_m^0$  from redshift-space distortions in  $z \sim 3$  galaxy surveys, *Monthly Notices of the Royal Astronomical Society*, **361** (2005b), 879–886.
- [213] J. Dunkley et al., Five-Year Wilkinson Microwave Anisotropy Probe Observations: Likelihoods and Parameters from the WMAP Data, *Astrophysical J.*

*Supplement*, **180** (2009), 306–329.

- [214] Planck Collaboration; P. A. R. Ade et al., Planck 2013 results. XVI. Cosmological parameters, *arXiv*: 1303.5076 (2013).
- [215] C. L. Bennett et al., Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results, *Astrophysical Journal Supplement*, **208** (2013), 20:1–20:54.
- [216] D. N. Spergel et al., Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology, *Astrophysical Journal Supplement Series*, **170** (2007), 377–408.
- [217] C. L. Reichardt et al., High-Resolution CMB Power Spectrum from the Complete ACBAR Data Set, *Astrophys. J.*, **694** (2009), 1200–1219.
- [218] E. Komatsu et al., Five-Year Wilkinson Microwave Anisotropy Probe Observations: Cosmological Interpretation, *Astrophys. J. Suppl.*, **180** (2009), 330–376.
- [219] E. Komatsu et al., Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation, *Astrophysical J. Suppl.*, **192** (2011), 18:1–18:47.
- [220] Y. Wang, Model-independent distance measurements from gamma-ray bursts and constraints on dark energy, *Physical Review D*, **78** (2008), 123532:1–123532:8.
- [221] N. Liang, W. K. Xiao, Y. Liu and S. N. Zhang, A Cosmology-Independent Calibration of Gamma-Ray Burst Luminosity Relations and the Hubble Diagram, *Astrophys. J.*, **685** (2008), 354–360.
- [222] B. E. Schaefer, The Hubble Diagram to Redshift > 6 from 69 Gamma-Ray Bursts, *Astrophys. J.*, **660** (2007), 16–46.
- [223] B. D. Sherwin et al., Evidence for Dark Energy from the Cosmic Microwave Background Alone Using the Atacama Cosmology Telescope Lensing Measurements,

*Physical Review Letters*, **107** (2011), 021302:1–021302:5.

- [224] D. Pietrobon, A. Balbi and D. Marinucci, Integrated Sachs-Wolfe effect from the cross correlation of WMAP 3year and the NRAO VLA sky survey data: New results and constraints on dark energy, *Physical Review D*, **74** (2006), 043524:1–043524:9.
- [225] D. Spergel, R. Flauger and R. Hlozek, Planck Data Reconsidered, *arXiv*: 1312.3313 (2013), 1–15.
- [226] J. L. Sievers et al., The Atacama Cosmology Telescope: cosmological parameters from three seasons of data, *Journal of Cosmology and Astroparticle Physics*, **10** (2013), 060:1–060:26.
- [227] D. Larson et al., Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Power Spectra and WMAP-derived Parameters, *Astrophysical Journal Supplement*, **192** (2011), 16:1–16:19.
- [228] C. M. Will, The Confrontation between General Relativity and Experiment, *Living Reviews in Relativity*, **9** (2006), 3:1–3:100.
- [229] N. Yunes and X. Siemens, Gravitational-Wave Tests of General Relativity with Ground-Based Detectors and Pulsar-Timing Arrays, *Living Reviews in Relativity*, **16** (2013), 9:1–9:124.
- [230] J. R. Gair, M. Vallisneri, S. L. Larson and J. G. Baker, Testing General Relativity with Low-Frequency, Space-Based Gravitational-Wave Detectors, *Living Reviews in Relativity*, **16** (2013) 7:1–7:109.
- [231] K. S. Thorne and J. B. Hartle, Laws of motion and precession for black holes and other bodies, *Rev. Mod. Phys. D*, **31** (1985), 1815–1837.
- [232] K. S. Thorne, R. M. Price and D. A. Macdonald, *Black Holes: The Membrane Paradigm*, (Yale Univ. Press, New Haven, 1986).
- [233] N. Wex and S. M. Kopeikin, Frame dragging and other precessional effects in black hole pulsar binaries, *Astrophys. J.*, **514** (1999), 388–401

- [234] C. M. Will, Testing the general relativistic “no-hair” theorems using the Galactic Center black hole Sagittarius A\*, *Astrophys. J.*, **674** (2008), L25–L28.
- [235] A. Sillanpää, S. Haarala, M. J. Valtonen, B. Sundelius and G. G. Byrd, OJ287 – Binary pair of supermassive black holes, *Astrophys. J.*, **325** (1988), 628–634.
- [236] M. J. Valtonen et al., Predicting the next outbursts of OJ287 in 2006–2010, *Astrophys. J.*, **646** (2006), 36–48.
- [237] P. Pihajoki, M. Valtonen and S. Ciprini, Short timescale periodicity in OJ287, *Mon. Not. RAS*, **434** (2013), 3122–3129.
- [238] M. J. Valtonen, S. Mikkola, H. J. Lehto, A. Gopakumar, R. Hudec and J. Polednikova, Testing the Black Hole No-hair Theorem with OJ287, *Astrophys. J.*, **742** (2011a), 22–33.
- [239] L. Barack and C. Cutler, Using LISA extreme-mass-ratio inspiral sources to test off-Kerr deviations in the geometry of massive black holes, *Physical Review D*, **75** (2007), 042003:1–042003 :11.
- [240] A. E. Broderick, T. Johannsen, A. Loeb and D. Psaltis, Testing the No-Hair Theorem with Event Horizon Telescope Observations of Sagittarius A\*, *arXiv*: 1311.5564 (2013)
- [241] T. Johannsen and D. Psaltis, Testing the No-hair Theorem with Observations in the Electromagnetic Spectrum. I. Properties of a Quasi-Kerr Spacetime, *Astrophys. J.*, **716** (2010a), 187–197.
- [242] T. Johannsen and D. Psaltis, Testing the No-hair Theorem with Observations in the Electromagnetic Spectrum. II. Black Hole Images, *Astrophys. J.*, **718** (2010b), 446–454.
- [243] T. Johannsen and D. Psaltis, Testing the No-hair Theorem with Observations in the Electromagnetic Spectrum. III. Quasiperiodic Variability, *Astrophys. J.*, **726** (2011), 11–20.

- [244] T. Johannsen and D. Psaltis, Testing the No-hair Theorem with Observations in the Electromagnetic Spectrum. IV. Relativistically Broadened Iron Lines, *Astrophys. J.*, **773** (2013), 57-65.
- [245] C. Bambi and E. Barausse, Final stages of accretion onto non-Kerr compact objects, *Physical Review D*, **84** (2011), 084034:1-084034:13.
- [246] C. Bambi and E. Barausse, Constraining the Quadrupole Moment of Stellar-mass Black Hole Candidates with the Continuum Fitting Method, *The Astrophysical Journal*, **731** (2011b), 121:1-121 :12.
- [247] C. Bambi, Towards the use of the most massive black hole candidates in active galactic nuclei to test the Kerr paradigm, *Physical Review D*, **85** (2012a), 043001:1-043001:11.
- [248] C. Bambi, A Code to Compute the Emission of Thin Accretion Disks in Non-Kerr Spacetimes and Test the Nature of Black Hole Candidates, *The Astrophysical Journal*, **761** (2012b), 174:1-174 :9.
- [249] C. Bambi, Probing the spacetime geometry around black hole candidates with the resonance models for high-frequency QPOs and comparison with the continuum-fitting method, *Journal of Cosmology and Astroparticle Physics*, **09** (2012c), 014:1-014:17.
- [250] C. Bambi, Testing the spacetime geometry around black hole candidates with the analysis of the broad K $\alpha$  iron line, *Physical Review D*, **87** (2013), 023007:1-023007:12.
- [251] H. Krawczynski, Tests of General Relativity in the Strong-gravity Regime Based on X-Ray Spectropolarimetric Observations of Black Holes in X-Ray Binaries, *The Astrophysical Journal*, **754** (2012), 133:1-133:12.
- [252] M. J. Valtonen, M. Kidger, H. Lehto and G. Poyner, The structure of the October/November 2005 outburst in OJ287 and the precessing binary black hole model, *Astronomy and Astrophysics*, **477** (2008a), 407-412.

- [253] M. J. Valtonen, OJ287: a binary black hole system, *Rev. Mex. Astron. y Astrofis. (Ser. Conf., The Nuclear Region, Host Galaxy and Environment of Active Galaxies, Huatulco, Mexico, April 18-20, 2007)*, **32** (2008), 22-24.
- [254] M. J. Valtonen et al., A massive binary black-hole system in OJ287 and a test of general relativity, *Nature*, **452** (2008b), 851-853.
- [255] H. J. Lehto and M. J. Valtonen, OJ287 outburst structure and a binary black hole model, *Astrophys. J.*, **460** (1996), 207-213.
- [256] M. J. Valtonen and P. Pihajoki, A helical jet model for OJ287, *Astronomy and Astrophysics*, **557** (2013), A28:1-A38:5.
- [257] B. Sundelius, M. Wahde, H. J. Lehto and M. J. Valtonen, A numerical simulation of the brightness variations of OJ287, *Astrophys. J.*, **484** (1997), 180-185.
- [258] P. B. Ivanov, I. V. Igumenshchev and I. D. Novikov, Hydrodynamics of Black Hole-Accretion Disk Collision, *Astrophys. J.*, **507** (1998), 131-144.
- [259] M. J. Valtonen, S. Ciprini and H. J. Lehto, On the masses of OJ287 black holes, *Mon. Not. RAS*, **427** (2012), 77-83.
- [260] H. Seta, et al., Multi-Wavelength Observations of OJ 287 during the Periodic Optical Outburst in 2007, *Pub. Ast. Soc. Japan*, **61** (2009), 1011-1022.
- [261] M. J. Valtonen, New orbit solutions for the precessing binary black hole model of OJ287, *Astrophys. J.*, **659** (2007), 1074-1081.
- [262] N. I. Shakura and R. A. Sunyaev, Black holes in binary systems. Observational appearance, *Astronomy and Astrophysics*, **24** (1973), 337-355.
- [263] P. J. Sakimoto and F. V. Corotini, Accretion disk models for QSOs and active galactic nuclei -The role of magnetic viscosity, *Astrophys. J.*, **247** (1981), 19-31.
- [264] A. R. King, J. E. Pringle and M. Livio, Accretion disc

- viscosity: how big is alpha?, *Mon. Not. RAS*, **376** (2007), 1740–1746.
- [265] M. J. Valtonen, H. J. Lehto, L. O. Takalo and A. Sillanpää, Testing the 1995 Binary Black Hole Model of OJ287, *Astrophys. J.*, **729** (2011b), 33–38.
- [266] M. J. Valtonen and K. Wiik, Optical polarization angle and VLBI jet direction in the binary black hole model of OJ287, *Mon. Not. RAS*, **421** (2012), 1861–1867.
- [267] H. Baumgardt, A. Gualandris and S. Portegies Zwart, Ejection of hypervelocity stars from the Galactic Centre by intermediate-mass black holes, *Mon. Not. RAS*, **372** (2006), 174–182.
- [268] T. Matsabayashi, J. Makino and T. Ebisuzaki, Orbital evolution of an IMBH in the Galactic Nucleus with a Massive Central Black Hole, *Astrophys. J.*, **656** (2007), 879–896.
- [269] M. Iwasawa, S. An, T. Matsabayashi, Y. Funato and J. Makino, Eccentric Evolution of Supermassive Black Hole Binaries, *Astrophys. J. Lett.*, **731** (2011), L9–L13.
- [270] M. J. Valtonen et al., Measuring the Spin of the Primary Black Hole in OJ287, *Astrophys. J.*, **709** (2010a), 725–732.
- [271] L. E. Kidder, Coalescing binary systems of compact objects to (post)5/2-Newtonian order. V. Spin effects, *Physical Review D*, **52** (1995), 821–847.
- [272] T. Mora and C. M. Will, Post-Newtonian diagnostic of quasiequilibrium binary configurations of compact objects, *Physical Review D*, **69** (2004), 104021:1–104021:25.
- [273] S. Kumar and J. E. Pringle, Twisted accretion discs – The Bardeen-Petterson effect, *Mon. Not. RAS*, **213** (1985), 435–442.
- [274] G. Lodato and J. E. Pringle, The evolution of misaligned accretion discs and spinning black holes, *Mon. Not. RAS*, **368** (2006), 1196–1208.
- [275] R. Hudec, M. Basta, P. Pihajoki and M. Valtonen,

The historical 1900 and 1913 outbursts of the binary blazar candidate OJ287, *Astronomy and Astrophysics*, **559** (2013), A20:1-A20:9.

- [276] M. J. Valtonen, The OJ287 binary model and the expected outburst in November 1995, *Tuorla Obs. Rep. (Workshop on Two Years of Intensive Monitoring of OJ287 and 3C66A, Oxford, England, September 11-14, 1995)*, **176** (1996), 64-72.
- [277] M. J. Valtonen et al., Tidally Induced Outbursts in OJ 287 during 2005-2008, *Astrophys. J.*, **698** (2009), 781-785.
- [278] A. Einstein, Näherungsweise Integration der Feldgleichungen der Gravitation, *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, Berlin (1916), 688-696.
- [279] A. Einstein, Über Gravitationswellen, *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, Berlin (1918), 154-167.

# General Relativity and dragging of inertial frames

**Abstract:** Dragging of inertial frames, or frame-dragging, is an intriguing and fascinating phenomenon predicted by Einstein's theory of General Relativity with relevant astrophysical effects in the vicinities of rotating black holes. The first part of this paper is devoted to the theory of this phenomenon and the second part is devoted to its experimental confirmations. We first discuss the relation of frame-dragging to the riddle of the origin of inertia and inertial forces ([Section 1.1](#)). Frame-dragging can be described, in a weak gravitational field with slow motions, by a useful *formal* analogy with electrodynamics and is for this reason often called gravitomagnetism ([Sections 1.2](#) and [1.3](#)). To shed light on the riddle of the origin of inertia, it is interesting to study the behaviour of test particles, gyroscopes, and light and clocks inside a rotating sphere ([Section 1.4](#)). The phenomena caused by a rotating body or by a current of mass on the flow of time and clocks are especially fascinating ([Section 1.5](#)). Time-delay in the propagation of photons inside a rotating sphere might, in principle, be observed by gravitational lensing and by the rotation of an external cluster of galaxies ([Sections 1.6](#) and [1.6.1](#)). In principle, if measured, that effect might give information about the nature of dark matter and dark energy. In a strong gravitational field, frame-dragging may be defined with the use of curvature invariants ([Section 1.7](#)). Then, after briefly describing the past experimental triumph of General Relativity ([Section 2.1](#)), we describe the need to further test General Relativity, alternative theories and unified theories, including gravitational interaction, and the enigma of dark energy ([Sections 2.2](#) and [2.3](#)). We finally present the limits set on string theories, of Chern-Simons type, by using the current

measurements of frame-dragging with the LAGEOS and LAGEOS 2 satellites, and the future improvement of these limits with the LARES satellite ([Section 2.4](#)). In the second part we review the previous experimental tests of frame-dragging which were obtained using the LAGEOS satellites and with the dedicated Gravity Probe B space mission which reported an accuracy of about 10% and 19% in their measurements in 2004–2011 and 2011, respectively ([Section 3](#)). To confirm the robustness of the results obtained with the LAGEOS satellites, we report similar results for the measurement of frame-dragging which were obtained *independently* by three different groups: the Universities of Salento, Maryland and Rome “Sapienza”; the University of Texas at Austin; and GFZ Potsdam. The three groups used in their analyses the three independent orbital programs GEODYN, UTOPIA and EPOS-OC, respectively. Finally, we present the LARES space experiment, which together with LAGEOS, LAGEOS 2, and the determinations of the Earth’s gravitational field by GRACE will improve the accuracy

**Ignazio Ciufolini:** Dip. Ingegneria dell’Innovazione, Università del Salento, Lecce, and Centro Fermi, Rome, Italy

of the test of frame-dragging by approximately an order of magnitude and the limits on the string theories of Chern-Simons type using laser-ranged satellites ([Section 4](#)). We describe the space mission with the LARES satellite designed to minimise the orbital effects of non-gravitational perturbations ([Section 4.1](#)). The first few months of LARES observations and orbital analyses confirmed that, due to its special design, the orbit of LARES is closer to the theoretical motion of a test particle predicted by General Relativity, *i.e.* geodesic motion, than that of any other artificial satellite ([Sections 4.2 and 4.3](#)). Finally, we briefly report the results of extensive error analyses and of one hundred Monte Carlo simulations confirming an error budget of approximately 1% in the test of frame-dragging with the LARES

space experiment ([Section 4.4](#)). The section headings in square brackets cover technical topics in General Relativity, whereas the other sections are more general.

## 1 Frame-dragging: the theory

### 1.1 Dragging of inertial frames and the origin of inertia

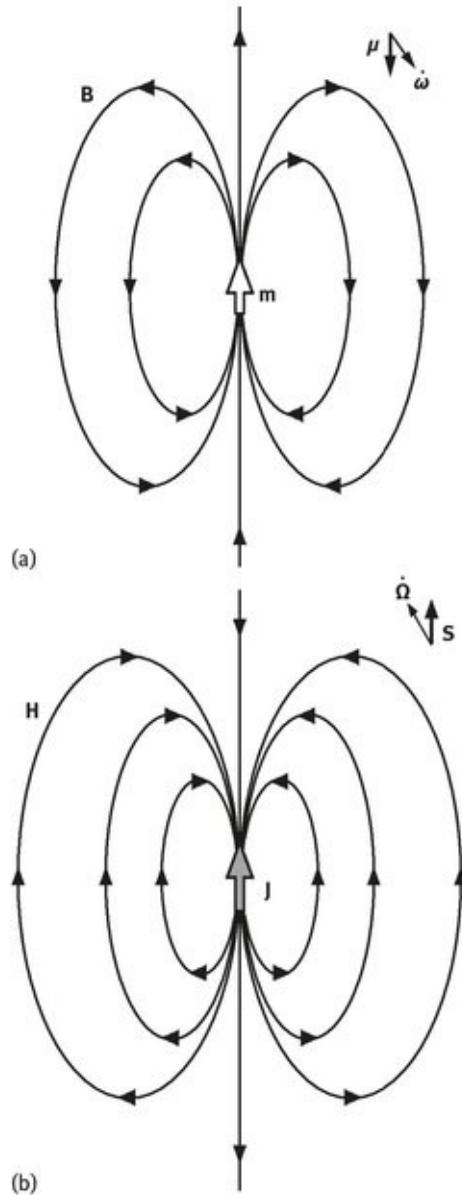
Frame-dragging is a fascinating and intriguing phenomenon predicted by Einstein's theory of General Relativity [1-3]. Inertial frames are systems in which any isolated body, not subject to any interaction, is at rest or moving with constant velocity. Inertial and centrifugal forces, such as the pull and push which we feel when our vehicle accelerates, brakes and turns, arise because of changes in velocity relative to uniformly moving inertial frames. However, what determines an inertial frame? And, in general, what is the origin of inertia? In the mechanics of Galileo Galilei and Isaac Newton, an inertial frame has an absolute existence, uninfluenced by the matter in the Universe. Therefore, the inertial forces arise because of acceleration relative to some absolute frame independent of the cosmological matter. An opposite view related inertial forces to acceleration relative to the cosmological masses; this is known as Mach's principle [3, 4]: Mach thought that centrifugal and inertial forces are due to rotations and accelerations with respect to all the masses and "fixed stars" in the Universe. An echo and partial realisation of the latter idea can be found in Einstein's theory of General Relativity [1-3], which predicts that a spinning mass will 'drag' inertial frames along with it. In Einstein's gravitational theory, the inertial and centrifugal forces are due to our accelerations and rotations with respect to the so-called local inertial frames which, in turn, are determined, influenced and dragged by the distribution and flow of mass-energy in the Universe. In particular, they are dragged by the motion and rotation of nearby matter; this general relativistic phenomenon is called frame-dragging ([7], and see page 544 of

[1] for example) and represents in Einstein's gravitational theory the remnant of the ideas of Ernst Mach on the origin of inertia [8].

## **1.2 Dragging of inertial frames and the gravitomagnetic analogy**

Frame-dragging phenomena, which are due to mass currents and mass rotation, have been called gravitomagnetism [3, 9] because of a *formal* analogy of electrodynamics and General Relativity in the weak gravitational field and slow motion approximation. In General Relativity, freely falling test-gyroscopes, *i.e.* sufficiently small and accurate spinning tops, determine the axes of the local, non-rotating, inertial frames [1–3], where the equivalence principle holds – *i.e.* where the gravitational field is locally ‘unobservable’ and all the laws of physics are the laws of Special Relativity theory [2]. Therefore, if we rotate with respect to these gyroscopes, we feel centrifugal forces, even though we may not rotate at all with respect to “distant stars”, contrary to our everyday intuition. Indeed, a gyroscope is dragged by spinning masses, *i.e.* its orientation changes with respect to “distant stars”. Whereas an electric charge generates an electric field and a current of electric charge generates a magnetic field, in Newtonian gravitational theory the mass of a body generates a gravitational field but a current of mass, for example the rotation of a body, would not generate any additional gravitational field. On the other hand, Einstein's gravitational theory predicts that a current of mass would generate a gravitomagnetic field which would exert an additional force on surrounding bodies and would change the structure of spacetime by generating additional curvature [3, 10]. A gravitomagnetic field generates the frame-dragging of a gyroscope in a similar way to a magnetic field producing the change of the orientation of a magnetic dipole (magnetic needle). Indeed, in General Relativity, a current of mass in a loop (*i.e.* a gyroscope) behaves in a formally similar way to that of a magnetic dipole, which is made of an electric current in a

loop ([Figure 1](#)), in electrodynamics. [Figure 1 \(a\)](#) shows the magnetic field  $\mathbf{B}$  generated by a magnetic dipole  $\mathbf{m}$  and a test magnetic dipole  $\boldsymbol{\mu}$ , i.e. a magnetic needle, and its precession. [Figure 1 \(b\)](#) shows the gravitomagnetic field  $\mathbf{H}$  [9] generated by the spin  $\mathbf{J}$  of a central body and the frame-dragging  $\dot{\Omega}$  of a test gyroscope  $\mathbf{S}$ .



**Fig. 1. Magnetism and gravitomagnetism.** (a) – The magnetic field  $\mathbf{B}$  generated by a central electric current distribution with magnetic moment  $\mathbf{m}$  and its effect on a magnetic needle (magnetic dipole)  $\boldsymbol{\mu}$ . (b) – The gravitomagnetic field  $\mathbf{H}$  generated by the angular momentum  $\mathbf{J}$  of a central body and its effect on a test gyroscope  $\mathbf{S}$ .

## 1.3 The gravitomagnetic formal analogy of General Relativity with electrodynamics

In electrodynamics, the Maxwell-Ampère equation for a stationary current distribution can be written:  $\Delta \mathbf{A} = -\frac{4\pi}{c}\mathbf{j}$ , in the Coulomb gauge:  $\nabla \cdot \mathbf{A} \equiv A_{,i}^i = 0$ , where  $\mathbf{A}$  is the vector potential and  $\mathbf{j}$  the electric current density. For a localised, stationary electric current distribution  $\mathbf{j}$ , far from the current, the lowest non-vanishing term of  $\mathbf{A}$  is then:  $\mathbf{A}(\mathbf{x}) \cong \frac{\mathbf{m} \times \mathbf{x}}{|\mathbf{x}|^3}$ , where  $\mathbf{m}$  is the magnetic moment of the current distribution. Therefore, for a localised current distribution, the lowest non-vanishing term of  $\mathbf{B}$  is the field of a magnetic dipole with dipole moment  $\mathbf{m}$ :

$$\mathbf{B} = \nabla \times \mathbf{A} \cong \frac{3\hat{\mathbf{x}}(\hat{\mathbf{x}} \cdot \mathbf{m}) - \mathbf{m}}{|\mathbf{x}|^3},$$

where  $\mathbf{m}$  is the magnetic dipole moment of the current distribution. Then, from the Lorentz force equation, given a current distribution with current density  $\mathbf{j}$ , the force on the current distribution due to  $\mathbf{B}$  is:  $\mathbf{F} = \frac{1}{c} \int \mathbf{j}(\mathbf{x}) \times \mathbf{B}(\mathbf{x}) d^3x$ . The torque on it due to  $\mathbf{B}$  is:  $\boldsymbol{\tau} = \frac{1}{c} \int \mathbf{x} \times (\mathbf{j} \times \mathbf{B}) d^3x$  and the lowest non-vanishing term of the torque on a localised, stationary current distribution is then the torque on a magnetic dipole with dipole moment  $\mu$ :  $\boldsymbol{\tau} = \mu \times \mathbf{B}(0)$ . For a classical electric current loop of particles of charge  $q$  and mass  $m$ , the dipole moment has then a precession:  $\dot{\omega} = -\frac{q\mathbf{B}}{2mc}$ ; whereas a different relation holds in quantum mechanics.

In the mechanics of Galilei and Newton, the external field of a rotating, uncharged spherical body is only given by its mass; in General Relativity the external field of such a body is instead given by its mass  $M$  and by its angular momentum  $\mathbf{J}$ . In the weak gravitational field and slow motion approximation of the Kerr metric, the spacetime geometry is then given in Boyer-Lindquist coordinates by [1, 3, 11]:

$$ds^2 \cong -\left(1 - \frac{2M}{r}\right)dt^2 + \left(1 - \frac{2M}{r}\right)^{-1}dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) - \frac{4J}{r}\sin^2\theta d\phi dt,$$

(1)

where in General Relativity we use geometrized units, i.e.  $G = c = 1$ .

The  $g_{\varphi t}$  component of the metric is called the gravitomagnetic [3, 9] potential.

In General Relativity [3], in the weak gravitational field and slow motion approximation, for a stationary, localised mass-energy distribution, the (0i) components of the Einstein field equation can be written in the Lorentz gauge:

$$\Delta h_{0i} \equiv 16\pi\rho v^i \quad (2)$$

(where in electrodynamics was:  $\Delta A^i = -\frac{4\pi}{c}\rho_e v^i$ ). With solution:

$$h_{0i}(\mathbf{x}) \equiv -4 \int \frac{\rho(\mathbf{x}')v^i(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} d^3x' \quad (3)$$

Then, far from the source [16, 17],  $\mathbf{h} \equiv (h_{01}, h_{02}, h_{03})$ , can be rewritten:

$$\mathbf{h}(\mathbf{x}) \equiv -2 \frac{\mathbf{J} \times \mathbf{x}}{|\mathbf{x}|^3} \quad (4)$$

For a slowly rotating sphere with angular momentum:  $\mathbf{J} \equiv (0, 0, J)$ , we have in spherical coordinates:

$$h_{0\phi} \equiv -\frac{2J}{r} \sin^2\theta \quad (5)$$

i.e. the  $g_{0\phi}$  component of the Kerr metric, in the weak field and slow motion approximation.  $\mathbf{h}$  is the gravitomagnetic potential. One can then define a gravitomagnetic field ([Figure 1](#)) [3, 9]:  $\mathbf{H} = \nabla \times \mathbf{h}$ :

$$\mathbf{H} = \nabla \times \mathbf{h} \equiv 2 \left[ \frac{\mathbf{J} - 3(\mathbf{J} \cdot \hat{\mathbf{x}})\hat{\mathbf{x}}}{|\mathbf{x}|^3} \right]$$

(6)

From these equations we see that, in General Relativity, in the weak field and slow motion limit, the angular momentum  $\mathbf{J}$  of a stationary, localised mass-energy current has a role formally similar to the magnetic dipole moment  $\mathbf{m}$  of a stationary, localised charge current in electrodynamics (the difference between electromagnetism and weak field General Relativity is an extra  $-4$  factor in the equations of General Relativity). Then, by using the geodesic equation  $\frac{Du^\alpha}{ds} = 0$ , in the weak field and slow motion limit one has:

$$m \frac{d^2 \mathbf{x}}{dt^2} \equiv m \left( \mathbf{G} + \frac{d\mathbf{x}}{dt} \times \mathbf{H} \right)$$

(7)

where  $\mathbf{G} \equiv -\frac{M}{|\mathbf{x}|^2} \hat{\mathbf{x}}$  is the standard Newtonian acceleration and  $\mathbf{H}$  is the gravitomagnetic field. Furthermore, as for electromagnetism, and so for General Relativity [3, 9], the torque acting on a gyroscope with angular momentum  $\mathbf{S}$ , in the weak field and slow motion approximation is:

$$\boldsymbol{\tau} \equiv \frac{1}{2} \mathbf{S} \times \mathbf{H} = \frac{d\mathbf{S}}{dt} \equiv \dot{\Omega} \times \mathbf{S}$$

(8)

Therefore, the gyroscope precesses with respect to an asymptotic inertial frame with angular velocity  $\dot{\Omega}$ :

$$\dot{\Omega} = -\frac{1}{2} \mathbf{H} = \frac{-\mathbf{J} + 3(\mathbf{J} \cdot \hat{\mathbf{x}})\hat{\mathbf{x}}}{|\mathbf{x}|^3}$$

(9)

where  $\mathbf{J}$  is the angular momentum of the central object. This is the dragging of gyroscopes or dragging of inertial frames, of which the gyroscopes define the axes. As for electromagnetism, the force exerted on the gyroscope by the gravitomagnetic field  $\mathbf{H}$  is:

$$\mathbf{F} = \left( \frac{1}{2} \mathbf{S} \cdot \nabla \right) \mathbf{H}$$

(10)

Finally, a central object with angular momentum  $\mathbf{J}$ , due to the second term in the force, drags the orbital plane (and the orbital angular momentum) of a test particle (which can be thought of as an enormous gyroscope) in the sense of rotation of the central body. This dragging of the whole orbital plane is described by the formula for the rate of change of the longitude of the nodal line (the intersection between the orbital plane of the test particle and the equatorial plane of the central object) which was published by Lense and Thirring in 1918 [12]:

$$\dot{\Omega}^{\text{Lense-Thirring}} = \frac{2\mathbf{J}}{a^3(1-e^2)^{3/2}}$$

(11)

where  $a$  is the semimajor axis of the test particle,  $e$  its orbital eccentricity and  $\mathbf{J}$  the angular momentum of the central body. Similarly, by integrating the equation of motion of the test particle, we find the formulas derived by Lense-Thirring [12] for the secular rates of change of the longitude of the pericentre  $\dot{\omega}$ , (determined by the Runge-Lenz vector) and of the mean orbital longitude  $\dot{L}_0$

$$\dot{\omega}^{\text{Lense-Thirring}} = \frac{2J}{a^3(1-e^2)^{\frac{3}{2}}} (\hat{\mathbf{J}} - 3 \cos I \hat{\mathbf{i}})$$

(12)

where  $\cos I \equiv \hat{\mathbf{J}} \cdot \hat{\mathbf{i}}$  and  $I$  is the orbital inclination (angle between the orbital plane and the equatorial plane of the central object). Therefore, the term containing  $\cos I$  drags the pericentre in the same sense as  $\hat{\mathbf{i}}$  when  $\hat{\mathbf{J}} = -\hat{\mathbf{i}}$ , and in the opposite sense when  $\hat{\mathbf{J}} = \hat{\mathbf{i}}$ , and:

$$\dot{L}_0^{\text{Lense-Thirring}} = \frac{2J}{a^3(1-e^2)^{\frac{3}{2}}} (1 - 3 \cos I)$$

(13)

Despite the beautiful and illuminating analogies between General Relativity and classical electrodynamics, the two theories are, of course, fundamentally different. Their fundamental difference is valid even for weak field, linearised gravitation, as it is unambiguously displayed by the equivalence principle.

## 1.4 Dragging of inertial frames inside a hollow sphere

According to the Mach principle, inertial and centrifugal forces arise because we accelerate and rotate with respect to all the masses in the universe and vice versa we should feel inertial and centrifugal forces if all the masses of the universe were rotating with respect to our local gyroscopes. So an interesting related question is what happens in General Relativity inside a rotating shell.

In General Relativity, inside a hollow, static, spherically symmetric distribution of matter, in a vacuum, we have the flat metric  $\eta_{\alpha\beta}$ . Therefore, in the weak field and slow motion approximation the metric inside a slowly rotating massive shell may be written:  $g_{\alpha\beta} \cong \eta_{\alpha\beta} + h_{\alpha\beta}$ , and we may then apply the results (2) and (3). Inside a thin shell of total mass  $M$  and radius  $R$ , rotating with angular velocity  $\boldsymbol{\omega} = (0, 0, \omega)$ , we thus have the field (3); furthermore, by integrating (3) inside a thin rotating shell with mass density:  $\rho(\mathbf{x}') = \frac{M}{4\pi R^2} \delta(R - r')$ , we have inside the shell:

$$\mathbf{h} \equiv (h_{0x}, h_{0y}, h_{0z}) = -\frac{4}{3} \frac{M}{R} \boldsymbol{\omega} \times \mathbf{x} = \left( \frac{4M}{3R} \omega y, -\frac{4M}{3R} \omega x, 0 \right)$$

[\(14\)](#)

By substituting the components of  $h_{\alpha\beta}$  inside the slowly rotating shell in the geodesic equation we then find the 1917 Thirring result [3, 13] for the acceleration of a test particle inside a rotating shell due to the rotation of the shell:

$$\begin{aligned}
\ddot{x} &= -\frac{8}{3} \frac{M}{R} \omega \dot{y} + \frac{4}{15} \frac{M}{R} \omega^2 x \\
\ddot{y} &= \frac{8}{3} \frac{M}{R} \omega \dot{x} + \frac{4}{15} \frac{M}{R} \omega^2 y \\
\ddot{z} &= -\frac{8}{15} \frac{M}{R} \omega^2 z
\end{aligned}$$

(15)

where the  $\omega^2$  terms are due to the other components of  $h_{\alpha\beta}$  and maybe thought of as a change in the inertial, and gravitational, mass of the shell due to the velocity  $\omega R$ . Therefore, due to the rotation of the shell, a test particle is affected by forces formally similar to the Coriolis and the centrifugal forces.

Furthermore, from Eqs. (6) and (9), we have that the axes of the local inertial frames, *i.e.* the gyroscopes, are dragged by the rotating shell, according to:

$$\dot{\Omega} \equiv -\frac{1}{2} \mathbf{H} = -\frac{1}{2} \nabla \times \mathbf{h} = \frac{4}{3} \frac{M}{R} \boldsymbol{\omega}$$

(16)

The 1966 Brill-Cohen solution [14] describes the metric inside a shell with arbitrary mass at the lowest order in the angular velocity. It is a lowest order series expansion in the angular velocity  $\omega$  of the shell on the Schwarzschild background of a spherical mass shell of arbitrary mass  $M$ , valid both inside and outside the shell. The Brill-Cohen metric is:

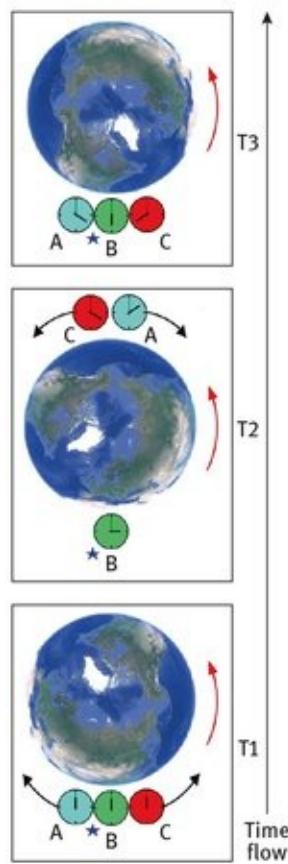
$$ds^2 = -H(r)dt^2 + J(r)[dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta(d\phi - \omega(r)dt)^2]$$

(17)

where  $H(r) = [(r - \alpha)/(r + \alpha)]^2$  for  $r > R$ ;  $H(r) = [(R - \alpha)/(R + \alpha)]^2$  for  $r \leq R$ ;  $J(r) = (1 + \alpha/r)^4$  for  $r > R$ ;  $J(r) = (1 + \alpha/R)^4$  for  $r \leq R$ ; and where  $\alpha = 2M$ , and  $R$ ,  $\omega$ , and  $M$  are respectively the radius, angular velocity and mass of the shell. For further details on this solution, see [5] and related papers in [4]. For other solutions inside a rotating shell to higher order in the angular velocity  $\omega$  and for detailed discussions on the interpretation of the forces inside a rotating shell, see [2, 4–6, 14, 15].

## 1.5 Frame-dragging phenomena on clocks and photons

In [Sections 1.2](#) and [1.3](#) we have seen that General Relativity predicts the occurrence of peculiar phenomena on gyroscopes and test-particles in the vicinities of a spinning body, caused by its rotation. However, Einstein's gravitational theory also predicts intriguing phenomena on clocks and photons around a spinning body.



**Fig. 2. Flow of time and frame-dragging.** Frame-dragging effects on clocks by a rotating mass. If two twins A and C fly around a spinning body, even very slowly, and a third twin, B, awaits them at the starting point, fixed relative to the “distant stars” (a “fixed star” is shown in blue, and T1, T2 and T3 are three consecutive instants of time), when they meet again, A, travelling in the direction opposite to the rotation of the central body, is younger relative to B. On the other hand, C, travelling in the direction of rotation of the body, is older with respect to B and to A [8, 16, 17] (Earth image: IBCAO and Landsat).

Frame-dragging indeed has an intriguing influence on the flow of time around a spinning body ([Figure 2](#)). The synchronisation

of clocks around a closed path near a spinning body is not possible [16, 17] in any rigid frame which is not rotating relative to the “fixed stars”, because light co-rotating around a spinning body would take less time to return to a starting point (fixed relative to the “distant stars”) than light rotating in the opposite direction [8, 16, 17].

On the one hand, the orbital period of a test particle orbiting around a spinning body in the same direction as the rotation of the body, *i.e.* “co-rotating” with the central object, is longer than the orbital period of a particle orbiting at the same distance but in the opposite direction, *i.e.* “counter-rotating” with respect to the spin of the central object. The difference between co-rotating and counter-rotating orbital periods is:

$$\Delta\tau = 4\pi \frac{J}{M}$$

(18)

Nevertheless, a photon co-rotating around a spinning body takes less time to return to a “fixed point” (with respect to “distant stars”) than a photon rotating in the opposite direction; in the Kerr metric [11], characterised by the mass and angular momentum of the central object, a fixed point can be determined by constant Boyer-Lindquist spatial coordinates, *i.e.* by constant spatial coordinates of the weak field and slow motion metric (1) of [Section 1.3](#); operationally a fixed point can be realised with a small telescope always pointing toward the same distant star, always oriented toward the centre of the spinning body and at the same distance from it, by using gradiometers and rockets attached to the telescope. For example, around the spinning Earth the difference between the travel-time of two pulses of electromagnetic radiation counter-propagating in the same circuit would be  $\sim 10^{-16}$  seconds [18]. Since light rays are used to synchronise clocks, the different travel-time of co-rotating and counter-rotating photons implies the impossibility of the synchronisation of clocks around a closed path around a spinning body; the behaviour of light rays

and the behaviour of clocks around a spinning body are intimately connected. Let us then briefly analyse the behaviour of clocks around a spinning object.

In [Figure 2](#) we show the frame-dragging effects on clocks by a rotating mass. For this time dilation, which is due to the spin of the central body, to occur, the clocks, or twins, would not need to move near the speed of light (as in the case of the well-known ‘twin-paradox’ of Special Relativity). For example, when two such twins meet again after a full revolution, having flown arbitrarily slowly around the whole Earth in opposite directions on the equatorial plane and exactly at the same altitude, the difference in their ages due to the Earth’s spin would be approximately  $10^{-16}$  seconds – for an altitude of about 6000 km (see the formula here below) – which would in principle be detectable if not for the other, much larger, relativistic clock effects. Nevertheless, frame-dragging does produce relevant effects on light and matter around a rotating black hole [9].

In several papers the “frame-dragging clock effect” around a spinning body has been estimated and space experiments have been proposed to test it [18–21]. To synchronise clocks around a path in a stationary field, we can use light rays or even very slowly moving clocks (so that the special relativistic time-dilation is a higher order effect) which are always at the same distance from the central spherically symmetric spinning body (so that the mass time-dilation is the same for both clocks). Thus, when a clock, co-rotating very slowly (using rockets) around a spinning body and at a constant distance from it, returns to its starting point, it finds itself advanced relative to a clock kept there at “rest” (with respect to “distant stars”). Similarly, a clock, counter-rotating arbitrarily slowly and at a constant distance around the spinning body, finds itself retarded relative to the clock at rest at its starting point. For example, when a clock which co-rotates very slowly around the spinning Earth at an altitude of  $\sim 6000$  km returns to its starting point, it finds itself advanced relative to a clock kept there at “rest” (with respect to “distant stars”) by  $\oint \frac{g_{0i}}{g_{00}} dx^i \sim \frac{4\pi f R}{r} \sim 5 \times 10^{-17}$  seconds, where

$g_{0i} \sim \frac{2J_\oplus}{r^2}$  is the Earth gravitomagnetic field and  $J_\oplus \approx 145 \text{ cm}^2$  is the Earth angular momentum in geometrised units, *i.e.* with the gravitational constant,  $G$ , and the speed of light,  $c$ , equal to one:  $G \equiv c \equiv 1$ . Similarly, a clock, which counter-rotates very slowly around the spinning Earth, finds itself retarded relative to a clock kept there at “rest” by the same amount. Nevertheless, a larger clock effect, of the order of  $10^{-7}$  seconds, has been estimated in [19]; let us explain this apparent disagreement. The orbital period of a particle, or clock, freely co-orbiting (along a geodesic) around a spinning body is longer than the orbital period of a particle, or clock, freely counter-orbiting on the same path [19], see formula (18). The difference between the two orbital periods, *i.e.* the difference between the two times read by a clock at a fixed point (with respect to “distant stars”, see above) when the two counter-rotating particles come back at this point after one revolution, is  $\sim 4\pi J/M$ , *i.e.* around the spinning Earth, is  $\sim 1.4 \times 10^{-7} \text{ s}$  [19, 20]; this is basically the effect derived in [19]. Nevertheless, the difference between the time read by the two clocks when they meet again after a whole revolution is still  $\sim 10^{-16}$  [18, 20].

In Einstein’s general theory of relativity, all these phenomena, on test particles, gyroscopes, photons, and clocks, are the result of the rotation of the central mass.

## 1.6 Frame-dragging, time-delay and gravitational lensing

The time-delay in the propagation of photons gravitationally lensed by a central spinning body was treated in [18, 22].

Null geodesics in the Kerr metric, also in regard to gravitational lensing and the image’s position, polarisation and intensification, distortion, and optical caustic, have been studied in several papers (see: [23, 24] and references therein).

By assuming that the source of electromagnetic radiation and the observer are very far from the deflecting body, so that their distance from the central body is much higher than the impact parameter,  $b$ , *i.e.* of the distance of closest approach of the

photons to the spinning deflecting body, the relative time-delay by frame-dragging,  $\Delta t_J$ , in the arrival time of photons gravitationally lensed by a central body with angular momentum  $J$  is then [18]:

$$\Delta t_J \sim \frac{4J}{b}$$

(19)

where the precise amount of this time-delay depends on the geometry of the paths followed by the photons with respect to the orientation of the spin of the central deflecting body (see: [18]).

Whereas, the deflection of the photons' path due to the spin of the deflecting body is:

$$\delta_J \sim \frac{4J}{b^2}$$

(20)

where the precise amount of deflection also depends on the geometry of the system (see: [18]).

The time-delay of photons by frame-dragging of a deflecting spinning body could in principle be measured if other biasing effects could be taken into account. For example, in the case of an impact parameter  $b$  nearly equal to the Sun equatorial radius, the time-delay due to the Sun's angular momentum could amount to  $\sim 10^{-11}$  sec. The time-delay due to the Sun's angular momentum could then, in principle, be detected using a laser-interferometer around the Sun; this would consist of a source and a detector on opposite sides of the Sun. The source, a laser, would emit, at the same time, photons toward the Sun but with slightly different angles so that they would travel on opposite sides of the Sun (i.e. photons co-rotating and counter-rotating with respect to the Sun's rotation). Then, there would be a relative time-delay in the arrival time of the photons due to the Sun's rotation. Of course all the other travel time-delays should be modelled and removed from the observed delays, in

particular the time-delay due to the dispersion of electromagnetic waves by the solar plasma.

The time-delay due the lensing galaxy of the Einstein cross Q2237+031, [25, 26], by assuming a simple model for the rotation and shape of the central body, might amount to a few minutes. In the case of a system with a tiny angular separation between the images such as B0218+357, it is measured a tiny relative time-delay between the images of about 10.5 days with an uncertainty of less than 10 hours. At least in principle, for such a system, one could measure the time-delay due to the angular momentum of the lensing galaxy. Of course, from the observed delays between the images one should be able to accurately enough model and remove all the other time-delays due to other physical effects, including the geometrical time-delay and that due to the quadrupole moment of the central deflecting body (see: [18]).

As a third example, the relative time-delay of photons due to the spin of a typical cluster of galaxies, depending on the geometry of the system and on the path followed by the photons, has been calculated to range from a few minutes to several days [18].

### **1.6.1 Time-delay inside a slowly rotating massive shell**

In [Section 1.4](#) we have described the frame-dragging effects on gyroscopes inside a rotating shell. However there is also a spin-time-delay, and an additional deflection, due to the rotation of the external mass [27].

Inside a thin shell of mass  $M$  and radius  $R$ , rotating with slow angular velocity  $\boldsymbol{\omega} = (0, 0, \omega)$ , the  $g_{0i} \cong h_{0i}$  components of the metric tensor are given by [Eq. \(14\)](#). Therefore, inside a rotating shell it is not possible to synchronise clocks around a closed path. Indeed, if we consider a clock co-rotating very slowly along a circular path with radius  $r$  ( $r < R$ ), when it comes back to its starting point it is advanced with respect to a clock kept there at rest (with respect to “distant stars”). The difference between the time read by the co-rotating clock and the clock at

rest is given in [18]. Indicating with  $\hat{\mathbf{r}}_1$  and  $\hat{\mathbf{r}}_2$  the unit vectors from the centre of a spherical shell to the two points on the shell where a light ray respectively enters and leaves the sphere, we have the spin-time-delay due to the rotation of the shell [18]:

$$\Delta_{GM} = -\frac{4MR}{3}\boldsymbol{\omega} \cdot (\hat{\mathbf{r}}_1 \times \hat{\mathbf{r}}_2), \quad (21)$$

A general expression of the relative spin-time-delay due the rotation of an external mass between two photons travelling inside the mass (a) with different impact parameters,  $b_1$  and  $b_2$ , and (b) for any finite thickness of the external shell, is given in [18].

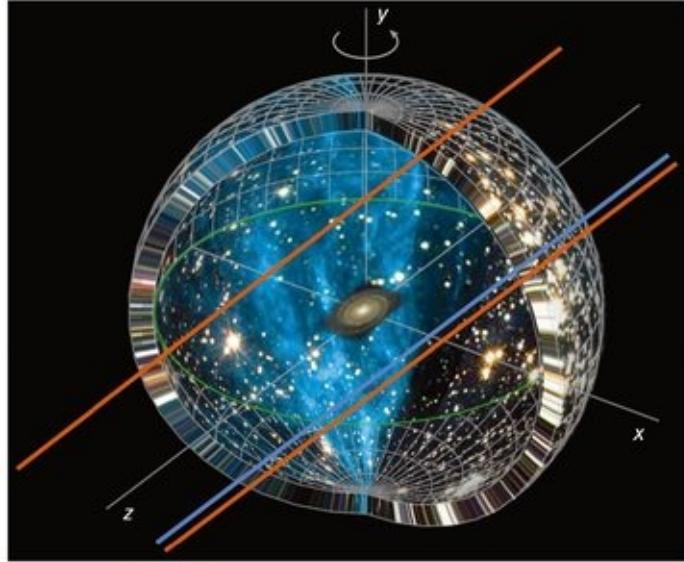
For a lensing galaxy of the type of the lens of the “Einstein cross” [25], the relative time-delay, due to the rotation of the external mass, between two photons travelling at a distance of  $b_1$  ( $b_1 \approx 650$  pc) and  $b_2$  ( $b_2 \approx -650$  pc) from the centre has been estimated to be of the order of several minutes [18].

If we consider two light rays deflected by a lensing galaxy which is inside a rotating cluster, or super-cluster of galaxies, we can also estimate the amount of time-delay due to the spin of the mass rotating around the deflecting galaxy. To get an order of magnitude of the time-delay, we can use typical super-cluster parameters [28]. By considering a lensing galaxy in the centre of the cluster and light rays with impact parameters  $b_1 \approx 15$ kpc and  $b_2 \approx -15$ kpc (of the order of the Milky Way radius) and by assuming for simplicity a constant mass density the spin time-delay has been estimated to be of the order of a few days.

In general, it has been estimated [18] that if the lensing galaxy is not at the centre of the cluster, the relative spin-time-delay between two photons deflected by the galaxy, which are propagating inside a rotating cluster or super-cluster of galaxies, may, under special conditions, be as large as several years.

In conclusion, the “spin-time-delay” due to the angular momentum of a body and experienced by the photons of two or

more images of a source observed at a far point by gravitational lensing, is caused by the propagation of the photons in opposite directions with respect to the direction of the spin of the rotating body. The spin-time-delay can be caused by a central rotating mass or by the external mass of a rotating shell.



**Fig. 3.** Time-delay and gravitational lensing. The path of photons propagating inside a rotating mass-energy distribution, such as a cluster of galaxies. The red and blue lines represent photons which, observed by a far detector by gravitational lensing, may have a relative delay in the arrival time at the observer due to frame-dragging by the external rotating mass.

In summary:

The spin-time-delay may need to be taken into account in the modelling of the relative time-delays between images observed by gravitational lensing, in addition to other time-delays such as the geometrical time-delay and the delay due to the quadrupole moment of the lensing body.

If other smaller time-delays could be modelled accurately enough and removed from the observations, the large relative delay due to the quadrupole moment of the lensing body could be removed for some configurations of the images by using special combinations of observables [18, 27, 29]. With this method, one could measure the spin-time-delay due to the rotation of a mass.

The measurement of the spin-time-delay might, in principle,

be a new observable for the determination of the total mass of a rotating body, *i.e.* of the dark matter content of objects such as galaxies, and clusters and super-clusters of galaxies [18]. It might, in principle, also be useful to get some insight on dark energy.

Depending on the geometry of the system, the relative spin-time-delay can be quite a large effect and may then be detected on Earth, in particular in systems with a tiny angular separation and a tiny relative time-delay between the images, such as asas B0218+357. Indeed, this system has a time-delay between the images of about 10.5 days with an estimated uncertainty of less than 10 hours [18, 30, 31].

## **1.7 [An invariant characterisation of frame-dragging]**

In electromagnetism, in a frame where a test-particle with an electric charge is at rest we only observe an electric field  $E^k$  but no magnetic field, however, in a frame which is moving relative to the charge we also measure a magnetic field  $B^k$ . In General Relativity, in a similar way, in a suitable frame where a non-rotating mass is at rest, the components of the gravitomagnetic vector potential  $h^k$  are zero. Nevertheless, if we consider an observer moving relative to the mass, in a local frame moving with the observer, the components of the gravitomagnetic vector potential  $h^k$  can be non-zero. Indeed, in a frame where a non-rotating mass  $M$  is at rest, the only components of the Schwarzschild metric  $g_{\alpha\beta}$  different from zero (written in standard Schwarzschild coordinates) are:

$g_{00} = -g_{rr}^{-1} = -(1 - \frac{2GM}{c^2 r})$ ,  $g_{\theta\theta} = r^2$  and  $g_{\phi\phi} = r^2 \sin^2 \theta$  and the three non-diagonal components of the metric  $g_{0k}$ , *i.e.* the components of the “gravitomagnetic vector potential”  $h^k$ , are zero. Nevertheless, if we perform a local Lorentz transformation with velocity  $v^k$  relative to the mass  $M$ , the components of the gravitomagnetic vector potential  $g_{0k}$  are in general non-zero in the new frame. The orbital effects of this gravitomagnetic vector potential, arising from the motion of the Earth-Moon system relative to

the Sun's mass, have been observed by Lunar Laser Ranging (LLR) since the first measurements of the geodetic precession of the Moon orbit, *i.e.* of the Earth-Moon “gyroscope” moving around the Sun [32]; we have proposed that these be called: “frame-dependent gravitomagnetic effects” [10]. On the other hand, the angular momentum  $\mathbf{J}$  of a body generates a gravitomagnetic field and produces spacetime curvature which cannot be eliminated by a simple change of frame of reference or by a coordinate transformation. This gravitomagnetic field generates the Lense-Thirring effect on the orbit of the LAGEOS satellites; we have proposed that this be called: “intrinsic gravitomagnetic effect” [10, 33].

In order to distinguish between “intrinsic” gravitomagnetic effects (such as the Lense-Thirring effect) and “frame-dependent” ones (such as the geodetic precession), we have proposed that spacetime curvature invariants be used [3, 33]. In general, one cannot derive intrinsic gravitomagnetic effects from frame-dependent ones unless making additional theoretical hypotheses, such as the linear superposition of the frame-dependent gravitomagnetic effects; for example, the magnetic field generated by the intrinsic magnetic moment (Bohr magneton) is an intrinsic phenomenon due to the intrinsic spin of a particle which cannot be explained and derived as a frame-dependent effect by any Lorentz and frame transformation.

In electromagnetism, in order to characterise the electromagnetic field, using the electromagnetic field Lorentz-tensor  $F_{\alpha\beta}$  we can build the scalar Lorentz-invariant

$*\mathbf{F} \cdot \mathbf{F} \equiv \frac{1}{4} F_{\alpha\beta}^* F^{\alpha\beta} = \mathbf{E} \cdot \mathbf{B}$ , where  $*F^{\alpha\beta}$  is the dual of  $F^{\alpha\beta}$ , defined as:  $*F^{\alpha\beta} = \frac{1}{2} \epsilon^{\alpha\beta\mu\nu} F_{\mu\nu}$  and  $\epsilon^{\alpha\beta\mu\nu}$  is the Levi-Civita pseudotensor (that is equal to  $\pm \sqrt{-g}$ , *i.e.* plus the square root of minus the determinant,  $g$ , of the metric, if the indices are even permutations of  $(0,1,2,3)$ ,  $-\sqrt{-g}$  for odd permutations of  $(0,1,2,3)$  and 0 if any indices are repeated).  $*\mathbf{F} \cdot \mathbf{F}$  is an invariant for Lorentz transformations (precisely a pseudo-invariant under coordinate reflections), *i.e.* it is either null or not

in every inertial frame. For example, in the rest frame of a test-particle with charge  $q$  we have an electric field only and no magnetic field, and this invariant is zero—therefore even in a frame moving relative to  $q$ , where both  $\mathbf{B} \neq \mathbf{0}$  and  $\mathbf{E} \neq \mathbf{0}$ , this invariant is zero. However, in a frame where a charge  $q$  and a magnetic dipole  $\mathbf{m}$  are at rest, we have in general  ${}^*\mathbf{F} \cdot \mathbf{F} \neq 0$  and therefore this invariant is non-zero in any other inertial frame.

In General Relativity, the gravitomagnetic “vector” potential  $h^k$  can be zero or not depending on the frame where it is calculated. Nevertheless, the spacetime curvature of a manifold is a coordinate-independent quantity [1–3]. Therefore, in order to test for intrinsic gravitomagnetic effects, *i.e.* independent of the coordinate system (and not eliminable with a coordinate transformation) we have to use the Riemann curvature tensor  $R_{\alpha\beta\mu\nu}$ . The Riemann tensor can be split in a magnetic-like part and an electric-like part, however, the magnetic part of the Riemann tensor is “frame-dependent”, for example in empty space one can find observers for which the magnetic part of the Riemann tensor (*i.e.* of the Weyl tensor in a vacuum) is null and observers for which it is different from zero—see, *e.g.* the treatment of gravitomagnetism in [34]. Therefore, since the magnetic components of the Riemann tensor are frame-dependent, we have proposed the use of curvature invariants built out of the Riemann tensor [3, 33].

Given a metric  $g_{\alpha\beta}$  in a coordinate system (with or without the so-called “magnetic” components  $g_{0k}$ ), in a way similar to electromagnetism, using the Riemann curvature tensor  $R_{\alpha\beta\mu\nu}$  we can build the spacetime curvature invariant  ${}^*\mathbf{R} \cdot \mathbf{R} \equiv {}^*R^{\alpha\beta\mu\nu}R_{\alpha\beta\mu\nu}$ , called the Pontryagin invariant, where  ${}^*R^{\alpha\beta\mu\nu} \equiv \frac{1}{2}\epsilon^{\alpha\beta\sigma\rho}R_{\sigma\rho}^{\mu\nu}$  is the dual of  $R_{\alpha\beta\mu\nu}$  [3].

The Pontryagin invariant plays an key role in the Chern-Simons gravitational theory and in String theories of Chern-Simons type. The Chern-Simons theory is not described by the standard Post-Newtonian Parametrized parameters and predicts frame-dragging effects additional to General Relativity. See

below, [Section 2.4](#), for some limits put on the String theories of Chern-Simons type using LAGEOS, Gravity Probe B and LARES.

In [3] the exact explicit expression of the Riemann curvature invariant  $*\mathbf{R} \cdot \mathbf{R}$  is given for some spacetime solutions of the Einstein field equation, *i.e.* in the case of the Kerr metric generated by the angular momentum  $J$  and mass  $M$  of a rotating body and in the case of the Schwarzschild metric generated by the mass only of a non-rotating body with  $J = 0$ . It turns out that  $*\mathbf{R} \cdot \mathbf{R}$  is in general different from zero if and only if  $J \neq 0$ , *e.g.* it is different from zero in the case of the Kerr metric with  $J \neq 0$  but it is zero in the case of the Schwarzschild metric, with  $J = 0$ . In the case of Earth with angular momentum  $J_{\oplus}$ , the invariant  $*\mathbf{R} \cdot \mathbf{R}$  is at the lowest order:  $*\mathbf{R} \cdot \mathbf{R} \sim \frac{G^2 J_{\oplus} M_{\oplus}}{c^5 r^5}$ , thus the Lense-Thirring effect on the LAGEOS satellites is an intrinsic gravitomagnetic effect [3, 33] which cannot be eliminated by a change of frame of reference—indeed, as measured in any frame, the invariant  $*\mathbf{R} \cdot \mathbf{R}$  is different from zero.

## 2 The need to further test General Relativity

### 2.1 The universe and the triumph General Relativity

The current study of the universe and nature has on the one hand allowed us to understand some of the basic laws governing the infinitely small, down to spatial distances of the order of  $10^{-16}$  cm or less corresponding to quarks, and on the other hand to observe and in part understand the evolution of the universe up to spatial distances of the order of  $10^{28}$  cm corresponding to the position of some quasars and near the so-called big-bang. With regard to the time scale, the current study of the universe ranges from a few instants after the big-bang up to the present time, approximately 14 billion years later. General Relativity has a key role in the study of the universe. The gravitational interaction and the evolution of the universe are indeed successfully described, both theoretically and

experimentally, by General Relativity [1]. Gravitation is described in Einstein's theory as dynamical curvature of spacetime and the motion of any test-particle is just a spacetime geodesic. All test-particles with the same initial conditions follow the same spacetime geodesic as it is implied by the Equivalence Principle, which states that locally, in a freely falling frame, it is possible to eliminate the observable effects of gravitation.

During the past one hundred years General Relativity has achieved an experimental triumph [3, 35, 36]. On the one hand, a number of key predictions of Einstein's gravitational theory have been experimentally confirmed with impressive accuracy. The current experimental tests of Einstein's gravitational theory include solar system measurements, among which are redshift and clock measurements, light deflection, pericentre advance, time-delay of electromagnetic waves, Lunar Laser Ranging (LLR), geodetic precession and frame-dragging measurements by Satellite Laser Ranging and dedicated spacecraft. However, current observational tests also include binary pulsars and cosmological observations [101] (see the preceding chapter in this book). General Relativity is indeed a basic ingredient for the understanding of the dynamics of binary systems of neutron stars and of the expanding universe. Einstein's theory predicts and describes the observed expansion of the universe and the Hubble law, *i.e.* the relation between the distance of a galaxy and its velocity away from us measured by the observed redshift of its electromagnetic radiation. On the other hand, General Relativity today has practical applications in space research, geodesy, astronomy and navigation in the Solar System, from the Global Navigation Satellite Systems (GNSS) [102] (see the next chapter in this book) to the techniques of Very Long Baseline Interferometry (VLBI) and Satellite Laser Ranging (SLR).

## 2.2 The riddle of dark energy and dark matter

One of the greatest enigmas and riddles of all time in science is

the composition of most of the universe in which we live, *i.e.* the nature of dark energy and dark matter. The study of distant supernovae in 1998 led to the discovery that they accelerate in their expansion away from us. The discovery of the accelerated universe [37, 38] is one of the outstanding events in science today. Dark energy, or ‘quintessence’, is regarded as a new exotic physical substance which is accelerating the expansion of the universe. Dark energy and dark matter together are thought to constitute approximately 95% of the mass-energy of the universe in an unexplained form [39–41].

## **2.3 Unified theories, alternative gravitational theories and some limits of General Relativity**

In spite of its experimental and theoretical triumph, General Relativity has its limits. Einstein’s gravitational theory is a classical theory which does not include Quantum Mechanics. Attempts to come to a quantised version of General Relativity include Loop Quantum Gravity [103] and String Theory. Among its theoretical problems, General Relativity predicts the occurrence of spacetime singularities [42], events in which every known physical theory ceases to be valid, the spacetime curvature diverges and time ends. Furthermore, in observational cosmology, as explained in the previous [section \(2.2\)](#), the accelerated expansion of the universe is an unexplained mystery.

Theories, not yet experimentally verified, such as String and Brane-World theories, try to unify gravitation with the other three interactions and to unify the two physical theories of General Relativity and Quantum Mechanics. The goal is the unification of the four interactions of nature in a theory which can be experimentally tested. Combining gravity with quantum field theory might help to solve one of the biggest mysteries of science, the composition of most of the universe in which we live, *i.e.* the nature of dark energy and dark matter, the riddle of the value of dark energy and whether it might be related to dark matter.

However, even though a breakdown of General Relativity should occur at the quantum level, some viable modifications of Einstein's theory already give different predictions at the classical level and might explain the riddle of dark energy. The current observations support its interpretation as the cosmological constant introduced by Einstein. However, modifications of Einstein's theory on cosmological scales, for instance the so called  $f(R)$  theories, with higher order curvature terms in the action, see the next [section \(2.4\)](#), have been proposed to explain the acceleration of the universe without dark energy [43].

General Relativity is a particular case of metric theory, *i.e.* a theory based on a spacetime with a dynamical geometry described by a metric tensor field which determines the spacetime distances. To distinguish the theoretical predictions of General Relativity from those of some other metric theories of gravitation, one generally uses ten parameters, testing the so-called Parametrized Post-Newtonian (PPN) approximation.

Non-metric theories of gravitation (those whose spacetime has, in addition to the metric tensor field, other structures, such as a non-degenerate differential 2-form or a connection 2-form, *i.e.* torsion) have also been proposed to describe the gravitational interaction and also to explain the accelerated expansion of the universe. These theories are not described by the standard ten PPN parameters. Some variants of Strings and Brane-World theories, possibly related to dark energy, predict deviations from General Relativity in the value of some PPN parameters, while others predict different results for the effect of frame-dragging and gravitomagnetism [44] (see the next [Section 2.4](#)). Others predict deviations from the principle of equivalence at the foundations of General Relativity and other metric theories of gravitation. Brane-World theories may have observable effects on the pericentre of the Moon and the orbit of bodies in the solar system (see, e.g. [45]).

Then, new effects are being sought on very small length scales, on very large scales and even in the Solar System for what they might reveal about theories such as quantum gravity,

and String and Brane-World models of the universe. Measurable effects on the orbital elements of Earth's satellites are predicted by String theories of the type of Chern-Simons gravity (see next section and [44]). Today, in experimental gravitation, among some of its main challenges, we have the direct detection of gravitational waves, the improved measurement of the PPN parameters testing General Relativity versus alternative gravitational theories, the accurate measurement of gravitomagnetism and frame-dragging generated by mass-energy currents, e.g. by the angular momentum of a body, and the possible experimental test of non-metric theories of gravitation. Gravitational radiation can be observed using space telescopes which would allow the study of the universe via new phenomena, such as those in the very strong gravitational field near black holes. Dragging of inertial frames, or frame-dragging, was measured by Satellite Laser Ranging with approximately 10% accuracy [46, 47]. The Gravity Probe team reported a test of frame-dragging with approximately 19% accuracy through an analysis of the data of the Gravity Probe B spacecraft. A measurement of frame-dragging with much improved accuracy is expected from the analyses of the LARES satellite (see Sections 4.1–4.4).

## **2.4 [Frame-dragging, Chern-Simons gravity and String theory]**

String and Brane-World theories unify gravitation with the other three forces of nature and unify the two great physical theories of General Relativity and Quantum Mechanics. However, String and Brane-World theories have so far no experimental verification.

Chern-Simons gravity is a modification of General Relativity. It is used in the interpretation of a number of cosmological and astrophysical phenomena, such as dark energy and 'quintessence', the dynamics of binary pulsars, gravitational wave emission by binary black holes, the properties of the jets observed in active galactic nuclei and other properties of the

accretion into massive black holes; see [44, 48–52].

Chern-Simons gravity is defined by adding to the standard action of General Relativity second order terms in the curvature, *i.e.* adding terms containing the Riemann tensor “squared”, which may represent high energy quantum gravity corrections. One such term is the Pontryagin pseudo-invariant introduced in [Section 1.7](#) to characterise the frame-dragging effect, *i.e.*

${}^* \mathbf{R} \cdot \mathbf{R} \equiv {}^* R^{\alpha\beta\mu\nu} R_{\alpha\beta\mu\nu}$ , where  ${}^* R^{\alpha\beta\mu\nu} \equiv \frac{1}{2} \epsilon^{\alpha\beta\sigma\rho} R_{\sigma\rho}^{\mu\nu}$  is the dual of  $R_{\alpha\beta\mu\nu}$  [3].

The modified action of Chern-Simons theory is then:

$$S_{CS} = \int d^4x \sqrt{-g} \left[ \frac{1}{16\pi} R - \frac{l}{12} \theta {}^* \mathbf{R} \cdot \mathbf{R} - \frac{1}{2} (\partial\theta)^2 - V(\theta) + L_{mat} \right] \quad (22)$$

where geometrical units  $c = G = 1$  are used,  $g \equiv \det g_{\alpha\beta}$  is the determinant of the metric,  $L_{mat}$  is the Lagrangian density for matter,  $R$  is the Ricci scalar,  $\theta$  is a dynamical scalar field and  $l$  is a new length scale parameter of the theory. Here both the Riemann tensor and the Pontryagin pseudo-invariant are defined with opposite sign with respect to ref. [44].

The dynamical equation for the scalar field  $\theta$  is:

$$\square\theta = \frac{dV}{d\theta} + \frac{1}{12} l {}^* \mathbf{R} \cdot \mathbf{R}. \quad (23)$$

where  $\square$  is the d'Alembertian operator. The modified gravitational field equations are then:

$$G_{\alpha\beta} - \frac{16\pi}{3} l C_{\alpha\beta} = 8\pi T_{\alpha\beta}, \quad (24)$$

where  $G_{\alpha\beta}$  is the Einstein tensor,  $T_{\alpha\beta}$  is the stress-energy tensor for the scalar field and matter Lagrangian, and  $C_{\alpha\beta}$  is the Cotton-York tensor:

$$C^{\alpha\beta} = \frac{1}{2} \left[ (\partial_\sigma \theta) \left( \epsilon^{\sigma\alpha\mu\nu} \nabla_\mu R_\nu^\beta + \epsilon^{\sigma\beta\mu\nu} \nabla_\mu R_\nu^\alpha \right) + \nabla_\rho (\partial_\sigma \theta) \left( {}^* R^{\rho\alpha\sigma\beta} + {}^* R^{\rho\beta\sigma\alpha} \right) \right] \quad (25)$$

Then, in the weak field and slow motion approximation of [Section 1.3](#), for a stationary mass distribution, we get the modified Maxwell-Ampère equation:

$$\Delta h_{0i} + \frac{1}{m_{cs}} \square H_i \cong 16\pi\rho v^i \quad (26)$$

Which, with respect to the analogous [equation \(2\)](#) of General Relativity, contains the term  $\frac{1}{m_{cs}} \square H_i$ , where  $H^i = \epsilon^{0ijk} h_{jk} =$ , that is:  $\mathbf{H} = \nabla \times \mathbf{h}$ . From this modified Maxwell-Ampère equation, the gravitomagnetic field,  $H^i$ , for a homogeneous sphere with mass density  $\rho$  of radius  $R$  rotating with angular velocity  $\omega^i$ , can be derived, including the Chern-Simons contribution [44].

Outside the sphere the gravitomagnetic field  $\mathbf{H}$  is then [44]:

$$\mathbf{H} = \mathbf{H}_{GR} + \mathbf{H}_{CS} \quad (27)$$

where  $\mathbf{H}_{GR}$  and  $\mathbf{H}_{CS}$  are the contributions of General Relativity, [Equation \(6\)](#), and Chern-Simons theory, respectively:

$$\mathbf{H}_{GR} = \frac{-16\pi\rho R^5}{15r^3} [2\boldsymbol{\omega} + 3\hat{\mathbf{r}} \times (\hat{\mathbf{r}} \times \boldsymbol{\omega})] \quad (28)$$

and:

$$\begin{aligned} \mathbf{H}_{CS} = & -16\pi\rho R^2 \{ D_1(r)\boldsymbol{\omega} + D_2(r)\hat{\mathbf{r}} \times \boldsymbol{\omega} \\ & + D_3(r)\hat{\mathbf{r}} \times (\hat{\mathbf{r}} \times \boldsymbol{\omega}) \} \end{aligned} \quad (29)$$

where:

$$D_1(r) = \frac{2R}{r} j_2(m_{CS}R) y_1(m_{CS}r),$$

(30)

$$D_2(r) = m_{CS}R j_2(m_{CS}R) y_1(m_{CS}r),$$

(31)

$$D_3(r) = m_{CS}R j_2(m_{CS}R) y_2(m_{CS}r),$$

(32)

and where  $m_{CS} = -3/(8\pi l\dot{\theta})$ , is a Chern-Simons parameter, called the Chern-Simons mass [44], and  $j_\ell(x)$  and  $y_\ell(x)$  are respectively the first-and the second-kind spherical Bessel functions. Finally, by integrating the gravitomagnetic Lorentz force equation for a test-particle, Eq. (7),  $m \frac{d^2x^i}{dt^2} \equiv m(\ddot{G}^i + \frac{dx^i}{dt} \times H^i)$ , the ratio of the nodal drag of Chern-Simons gravity to that predicted by General Relativity is [44]:

$$\frac{\dot{\Omega}_{CS}}{\dot{\Omega}_{GR}} = 15 \frac{a^2}{R^2} j_2(m_{CS}R) y_1(m_{CS}a),$$

(33)

where  $a$  is the semimajor axis of the orbit of the test-particle. Then, by using the measurement of the nodal drag obtained from the LAGEOS satellites, which agrees with General Relativity with an uncertainty of about 10% (see Section 3 below), Smith *et al.* [44] have set the lower limit of  $0.001 \text{ km}^{-1}$  [44] for the Chern-Simons mass  $m_{CS}$ .

In [44] it is shown that Chern-Simons gravity is equivalent to a heterotic and type II String theory. Then the limit on the Chern-Simons mass is also a weak limit on String theories of that type.

Since the LARES space experiment (described below in Sections 4.1–4.4) would improve the accuracy of the frame-dragging measurement, it would also improve the constraint on the Chern-Simons mass and on heterotic and type II String theories using laser-ranged satellites.

### **3 The holy grail of experimental General Relativity and its observation with the LAGEOS satellites and Gravity Probe-B**

The search for frame-dragging has been the holy grail of experimental General Relativity since before this theory was conceived. Frame-dragging was indeed hypothesised in 1896. Researchers, influenced by the ideas of Ernst Mach, attempted to measure frame-dragging effects generated by the rotation of Earth on torsion balances [53] and ground gyroscopes [54]. The idea was right but the experimental accuracy reached with the gyroscopes should have been at least a factor  $10^8$  better. The first quantitative derivation was obtained by de Sitter in 1916 [55], on the basis of Einstein's General Relativity, which calculated the precession of the perihelion of Mercury due to the angular momentum of the Sun. In 1918, Lense and Thirring [12] derived the formula describing, in the case of weak field and slow motion, the frame-dragging orbital effects on a test particle orbiting a rotating body, which is now known as the Lense-Thirring effect (see [Section 1.3](#)), and they estimated its effect on the satellites of Jupiter. Many other experiments were then proposed to measure frame-dragging – for a description of some of these suggested experiments and attempts to detect frame-dragging, see [3].

In 1959 and 1960, a space experiment (Gravity Probe B) was proposed to test the frame-dragging of orbiting gyroscopes [56, 57]. General Relativity predicts that, at the altitude of the Gravity Probe spacecraft, the frame-dragging secular precession of the four gyroscopes, owing to the rotation of the Earth, is approximately 39 milliarcsec per year (i.e. about 0.000 011 degrees per year) around an axis contained in the polar orbital plane of Gravity Probe B. On 20 April 2004, after more than 40 years of preparation, the experiment was launched into a polar orbit at an altitude of about 642 km. Gravity Probe B [58] (see <http://einstein.stanford.edu/>) consisted of a spacecraft carrying four gyroscopes and a telescope pointing towards the guide star IM Pegasi (HR8703). It was designed to measure the drifts of the

gyroscopes predicted by General Relativity (frame-dragging and geodetic precession) with respect to the very distant “fixed stars”: frame-dragging should have been measured with an accuracy of less than 1%. On 14 April 2007, after approximately 18 months of data analysis the first results of Gravity Probe B were presented: the Gravity Probe B experiment was suffering from unexpectedly large drifts of the rotation axes of the gyroscopes produced by unexpected classical torques. The Gravity Probe B team [59] (see also [60]) ascribed these large drifts to electrostatic patches on the surface of the sphere of the gyroscopes and supports, and estimated the non-modelled systematic errors to be of the order of 100 milliarcsec per year, corresponding to an uncertainty of more than 250% of the frame-dragging induced by the Earth’s rotation. In 2011, after approximately four years of data analysis, the Gravity Probe B team claimed that, by including in the estimation the modelling of the systematic errors due to the effect of the charge patches, the uncertainty in the measurement of frame-dragging had been reduced to about 19% [58].

An alternative approach to the measurement of the dragging of inertial frames is the use of laser-ranged satellites. The use of laser-ranged satellites to measure the Lense-Thirring effect was proposed in [61] (however, the calculated value of the Lense-Thirring effect for the LAGEOS satellite was in [61] at variance with the prediction of General Relativity). In 1984, the use of two LAGEOS-type satellites was proposed to measure the Lense-Thirring effect; two LAGEOS-type satellites were proposed to eliminate the large errors due to the non-sphericity of the Earth’s gravitational field [62, 63]. Since frame-dragging is extremely small for solar system bodies and Earth’s satellites, it is necessary to measure the position of a spacecraft with high accuracy. This is possible thanks to the technique of laser-ranging, which is the most accurate technique for measuring distances to the Moon and to artificial satellites such as LAGEOS (LAser GEodynamics Satellite) and LAGEOS 2 [64]. Short laser pulses emitted by ground stations and reflected back from retro-reflectors placed on the Moon or on artificial satellites

allow, through the measurement of their time of flight, an extremely precise determination of the distance of the retro-reflectors from the ground station. The positioning of the LAGEOS type satellites can reach an accuracy of a few millimetres [65]. The combined use of ranging measurements from a number of stations around the globe, allows the accurate reconstruction of the orbit of these satellites and in particular of the nodal longitude of LAGEOS and LAGEOS 2, which can be determined with an accuracy of a fraction of a milliarcsec per year [66-68].

The LAGEOS satellites are heavy brass and aluminium satellites, about 406 kg in weight, completely passive and covered with retro-reflectors, orbiting at an altitude of almost 6000 km above the surface of Earth. LAGEOS, launched in 1976 by NASA, and LAGEOS 2, launched by NASA and ASI (the Italian Space Agency) in 1992, have an essentially identical structure but they have different orbits. The semimajor axis of LAGEOS is  $a \approx 12\ 270$  km, the period  $P \approx 3.758$  hr, the eccentricity  $e \approx 0.004$  and the inclination  $I \approx 109.9^\circ$ . The semimajor axis of LAGEOS 2 is  $a_{||} \approx 12\ 163$  km, the period  $P \approx 3.708$  hr, the eccentricity  $e_{||} \approx 0.014$  and the inclination  $I_{||} \approx 52.65^\circ$ . The frame-dragging effect on the nodal longitude of the LAGEOS satellites, called the Lense-Thirring effect (see [Section 1.3](#)), is about 31 milliarcsec/yr for LAGEOS and about 31.5 milliarcsec/yr for LAGEOS 2.

Since using laser-ranging makes it possible to determine the overall orbits of the LAGEOS satellites with an accuracy of within a few centimetres [66-68] and their nodal rate within an accuracy of a fraction of milliarcsec per year, provided that all their orbital perturbations are modelled well enough [63, 66, 69], it would then be possible to measure the Lense-Thirring effect and frame-dragging [63]. The LAGEOS satellites are indeed solid metal spheres with low surface-to-mass ratios, so that atmospheric particles and photons can only slightly perturb their orbits and in particular can only very slightly change the orientation of their orbital planes [66, 69-71].

Nevertheless, the main secular displacement of their orbital

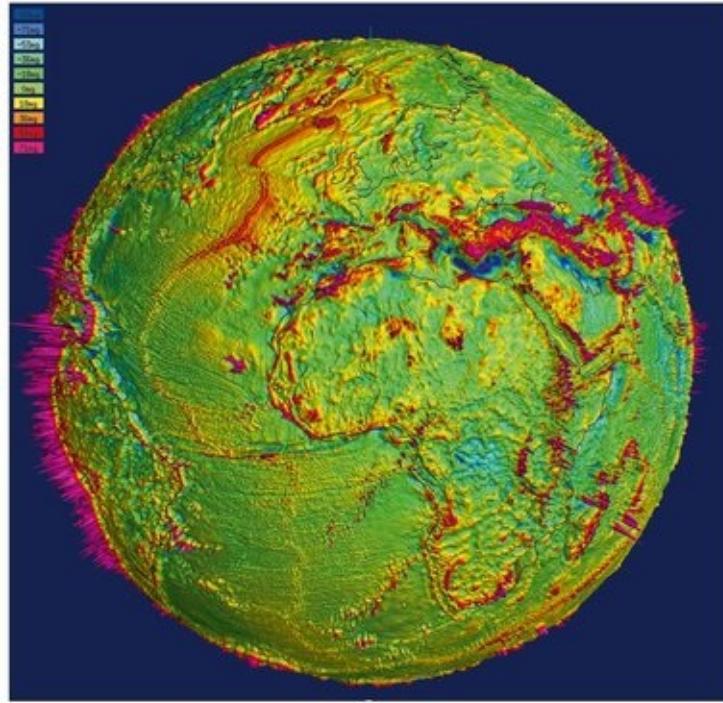
planes is by far that from the Earth's deviations from spherical symmetry and in particular that from the Earth's even zonal spherical harmonics, which are the main source of error in the measurement of frame-dragging [62, 63, 72]. The gravitational field of the Earth and its gravitational potential can be expanded in spherical harmonics and the even zonal harmonics are those of even degree and zero order, *i.e.* they are those deviations from spherical symmetry of the gravitational potential of Earth which are axially symmetric and which are also symmetric with respect to its equatorial plane. They produce large secular drifts of the nodes of the LAGEOS satellites. The coefficients which measure the size of the even zonal harmonics are denoted by  $J_{2n}$ , where  $2n$  is their degree. In particular, the flattening of the gravitational potential of the Earth, corresponding to the second degree zonal harmonic,  $J_2$ , describes the quadrupole moment of the Earth and is by far the largest source of error in the measurement of frame-dragging since it produces the highest secular perturbation of the nodes of the LAGEOS [63, 73] and LARES satellites (for the LARES satellite see Sections 4.1–4.4 below). However, thanks to the observations of the geodetic satellites, the Earth's gravitational field is extremely well known. For example, the flattening of the gravitational potential of Earth is currently measured [74] with an uncertainty of only about one part in  $10^7$ , which is, however, still not sufficient for testing frame-dragging. To eliminate the orbital uncertainties due to the errors in the models of the Earth's gravitational field, the use of both the LAGEOS and LAGEOS 2 satellites was proposed in [69] (see also [73]): “*a solution would be to orbit several high-altitude, laser-ranged satellites, similar to LAGEOS, to measure  $J_2, J_4, J_6$  etc, and one satellite to measure  $\dot{\Omega}_{\text{Lense-Thirring}}$* ”. Nevertheless, it was not easy to evaluate the accuracy of some previous measurements [75] of the Lense-Thirring effect with the LAGEOS satellites, given the limitations in the knowledge of the real uncertainty of the Earth's gravitational models available in 1998.

In March 2002, the problem of the uncertainties in the Earth's

gravitational field was overcome with the launch of two twin GRACE (Gravity Recovery and Climate Experiment) satellites [76, 77] of NASA and DLR, 200–250 km apart, into a polar orbit at an altitude of about 400 km. The spacecraft range to each other by radar and they also track the Global Positioning System (GPS) satellites. The GRACE satellites have greatly improved our knowledge of the Earth’s gravitational field.

The method which we proposed [69, 73] to measure the Lense-Thirring effect was to use the two observables provided by the two nodes of the LAGEOS satellites for the two unknowns: the Lense-Thirring effect and uncertainty in the Earth’s quadrupole moment  $J_2$ . Then, by combining the Earth’s gravitational determinations obtained through GRACE with the observations of the two LAGEOS satellites, it would be possible to measure the Lense-Thirring effect.

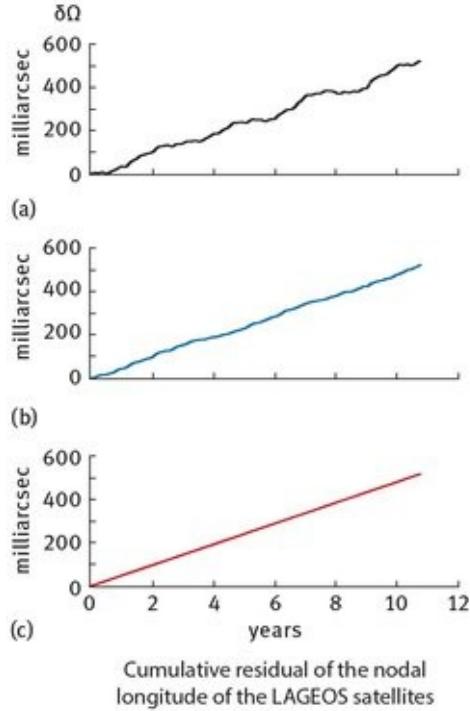
In 2004, using that method, nearly eleven years of laser ranging data were analysed and this analysis allowed the measurement of the Lense-Thirring effect with an accuracy [8, 46, 47, 78] of about 10%. The largest uncertainty in the gravitational field of the Earth, owing to its quadrupole moment  $J_2$ , was in fact eliminated with the use of the two LAGEOS satellites, see [Figures 5 and 6](#). However, that measurement was still affected by the uncertainties of the Earth’s even zonal harmonics of degree greater than two, and especially of the zonal harmonic of degree four, *i.e.*  $J_4$ . After 2004, more accurate determinations of the Earth’s gravitational field by GRACE were then published.



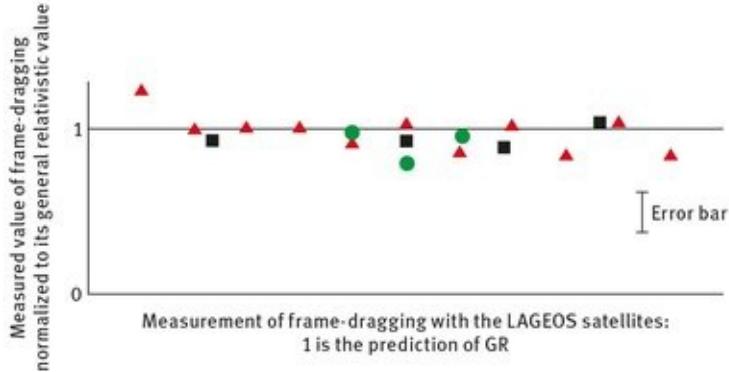
**Fig. 4.** Maps of the anomalies of the Earth's gravity field of the recent EIGEN-6C3stat model [100]. The Earth's gravity field anomalies show the differences between the Earth's actual gravity field and the gravity field of a simplified model with a smooth and featureless Earth. The anomalies display the variations in the strength of the gravitational acceleration over the surface of the Earth. The areas in yellow, orange, red and violet are those where the actual Earth's gravity field is larger than the one of the smooth Earth's model. The areas in green and blue are those where the actual Earth's gravity is smaller. The units are milligal (mg), i.e.  $10^{-6} \text{ cm/s}^2$  (courtesy of Rolf Koenig and GFZ-Potsdam).

The analyses of the LAGEOS orbital data were repeated *independently* using the new gravitational field determinations for a longer period of observation and by using three different orbital programs: GEODYN, developed by NASA Goddard, UTOPIA by the University of Texas at Austin [79, 80], and EPOS-OC by the GFZ, German Research Centre for Geosciences of Potsdam [81]; see [Figure 6](#), which displays the measurements of frame-dragging by the three groups. The recent measurements of frame-dragging [47, 78–81] carried out by a team of universities (in Lecce, Rome, Maryland, and Texas) and research centres (NASA Goddard and GFZ Potsdam) have confirmed the previous determination of the Lense-Thirring effect which were obtained in 2004 with an uncertainty of

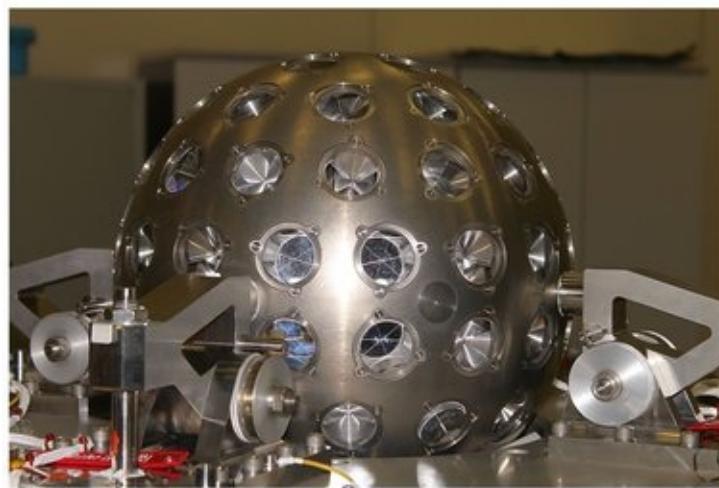
approximately 10%. Error analyses of the measurements of frame-dragging with the LAGEOS satellites were carried out in a large number of papers [47, 63, 66–69, 71, 73, 78–83, 86, 87]. The predictions of General Relativity were consistent with these measurements of frame-dragging carried out with the two LAGEOS satellites.



**Fig. 5.** The 2004 measurement of frame-dragging with the LAGEOS satellites. This measurement was obtained in 2004 using the LAGEOS and LAGEOS 2 satellites, the Earth's gravitational field by GRACE [46, 82] and GEODYN. The figure shows the cumulative residual of the nodal longitudes,  $\delta\Omega$ , of the LAGEOS satellites combined in such a way as to eliminate the uncertainty due to the quadrupole moment of Earth. In black (a) is the raw, observed, cumulative residual of the combined nodal longitudes of the LAGEOS satellites without removal of any signal, whereas in blue (b) is the observed cumulative residual of the combined nodal longitudes of the LAGEOS satellites after removal of six known periodic signals. The best fit line through this observed residual nodal longitude has a slope of 47.9 milliarcsec per year. The theoretical prediction of frame-dragging for the combined nodal longitudes of the LAGEOS satellites is plotted in red (c): its slope is 48.2 milliarcsec per year (from [46]).



**Fig. 6.** The 2004–2012 independent measurements of frame-dragging with the LAGEOS satellites. These measurement of frame-dragging were obtained independently in 2004–2012 using the LAGEOS and LAGEOS 2 satellites by three international teams using three independent orbital estimators and a number of different determinations of the Earth’s gravitational field using the GRACE space mission. The red triangles show the 2008 results by the CSR team of the University of Texas at Austin using UTOPIA and a number of different GRACE models. The average value of frame-dragging measured by Ries *et al.* using these models is 0.99, where 1 is the value predicted by General Relativity. The total error budget of the CSR-UT team is about 12% of the frame-dragging effect (adapted from [79, 80]). The black squares show the measurement of frame-dragging with the LAGEOS and LAGEOS 2 satellites obtained independently in 2012 by the German GFZ team of Potsdam-Munich using EPOS-OC and a number of different GRACE models. The average value of frame-dragging measured by Koenig *et al.* using these models is 0.95 (adapted from [81]). The green circles show some of the measurements obtained by the team from the University of Salento, University of Rome and Maryland University using the orbital estimator GEODYN II and a number of different GRACE models. The average value of frame-dragging measured by Ciufolini *et al.* using these models is 0.97 (adapted from [78]).



**Fig. 7.** The LARES satellite on the separation mechanism before launch.

## 4 The LARES space experiment

With the launch of the LAser RElativity Satellite, LARES, by the Italian Space Agency (ASI), see [Figure 7](#), a new orbital observable is available which will improve the accuracy of the measurement of frame-dragging according to the technique [69, 73] outlined in the previous Section 3. LARES was successfully launched on 13 February 2012 by the maiden flight of the VEGA launcher of the European Space Agency (ESA), which was developed by ELV (Avio-ASI) [83, 84]. LARES is well observed by the global network of laser stations of the ILRS [85]. The LARES orbital elements are: semimajor axis  $\approx 7820$  km, orbital eccentricity  $\approx 0.0008$ , and orbital inclination  $\approx 69.5^\circ$ .

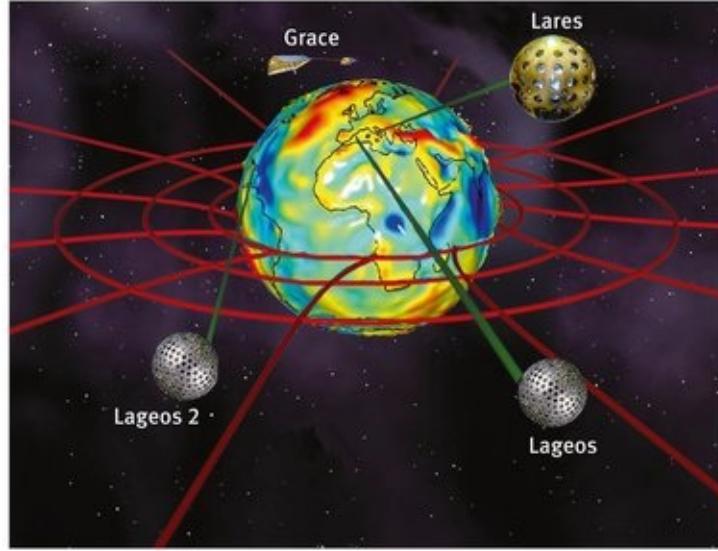
The availability of this new, additional observable, along with those provided by LAGEOS and LAGEOS 2, will enable the elimination of the error source due to the Earth's even zonal harmonic of degree four,  $J_4$ , which introduces an uncertainty which can be as large as 10% in the measurement of frame-dragging when only the LAGEOS and LAGEOS 2 satellites are used [86]. Thus, the uncertainty in the measurement of frame-dragging using LARES, plus LAGEOS and LAGEOS 2, and the determination of the Earth's gravitational field by the GRACE space mission, can be as low as approximately one percent [47, 86, 87]; see [Section 4.4](#). Indeed, the GRACE space mission [76, 77] provides very accurate determinations of the Earth's gravitational field, which are required to be able to test frame-dragging due to the Earth's rotation with an uncertainty of a few percent. The Lense-Thirring drag of the orbital plane of LARES, theoretically predicted by General Relativity, is approximately 118 milliarcsec per year, corresponding, at the altitude of LARES, to about 4.5 m/year. The combination of the three satellites will also allow other tests of fundamental physics [47, 86, 87], see [Section 2.4](#). [Figure 8 shows](#) an artistic representation of the LARES space experiment.

### 4.1 The LARES satellite, its structure and its orbit

LARES is spherical and covered with 92 retro-reflectors. It has a radius of 182 mm and a total mass of 386.8 kg. It has the highest average density of any single known object in orbit in the solar system [88].

The LAGEOS satellites, which have a high sensitivity in order to detect frame-dragging, have an altitude of about 5900 km. Nevertheless, the qualification launch of VEGA could not inject the LARES satellite at an altitude higher than 1500 km. Therefore, the satellite was designed to minimise the orbital perturbations due to the denser atmosphere at that altitude.

The ratio of cross-sectional area to mass was designed for LARES to be about 2.6 times smaller than that of the LAGEOS satellites. Before LARES, the LAGEOS satellites had the lowest cross-sectional to mass ratio of any other artificial satellite. This ratio is a key parameter because the orbital effects due to the non-gravitational perturbations, which have to be minimised to accurately measure frame-dragging, are proportional to it. Consequently, not only the atmospheric drag but also all the other non-gravitational orbital perturbations of the LARES satellite were minimised, including radiation pressure and thermal thrust. The thermal thrust, or Yarkovsky effect, on a spinning satellite is a thermal acceleration resulting from the anisotropic flux of radiation from the satellite due to the anisotropic temperature distribution on its surface by solar heating. A variation of this effect, due to the infrared radiation of the Earth absorbed by the retro-reflectors of a satellite, is the Earth-Yarkovsky effect or Yarkovsky-Rubincam effect [89, 90].



**Fig. 8.** Concept of the LARES space experiment. Shown in the figure are the LARES, LAGEOS, LAGEOS 2 and GRACE satellites which play a key role in the experiment. The spacetime distortion owed to frame-dragging induced by the Earth's rotation is represented by the distortion of the radial curves. The Earth's colours represent the EIGEN-GRACE02S anomalies (see caption of [Figure 4](#)) of the Earth's gravitational field obtained with GRACE.

Therefore, the satellite was designed to be a single spherical piece of a very dense tungsten alloy with a density equal to about  $18\ 000\ \text{kg/m}^3$ . The details of the satellite construction are given in [88]. Thanks to the extremely low value of the cross-sectional to mass ratio for LARES, equal to only about  $0.00027\ \text{m}^2/\text{kg}$ , and to its particular structure, a single spherical piece with high thermal conductivity, and despite the lower altitude of LARES compared to that of LAGEOS, the non-gravitational orbital perturbations, which are unmodelled or not modelled at all, are extremely small. Indeed, the non-gravitational perturbations which were observed during a few months of laser ranging observations [83] are smaller than those of any other satellite.

## 4.2 The LARES satellite, General Relativity and geodesic motion

In General Relativity, spacetime curvature, which describes the gravitational interaction, is generated by the distribution and

the currents of mass-energy via the Einstein field equations [1, 3, 91]. A test particle then moves along a spacetime geodesic [92]. A time-like geodesic path in spacetime's Lorentzian geometry (i.e. with a metric which can locally be reduced to the Minkowski metric) is one which locally maximises proper time, in analogy with the length-minimising property of Euclidean straight lines. In General Relativity a test particle is defined to be electrically neutral, to have negligible gravitational binding energy compared to its rest mass, to have negligible angular momentum and to be small enough that inhomogeneities of the gravitational field within its volume have negligible effect on its motion.

A star, a planet, or a satellite can approach the behaviour of a small test particle. For example, the gravitational pull of the Earth on its Moon or on artificial satellites is explained by General Relativity through the spacetime curvature generated by the mass of the Earth. These Earth satellites can be considered as test particles whose orbits describe spacetime geodesics with deviations from ideal geodesic paths due to their finite size and to the non-gravitational forces acting on them. Geodesic motion is also used to calculate the advance of the perihelion of a planet's orbit, such as Mercury, and the dynamics of a binary pulsar system.

Since geodesic motion is at the foundations of General Relativity and of other theories in which the gravitational interaction is described by spacetime curvature dynamically generated by mass-energy, the realisation of the best possible approximation of the free motion of a test particle, a spacetime geodesic, is an outstanding result for the experiments dedicated to analysing the spacetime geometry in the vicinities of a body. Such experiments would provide high-precision tests of General Relativity and constraints on alternative gravitational theories. Given the extreme weakness of the gravitational interaction compared to the other interactions of nature, the space environment is the ideal laboratory to test gravitational and fundamental physics. However, in order to test gravitational physics, a satellite must behave as nearly as possible as a test

particle and must be as little as possible affected by non-gravitational perturbations, such as radiation pressure and atmospheric drag, which shift the orbit of the satellite away from an ideal spacetime geodesic.

There are basically two ways to minimise the effects of these perturbations: a drag-free satellite or a small satellite with a high density and an extremely small surface-to-mass ratio. In the case of the Gravity Probe B drag-free satellite, an average residual acceleration of approximately  $40 \times 10^{-12}$  m/s<sup>2</sup> was achieved [58]. For a passive satellite (with no drag-free system), the key feature which determines the level of attenuation of the non-gravitational perturbations is the ratio of its cross-sectional area to its mass. In addition, it is necessary to determine the position of the satellite with extreme precision: on bodies of extended dimensions, one must follow the trajectory of a precise point of the orbiting object, for example, the geometric centre or the centre of mass. All these considerations have influenced the design of most laser-ranged satellites and in particular that of LARES, which, at present, appears to be the best realisation around Earth of an orbiting test particle. Moreover, the retro-reflectors on its surface allow the measurement of the instantaneous distance to the satellite within an accuracy of a few millimetres by measuring the round-trip time of a laser pulse.

In [83] we discuss the approximation of the motion of a satellite to the geodesic motion followed by a test particle in General Relativity and in the next section we show that LARES provides the best available test particle in the solar system with which to test General Relativity and gravitational physics, e.g. to accurately measure frame-dragging, and after modelling its known non-gravitational perturbations its orbit shows the best agreement, among those of all other satellites, with the geodesic motion predicted by General Relativity.

### **4.3 The LARES satellite and its preliminary orbital results**

The laser return signals from the LARES satellite have been

observed since the 17th of February 2012. Since then, we have performed preliminary data analyses and studied the behaviour of LARES. The orbital analysis and data reduction was performed using UTOPIA of UT/CSR (Center for Space Research at the University of Texas at Austin), GEODYN II of NASA Goddard, and EPOS-OC of GFZ (Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences) [93]. These orbital programs included the state-of-the-art orbital dynamical models. In particular, they included the models of the Earth's gravitational field based on GRACE data [76, 77], the best updated models for oceans and solid Earth tides, as well as solar radiation pressure, Earth's albedo, and atmospheric drag [70, 94, 95] with the post-Newtonian corrections of General Relativity. However, in our preliminary analyses, no model of thermal thrust was used [89, 90]. The laser data of LARES were processed in arcs of 15 days.

For the 105 days analysed, GEODYN, UTOPIA and EPOS-OC independently determined that the residual along-track accelerations for LARES were only about  $0.4 \times 10^{-12} \text{ m/s}^2$ , whereas for the two LAGEOS satellites, the acceleration residuals were in the range of  $1-2 \times 10^{-12} \text{ m/s}^2$ . This is an outstanding result since LARES is in an orbit lower by far than that of LAGEOS. As a comparison, the residual along-track acceleration of the laser-ranged STARLETTE satellite is at a level of about  $40 \times 10^{-12} \text{ m/s}^2$ . The residual acceleration along the orbit of a satellite is a measure of the level of minimization of its non-gravitational perturbations: atmospheric drag, solar and terrestrial radiation pressure and the effects of thermal thrust. The atmospheric drag acts mainly along the velocity vector of the satellite, while the solar radiation pressure, the pressure of terrestrial radiation (visible and infrared radiation from the Earth), and the effects of thermal thrust will have both along-track and non along-track contributions.

It has to be pointed out that assuming an out-of-plane residual acceleration, constant in direction, of the same order of magnitude as the non-modelled along-track acceleration observed in the orbit of LARES, this acceleration will produce a

very small secular variation of the longitude of the LARES node, *i.e.* of its orbital angular momentum. For example, by assuming an out-of-plane acceleration with an amplitude of  $0.4 \times 10^{-12}$  m/s<sup>2</sup>, constant in direction, its effects on the node of LARES would be many orders of magnitude smaller than the tiny secular drift of the node of LARES due to frame-dragging [47] of about 118 milliarcsec/year. Therefore, LARES, along with the LAGEOS satellites, and with the determination of the Earth's gravitational field obtained by the GRACE mission, can be used to accurately measure the frame-dragging effect predicted by General Relativity, allowing an improvement of about one order of magnitude over the accuracy of previous measurements of frame-dragging through the use of only the two LAGEOS satellites [46, 47, 78].

#### **4.4 [LARES error analyses]**

The major uncertainties in the measurement of frame-dragging using a test satellite of the Earth are due to the Earth's lowest degree even zonal harmonics. However, the largest  $n - 1$  uncertainties, due to the lowest degree even zonal harmonics, can be eliminated by using  $n$  test satellites ([69, 73]), thus allowing a measurement of frame-dragging not affected by these  $n - 1$  uncertainties (see Section 3). In particular, using three observables, namely the three nodes of LARES, LAGEOS and LAGEOS 2, one can eliminate the two largest uncertainties due to the two lowest degree even zonal harmonics,  $J_2$  and  $J_4$ , and measure frame-dragging with an uncertainty of only a few percent (see below).

The inclination of the orbit of LARES (set at 69.5°) was chosen for the minimisation of the uncertainties from the even zonal harmonics of degree strictly larger than four, *i.e.* from  $\delta J_{2n}$  with  $2n > 4$ , compatibly with the allowed orbits of the VEGA launcher. The final chosen inclination was anyway close to the optimal one (which should have been 70°) for the measurement of frame-dragging. In [47, 86] the errors induced by each even zonal harmonic up to degree 70 are analysed. As to the other

gravitational perturbations, due to temporal variations of the Earth's gravitational field, and especially due to terrestrial tides, we stress that the major tidal effects on the node of LARES are due to the zonal tides with  $l = 2$  and  $m = 0$  induced by the Moon's node, and to the  $K_1$  tide with  $l = 2$  and  $m = 1$  (tesseral tide). However, the error caused by the medium and long period zonal tides with  $l = 2$  and  $m = 0$  will be eliminated, along with the static errors due to  $J_2$  (and  $J_4$ ), using the combination of the three nodes. Also, the uncertainties in the time-dependent secular variations  $J_2, J_4$  will be cancelled using this combination of three observables. Furthermore, the tesseral tide  $K_1$  produces a periodical nodal perturbation of LARES with the period of its node, therefore this tide will be fitted over a period which is a multiple of the LARES nodal period (see [66] and Chapter 5 of [68]) and will then introduce only a small uncertainty in our combination.

For what concerns the effects of the non-gravitational perturbations, such as solar radiation pressure, albedo, thermal thrust and atmospheric drag we refer to [66, 69, 78, 82, 86]. Here we only stress that the secular nodal shift of LARES due to atmospheric drag is negligible. This is due to the almost circular orbit of LARES, which has an orbital eccentricity  $e \approx 0.0008$ , and to the special structure of LARES with an extremely small ratio of its cross-sectional area to its mass. In fact, even assuming that the exosphere were co-rotating with the Earth at the altitude of a satellite, in the case of zero orbital eccentricity,  $e = 0$ , the total nodal shift of the satellite would be zero, as calculated in [69]. In the case of a very small orbital eccentricity, the total nodal shift is proportional to the orbital eccentricity and thus for LARES it is a negligible effect [69] owing to its almost zero orbital eccentricity, in addition to the very small ratio of its cross-sectional area to its mass. Detailed examinations of all the error sources, both gravitational and non-gravitational, in the measurement of frame-dragging using the LAGEOS and LARES satellites have been published in a large number of papers including [47, 66–69, 73, 78, 82, 86, 87, 96, 97]. These studies have shown that the uncertainties in the

knowledge of the even zonal harmonics of degree strictly higher than four are the main source of error in the LARES experiment and that this error due to the even zonal harmonics is approximately 1% of frame-dragging. An additional error analysis was recently carried out using a number of Monte Carlo simulations [87].

In [47, 86], we analysed the error due to each even zonal harmonic up to degree 70 in the measurement of frame-dragging using LARES, LAGEOS, LAGEOS 2 and the Earth's gravitational field determined by GRACE. We concluded that the error due to each even zonal harmonic with degree greater than 4 is much less than 1% and, in particular, that the error is practically zero for even zonal harmonics of degree higher than or equal to 26. We also concluded that, using the older GRACE model EIGEN-GRACE02S, the total error in the LARES measurement of frame-dragging due to the even zonal harmonics is 1.4%, while, to evaluate an upper bound for this error, using as uncertainty the difference between the same even zonal coefficient in the two different GRACE models EIGEN-GRACE02S and GGM02S (today both outdated), the total error is 3.4%. The evaluation of the upper bound of the error in each harmonic coefficient by using the difference between the values for that coefficient in two different models is routinely used in space geodesy to estimate the reliability of the published errors of the harmonic coefficients of a model - however, one can apply this technique to estimate the real errors in the harmonic coefficients of two models only if these two models are comparable, *i.e.* if they have comparable accuracies. Then, the upper bound of the error in the measurement of frame-dragging with EIGEN-GRACE02S, *i.e.* 3.4%, is about two or three times larger than the error obtained using its published uncertainties, *i.e.* 1.4%. However, one has to consider that EIGEN-GRACE02S was just a preliminary 2004 GRACE determination of the Earth's gravitational field based on less than 365 days of observations after the 2002 launch of GRACE. Today's (2014) determinations of the Earth's gravitational field are much more accurate than EIGEN-GRACE02S, especially thanks to the much longer period

of observations of GRACE. Therefore, there will be a substantial improvement in the accuracy of the frame-dragging measurements obtained using the new GRACE gravitational models available now and in the next few years, *i.e.* during the period of the first LARES data analyses to measure frame-dragging. In addition, other space missions, such as GOCE [98, 99], have further improved the knowledge of the Earth's gravitational field over the models from 2004.

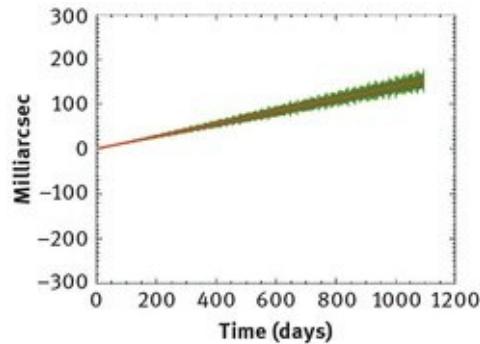
In conclusion, the error in the determination of frame-dragging owing to the uncertainties in the Earth's gravitational field is greatly minimised by the use of three satellites and by the GRACE determinations of the Earth's gravitational field. On the other hand, a minimisation of the non-gravitational orbital perturbations was possible by optimising the design of the three satellites; the LARES and LAGEOS satellites are indeed spherical and very dense, designed with a very small surface-to-mass ratio to minimise the non-gravitational perturbations [69]. In particular, thanks to the LARES design, the orbital effects of the non-modelled perturbations are smaller for LARES, in spite of its lower orbit, than for the LAGEOS satellites, as confirmed by the preliminary orbital analyses shown above in [Section 4.3](#) and in ref. [83].

Finally, as mentioned above, in order to obtain an extremely solid error estimate of the forthcoming LARES test of frame-dragging and to eventually confirm its previous error analyses, we used a completely independent method of analysis, *i.e.* a number of Monte Carlo simulations of the LARES experiment. In [87] the detailed results of these simulations are reported, and they are summarised below.

One hundred simulations were designed and performed to take into account the uncertainties in the physical parameters of the LARES experiment and to quantitatively evaluate the effect of these uncertainties on the final measurement of frame-dragging. We first selected the main physical parameters whose uncertainties have a critical impact on the accuracy of the measurement of frame-dragging using LARES, LAGEOS and LAGEOS 2, *i.e.* the main parameters describing the Earth's

gravitational field and the parameters related to the orbital perturbations due to radiation pressure. These parameters are:  $GM$ , i.e. the gravitational constant times the mass of Earth, the five largest even zonal harmonics,  $J_2$ ,  $J_4$ ,  $J_6$ ,  $J_8$  and  $J_{10}$ , the secular rate of change of the two largest even zonal harmonics  $J_2$  and  $J_4$  and the radiation pressure coefficients of the LARES, LAGEOS and LAGEOS 2 satellites. Then, using the orbital estimator EPOS-OC, we simulated the orbits of LARES, LAGEOS and LAGEOS 2 one hundred times by generating by chance the values of  $GM$ ,  $J_2$ ,  $J_4$ ,  $J_6$ ,  $J_8$ ,  $J_{10}$ , of  $J_2$  and  $J_4$ , and of the radiation pressure coefficients of the LARES, LAGEOS and LAGEOS 2 satellites. Frame-dragging was always kept equal to its value predicted by the theory of General Relativity. Finally, we performed an analysis of their simulated laser-ranging observations.

The results of the 100 simulations of the LARES space experiment confirmed the previous extensive analyses. In fact, the standard deviation of the 100 simulated measurements of frame-dragging, obtained corresponding to each of the 100 Monte Carlo simulations, was equal to 1.4% of the frame-dragging effect predicted by General Relativity. The mean value of the 100 simulated measurements of frame-dragging was equal to 100.24% of its relativistic value (see [Figure 9](#)). Thus, the Monte Carlo simulations have confirmed an error budget of approximately 1% in the forthcoming measurement of frame-dragging using LARES, LAGEOS and LAGEOS 2, and the Earth's gravitational field determined by GRACE.



[\*\*Fig. 9.\*\*](#) Monte Carlo simulations of the LARES experiment. This figure shows the

spread of the one hundred Monte Carlo simulations and the corresponding estimated error (standard deviation) of 1.4% in the measurement of frame-dragging with LARES, LAGEOS and LAGEOS 2 and the GRACE Earth's gravitational field. The red line is the theoretical prediction of General Relativity.

## 5 Conclusions

We have described the main theoretical and experimental aspects of the intriguing phenomenon of dragging of inertial frames, or frame-dragging, predicted by General Relativity; the effect of frame-dragging on time and clocks is especially fascinating and intriguing.

We have shown that the measurement of frame-dragging may shed light on some of the biggest open problems of physics today: dark energy and dark matter, and unified theories of the fundamental interactions, such as String theories.

We have then presented the measurements of frame-dragging obtained with the LAGEOS satellites and with the Gravity Probe B spacecraft, with reported accuracies of about 10% and 19%, respectively. The LARES satellite was launched in 2012 to improve these measurements of frame-dragging. Extensive error analyses and Monte Carlo simulations have shown that the LARES experiment will be able to test frame-dragging with an accuracy of approximately 1%.

**Acknowledgment:** I gratefully acknowledge the support of the Italian Space Agency, grants I/043/08/0, I/016/07/0, I/043/08/1 and I/034/12/0 and the International Laser Ranging Service for providing high-quality laser ranging tracking of the LARES satellites.

## Bibliography

- [1] C. W. Misner, K. S. Thorne and J. A. Wheeler, *Gravitation*, Freeman, San Francisco, (1973).
- [2] S. Weinberg, *Gravitation and Cosmology: Principles*

*and Applications of the General Theory of Relativity*, Wiley, New York, (1972).

- [3] I. Ciufolini and J. A. Wheeler, *Gravitation and Inertia*, Princeton Univ. Press, (1995).
- [4] H. Pfister, in: *Mach's Principle. From Newton's Bucket to Quantum Gravity*, eds. J. Barbour and H. Pfister, Birkhäuser, Boston, (1995), 315-331.
- [5] H. Pfister and K. Braun, A mass shell with flat interior cannot rotate rigidly. *Class. Quantum Grav.* **3**, 335-345 (1986).
- [6] H. Pfister, Rotating mass shells with flat interiors. *Class. Quantum Grav.* **6**, 487-503 (1989).
- [7] A. Einstein, Letter to Ernst Mach. Zurich, 25 June 1913, in ref. [3], p. 544.
- [8] I. Ciufolini, Dragging of Inertial Frames. *Nature* **449**, 41-47 (2007).
- [9] K. S. Thorne, R. H. Price and D. A. Macdonald, *The Membrane Paradigm*, Yale Univ. Press, New Haven, (1986).
- [10] I. Ciufolini, Frame-Dragging, Gravitomagnetism and Lunar Laser Ranging. *New Astronomy* **15**, 332-337 (2010).
- [11] R. P. Kerr, Gravitational field of a spinning mass as an example of algebraically special metrics. *Phys. Rev. Lett.* **11**, 237-238 (1963).
- [12] J. Lense and H. Thirring, Über den Einfluss der Eigenrotation der Zentralkörper auf die Bewegung der Planeten und Monde nach der Einsteinschen Gravitationstheorie. *Phys. Z.* **19**, 156-163 (1918). See also English translation by B. Mashhoon, F. W. Hehl, D. S. Theiss, *Gen. Relativ. Gravit.* **16**, 711-750 (1984).
- [13] H. Thirring, Über die Wirkung rotierender ferner Massen in der Einsteinschen Gravitationstheorie. *Z. Phys.* **19**, 33-39 (1918). English translation by B. Mashhoon, F. W. Hehl, D. S. Theiss, *Gen. Relativ. Gravit.* **16**, 711-750 (1984).
- [14] D. R. Brill and J. M. Cohen, Rotating Masses and their

Effect on Inertial Frames. *Phys. Rev.* **143**, 1011-1015 (1966).

- [15] L. Bass and F. A. E. Pirani, On the Gravitational Effects of Distant Rotating Masses. *Phil. Mag.* **46**, 850-856 (1955).
- [16] Ya. B. Zeldovich and I. D. Novikov, *Relativistic Astrophysics, Vol. I Stars and Relativity*, Univ. Chicago Press, Chicago, (1971).
- [17] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 3rd rev. English edn., Pergamon, London, (1971).
- [18] I. Ciufolini and F. Ricci, Time delay due to spin and gravitational lensing. *Class. and Quantum Grav.* **19**, 3863-3874 (2002).
- [19] J. M. Cohen and B. Mashhoon, Standard clocks, interferometry, and gravitomagnetism. *Phys. Lett. A* **181**, 353-358 (1993).
- [20] A. Tartaglia, Detection of the gravitomagnetic clock effect. *Class. Quantum Grav.* **17**, 783-792 (2000).
- [21] A. Tartaglia, Gravitomagnetism, clocks and geometry. *Eur. J. Phys.* **22**, 105-111 (2001).
- [22] I. G. Dymnikova, Gravitational time delay of signals in the Kerr metric in Relativity, in: *Celestial mechanics and astrometry: High precision dynamical theories and observational verifications*, Proc. of the Symposium, Leningrad, USSR, May 28-31, 1985; D. Reidel Publishing Co, Dordrecht, (1986), 411-416.
- [23] S. Pineault and R. C. Roeder, Applications of Geometrical Optics to the Kerr Metric. 11. Numerical Results. *Astron. J.* **213**, 548-557 (1977).
- [24] K. P. Rauch and R. D. Blandford, Optical Caustics in a Kerr Spacetime and the Origin of Rapid X-ray Variability in AGN. *Astron. J.* **421**, 46-68 (1994).
- [25] J. Huchra, M. Gorenstein, S. Kent, I. Shapiro, G. Smith, E. Horine and R. Perley, 2237 + 0305: A new and unusual gravitational lens. *Astron. J.* **90**, 691-696 (1985).

- [26] R. Racine, Fifth image and photometric variability in 2237 + 0305 ('Einstein Cross'). *Astron. J.* **102**, 454–460 (1991).
- [27] I. Ciufolini and F. Ricci, Time delay due to spin inside a rotating shell. *Class. and Quantum Grav.* **19**, 3875–3881 (2002).
- [28] G. O. Abell, in: *Stars and Stellar Systems, IX Galaxies and the Universe*, eds. A. Sandage, M. Sandage and J. Kristian, Univ. of Chicago Press, Chicago, (1975) p. 601.
- [29] I. Ciufolini, F. Ricci, S. Kopekin and B. Mashhoon, On the Gravitomagnetic Time Delay. *Phys. Lett. A* **308**, 101–109 (2003).
- [30] A. D. Biggs, I. W. A. Browne, P. Helbig, L. V. E. Koopmans, P. N. Wilkinson, R. A. Perley, Time delay for the gravitational lens system B0218+357. *Mon. Not. R. Astron. Soc.* **304**, 349–358 (1999).
- [31] I. Ciufolini and F. Ricci, Gravitational lensing and gravitomagnetic time delay. To be published (2014).
- [32] N. Ashby and B. Shahid-Saless, Geodetic Precession or Dragging of Inertial Frames? *Phys. Rev. D* **42**, 1118–1122 (1990).
- [33] I. Ciufolini, Gravitomagnetism and status of the LAGEOS III experiment. *Class. Quantum Grav.* **11**, A73–A81 (1994).
- [34] L. F. Costa and C. A. R. Herdeiro, A gravitoelectromagnetic analogy based on tidal tensors. *Phys. Rev. D* **78**, 024021 (2008).
- [35] S. Turyshev, Experimental Tests of General Relativity: Recent Progress and Future Directions. *Physics-Uspekhi* **52**, 1–27 (2009).
- [36] C. M. Will, *Theory and Experiment in Gravitational Physics*, 2nd edn., Cambridge Univ. Press, Cambridge, UK, (1993).
- [37] A. Riess et al., Observational evidence from supernovae for an accelerating universe and a cosmological constant. *Astron. J.* **116**, 1009–1038 (1998).

- [38] S. Perlmutter et al., Measurements of  $\Omega$  and  $\Lambda$  from 42 High-Redshift Supernovae. *Astrophys. J.* **517**, 565-586 (1999).
- [39] S. Perlmutter, Supernovae, Dark Energy, and the Accelerating Universe. *Phys. Today* **56**, 53-60 (2003).
- [40] R. Caldwell, Dark Energy. *Physics World* **17**, 37-42 (2004).
- [41] Planck Collaboration, Planck 2013 results, to appear in *Astronomy and Astrophysics* (2014).
- [42] R. Penrose, Gravitational Collapse and Space-Time Singularities. *Phys. Rev. Lett.* **14**, 57-59 (1965).
- [43] A. De Felice and S. Tsujikawa, f(R) Theories. *Living Rev. Relativ.* **13** (2010).
- [44] T. L. Smith, A. Erickcek, R. Caldwell and M. Kamionkowski, Effects of Chern-Simons gravity on bodies orbiting the Earth. *Phys. Rev.D* **77**, 024015 (2008).
- [45] G. Dvali, Filtering Gravity: Modification at Large Distances? Infrared Modification of Gravity, in: *Nobel Symp. on Cosmology and String Theory and Cosmology*, Proc. of Nobel Symposium 127, Sigtuna, Sweden, August 2003; eds. U. Danielsson, A. Goobar and B. Nilsson World Scientific, Singapore, (2005).
- [46] I. Ciufolini and E. C. Pavlis, A confirmation of the general relativistic prediction of the Lense-Thirring effect. *Nature* **431**, 958-960, (2004).
- [47] I. Ciufolini, A. Paolozzi, E. C. Pavlis, J. Ries, R. Koenig, R. Matzner, G. Sindoni and H. Neumayer, Testing Gravitational Physics with Satellite Laser Ranging. *The European Physical Journal Plus* **126**, 72 (2011).
- [48] B. A. Campbell, M. J. Duncan, N. Kaloper and K. A. Olive, Gravitational dynamics with Lorentz Chern-Simons terms. *Nuclear Physics B* **351**, 778-792 (1991).
- [49] K. Yagi, N. Yunes and T. Tanaka, Slowly Rotating Black Holes in Dynamical Chern-Simons Gravity: Deformation Quadratic in the Spin. *Phys. Rev.D* **86**, 044037 (2012).

- [50] S. Alexander, A. Marciano and D. Spergel, Chern-Simons Inflation and Baryogenesis. *JCAP* **1304**, 046 (2013).
- [51] T. Harko, Z. Kovács, F. S. N. Lobo, Thin accretion disk signatures in dynamical Chern-Simons modified gravity. *Class. Quant. Grav.* **27**, 105010 (2010).
- [52] K. Yagi, N. Yunes and T. Tanaka, Gravitational Waves from Quasicircular Black-Hole Binaries in Dynamical Chern-Simons Gravity. *Phys. Rev. Lett.* **109**, 251105 (2012).
- [53] B. and I. Friedländer *Absolute und Relative Bewegung?*, Simion-Verlag, Berlin, (1896).
- [54] A. Föppl, Über einen Kreiselversuch zur Messung der Umdrehungsgeschwindigkeit der Erde. *Sitzb. Bayer. Akad. Wiss.* **34**, 5–28 (1904), *Phys. Z.* **5**, 416; see also A. Föppl, Über Absolute und Relative Bewegung. *Sitzb. Bayer. Akad. Wiss.* **34**, 383–395 (1904).
- [55] W. de Sitter, On Einstein's Theory of Gravitation and its Astronomical Consequences. *Mon. Not. Roy. Astron. Soc.* **76**, 699–728 (1916).
- [56] G. E. Pugh, Proposal for a Satellite Test of the Coriolis Prediction of General Relativity. *Weapons Systems Evaluation Group Research Memorandum N. 11*. The Pentagon, Washington, (1959).
- [57] L. I. Schiff, Motion of a Gyroscope According to Einstein's Theory of Gravitation. *Proc. Nat. Acad. Sci.* **46**, 871–882 (1960); Possible New Test of General Relativity Theory. *Phys. Rev. Lett.* **4**, 215–217 (1960).
- [58] C. W. Everitt et al., 2011 Gravity Probe B: Final Results of a Space Experiment to Test General Relativity. *Phys. Rev. Lett.* **106**, 22110 (2011).
- [59] D. K. Gill and S. Buchman, Evidence for Patch Effect Forces On the Gravity Probe B Gyroscopes, Stanford Univ., Stanford, (April 2007); poster at [http://einstein.stanford.edu/content/aps\\_posters/Evidence\\_for\\_Patch\\_Effect\\_Forces\\_on\\_GPB\\_Gyroscopes.pdf](http://einstein.stanford.edu/content/aps_posters/Evidence_for_Patch_Effect_Forces_on_GPB_Gyroscopes.pdf)
- [60] B. M. Barker and R. F. O'Connel, The gyroscope test of General Relativity. *Nature* **312**, 314 (1984).

- [61] L. Cugusi and E. E. Proverbio, Relativistic Effects on the Motion of Earth's Artificial Satellites. *Astronomy and Astrophysics* **69**, 321–325 (1978).
- [62] I. Ciufolini, *Theory and Experiments in General Relativity and other Metric Theories*, PhD Dissertation, Univ. of Texas, Austin, Pub. Ann Arbor, Michigan, (1984).
- [63] I. Ciufolini, Measurement of the Lense-Thirring drag on high-altitude laser-ranged artificial satellites. *Phys. Rev. Lett.* **56**, 278–281 (1986).
- [64] S. C. Cohen and P. J. Dunn (eds.), LAGEOS Scientific Results, *J. Geophys. Res.* **90** (B11), 9215 (1985).
- [65] *13th International Workshop on Laser Ranging*, Proceedings From the Science Session and Full Proceedings CD-ROM; eds. R. Noomen, S. Klosko, C. Noll and M. Pearlman, NASA CP 2003-212248, NASA Goddard, Greenbelt, MD, (2003).
- [66] B. Tapley, J. C. Ries, R. J. Eanes and M. M. Watkins, *NASA-ASI Study on LAGEOS III*, CSR-UT publication n. CSR-89-3, Austin, Texas (1989); I. Ciufolini et al., *ASI-NASA Study on LAGEOS III*, CNR, Rome, Italy (1989). See also: I. Ciufolini et al., *INFN study on LARES/WEBER-SAT* (2004).
- [67] J. C. Ries, *Simulation of an experiment to measure the Lense-Thirring precession using a second LAGEOS satellite*, PhD Dissertation Univ. of Texas, Austin, (1989).
- [68] G. E. Peterson, *Estimation of the Lense-Thirring Precession Using Laser-Ranged Satellites*, PhD Dissertation Univ. of Texas, Austin, (1997).
- [69] I. Ciufolini, A comprehensive introduction to the Lageos gravitomagnetic experiment: from the importance of the gravitomagnetic field in physics to preliminary error analysis and error budget. *Int. J. Mod. Phys. A* **4**, 3083–3145 (1989).
- [70] D. P. Rubincam, On the secular decrease in the semimajor axis of Lageos's orbit. *Celest. Mech.* **26**,

361-382 (1982).

- [71] D. M. Lucchesi, Reassessment of the error modelling of non-gravitational perturbations on LAGEOS 2 and their impact in the Lense-Thirring determination. Part I. *Planet. Space Sci.* **49**, 447-463 (2001).
- [72] W. M. Kaula, *Theory of Satellite Geodesy*, Blaisdell, Waltham, (1966).
- [73] I. Ciufolini, On a new method to measure the gravitomagnetic field using two orbiting satellites. *Nuovo Cimento A* **109**, 1709-1720 (1996).
- [74] G. G Petit and B. Luzum (eds.), 2010 *IERS Conventions*, Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main, 179 pp., ISBN 3-89888-989-6.
- [75] I. Ciufolini, E. C. Pavlis, F. Chieppa, E. Fernandes-Vieira and J. Perez-Mercader, Test of general relativity and measurement of the Lense-Thirring effect with two Earth satellites. *Science* **279**, 2100-2103 (1998).
- [76] Ch. Reigber, F. Flechtner, R. Koenig, U. Meyer, K. Neumayer, R. Schmidt, P. Schwintzer and S. Zhu, GRACE Orbit and Gravity Field Recovery at GFZ Potsdam - First Experiences and Perspectives. *Eos. Trans. AGU* **83**(47), Fall Meet. Suppl., Abstract G12B-03 (2002).
- [77] B. D. Tapley, The GRACE Mission: Status and Performance Assessment, *Eos. Trans. AGU* **83**(47), Fall Meet. Suppl., Abstract G12B-01 (2002).
- [78] I. Ciufolini, E. C. Pavlis, J. Ries, R. Koenig, G. Sindoni, A. Paolozzi and H. Newmayer, Gravitomagnetism and its Measurement with Laser Ranging to the LAGEOS satellites and GRACE Earth Gravity Models, in: *John Archibald Wheeler and General Relativity*, eds. I. Ciufolini and R. Matzner, Springer Verlag, (2010), 371-434.
- [79] J. C. Ries, R. J. Eanes and M. M. Watkins, Confirming the Frame-Dragging Effect with Satellite Laser Ranging, *16th International Workshop on Laser Ranging*, 13-17

October 2008, Poznan, Poland.

- [80] J. C. Ries, Relativity in Satellite Laser Ranging, *American Astronomical Society, IAU Symposium 261. Relativity in Fundamental Astronomy: Dynamics, Reference Frames, and Data Analysis*, 27 April – 1 May 2009, Virginia Beach, VA, USA.
- [81] R. Koenig, B. Moreno Monge and G. Michalak, Some aspects and perspectives of measuring Lense-Thirring with GNSS and geodetic satellites, *Second International LARES Science Workshop*, September 2012, Accademia dei Lincei, Rome, Italy.
- [82] I. Ciufolini, E. C. Pavlis and R. Peron, Determination of frame-dragging using Earth gravity models from CHAMP and GRACE. *New Astronomy* **11**, 527–550 (2006).
- [83] I. Ciufolini, A. Paolozzi, E. C. Pavlis, J. Ries, V. Gurzadyan, R. Koenig, R. Matzner, R. Penrose and G. Sindoni, Testing General Relativity and gravitational physics using the LARES satellite. *The European Physical Journal Plus* **127**, 1–7 (2012).
- [84] A. Paolozzi and I. Ciufolini, LARES successfully launched in orbit: Satellite and mission description. *Acta Astronautica* **91**, 313–321 (2013).
- [85] M. R. Pearlman, J. J. Degnan and J. M. Bosworth, The International Laser Ranging Service. *Advances in Space Research* **30**, 135–143 (2002).
- [86] I. Ciufolini, A. Paolozzi, E. C. Pavlis, J. Ries, R. Koenig, R. Matzner and G. Sindoni, The LARES Space Experiment: LARES Orbit, Error Analysis and Satellite Structure, in: *John Archibald Wheeler and General Relativity*, eds. I. Ciufolini and R. Matzner, Springer Verlag, (2010), 371–434.
- [87] I. Ciufolini, B. Moreno Monge, A. Paolozzi, R. Koenig, G. Sindoni and G. Michalak, Monte Carlo Simulations of the LARES space experiment to test General Relativity and fundamental physics. *Class. Quantum Grav.* **30**, 235009 (2013).

- [88] A. Paolozzi, I. Ciufolini, C. Vendittozzi, Engineering and scientific aspects of LARES satellite. *Acta Astronautica* **69**, 127–134 (2011).
- [89] D. P. Rubincam, Yarkovsky Thermal Drag on LAGEOS. *J. Geophys. Res. B* **93**, 13 805–13 810 (1988).
- [90] D. P. Rubincam, Drag on the LAGEOS satellite. *J. Geophys. Res.* **95** (B11), 4881–4886 (1990).
- [91] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time*, Cambridge University Press, (1975).
- [92] I. Ciufolini, Generalized geodesic deviation equation. *Physical Review D* **34**, 1014–1017 (1986).
- [93] S. Zhu, Ch. Reigber and R. Koenig, Integrated Adjustment of CHAMP, GRACE, and GPS Data. *Journal of Geodesy* **78**, 103–108 (2004).
- [94] D. E. Pavlis et al., *GEODYN operations manuals*, Contractor Report, Raytheon, ITSS, Landover, MD, (1998).
- [95] C. F. Martin and D. P. Rubincam, Effects of Earth albedo on the LAGEOS I satellite. *J. Geophys. Res. B* **101**, 3215–3226 (1996).
- [96] J. Ries, I. Ciufolini, E. Pavlis, A. Paolozzi, R. Koenig, R. Matzner, G. Sindoni and H. Neumayer, The Earth's frame-dragging via laser-ranged satellites: a response to “some considerations on the present-day results for the detection of frame-dragging after the final outcome of GP-B” by L. Iorio. *EPL* **96**, 30002 (2011).
- [97] V. G. Gurzadyan, I. Ciufolini, S. Sargsyan, G. Yegorian, S. Mirzoyan and A. Paolozzi, Satellite probing General Relativity and its extensions and Kolmogorov Analysis. *EPL* **102**, 60002 (2013).
- [98] M. R. Drinkwater, R. Floberghagen, R. Haagmans, D. Muzi and A. Popescu, GOCE: ESA's first Earth Explorer Core mission, in Earth Gravity Field From Space – From Sensors to Earth Science. *Space Sci. Ser. ISSI* **18**, eds. G. Beutler et al., Kluwer Acad., Dordrecht, Netherlands, (2003), 419–432.

- [99] R. Pail et al., Combined satellite gravity field model GOCE01S derived from GOCE and GRACE. *Geophys. Res. Lett.* **37**, L20314 (2010).
- [100] C. Foerste, S. Bruinsma, J.-C. Marty, F. Flechtner, O. Abrikosov, C. Dahle, J.-M. Lemoine, K.H. Neumayer, R. Biancale, F. Barthelmes and R. Koenig: EIGEN-6C3stat – the newest High Resolution Global Combined Gravity Field Model based on the 4th Release of the GOCE Direct Approach, <http://icgem.gfz-potsdam.de/ICGEM/documents/Foerste-et-al-EIGEN-6C3stat.pdf>, 20 October 2014.
- [101] G. G. Byrd, A. Chernin, P. Teerikorpi and M. Valtonen, Observations of General Relativity at strong and weak limits, in this volume: *General Relativity: The most Beautiful of Theories*, ed. C. Rovelli, De Gruyter, Berlin, (2015).
- [102] N. Ashby, GNSS and other applications of General Relativity, in this volume: *General Relativity: The most Beautiful of Theories*, ed. C. Rovelli, De Gruyter, Berlin, (2015).
- [103] C. Rovelli, *Quantum Gravity*, Cambridge Univ. Press, Cambridge, UK, (2004).

Neil Ashby

# **GNSS and other applications of General Relativity**

**Abstract:** Global Navigation Satellite Systems (GNSS) rely on networks of synchronized orbiting and ground-based clocks, to provide navigation information to users. Accounting for many relativistic effects in GNSS systems is necessary in order that these systems function. Relativistic concepts and effects include the principle of equivalence, constancy of the speed of light, relativity of simultaneity, Sagnac effect, proper time, coordinate time, time dilation, gravitational frequency shifts, relations between non-rotating and rotating reference frames, and relations between time scales based on clocks at different locations and in different states of motion. This article discusses applications of General Relativity (GR) in GNSS and in internationally defined time synchronization systems.

## **1 Introduction**

“ ...no doubt many a relativist looks forward to the day when  
governments will seek  
his opinion on important questions.”  
-J. L. Synge [1]

The United States' Global Positioning System (GPS) was the first GNSS system to be fully deployed. It was conceived as a military navigation system, but has had such success in numerous civilian and scientific applications, that the US has committed to providing free, worldwide navigation service through GPS in the foreseeable future, except perhaps in theaters of conflict. Other governments and international organizations, not entirely

trusting US control of such a valuable asset, are creating and deploying their own systems; these include GLONASS (Russia), GALILEO (European Union), BEIDOU (China). There are also a number of augmentation systems that use geosynchronous satellites to transmit or transpond GPS-like signals. Nationalities deploying augmentation systems include Europe (EGNOS), Japan (QZSS), the US (WAAS), and India (GAGAN).

Global Navigation Satellite Systems (GNSS) are essentially time distribution systems, which rely on networks of coordinated, synchronized atomic clocks that transmit their position and time to millions of receivers. To users interested in navigation (position and velocity), time that is accurate to better than 100 nanoseconds (ns) is of minor interest so many significant digits of the time data are first used to solve navigation

**Neil Ashby:** National Institute of Standards and Technology, Boulder, CO 80305, USA equations, but are then dropped. Most GNSS receivers only display time to 1 second accuracy.

Concepts of time and its measurement with clocks have evolved over millennia and are still changing; a recapitulation of this history is appropriate. A clock (by definition) is an object that exhibits a repeated periodic phenomenon, together with a count of the number of repetitions (cycles) or some number that is simply related to the number of cycles. The apparently uniform rotation of the earth, as seen for example by observing motions of the stars at night, and the alternation of day and night, provide one such phenomenon that was important to ancient peoples. Other obvious periodic natural phenomena include the cycle of the seasons and the phases of the moon, although their periods are incommensurable and this leads to additional scales of time. The Egyptians divided a day into 12 hours of daytime and 12 hours of night-time; hours did not have fixed length until after Hipparchus proposed dividing the day

equally into 24 hours, but many people continued to use seasonally varying hours until mechanical clocks were invented in the 14th century CE.

Astronomical observatories such as Stonehenge (3000–2000 BCE), or Chankillo in Peru [2], permit the prediction of equinoxes and solstices; however such objects were probably not clocks because if cycle counting existed when these observatories were actually in use, no evidence for such tallies has survived. Counting was invented by the Sumerians about 4000 BCE. Their base-60 number system is with us today when we tell time with 60 minutes per hour and 60 seconds per minute. Counting days, seasons and years leads to the development of calendars; one of the oldest known calendars is the lunisolar calendar of Shulgi of Umma (ca. 21st century BCE), but calendars take various forms depending on whether they are based on lunar or solar cycles or on a combination of the two.

The following descriptions of time scales and reference frames used for calculation in GNSS are approximate because accurate calculations, based now on many technologically advanced theories and observations, are very complicated. Two calendars are of particular current relevance. These are the Julian calendar, introduced by Julius Caesar in 46 BCE, which began its day count at noon, January 1, 4713 BCE and assigns extra “leap” days to February every four years. A Julian date is a day number that may have many significant digits after the decimal point representing divisions of the day into smaller parts; this is used in many astronomical calculations. For example, an important Julian date for GNSS calculations is denoted by “J2000,” the Julian date JD 2451545.0 = noon January 1, 2000 CE. The Julian calendar was based on the tropical year, which is determined by the position of the sun on the celestial sphere as seen from earth. However a slight lack of agreement between the Julian calendar year and the true value of the tropical year led to the Gregorian reform of 1582 CE, which reduced the number of leap years to those currently used worldwide, and eliminated ten calendar days, restoring the date of the vernal equinox to March 21. To work with smaller

numbers than the current values of JD, it is common to use the Modified Julian Date (MJD), defined by  $MJD = JD - 2400000.5$ . The MJD for J2000 is then 51544.5.

Earth's rotation is currently measured by the "Universal Time" scale UT1, measured in principle by observation of distant stars crossing the Greenwich meridian every day. UT1 is proportional to the rotation angle of the earth with respect to the International Celestial Reference Frame (ICRF), which is a reference frame determined by VLBI observations of extragalactic sources. The origin of the ICRF is at the solar system barycenter, and the  $x$ ,  $y$  axes lie close to the mean equator of the earth at J2000, with the  $x$ -axis pointing to the vernal equinox. UT1 neglects many small variations in earth's rotation due for example to tides, or changes correlated with the seasons. UT1 and the earth rotation angle (ERA) are related by

$$ERA = 2\pi(0.7790572732640 + 1.00273781191135448T_u) \text{ radians ,} \quad (1)$$

where  $T_u = (\text{Julian UT1 date} - 2451545.0)$ .

There are several other versions of Universal Time that incorporate or neglect small effects on earth's rotation.

For precise navigation on earth's surface, it is important to refer positions to a well-defined reference frame fixed in the earth. The International Astronomical Union (IAU) has defined such a frame (the ITRF or International Terrestrial Reference Frame) [3], but the definition is continually being updated because the earth is constantly changing shape (due to crustal motions, post-glacial rebound, etc). The ITRF is constructed as though the geometry near earth were Euclidean, with coordinates of a number of fiducial station points precisely specified in a particular coordinate system. Up to 29 years of observations with VLBI, satellite laser ranging (SLR), Doppler Orbitography and Radio Positioning Integrated by Satellite (DORIS), and GPS observations involving over 900 observing stations have been processed to provide these coordinates.

Different GNSS systems may realize an earth-fixed reference frame using surveys that are relevant for their own regions. The reference system for GPS is denoted as WGS84(G1150), which specifies the coordinates of the GPS monitor stations. Updates to these coordinates have been less than about 30 cm but further refinements can be expected. For many practical purposes this coordinate system is equivalent to the ITRF. The origin is at the center of mass of the earth, including oceans and atmosphere. It is a right-handed Cartesian coordinate system with z-axis in the direction of the International Earth Rotation Service (IERS) reference pole and the x-axis passing through the IERS reference meridian; these are at epoch 1984.0 CE and have uncertainties of about 5 milliarcseconds. (On earth's surface  $.001 \text{ arcsec} \approx 30 \text{ cm}$ .) An ellipsoidal reference surface is determined by adopting values for the equatorial radius, flattening, angular rotation rate, and a value of the product  $GM$  where  $G$  is the Newtonian gravitational constant and  $M$  is earth's total mass. The adopted values for GPS are listed in [Table 1](#). [5-7]

An important additional parameter is the quadrupole moment coefficient of the earth,  $J_2 = 1.082629\ 821 \times 10^{-3}$ . The polar radius is  $b = a(1 - f)$  and the reference surface is an ellipsoid of revolution that approximates the actual figure of the earth. GPS receivers then determine altitudes as distances measured from the reference ellipsoid in the direction normal to it. In the GALILEO system, the angular rotation rate of the earth is specified with additional digits; in GLONASS, a slightly different value for the flattening is specified, entailing a slightly different reference ellipsoid; these differences are negligible except for the most precise work.

[Table 1](#). WGS84 values for earth parameters used to define the GPS reference surface.

Parameter	Description	Value
$GM$	$M$ is total earth	$3.986004418 \times 10^{14} \text{ m}^3/\text{s}^2$

mass		
$a_e$	Equatorial radius	6378 137.0 m
$f$	Flattening	1/298.257223 563
$\omega$	Angular velocity	7.292 115 × 10 <sup>-5</sup> radian/s

---

For navigation with satellite systems, there are four important earth rotation effects to be considered. It is simplest to describe these effects from the point of view of the International Celestial Reference Frame (ICRF). The first of these is precession, discovered by Hipparchus (ca. 190–129 BCE). Precession is the slow rotation of the pole of earth's rotation axis on the celestial sphere, having a period of 26 000 years. Second is nutation, rocking motions of the pole that were discovered in 1728 by Bradley. Dynamical theories of nutation are based on tidal forces due to other solar system bodies and internal motions of the earth, and give rise to large numbers of contributions. The largest is due to the moon and has a period of 18.6 years. Third is earth's rotation angle. Fourth is polar motion, a small deviation in earth's rotation axis relative to the solid earth that is apparently due to non-rigidity of earth's mass. The angular amplitude of this motion (including the Chandler wobble) is a few tenths of an arcsecond and is quite variable. Fluctuations in earth's rotation rate were not observed until after atomic clocks could provide a reference more stable than that of earth's rotation itself. For GNSS the net effect is a rotation matrix that transforms to the “Earth-Centered-Earth-Fixed” (ECEF) reference frame from the ECI frame at the epoch J2000. The rotation of the earth is monitored by the IERS, which provides products (available at <http://www.iers.org>) from which the net rotation can be calculated. Until recently within the GPS community the rotation was specified by an “ABCD” rotation matrix where “A” represents polar motion, “B” represents earth rotation, “C” represents nutation and “D” represents precession [5]. Since 2012, the GPS [8] specifies that these rotations be calculated by the SOFA (Standards of Fundamental Astronomy)

library of routines that is published and maintained by the IAU [9]. To discuss the impact of relativistic concepts, we need to consider a simple rotation matrix relating ECEF and ECI reference frames for a short period of time.

Accurate time measurement has advanced over the centuries to become a highly sophisticated science. Efforts to achieve measures that realize a uniform flow of time, in the absolute sense of Newton, have resulted in different non-relativistic time scales. For example, mean solar time is a reckoning of time based on the Sun's position in the sky. Sidereal time is a similar time scale based on apparent motions of the stars; both of these have difficulties due to the earth's orbital eccentricity and the obliquity of its rotation axis with respect to the plane of the orbit (ecliptic plane). Ephemeris time (ca. 1952–1970) used Newtonian theory to compensate for variations in earth's orbital motion. After the Cs atomic beam clock was invented, an atomic time scale was developed that was more uniform than ephemeris time; this has evolved into terrestrial time (TT) or International Atomic Time (TAI). Several relativistic effects enter into the realization of TAI; these will be discussed in the next sections. Universal Coordinated Time (UTC) differs from TAI by an integral number of leap seconds. Ensembles of atomic clocks such as those in the GPS constellation, or the GLONASS constellation, can be appropriately averaged and weighted to create independent time scales. For example, GPS time (except for leap seconds) is steered to UTC to within about 10 ns.

## 2 Relativity principles

To discuss relativistic relations between time scales, we summarize here the basic principles and results needed from the theory of relativity. The speed of light,  $c = 299\,792\,458$  m/s, is a universal constant, relative to any source of radiation in an inertial frame. Second, solutions of the field equations of General Relativity (GR) are represented by metric tensor functions  $g_{\mu\nu}(X^\alpha)$ , where the four coordinates that the metric

depends on are denoted by a Greek superscript taking on values (0, 1, 2, 3);  $X^0 = cT$ . Latin indices take on the values (1, 2, 3). (We use Einstein's summation convention.) The fundamental scalar invariant of the theory is

$$-ds^2 = g_{\mu\nu} dX^\mu dX^\nu.$$

(2)

The negative sign on the left side of [equation \(2\)](#) is chosen to correspond to positive proper time; we employ the metric signature  $(-1, 1, 1, 1)$ . We use capital letters to denote coordinates in a reference frame unattached to the geocenter. The proper time elapsed on an ideal clock moving along an arbitrary path from endpoint A to endpoint B is

$$\Delta\tau = \frac{1}{c} \int_A^B ds = \frac{1}{c} \int_A^B \sqrt{-g_{\mu\nu} dX^\mu dX^\nu}.$$

(3)

The universality of free fall is expressed by stating that the trajectory of an arbitrary freely falling test particle passing through endpoints A and B is a geodesic, which is an extremum of the integral in [equation \(3\)](#). In addition, the geodesic trajectory of a light ray satisfies  $d\tau = ds/c = 0$  at each point.

Excepting precise orbit determination and frame-dragging studies, most relativistic effects in the solar system can be described by keeping only the leading post-Newtonian terms in the solutions of the field equations. To order  $c^{-2}$ , the solutions in isotropic coordinates in vacuo are

$$g_{00} = -\left(1 + \frac{2\Phi}{c^2}\right); \quad g_{0i} = 0; \quad g_{ij} = \delta_{ij} \left(1 - \frac{2\Phi}{c^2}\right),$$

(4)

where  $\Phi$  is the Newtonian gravitational potential and  $\delta_{ik}$  is Kronecker's delta-function, zero unless the indices are equal. Inserting these solutions into [equation \(3\)](#),

$$\Delta\tau = \frac{1}{c} \int_A^B \left[ 1 + \frac{2\Phi}{c^2} - \left( 1 - \frac{2\Phi}{c^2} \right) \frac{dX^2 + dY^2 + dZ^2}{c^2 dT^2} \right]^{1/2} d(cT),$$

(5)

where we have extracted the common coordinate time factor  $d(cT)$ . The last terms can be more simply written in terms of the velocity of the clock since  $V^2 = (dX^2 + dY^2 + dZ^2)/dT^2$ . For a slowly moving object, the product  $\Phi V^2/c^4$  is negligible. Then expanding the square root in [equation \(5\)](#) to leading order,

$$\Delta\tau = \int_A^B \left( 1 + \frac{\Phi}{c^2} - \frac{1}{2} \frac{V^2}{c^2} \right) dT.$$

(6)

Thus an atomic clock at rest at an infinite distance away from any sources of potential would beat at the same proper rate as coordinate time  $T$ .

### 3 Astronomical and geocentric time scales

Equations presented in the previous section permit discussion of new time scales, and relationships between them, including relativistic effects. The definition of Geocentric Coordinate Time (TCG) depends heavily on the universality of free fall and the Principle of Equivalence. Let us separate the total gravitational potential into a part  $\phi$  due to the earth itself, and  $\Phi^{\text{ext}}$  due to external mass sources-other solar system bodies. Write the fundamental scalar invariant as

$$ds^2 = \left( 1 + \frac{2\phi}{c^2} + \frac{2\Phi^{\text{ext}}}{c^2} \right) (dX^0)^2 - \left( 1 - \frac{2\phi}{c^2} - \frac{2\Phi^{\text{ext}}}{c^2} \right) (dX^2 + dY^2 + dZ^2).$$

(7)

Let us transform this expression into the locally inertial, freely-falling frame with origin attached to earth's geocenter. Imagine a reference clock at the geocenter that experiences the external

potential but is not subject to the potential well of the earth itself. (In GR, arbitrary coordinate transformations are allowed as long as reasonable physical conditions are satisfied.) Appropriate transformations have been published to  $O(c^{-4})$ ; [10]; here we shall need only terms of  $O(c^{-2})$ .

The reference clock at the geocenter will be subject to time dilation due to its motion, and its rate will be affected by the external potential evaluated at the geocenter, which we denote by  $\Phi_0$ . The proper time on the reference clock is  $x^0/c$ ; the clock position is denoted by  $X_0^k(x^0)$  and its velocity is  $v_0^k$ . We take the time transformation to  $k$  be

$$X^0 = \int_{\text{path}} dx^0 \left( 1 - \frac{\Phi_0}{c^2} + \frac{1}{2} \frac{V_0^2}{c^2} \right) + \frac{V_0^i x^j \delta_{ij}}{c}. \quad (8)$$

Lower case letters are used for coordinates in the freely-falling frame. The last term comes from the breakdown of simultaneity and expresses resynchronization of local clocks displaced by  $x^k$  from the geocenter, due to motion of the earth. The spatial part of the transformation we take to be [10]

$$X^k = X_0^k + x^k \left( 1 + \frac{\Phi_0}{c^2} - \frac{A^i x^j \delta_{ij}}{c^2} \right) + \frac{1}{2} \frac{x^i x^j \delta_{ij} A^k}{c^2} + \frac{1}{2} \frac{V_0^k V_0^i x^j \delta_{ij}}{c^2}. \quad (9)$$

The last term in [equation \(9\)](#) is included to account for Lorentz contraction. The additional terms allow for a change of length scale due to the external potential, and some subtle acceleration terms for enforcement of the equivalence principle. We then differentiate to obtain

$$dX^0 = \left( 1 - \frac{\Phi_0}{c^2} + \frac{1}{2} \frac{V_0^2}{c^2} \right) dx^0 + \frac{A^i x^j \delta_{ij}}{c^2} dX^0 + \frac{V_0^i dx^j \delta_{ij}}{c} \quad (10)$$

$$dX^k = \left( 1 + \frac{\Phi_0}{c^2} - \frac{1}{2} \frac{A^i x^j \delta_{ij}}{c^2} \right) dx^k + \frac{V_0^i}{c} dX^0 + \frac{1}{2} \frac{V_0^k V^i dx^j \delta_{ij}}{c^2};$$

$$- \frac{x^k A^i dx^j \delta_{ij}}{c^2} + \frac{A^k x^i x^j \delta_{ij}}{c^2}.$$

(11)

Time dependences of acceleration and velocity have been neglected where they would contribute in higher order. Also, to the order of the calculation on the right sides,  $dX^0$  can be replaced by  $dx^0$ . Transformation coefficients can be identified from these two expressions. Substituting into [Equation \(7\)](#), after many cancellations the result is

$$ds^2 = \left( 1 + \frac{2\phi}{c^2} + \frac{2(\Phi^{\text{ext}} - \Phi_0 + A^i x^j \delta_{ij})}{c^2} \right) (dx^0)^2$$

$$- \left( 1 - \frac{2\phi}{c^2} - \frac{2(\Phi^{\text{ext}} - \Phi_0 + A^i x^j \delta_{ij})}{c^2} \right) (dx^2 + dy^2 + dz^2).$$

(12)

Further cancellations in the external contributions can occur; we have not yet imposed the condition that the reference clock at the geocenter is in free fall. Developing a Taylor expansion for the external potential about the geocenter,

$$\Phi^{\text{ext}} = \Phi_0 + \Phi_{,i}^{\text{ext}} (X^i - X^i(X^0)) + \frac{1}{2} \Phi_{,ij}^{\text{ext}} (X^i - X^i(X^0))(X^j - X^j(X^0)).$$

(13)

The coefficients are evaluated at the geocenter; also, from [equation \(9\)](#), since the contributions already have a factor  $c^{-2}$ , the coordinate differences in this expansion may be replaced by local coordinate differences:

$$\Phi^{\text{ext}} = \Phi_0 + \Phi_{,i}^{\text{ext}} x^i + \frac{1}{2} \Phi_{,ij}^{\text{ext}} x^i x^j.$$

(14)

The quadratic terms are tidal terms. If these terms are neglected then there remain terms of the form

$$(\Phi_{,i}^{\text{ext}} + A^i) x^j \delta_{ij}. \quad (15)$$

However if the earth is in free fall,  $A^i = -\Phi_{,i}^{\text{ext}}$  and only tidal terms remain as contributions from the external potentials. In a locally inertial, freely-falling frame, the gravitational field strength at the origin is cancelled by the acceleration of the origin due to the external potentials. This enforces Einstein's Principle of Equivalence. The net metric is

$$ds^2 = \left( 1 + \frac{2\phi}{c^2} + \frac{\Phi_{,ij}^{\text{ext}} x^i x^j}{c^2} \right) (dx^0)^2 - \left( 1 - \frac{2\phi}{c^2} - \frac{\Phi_{,ij}^{\text{ext}} x^i x^j}{c^2} \right) (dx^2 + dy^2 + dz^2). \quad (16)$$

This discussion of the Principle of Equivalence is well-known, and perhaps obvious, to most students of relativity. It is included here to address one of the most common recurring incorrect criticisms of relativity in GPS, namely the argument: "The sun's potential on the day side of the earth is lower than it is on the night side, so satellite clocks on the day side should beat more slowly and this has not been accounted for in GPS." The argument overlooks the cancellation of the real solar gravitational field by the effective gravitational field induced by acceleration of the freely-falling earth.

Also incorrect, is the claim that position calculations performed by the receiver must be performed in the receiver's own rest frame. The Principle of General Covariance permits analysis in any admissible coordinate system.

The time coordinate defined by [equation \(16\)](#) is Geocentric Coordinate Time (TCG -Temps-Coordonnée Géocentrique). The unit of time is the SI second. TCG is intended to be used as the independent variable of time for calculations pertaining to precession, nutation, the Moon, and artificial satellites of the earth. It was defined in 1991 by the IAU, in Recommendation III of the XXIst General Assembly. [3, 4]. TCG is defined in the

context of general relativity.

The relationships between TCG and other relativistic time scales are defined with fully general relativistic metrics. For example, the relationship between TCG and Barycentric Coordinate Time (TCB) is defined by a relation similar to [equation \(8\)](#):

$$TCB = TGC + \frac{1}{c^2} \left( \int_{t_0}^t \left[ -\Phi_e^{\text{ext}} + \frac{V_e^2}{2} \right] dt + V_e \cdot (\mathbf{x} - \mathbf{x}_e) \right), \quad (17)$$

where the subscript ‘e’ refers to the geocenter. The difference between these two times at JD 2443144.5 is  $-6.55 \times 10^{-5}$  s [3]. The velocity and external potential at the geocenter are complicated by the many perturbations of earth’s orbit from other solar system bodies. In the resynchronization term, to the order of present calculations it is immaterial whether TCG or TCB spatial coordinates are used. To obtain better than 100 ns accuracy from [equation \(17\)](#), more than 100 terms of many different frequencies must be included in the integral. The leading terms in [equation \(17\)](#) can be calculated by modeling earth’s motion as an unperturbed Keplerian orbit with the sun the only external mass. In such an orbit, energy is conserved and

$$\frac{V_e^2}{2} + \Phi_e^{\text{ext}} = -\frac{GM_{\odot}}{2a}, \quad (18)$$

where  $a = 1.49597870\ 700 \times 10^{11}$  m is the orbit semi-major axis and the sun’s mass is  $M_{\odot} = 1.32712442\ 099 \times 10^{20}$  m<sup>3</sup> s<sup>-3</sup>. The heliocentric radius is

$$r = a(1 - e \cos E), \quad (19)$$

where  $e$  is the orbit eccentricity and  $E$  is the eccentric anomaly,

defined by solving the transcendental equation

$$E - e \sin E = \sqrt{\frac{GM_{\odot}}{a^3}}(t - t_p) , \quad (20)$$

where  $t_p$  is the time of perihelion passage. The integral in [equation \(17\)](#) is performed by changing the independent variable from  $t$  to  $E$ . Differentiating [equation \(20\)](#), we have  $dE(1 - e \cos E) = \sqrt{GM_{\odot}/a^3} dt$ . Then integration gives

$$\frac{1}{c^2} \int_{t_0}^t \left( \frac{V_e^2}{2} + \Phi_e^{\text{ext}} \right) dt = \frac{3GM_{\odot}}{2c^2 a}(t - t_0) + \frac{2\sqrt{GM_{\odot}a}e}{c^2}(\sin E - \sin E_0) .$$

(21)

The value of the coefficient in the first term is

$$L_C \approx \frac{3GM_{\odot}}{2c^2 a} = 1.48059 \times 10^{-8} , \quad (22)$$

whereas the more accurate value recommended by the IAU is  $L_C = 1.48082686741 \times 10^{-8}$ . [3] [Equation \(21\)](#) will arise later in connection with similar relativistic corrections in GNSS.

Barycentric time scales are important for astronomical timing studies since the specific frequency and time variations due to motions of the earth and its nearby satellites can be accounted for by reducing observations to the solar system barycenter. For example, measurements of the arrival time of photons from pulsars by x-ray observatories such as FERMI or CHANDRA have to be reduced to the barycenter. Some software systems that perform such reductions use the TDB time scale, which is related to TCG by

$$TCG - TDB = L_B(MJD - 43144) \times 86400 \text{ s} , \quad (23)$$

where  $L_B$  is a relativistic correction having the approximate value

$$L_B = 1.55051976 \times 10^{-8}.$$

(24)

The introduction of TDB was an attempt to introduce an average uniform time scale for ephemerides with respect to the solar system barycenter. However the solar system is not periodic, so no well-defined average exists. One can find different values of  $L_B$  incorporated into different barycentric reduction programs. TDB applies to the solar-system-barycentric reference frame, and was first defined in 1976 as a successor to the (non-relativistic) former standard of ephemeris time. In 2006, after a history of multiple time-scale redefinitions, a new definition of TDB was approved by the IAU. TDB according to the 2006 redefinition can now be treated as equivalent, for practical astronomical purposes, to the long-established JPL ephemeris time argument  $T_{\text{eph}}$  as implemented in JPL Development Ephemeris DE405 [11] (in use as the official basis for planetary and lunar ephemerides in the Astronomical Almanac, editions for 2003 and succeeding years). [12]

## 4 Earth-based time scales TT, TAI, UTC

If it were not for dynamical processes such as those arising from solar and lunar tides, crustal motions, etc., then in the earth-fixed rotating frame the surface of the ocean would assume a smooth, nearly ellipsoidal shape. This shape (approximating mean sea level) is a surface of effectively constant gravitational potential in the rotating frame, for if there were a (this already assumes equivalence) potential gradient component directed within the surface, fluid would have to flow. Starting in the ICRF (inertial frame of reference attached to earth's geocenter), consider a transformation to a reference frame fixed in the earth (ECEF). Denote the new spatial coordinates by  $(x', y', z')$ .

We consider a model in which the entire potential is that due to the earth, denoted by  $\phi$  and assumed axially symmetric, and we transform to a rotating frame with axis instantaneously pointing in the direction of earth's symmetry axis. The transformation is a rigid non-relativistic rotation with constant angular velocity:

$$\begin{aligned}x &= x' \cos \omega t - y' \sin \omega t; \\y &= x' \sin \omega t + y' \cos \omega t; \\z &= z'; \quad t' = t.\end{aligned}\tag{25}$$

No change in the time coordinate is made. The time  $t = x^0 / c$  in the rotating frame is the same as the time in the underlying inertial frame, and so is not a "proper" time. Neglect tidal terms for the moment. In the rotating frame the metric [equation \(16\)](#) becomes

$$\begin{aligned}ds^2 &= \left(1 + \frac{2\phi}{c^2} - \frac{\omega^2(x'^2 + y'^2)}{c^2}\right)(dx^0)^2 - \left(1 - \frac{2\phi}{c^2}\right)(dx'^2 + dy'^2 + dz'^2) \\&\quad - \frac{2\omega}{c}(x'dy' - y'dx')dx^0.\end{aligned}\tag{26}$$

An approximate expression for earth's potential can now be inserted. For the GNSS, it is usually sufficient to consider only the monopole and quadrupole contributions to the potential, which are time-independent:

$$\phi = -\frac{GM}{r'} \left(1 - \frac{J_2 a_e^2}{r'^2} \frac{3z'^2 - r'^2}{2r'^2}\right)\tag{27}$$

where  $r^2 = x'^2 + y'^2$ . For a fluid element at rest in the rotating frame,

$$ds^2 = \left(1 - \frac{2GM}{r'c^2} \left(1 - \frac{J_2 a_e^2}{r'^2} \frac{3z'^2 - r'^2}{2r'^2}\right) - \frac{\omega^2 r'^2}{c^2}\right)(dx^0)^2\tag{28}$$

The effective gravitational potential, divided by  $c^2$ , in the rotating frame is then given at the equator by

$$\frac{2\phi_{\text{eff}}}{c^2} = -\frac{2GM}{a_e c^2} \left(1 + \frac{J_2}{2}\right) - \frac{\omega^2 a_e^2}{c^2}$$

(29)

$$\frac{\phi_{\text{eff}}}{c^2} = -6.969\,283 \times 10^{-10}.$$

(30)

This is the numerical value of a model surface that is an effective gravitational equipotential in the rotating frame; this surface is not exactly an ellipsoid of revolution but it is conventional in GNSS to define an ellipsoidal reference surface that has a numerical value very close to that given in [equation \(30\)](#). In the GPS the ellipsoidal reference surface is defined in such a way that the potential on the surface is

$$\phi_{\text{GPS}}/c^2 = -6.969\,289\,657 \times 10^{-10}.$$

(31)

Clocks at rest on this surface beat more slowly than clocks at rest at infinity. The constant estimated in [equation \(30\)](#) is, to within the limitations of the potential model, the same as the constant  $L_G$  defined by the IAU as the value on earth's geoid:

$$L_G = |W_0|/c^2 \equiv 6.969\,290\,134 \times 10^{-10},$$

(32)

where  $W_0$  is earth's gravitational potential on the geoid, including the centripetal term [3]. The constant  $L_G$  takes account of higher multipolar contributions to earth's potential. The differences among these values are almost insensible at the current time; however optical clocks with stabilities of order  $10^{-18}$  are currently under development; such clocks when flown in space will require that close attention be paid to small

geophysical effects that contribute to these differences.

International Atomic Time (TAI) is defined by averaging the readings of many clocks in time standards laboratories on the geoid of the rotating earth. The relationship between the rates of TAI ( $t'$ ) and TCG ( $x^0/c$ ) is defined by:

$$d(TAI) = (1 - L_G)d(TCG). \quad (33)$$

## 5 Gravitational frequency shifts

We perform an additional transformation on the metric in the ECEF frame, [equation \(26\)](#), by transforming to TAI (now represented by  $t'$ ) by means of

$$dx^0 = cdt'(1 + L_G); \quad (34)$$

then to order  $c^{-2}$ ,

$$\begin{aligned} ds^2 = & \left( 1 + \frac{2\phi}{c^2} + 2L_G - \frac{\omega^2(x'^2 + y'^2)}{c^2} \right) (cdt')^2 - \left( 1 - \frac{2\phi}{c^2} \right) (dx'^2 + dy'^2 + dz'^2) \\ & - \frac{2\omega}{c} (x' dy' - y' dx') cdt'. \end{aligned} \quad (35)$$

Here the symbols ( $t'$ ,  $x'$ ,  $y'$ ,  $z'$ ) refer to coordinates fixed in the earth. When the time variable changes, the unit of length changes since  $c$  is a defined constant. For fluid elements at rest on earth, [equation \(35\)](#) becomes

$$ds^2 = \left( 1 + \frac{2\phi}{c^2} + 2L_G - \frac{\omega^2(x'^2 + y'^2)}{c^2} \right) (cdt')^2, \quad (36)$$

and earth's geoid is the surface

$$\phi_{\text{eff}} = \phi + c^2 L_G - \frac{1}{2} \omega^2 (x'^2 + y'^2) = 0. \quad (37)$$

An atom at rest on the rotating geoid will satisfy the equation  $ds = cd\tau$ , so TAI is a time determined by averaging the proper time readings of several hundred clocks at rest on earth's rotating geoid.

For a clock at a fixed geographic location, at a geographically vertical distance  $z$  above the geoid, the net effective gravitational potential can be obtained by a first-order Taylor expansion of the effective potential, yielding

$$ds^2 = \left(1 + \frac{2gz}{c^2}\right) (cdt')^2, \quad (38)$$

where  $g$  is the local vertical value of the acceleration of gravity. The direction of the acceleration of gravity will lie at an angle from the equatorial plane equal to the geographic latitude. Taking a square root, the proper time elapsed during coordinate time interval  $dt'$  is

$$d\tau = ds/c = \left(1 + \frac{gz}{c^2}\right) (dt'), \quad (39)$$

When  $z$  is above the geoid, the proper time will elapse more rapidly than that on a reference clock on the geoid. This can be interpreted as an increased frequency of the cyclic phenomenon that is counted by the clock. The fractional frequency increase is

$$\frac{\Delta f}{f} = \frac{gz}{c^2}. \quad (40)$$

Time standards laboratories that maintain an atomic clock at a height  $z$  above the geoid must reduce their frequency measurements to the geoid by applying a correction  $-gz/c^2$

before transmitting their frequency to the BIPM for averaging to determine TAI. For example, NIST (National Institute of Standards and Technology) in Boulder, CO, USA, is at an altitude of approximately 1650 m. The correction amounts to about  $-15.5$  ns per day.

Consider a laboratory with a cesium fountain clock at a fixed geographic location. Atoms in an atomic fountain are launched upwards through a cavity that prepares their state; they then rise upward to some height  $h$  above the starting point, and then fall back down through the cavity and are detected. While in flight their gravitational potentials and velocities change, so they suffer variations in gravitational frequency shift and second-order Doppler shift. By a series of translations and rotations, a coordinate system can be introduced with origin at the launch point of the fountain (this point is not necessarily on the geoid) and  $z$ -direction locally vertical, and negligible motion of the atoms in the horizontal direction. The second line of [equation \(35\)](#) is negligible, and the potential term multiplying the spatial part of the metric in [equation \(35\)](#) can also be neglected since atom velocities are slow. The scalar invariant takes the form

$$ds^2 = \left( 1 + \frac{2gz}{c^2} - \frac{v'^2}{c^2} \right) (cdt')^2 . \quad (41)$$

The proper time elapsed on an atom is then

$$d\tau = \left( 1 + \frac{gz}{c^2} - \frac{1}{2} \frac{v^2}{c^2} \right) dt' . \quad (42)$$

The trajectory of an atom with launch velocity  $v_0$  is given by

$$z = v_0 t - \frac{1}{2} g t^2 ; \quad v = v_0 - g t . \quad (43)$$

For an atom that rises to height  $h$ , the launch velocity is  $v_0 = \sqrt{2gh}$

and the total time of flight is  $t_f = \sqrt{8h/g}$ . Substituting into [equation \(42\)](#) and performing the integration, the proper time on the launched atom is found to be

$$\tau = \sqrt{\frac{8h}{g}} \left( 1 + \frac{gh}{3c^2} \right). \quad (44)$$

The proper time elapsed on a reference atom that remains at rest at the launch point is just  $\tau_{ref} = \sqrt{8h/g}$ . Therefore the fractional change in the elapsed proper time of the atom, which can be interpreted as an average fractional frequency shift, is

$$\frac{\tau - \tau_{ref}}{\tau_{ref}} = \frac{\delta f}{f} = \frac{1}{3} \frac{gh}{c^2}. \quad (45)$$

If  $h \approx 1$  m, this is a fractional frequency shift of a few parts in  $10^{-17}$ , and is a significant systematic effect that must be accounted for in the error budgets of modern cesium fountains. Thus the best atomic clocks in the world require a relativistic correction due to their size.

## 6 Sagnac effect; realizing coordinate time

It is already well-known that the Sagnac effect prevents self-consistent synchronization of clocks on the surface of the rotating earth by ordinary synchronization processes including transport of portable clocks or transmission of electromagnetic signals. [14] This effect arises from the last term in [equation \(35\)](#). Suppose an atomic clock is transported from one point to another on the rotating geoid, so slowly that the spatial quadratic terms in [equation \(35\)](#) can be neglected. This is equivalent to neglecting second-order Doppler terms such as those calculated in the previous section. Then we obtain the special case

$$ds^2 = \left(1 - \frac{2\omega}{c^2} \left(x' \frac{dy'}{dt'} - y' \frac{dx'}{dt'}\right)\right) (cdt')^2. \quad (46)$$

Taking a square root the proper time elapsed on the slowly moving clock is

$$\begin{aligned} d\tau &= \left(1 - \frac{\omega}{c^2} \left(x' \frac{dy'}{dt'} - y' \frac{dx'}{dt'}\right)\right) dt' \\ &= dt' - \frac{\omega}{c^2} (x' dy' - y' dx'). \end{aligned} \quad (47)$$

The second term in [equation \(47\)](#) depends on the path of the transported clock, and can contribute a finite amount even if the clock moves extremely slowly. Imagine a vector from earth's rotation axis to the clock that when projected on a plane parallel to the equatorial plane sweeps out an area  $dA'$ , where  $dA'$  is positive if the path is eastward and negative if the path is westward. Since

$$dA' = \frac{1}{2} (x' dy' - y' dx'), \quad (48)$$

the additional effect when integrated along the path is

$$-\frac{\omega}{c^2} \int_{\text{path}} (x' dy - y' dx) = -\frac{2\omega}{c^2} A'. \quad (49)$$

For example carrying a clock eastward once around the equator, the proper time is less than that on an atomic clock that remains at rest by about 204 ns. This is basically because the transported clock has to move faster than the clock that remains at rest on the earth so suffers more time dilation, while if the transported clock moves westward the effect has the other sign. The effect vanishes if the clock is transported along a meridian as the projected area is zero.

Clocks distributed on the rotating earth that beat at the rate of the SI second can be synchronized by means of transporting clocks by recognizing that TAI is a realization of the coordinate time variable  $t'$ . Solving [equation \(47\)](#) for the coordinate time increment,

$$dt' = d\tau + \frac{\omega}{c^2} \int_{\text{path}} (x' dy - y' dx) = d\tau + \frac{2\omega^2}{c} dA' .$$

[\(50\)](#)

thus one corrects the measured proper time increment for the Sagnac effect. [Equation \(50\)](#) can be integrated along the path followed by the clock. Such corrections are applied in the ECEF; they do not exist in the ECI reference frame.

Einstein synchronization is based on constancy of the speed of electromagnetic signals. In GNSS, right circularly polarized electromagnetic signals carry the timing information; timing pulses can be thought of as places in the transmitted wave trains where there is a particular phase reversal of the circularly polarized electromagnetic signals. At such places the electromagnetic field tensor passes through zero and therefore provides relatively moving observers with sequences of events that they can agree on in principle.

For such signals, the scalar invariant  $ds$  is zero at every point along the path. Then [equation \(35\)](#) can be approximately written as

$$(cdt')^2 = \left(1 - \frac{2(\phi + \phi_{\text{eff}})}{c^2}\right) d\ell'^2 + \frac{4\omega}{c} dA' cdt' ,$$

(51)

where  $d\ell'^2 = dx'^2 + dy'^2 + dz'^2$ . To a very good approximation,  $dt' = d\ell'/c$ . Then taking a square root and keeping only correction terms of order  $c^{-2}$ ,

$$cdt' = \left(1 - \frac{(\phi + \phi_{\text{eff}})}{c^2}\right) d\ell' + \frac{2\omega dA'}{c} \frac{cdt'}{d\ell'}$$

$$dt' = \frac{1}{c} \left(1 - \frac{(\phi + \phi_{\text{eff}})}{c^2}\right) d\ell' + \frac{2\omega}{c^2} dA'.$$

(52)

Once again imagining a vector from earth's rotation axis to the position of the electromagnetic pulse, the coordinate time required for the light to traverse a synchronization path must be corrected by an amount that can be described by the same words used to describe synchronization by transporting an atomic clock. After such corrections are applied the coordinate clocks will be self-consistently synchronized in the ECEF. A coordinate clock will read the same time as a hypothetical clock at rest at the same instantaneous position in the underlying inertial frame, which has been synchronized using relativistic principles applied in the inertial frame. The potential terms in [equation \(52\)](#) give very small corrections due to coordinate slowing of the electromagnetic synchronization pulses, and can be neglected for most uses of electromagnetic signals near earth.

International Committees have recognized the need to apply such corrections. The International Radio Consultative Committee (CCIR) and the Committee for the Definition of the Second have recommended adoption of such procedures and indeed they have been adopted worldwide. [15, 16]

Remotely situated time standards are routinely compared by methods utilizing GNSS satellites. A single satellite transmits a coded train of pulses, which are received concurrently by two distant receivers whose time bases are supplied by clocks in their respective standards laboratories. This is called a "common-view" clock comparison. The signals basically transmit the satellite's clock time to the earth-based clocks. After accounting for propagation delays and then applying the Sagnac corrections, which can be hundreds of nanoseconds, the earth-based clocks can be compared since the satellite clock time drops out of the comparison. Another comparison method called "two-way" involves a transponder in a geosynchronous

satellite. Coded signals are transmitted concurrently from two ground-based laboratories, and relayed through the transponder to the other laboratory. Many common error sources are cancelled during the comparison process but the Sagnac corrections can be even larger because a geosynchronous satellite is further from the earth, consequently the projected area can be larger than in common-view.

## 7 Relativistic effects on orbiting clocks

Operating GPS satellites each contain an atomic clock that is synchronized in coordinate time. The synchronization is “on paper;” the navigation message specifies the corrections that must be applied to the transmitted time. It is most convenient to view these clocks from the ECI frame since then there are no Sagnac corrections and the orbits are Keplerian to a very good first approximation. Also, the speeds of these clocks are small (less than 7 km/s) so terms of order  $v^2 \phi/c^4$  can be neglected. We also neglect tidal perturbations due to external solar system bodies. The appropriate time scale is TAI, so the metric used to model the motion of the satellites is

$$ds^2 = \left(1 + \frac{2\phi}{c^2} + 2L_G\right)(cdt')^2 - dx^2 - dy^2 - dz^2. \quad (53)$$

For a slowly moving clock in a satellite, this equation can be solved for  $dt'$  by factoring out  $(cdt')^2$ , taking a square root, and writing  $dt'$  on the left side of the equation:

$$dt' = \left(1 - \frac{\phi}{c^2} - L_G + \frac{1}{2} \frac{v^2}{c^2}\right) d\tau. \quad (54)$$

This can be integrated along the path of the orbiting clock to obtain the elapsed coordinate time, using the actual proper time on the clock as the independent variable. Since the relativistic

term already has a factor  $c^{-2}$ , the increment  $d\tau$  can be replaced by  $dt$  without appreciable error. Then we can write

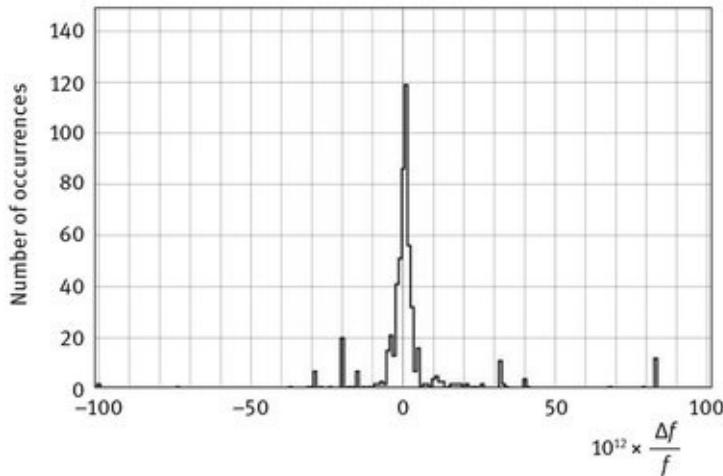
$$\Delta t = \Delta\tau + \int_{path} \left( -\frac{\phi}{c^2} - L_G + \frac{1}{2} \frac{v^2}{c^2} \right) dt.$$

(55)

For satellites sufficiently far from earth that the quadrupole and higher multipole moments of earth's gravitational potential can be neglected,  $\phi = -GM/r$  and energy (per unit mass) is conserved in the Keplerian orbit:

$$\frac{1}{2} v^2 - \frac{GM}{r} = -\frac{GM}{2a}.$$

(56)



**Fig. 1.** Histogram of factory frequency offsets of GPS satellite clocks. Values extracted from broadcast ephemerides for noon, January 1, for all years 1992 through 2014; 595 samples are plotted. Bins are of width  $10^{-12}$ .

Eliminating the velocity term from [equation \(55\)](#), the elapsed time on the coordinate clock can be expressed as

$$\Delta t = \Delta\tau + \int_{path} \left( \frac{3GM}{2ac^2} - L_G + \frac{2GM}{c^2} \left( \frac{1}{r} - \frac{1}{a} \right) \right) dt.$$

(57)

The form of the first term in this expression has already been seen: it is of the same form as the leading term in [equation \(21\)](#). In the GPS the first two constant rate correction terms have the values

$$\frac{3GM}{2ac^2} - L_G = +2.5046 \times 10^{-10} - 6.9693 \times 10^{-10} = -4.4647 \times 10^{-10}.$$

(58)

The resulting negative sign means that the orbiting atomic clock is beating too fast, since the gravitational blueshift from the high altitude is larger in magnitude than the redshift due to time dilation. The best GPS clocks have stabilities of order a few parts in  $10^{15}$  after averaging for one day. The effect calculated in [equation \(58\)](#) is four orders of magnitude greater than this. The GPS system is designed so that signals reaching the ground have frequency 10.23 MHz exactly, so the atomic clocks are adjusted before launch so their frequencies are

$$10.23 \times (1 - 4.4647 \times 10^{-10}) = 10.229\,999\,995\,43 \text{ MHz}.$$

(59)

Knowing the orbital period allows one to compute the offset required so that transmitted signals have the desired frequencies when they reach the ground. This design minimizes the relativistic corrections that must be applied by the receiver. For example, GLONASS satellites go around the earth 17 times during the time taken by GPS satellites to go around 16 times; GALILEO satellite ground tracks repeat every ten days, so the frequency offset corrections differ. Whereas these frequency offsets are specified explicitly in the ICD-200 Signal-in-Space documents for both GPS and GLONASS, no information is publicly available about such offsets for GALILEO or BEIDOU. The available documentation for GALILEO and BEIDOU does not mention relativity at all. Nevertheless the specified navigation messages are interoperable with GPS and GLONASS and allow for transmission of coefficients  $\alpha_0$ ,  $\alpha_1$ ,  $\alpha_2$  in the navigation

messages, just as in GPS and GLONASS. These coefficients represent corrections to the transmitted satellite clock time, clock frequency, and clock frequency drift, respectively. Frequency offsets for clocks in several satellite systems are estimated in [Table 2](#). In GPS it would be ideal to have the clock frequency correction exactly equal to zero; this would mean that the calculated frequency offset is correctly realized. [Figure 1 shows](#) a histogram of the so-called “factory frequency offset corrections” (the  $\alpha_1$  coefficients) of GPS clocks, taken from the GPS broadcast ephemerides for all the satellites at noon on January 1 each year for the years 1992 through 2014. Note that the units of  $\Delta f/f$  in [Table 2](#) are parts in  $10^{12}$ , while the bin sizes of the histogram in [Figure 1 are](#)  $10^{-12}$  so that if the frequency offsets had not been applied, the histogram peak would be several hundred units off to the side.

A correction whose purpose is to compensate for a relativistic effect necessarily has a sign opposite to that of the effect; the relativistic fractional frequency shift effect corresponding to the correction discussed in [equation \(58\)](#) is

$$\frac{\delta f}{f} = L_G - \frac{3GM}{2ac^2}. \quad (60)$$

If the orbit were adjusted by applying thrusts that change  $\alpha$  to  $\alpha + \Delta\alpha$ ,  $L_G$  does not change but the change in the frequency shift would be

$$\Delta\left(\frac{\delta f}{f}\right) = +\frac{3GM\Delta\alpha}{2a^2c^2}. \quad (61)$$

For a typical increase in height  $\Delta\alpha = 20$  km of a GPS satellite, the fractional frequency shift is  $+1.89 \times 10^{-13}$ . Such frequency shifts were observed for years but were not understood until explained as a relativistic effect [19, 20].

**Table 2.** Estimated values of SV fractional frequency clock offsets.

<b>GNSS System</b>	<b>a (km)</b>	<b><math>10^{12} \times \Delta f/f</math></b>
GLONASS	25 509.64	-436.144
GPS	26 562.76	-446.473
BEIDOU(MEO)	27 910.20	-458.538
GALILEO	29 601.31	-472.191
Geosynchronous	42 164.17	-539.151

## 8 The eccentricity effect

Due to earth's orbital eccentricity in its revolution about the sun, there is a contribution proportional to the eccentricity in the relation between TCG and TDB, [equation \(21\)](#). An analogous phenomenon arises from the last two terms in [equation \(57\)](#). For a Keplerian orbit about the earth, [equation \(20\)](#) gives in terms of the satellite's eccentric anomaly  $E$ ,

$$dt = \sqrt{\frac{a^3}{GM}} \frac{r}{a} dE, \quad (62)$$

so the last two terms in [equation \(57\)](#) give

$$\begin{aligned} & \int_{path} \frac{2GM(a-r)}{c^2 ar} \sqrt{\frac{a^3}{GM}} \frac{r}{a} dE \\ &= \frac{2\sqrt{GMae}}{c^2} \int \cos EdE = \frac{2\sqrt{GMae}}{c^2} (\sin E - \sin E_0). \end{aligned} \quad (63)$$

The constant term is lumped with other constant clock error corrections, while the first term, proportional to  $e \sin E$ , is corrected in the receiver. In the GPS, satellite eccentricity is

specified to be less than 0.030; for this eccentricity value the correction has an amplitude of 69 ns. The navigation message passes quantities such as  $e$  and  $\sqrt{a}$  to the receiver so that it can make this correction. This procedure was designed in the early days of GPS, but is a legacy that other GNSS systems adopt in order to be interoperable with GPS. Nowadays such an effect could be compensated by onboard software.

Some GNSS augmentation systems have satellites with large eccentricities and orbits designed to spend a considerable amount of time over some locality. For example, Japan's QZSS (Quasi-Zenith Satellite System) is designed to have four or more geosynchronous satellites with eccentricities  $0.075 \pm .015$ . Due to conservation of angular momentum, at apogee the satellites travel more slowly and the orbits can be designed so that more time is spent over a specific location. For an eccentricity of 0.08, the QZSS time correction has an amplitude of 231 ns. (100 ns error would lead to a navigation error of about 30 m.)

## 9 Navigation on the rotating earth

Consider the trajectory of an electromagnetic pulse as it travels from a GNSS transmitter at event  $t'_j, \mathbf{r}'_j$  to the receiver at event  $t, \mathbf{r}$ . It is simplest to view this process in the ECI reference frame, since complications such as the Sagnac effect can then be omitted. However, terms in the metric of order  $v^2 \phi/c^2$  can no longer be neglected, as they could when discussing slow motion of satellites. The appropriate metric is obtained from [equation \(16\)](#) where the time variable  $t'$  corresponds to TT or TAI, and  $\omega$  is set equal to zero. The scalar interval is null along the path so

$$ds^2 = 0 = \left(1 + \frac{2\phi}{c^2} + 2L_G\right)(cdt')^2 - \left(1 - \frac{2\phi}{c^2}\right)(dx^2 + dy^2 + dz^2). \quad (64)$$

Solving for the coordinate time increment  $dt'$ ,

$$dt' = \frac{1}{c} \left( 1 - \frac{2\phi}{c^2} - L_G \right) \sqrt{dx^2 + dy^2 + dz^2}. \quad (65)$$

The coordinate distance increment along the path is

$$d\ell = \sqrt{dx^2 + dy^2 + dz^2}. \quad (66)$$

We work entirely with coordinate distances and coordinate times, even though the proper time on a clock at rest on the geoid is the same as coordinate time. Using only the monopole contribution to earth's gravitational potential, the coordinate time elapsed along the trajectory is

$$\Delta t = (1 - L_G) \frac{\Delta\ell}{c} + \frac{2GM}{c^3} \int_{path} \frac{d\ell}{r}. \quad (67)$$

To a very good approximation, the integral in [equation \(67\)](#) can be taken along a straight line between the endpoints where it leads to a well-known expression for the Shapiro delay: [22]

$$\Delta t - (1 - L_G) \frac{\Delta\ell}{c} = \frac{2GM}{c^3} \ln \left( \frac{|\mathbf{r}| + |\mathbf{r}'| + |\mathbf{r}' - \mathbf{r}|}{|\mathbf{r}| + |\mathbf{r}'| - |\mathbf{r}' - \mathbf{r}|} \right). \quad (68)$$

For the earth,  $2GM/c^3 = 29$  ps, while the logarithm is of general order unity. For the  $\approx 100$  ms propagation time in the GNSS, the  $L_G$  correction in the above equation would be less than about 0.1 ps. Such corrections are of minor importance except in the most demanding of time comparison measurements. In the future a detailed consideration of coordinate distance scaling may be required. The proper distance corresponding to  $d\ell$  is  $d\sigma = \sqrt{1 - 2\phi/c^2} d\ell$ . The difference between the proper distance around earth's equator, and the coordinate distance, is less than 3 cm.

Consider now signals from several satellites to a single receiver. In the ECI frame, we may assume that light travels

with speed  $c$  and neglect coordinate slowing. The equations to be solved are:

$$c(t - t_j) = |\mathbf{r} - \mathbf{r}_j|, \quad j = 1, 2, 3\dots$$

(69)

where the minimum number of signals is four but there will be more in general since many satellites may be in view. We use unprimed symbols to denote ECI quantities. The information about the satellite position at transmission time  $t_j$  is contained in the navigation message. However, the transmitter position at this instant is specified by a set of parameters that provides positions  $\mathbf{r}'_j$  in the ECEF frame *rather than in the ECI frame*. Further, the distances between the satellites and the receivers are generally unequal, so the transmitter positions are specified at different times in the ECEF frame. This can create confusion, for it is incorrect to solve [equations \(69\)](#) in the ECEF frame. One way to deal with this might be to apply Sagnac corrections to the arrival time but this deepens the possibility of confusion because even though the receiver presents the signals at a particular time, the corrections would all be different. If this issue is not dealt with correctly, errors of many meters in the estimate of receiver position can result.

It is clearer to recognize that the positions of the satellites in the underlying ECI frame may be rotated by different amounts into their respective ECEF frames, because their transmission times differ. Imagine then an ECI frame fixed at some arbitrary time  $t_0$ . It would be desirable if this time is not too different from the transmission times from the various satellites. Let the rotation matrix that takes the vector  $\mathbf{r}'_j$  at time  $t_j$  to  $\mathbf{r}_j$  be  $R[t_j - t_0]$ . The position of the transmission event in the chosen ECI frame is

$$\mathbf{r}_j = R[t_j - t_0] \mathbf{r}'_j.$$

(70)

The navigation equations are then expressed in the chosen ECI frame as in [equation \(69\)](#). Such a system of nonlinear equations may be solved by iteration: a first approximation is guessed, the equations are linearized about the assumed solution, and a set of corrections is estimated. The guess is corrected and the process is repeated until the solution converges. For example, suppose the first guess is

$$t = t^{(0)} + \delta t, \quad \mathbf{r} = \mathbf{r}^{(0)} + \delta \mathbf{r}. \quad (71)$$

The linearized navigation equations can be put in the form

$$c\delta t - \mathbf{e}_j \cdot \delta \mathbf{r} = |\mathbf{r}^{(0)} - \mathbf{r}_j| - c(t^{(0)} - t_j), \quad (72)$$

where  $\mathbf{e}_j$  is a unit vector from the satellite position to the assumed received position:

$$\mathbf{e}_j = \frac{\mathbf{r}^{(0)} - \mathbf{r}_j}{|\mathbf{r}^{(0)} - \mathbf{r}_j|}. \quad (73)$$

[Equations \(72\)](#) then form a system of linear equations for the corrections to the guesses. Iteration will yield a solution when the right sides of [equations \(72\)](#) vanish. If there are more than four signals available, there will be more than four equations determining the position and time and the receiver; there are matrix methods available to accomplish this. For example, if there are  $n > 4$  equations and 4 unknowns represented by a column matrix  $X$ , then the linearized equations can be put in the form  $GX = M$ . Multiplying by the transpose of  $G$ ,  $G^T GX = G^T M$  and this will be a system of four equations for four unknowns, following as a necessary condition from the navigation equations provided the geometry of the transmitter configuration permits the existence of the inverse  $(G^T G)^{-1}$ .

Once a solution is found, the position of the receiver in the

ECEF frame will be simply  $R[t - t_0]^{-1}\mathbf{r}$ . There is no need to distinguish between a time  $t$  in the ECEF and the corresponding time  $t'$  in the ECI since the coordinate time used is identically the same in both frames.

## 10 Emission coordinates

In this article we described how GNSS works at the current time. GPS and GLONASS rely on a set of relativistic corrections, of order  $c^{-2}$ , in order to function. Dividing these corrections between the orbiting clocks and the receiver was intended to minimize the task of orbiting computers, which in the 1970's had very limited capability. The result is a legacy of conventions for the format of transmitted navigation messages, built into the hardware and software of GPS transmitters and receivers. Navigation messages in the GPS include the transmission time  $t'$ , the clock correction coefficients, and a host of parameters enabling a receiver to compute the satellite position at  $t'$  in the ECEF. Published specifications of the navigation messages for GALILEO and BEIDOU specify similar message formats. Interoperability of the different navigation systems is a significant international issue: improved safety, reliability, and accuracy of the systems as they work together is a highly desirable outcome.

A piecemeal approach to relativistic corrections is rather ungraceful; more elegant, completely relativistic, navigation systems based on "emission coordinates" have been proposed and studied within the last fifteen years by a number of researchers. [23-26, 28] An event in spacetime can be described by the reception of quantities  $(\tau^1, \tau^2, \tau^3, \tau^4, \tau^5\dots)$ , where  $\tau^j$  is the proper time of transmission of an electromagnetic pulse from the  $j^{\text{th}}$  freely-falling clock. The pulses from one such clock generate a set of labeled surfaces that can be visualized at one instant as providing one of the coordinates of spacetime. Four such sets of labelled surfaces

could coordinatize an entire region of space and time, and be used for navigation.

Algorithms for navigation with such coordinates in Minkowskian [27] and Schwarzschild [28] spacetimes have been developed and there have been studies of emission coordinates for rotating systems. In this picture there would be no frequency offsets of the orbiting clocks. A receiver uses solutions for the null geodesics of the signals in the given spacetime to find its position and time.

An adequate discussion of emission coordinates is beyond the scope of this article. However [Figure 1 illustrates](#) the hurdles to be faced if emission coordinates are ever to be implemented in GNSS. Receivers expect signals to arrive within the narrow frequency band allocated to the system. If no factory frequency offset were applied, the histogram peak would lie several hundred units off to the side, in fact it would lie near the limit of the numerical range currently allocated within the navigation message for the measured signal frequency correction. At the very least this would entail redesign of the navigation message and the signal in space, and complicate interoperability issues among systems.

## 11 JUNO and other missions

The JUNO mission will place a probe in polar orbit about Jupiter with high eccentricity. The angular momentum of Jupiter will cause significant frame-dragging, causing among other things a secular precession of the line of nodes. The main effect amounts to a precession rate of about  $9.8 \times 10^{-10}$  radians/s. Using an eccentricity  $e = 0.947$ , this can be visualized in terms of a rotation of the semiminor axis. The ends of the semiminor axis would move an estimated 85 meters during 32 polar orbits due only to this effect. The precession depends on the internal structure of Jupiter and is a rare example of a relativistic effect whose measurement can give significant geophysical information about Jupiter.

Plans are under way to place an ensemble of atomic clocks with exceptional stability and accuracy on the International Space Station [29]. Position and velocity determination with onboard GPS receivers will be important in testing relativity, worldwide time synchronization, and possible additional improvements in GNSS.

## 12 Summary

General Relativity is embedded in the conceptual design of GNSS and hence in their many applications. Demands are increasing for accurate worldwide synchronization and timing for faster communications, financial transaction time-stamping, smart roads, search and rescue, and scientific studies of all sorts. Applications in astronomy to black holes, degenerate stars, and many other topics are not discussed in this article. Without GR we would be at a loss to make sense out of recent measurements that show our universe is flat but expanding. [30] GNSS is but one of many topics that has been deeply permeated by GR.

## Bibliography

- [1] J. L. Synge, *Relativity, The General Theory*, North-Holland, Amsterdam, 1960, vii.
- [2] I. Ghezzi and C. Ruggles, *Chankillo: a 2300 year-old solar observatory in coastal Peru*, Science **315**, (2007), 1239–1243.
- [3] IERS Conventions (2010), International Earth Rotation Service, Technical Note No. 36, (ISSN: 1019-4568), Ch. 5, pp 43–78.
- [4] *Standardization of the units and origins of coordinate times used in astronomy; Geocentric Coordinate Time (TCG) and Barycentric Coordinate Time (TCB)*, Proc. XXIst General Assembly of the IAU, Buenos Aires,

Argentina, July 23-Aug 1 (1991) available from  
[http://www.iau.org/science/meetings/past/general\\_assembly/](http://www.iau.org/science/meetings/past/general_assembly/)

- [5] Interface specification IS-GPS-200D, Navstar GPS Space Segment/Navigation User Interfaces, [www.gps.gov/technical/icwg/IS-GPS-200D.pdf](http://www.gps.gov/technical/icwg/IS-GPS-200D.pdf), Dec. 7, (2004), 175.
- [6] World Geodetic System 1984, NIMA TR8350.2, 3rd edn., January 3 (2000), p. 2.2.
- [7] Addendum to National Imagery and mapping Agency Technical Report 8350.2, World Geodetic System 1984, Reference Frame G1150 (National Geospatial-Intelligence Agency YEAR).
- [8] Interface specification IS-GPS-705, Global Positioning Systems Directorate, [www.gps.gov/technical/icwg/IS-GPS-705C.pdf](http://www.gps.gov/technical/icwg/IS-GPS-705C.pdf), Sept. 5, (2012), 81.
- [9] The SOFA software collection may be found at the website <http://iau-sofa.hmnao.com>.
- [10] N. Ashby and B. Bertotti, *Phys. Rev. D* **34** (8), (1986), pp 2246–2259.
- [11] E. M. Standish, *JPL Planetary and Lunar Ephemerides, DE405/LE405*, JPL IOM 312.F-98-048 (1998).
- [12] *Astronomical Almanac for the year 2013*, US Government Printing Office, PO Box 979050, St. Louis, MO 63197-9000, 2013, iii.
- [13] N. Ashby and M. Weiss, *Global Positioning Receivers and Relativity*, NIST Technical Note Vol. 1385, US Government Printing Office, Washington D. C., 1999.
- [14] E. J. Post, *Revs. Mod. Phys.* **39**, (1967), 475–493.
- [15] *Recommendations and Reports of the CCIR*, 1982, XVth Plenary Assembly of the ITU, Geneva, Volume VII (1982), Report 439-3, pp 112–116.
- [16] *La Définition de la Seconde*, Consultative Committee for the Definition of the Second of the International Committee for Weights and Measures, 9e Session, 23–25 September 1980, pp 16–17.
- [17] *Interface Specification for QZSS (IS-QZSS)*, V1.5

Japan Aerospace Exploration Agency, March 27 (2013).

- [18] <http://www.isp.justthe80.com/space-applications/gagan>.
- [19] M. Epstein, E. Stoll and J. Fine, *Study of SVN43 clock from 6/18/00 to 12/30/00*, in Proc. 33rd Annual Precise Time and Time Interval Systems and Applications Meeting, Long Beach, Nov. 2001, L. Breakiron, ed., US Naval Observatory, Washington, DC
- [20] N. Ashby, *Using Lagrange perturbation theory to predict relativistic frequency changes in SV clocks due to earth's oblateness, and due to orbit changes*, in Proc. 33rd Annual Precise Time and Time Interval Systems and Applications Meeting, Long Beach, Nov. 2001, L. Breakiron, ed., US Naval Observatory, Washington, DC
- [21] N. Ashby and T. Allison, *Celestial Mechanics and Dynamical Astronomy* **57**, (1993), 537-585.
- [22] I. I. Shapiro, M. E. Ash, D. B. Campbell, R. B. Dyce, R. P. Ingalls, R. F. Jurgens and G. H. Pettengill, *Phys. Rev. Lett.* **26**, (1971), 27-30.
- [23] B. Coll, *Proc. 23rd Spanish Relativity Meeting, ERE-2000 on Reference Frames and Gravitomagnetism*, Singapore, World Scientific, 2001, 53-65.
- [24] B. Coll and J. M. Pozo, *Class. Quantum Grav.* **23**, (2006), 7395.
- [25] D. Bunandar, S. A. Caveny and R.A. Matzner, *Phys. Rev. D* **84**, (2011), 104005.
- [26] P. Delva, A. Cadez, U. Kostic and S. Carloni, *A relativistic and autonomous navigation satellite system*, <http://arXiv.org> 1106.3168v2 [gr-qc] (2011).
- [27] C. Rovelli, *Phys. Rev. D* **65**, (2002), 044017
- [28] D. Bini, A. Geralico, M. L. Ruggiero and A. Tartaglia, *Class. Quant. Grav.* **25**, (2008), 205011.
- [29] C. Salomon, L. Cacciapuoti and N. Dimarcq, *Atomic Clock Ensemble in Space: Fundamental Physics and Applications*, *Int. Jour. Mod. Phys. D* **16**, (2007), 2511.
- [30] D. N. Spergel, R. Bean, O. Doré, et al., *Three-Year*

*Wilkinson Microwave Anisotropy Probe (WMAP)  
Observations: Implications for Cosmology, ApJS **170**,  
(2007), 377–408.*

Carlo Rovelli

# The strange world of quantum spacetime

**Abstract:** Hundred years after the discovery of general relativity, consensus on a theory describing the quantum properties of gravity has not crystallized yet. But the progress has been substantial. Good tentative theories exist. Reasonable solutions of longstanding conceptual difficulties have been found. Furthermore, observations are increasingly closer to the Planck regime and have already ruled out alternatives. I describe the state of the art in the field, as I see it.

## 1 A world with no space

The central feature of quantum gravity is that it introduces a fundamental length scale into physics. This is the Planck length

$$L_P = \sqrt{\frac{\hbar G}{c^3}} . \quad (1)$$

This length sets the scale of the quantum gravitational phenomena, but plays also a role similar to the speed of light in special relativity, or the Planck constant in quantum theory: it determines a physical *limit*. The speed of light limits the velocity of propagation of interactions. The Planck constant determines the minimal phase space volume that a single state can take: it limits the amount of information we can obtain about a system.

The Planck length sets a lower limit to the divisibility of space. The existence of this limit is the core physical feature of quantum gravity, in the same sense in which maximal velocity

is a core feature of special relativity.

This fact was realised early by Bronstein [9, 10] and is supported today by a number of converging theoretical results. It is realised in the best developed theories that try to realise or incorporate quantum gravity, such as loop gravity or string theory. In my opinion, this is the most plausible result of quantum gravity to date.

Bronstein original argument can be recast as follows. To measure the properties of a small region, we must determine this region, for instance by putting something in it. If we take a particle and confine it in a small region, its momentum is large because of the Heisenberg relations. Therefore its energy is high. Therefore its gravitational mass is high. But a large mass in a small region gives rise to a black hole, and if a black hole forms, there is no way of marking a smaller region. Quantitatively, if we fold-in the Heisenberg uncertainty relations and the Schwarzschild radius, we find easily that the minimal region we can mark is of the order of the Planck length above.

**Carlo Rovelli:** CPT-CNRS, Case 907, Luminy, 13288  
Marseille, France

This argument is just a semiclassical hint. But today the same result can be obtained by much stronger arguments. I sketch here (leaving a few details aside) the full derivation of this result from classical GR and quantum theory, as obtained in loop gravity.

Our current information about gravity is encoded in the Einstein equations. Because of the existence of fermions, which can couple to tetrads but not the metric, these should be expressed, at the fundamental level, in terms of a tetrad field  $e^I$ ,  $I = 0, 1, 2, 3$ , a one-form with values in Minkowski space. An action that gives the Einstein equations, in first order form, is

$$\begin{aligned}
S[e, \omega] &= \frac{1}{8\pi G} \int \left[ (e^I \wedge e^J) \epsilon_{IJKL} + \frac{1}{y} (e_K \wedge e_L) \right] \wedge F^{KL}[\omega] \\
&= \frac{1}{8\pi G} \int \left[ (e \wedge e)^* + \frac{1}{y} (e \wedge e) \right] \wedge F[\omega],
\end{aligned} \tag{2}$$

where  $\omega$  is an  $SO(3, 1)$  connection,  $F[\omega]$  its curvature and the second line is a compact writing of the first. The equations for  $\omega$  gives the vanishing of the torsion

$$T^I \equiv de^I + \omega^I{}_J \wedge e^J = 0. \tag{3}$$

The second term of the action does not contribute to the equations of motion because when integrating by parts its bulk term is proportional to the torsion. But the boundary term remains and contains a term of the form

$$\frac{1}{y8\pi G} \int d(e \wedge de) \sim \frac{1}{y8\pi G} \int_{\partial M} e \wedge de \tag{4}$$

If a spacelike boundary crosses the initial time surface on a two-dimensional surface  $S$ , this term generates an  $e \wedge e$  surface term in the symplectic two form, which in turns implies that the symplectic structure contains a term purely in  $e$ . Going to the time gauge and to the triad variable  $i^i$ ,  $i = 1, 2, 3$ , which is a one form with values in  $R^3$  on the boundary, a brief calculation gives Poisson brackets

$$\{E^{ai}, E^{bj}\} = 8\pi y G \delta^{ab} \epsilon_k^{ik} E^k. \tag{5}$$

for the components of the two form

$$E^i \equiv \frac{1}{2} \epsilon_{jk}^i e^j \wedge e^k \equiv \frac{1}{2} E^{ai} \epsilon_{abc} dx^b \wedge dx^c. \tag{6}$$

This implies that in the quantum theory the corresponding quantities do not commute, and satisfy the local commutation relations

$$[E^{ai}, E^{bj}] = 8\pi i y \hbar G \delta^{ab} \epsilon^{ij}_k E^k .$$

(7)

This in turn implies that these operators are generators of unitary representations of  $SU(2)$ . This formal derivation is confirmed by two observations: first, these commutation relations are the standard ones among electric field components in lattice gauge theory. Second, the  $E$  fields is conjugate to the connection and generates the group action; therefore for geometrical consistency these operators must realise the gauge group algebra, which on the boundary in the time gauge is the algebra of the  $SU(2)$  group.

But now observe the area of a generic surface  $S$  is given, by standard geometry, by the surface integral of the norm of the 2-form  $E^i$

$$A[S] = \int_S \sqrt{^2 q} d^2 \sigma = \int_S \sqrt{\det q q^{ab} n_a n_b} d^2 \sigma = \int_S |E| .$$

(8)

where  $q$  is the 3-metric. In the quantum theory this is precisely  $8\pi y \hbar G$  times the Casimir of  $SU(2)$ . As the  $SU(2)$  Casimir's eigenvalues are  $\sqrt{j(j+1)}$  with  $2j$  a nonnegative natural number, it follows that a straightforward application of quantum theory to general relativity implies that the area of any surface is quantized, and the eigenvalues are (sums of) elementary areas of size

$$A = 8\pi y \hbar G \sqrt{j(j+1)} .$$

(9)

Similar results can be obtained for the spectra of other geometrical operator, such as volumes and angles. The

importance of this result is that it shows that there is a minimal non-vanishing area in nature (in the same sense in which there is a minimal non vanishing angular momentum.)

Quantum gravity is therefore the realisation that space is not continuum. There is no infinite tower of field modes all the way to infinite frequency. Space has a finite discrete granular structure. There is limit to the divisibility of space, like there is a limit to the divisibility of matter.

## 2 A world without time

Given the immense success of quantum theory and the complete lack of empirical hints of its limitations, it is reasonable to search a quantum theory of gravity within the general formalism of quantum theory. But there is at least one aspect of the conventional formulation that has to be generalised, in order to adapt it to gravitational physics: the way time is treated.

In non-relativistic physics, we describe change in terms of evolution with respect to an external time variable  $t$ , ideally measured by a clock dynamically independent from the system under consideration. This clock defines the independent variable  $t$  in terms of which the dynamics of the dependent variables  $q_n(t)$  that describe the system is given. The time variable  $t$  appears in the Schrödinger equation and the evolution in  $t$  is generated by the Hamiltonian. The same happens in conventional quantum field theory, where the Hamiltonian is incorporated as one of the generators of the Poincaré group, which acts unitarily on the state space of the theory.

In general relativistic physics, this formal structure does not work anymore. Instead, we must include the independent evolution parameter among the other variables of the system, and treat it on the same footing as them.<sup>12</sup> Accordingly, physics does not anymore describe the evolution of the variables “in time”, but rather the *relative* evolution of the variables, namely the evolution of the variables with respect to one another.<sup>13</sup>

The conceptual step is analogous to the step taken by describing a curve in the  $(x, y)$  plane in terms of a relation  $f(x, y) = 0$  rather than in the form  $y = y(x)$ . The first option is clearly more general than the second.

In the canonical language, this means that we must work with an  $N + 1$  dimensional *extended configuration space*  $\mathcal{E}$ , if  $N$  is the number of degrees of freedom, and the dynamics is not determined by a Hamiltonian, but by a hamiltonian constraint  $C$  on the corresponding phase space.

The dynamics of a finite dimensional system is compactly captured by the *Hamilton function*, which is a function on  $\mathcal{E} \times \mathcal{E}$  defined as the value of the action on a solution of the equation of motion interpolating between two given points in  $\mathcal{E}$ . For instance, the free particle dynamics is captured by the Hamilton function

$$S(x, t; x', t') = \frac{(x - x')^2}{2m(t - t')} . \quad (10)$$

The derivatives of  $S$  with respect to the two variables  $x$  and  $t$  (treated on equal footing) give the two momenta  $p_x$  and  $p_t$  and these satisfy the hamiltonian constraint

$$C(x, t, p_x, p_t) = p_t + \frac{p_x^2}{2m} = 0 . \quad (11)$$

In other words, the Hamilton function satisfies the Hamilton-Jacobi equation

$$C\left(x, t, \frac{\partial S}{\partial x}, \frac{\partial S}{\partial t}\right) = 0 . \quad (12)$$

The quantum version of this equation

$$c \left( x, t, \frac{\partial}{\partial x}, \frac{\partial}{\partial t} \right) \Psi(x, t) = 0$$

(13)

is called the Wheeler-deWitt equation. It reduces to the time-dependent Schrödinger equation in non relativistic systems, when we single out one coordinate on  $\mathcal{E}$  as the time variable. But its validity is more general.

A quantum theory is also defined by its transition amplitudes, which determine the relative probability of different *processes*. If spectra are continuous, transition amplitudes are functions on  $\mathcal{E} \times \mathcal{E}$ , like the Hamilton function. In fact, in the approximation where the Planck constant  $\hbar$  can be considered small, the transition amplitude satisfy

$$W \sim e^{\frac{i}{\hbar} S}.$$

(14)

When everything is well defined, the transition amplitudes are the matrix elements of an operator  $P \sim \delta(C)$  that “projects” on the solutions of the Wheeler-deWitt equation

$$W(x, t; x', t') = \langle x, t | P | x', t' \rangle.$$

[\(15\)](#)

The transition amplitudes can also be defined à la Feynman, by a functional integral over paths going from the first to the second point in  $\mathcal{E}$ , weighted by the exponential of  $\frac{i}{\hbar}$  times the classical action of the path. In this representation it is immediately to see why (14) holds: it is just the saddle point approximation of the path integral.

The dynamics of a quantum theory is defined by giving the Wheeler-deWitt equation, or by giving the path-integral representation of the transition amplitudes, or by directly giving the transition amplitudes, perhaps in terms of a perturbative expansion. The last option is that of the Feynman-rules definition of QED and is also the option used below for quantum

gravity. This option circumvents the mathematical complications of the Wheeler-deWitt equation or the full path integral.

The quantum dynamics determines the probability of the state  $|x, t\rangle$ , given the state  $|x', t'\rangle$ . Notice that the quantum states of the theory live on the *boundary* of the *process* considered.

This formal structure can be generalized to a field theory as follows. Consider a closed compact region in spacetime and let  $\Sigma$  be its boundary. The states of the theory live on  $\Sigma$  and describe the possible outcomes of any interaction at the boundary.<sup>14</sup>

The quantum amplitudes are functions of boundary data and determine the probability of a process defined by a given boundary state. The boundary states define a Hilbert space  $\mathcal{H}$  and the transition amplitudes can be thought, à la Feynman, as the path integral of the field in the bulk.<sup>15</sup>

In quantum gravity, the transition amplitudes describe the full process in the bulk including the gravitational phenomena, therefore there is no Riemannian manifold inside, as there is no trajectory of a quantum particle. The boundary data include the gravitational boundary data, and these amount to a specification of boundary *metric* quantities. Therefore information such as “the time lapsed” during the process or “the physical distance” between two boundary points are not specified externally: they are already contained in the boundary data about the boundary value of the gravitational field, that is, the metric. This is the beautiful and subtle manner in which time and space are reinterpreted in quantum gravity: as gravitational properties of the boundary data for *physical processes*.

Notice that in [equation \(15\)](#) the arguments of the transition amplitudes are coordinates on  $\Sigma$  only if the quantum spectrum of these is continuous. If it is discrete, the arguments of the transition amplitudes are the quantum numbers labeling the discrete spectrum. This is what will happen below.

Summarizing, to construct a quantum theory of gravity we need two ingredients: a boundary Hilbert space capable of

describing the possible outcome of interactions with (“measurements” of) the gravitational field and matter on the boundary of a process, and transition amplitudes for any given boundary states. The theory will have the appropriate classical limit if the transition amplitudes behave as (14) for small  $\hbar$ , where  $S$  is the Hamilton function of General Relativity.

This structure circumvents entirely the so-called “problem of time” of quantum gravity. The problem of time is resolved by this way of defining the quantum dynamics. The theory is about probabilities assigned to alternative processes. We may avoid talking about “time” altogether, we may forget the word “time”, and still fully and consistently describe change in the world.

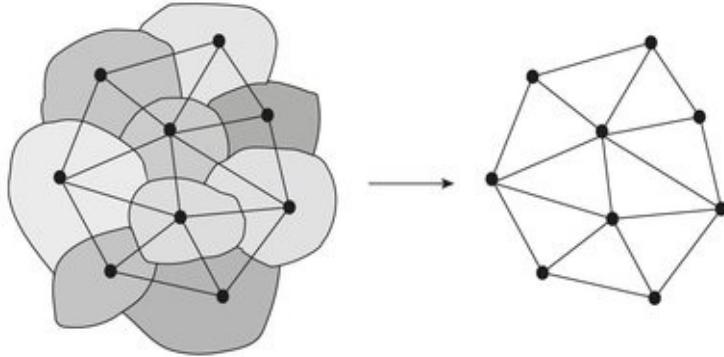
### 3 Loop gravity

In loop quantum gravity (see [29] for a recent introduction), the Hilbert space  $\mathcal{H}_\Gamma$  that represents the gravitational field on a given boundary, at fixed truncation of the theory, can be defined as follows. Let  $\Gamma$  be an oriented graph (defined solely by its combinatorial structure). Intuitively,  $\Gamma$  is the dual graph of a cellular decomposition of  $\Sigma$ . See [Figure 1](#). The graph determines the truncation. Refining the graph leads to a better approximation of the theory.

Associate to each link  $l$  of the graph an  $SU(2)$  element  $U_l$ . The states of the theory are given by square integrable functions  $\psi(U_l)$  invariant under the gauge transformations

$$\psi(U_l) \mapsto \psi(\Lambda_{s(l)} U_l \Lambda_{t(l)}^{-1}) \quad (16)$$

where  $s(l)$  and  $t(l)$  are the two nodes where the link  $l$  starts and ends and  $\Lambda_n \in SU(2)$  for any node  $n$  of the graph. Square-integrability is under the  $SU(2)$  Haar measure.



**Fig. 1.** The graph  $\Gamma$  is the dual of a cellular decomposition of the 3d boundary of the process.

The remarkable feature of this Hilbert space is that it describes a 3d curved metric geometry in the classical limit. In a sense, this should not be surprising: the main result of Ashtekar's formulation of general relativity [3] is that gravity can be described using the phase space of an  $SU(2)$  Yang–Mills theory, and the Hilbert space  $\mathcal{H}_\Gamma$  is precisely the Hilbert space of a lattice  $SU(2)$  Yang–Mills theory.

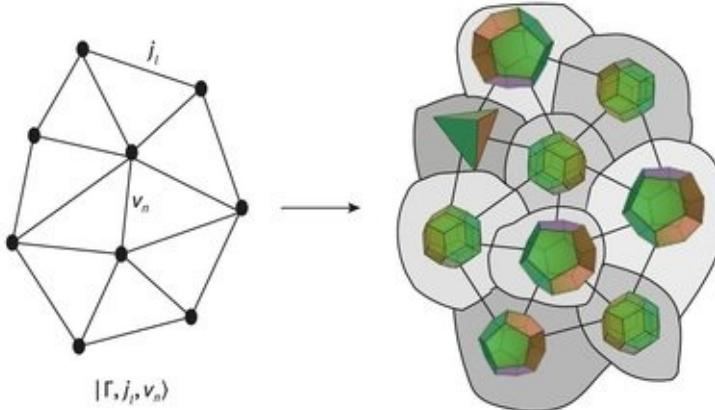
Therefore it describes precisely the gravitational field on the boundary of a process [8, 12, 17]. More precisely, at each given truncation  $\Gamma$ , semiclassical states describe the discrete geometry of a piecewise flat cellular decomposition of a curved metric space.

The key point is that there are natural derivative operators defined on the Hilbert space  $\mathcal{H}_\Gamma$ : the left-invariant vector fields at the nodes, along the links. These can be shown to satisfy algebraic properties that imply that they are in 1-1 correspondence with the quantities describing the metric geometry of a discrete space. In particular, the Casimir on each link is the area of the corresponding face bounding two cells. The scalar product of two links emerging from the same node determines the angle between the normals of the corresponding faces. Thus the expectation values of these operators define a piecewise flat geometry on the cellular decomposition. See [Figure 2](#).

A complete set of commuting observables in  $\mathcal{H}_\Gamma$  is provided by the areas of the faces and the volume of the 3d cells.

Accordingly, the Hilbert space admits a basis  $|\Gamma, j_i, v_n\rangle$ , called the spin network basis [35], labelled by three groups of quantum numbers: the graph  $\Gamma$  itself, which gives the connectivity of the cells, the spins  $j_i$  associated to the faces, that are the quantum numbers of the areas and, at each node  $n$ , the quantum number  $v_n$  of the volume of the corresponding cell. Since areas of surfaces and volume of cells do not fully determine the classical geometry, the rest of the geometry fluctuates. This is a situation analog to angular momentum theory, where only  $L^2$  and  $L_z$  can be diagonalized simultaneously.

An important result is that the spectrum of area and volume is discrete [4, 34]. This is the realization of the intuitive idea of the existence of a physical cut off at the Planck scale. Intuitively: the physical size of the polyhedra of [Figure 2](#) can never become “smaller than the Planck size”. This is a typical quantum phenomenon: the value of the angular momentum can never become smaller than  $\hbar/2$ . In quantum gravity, it is the reason of the ultraviolet finiteness of the transition amplitudes.



[Fig. 2](#). A spin network and the “quanta of space” it describes.

The transition amplitude associated to a given boundary state is defined as follows. First, pick a two-complex  $\mathcal{C}$  bounded by  $\Gamma$ . Call  $f$  its faces,  $e$  its edges and  $v$  its vertices<sup>[16](#)</sup>. Then

$$W_{\mathcal{C}}(U_I) = \int_{SU(2)} dh_{vf} \prod_f \delta(\prod_v h_{vf}) \prod_v A_v(h_{vf})$$

(17)

where the vertex amplitude is

$$A(h_f) = \int_{SL(2, C)} dg_e \prod_f \sum_j d_j \text{Tr}_j [h_f Y^\dagger g_e g_{e'} Y]. \quad (18)$$

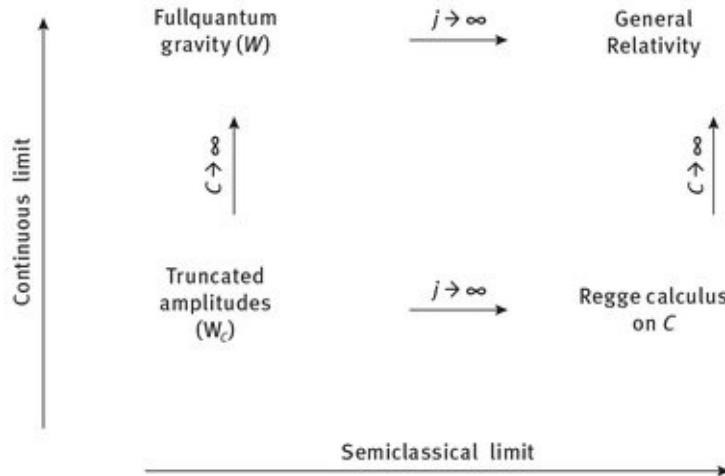
$d_j = 2j + 1$  and  $Y$  maps the  $SU(2)$  representation of spin  $j$  into the spin  $j$  subspace of the  $SL(2, C)$  unitary representation determined by the discrete spin  $k = j$  and the continuous parameter  $\rho = yj$ . The parameter  $y$  is a free parameter in the theory<sup>6</sup>. This is the full definition of a quantum theory of gravity.

This amplitude was derived in [13] building on [14, 16, 23, 26], and is sometimes denoted the EPRL amplitude. For details, full references and a derivation of these expressions from the classical action of GR see [33] and [27]. The extension of this amplitude that includes the cosmological constant using a quantum deformation of the groups is defined in [15, 20].

The key properties of these amplitudes are three

1. They define transition amplitudes for the Hilbert space  $\mathcal{H}_T$ , that have the correct degrees of freedom to describe the gravitational field.
2. They yield the Hamilton function of a truncation of Lorentzian General Relativity over a cellular decomposition dual to  $\mathcal{C}$ . More precisely, the vertex amplitude (18) yields to the exponential of the Regge action on the corresponding cell [6] and the quantum deformed amplitude yields the exponential of the Regge action with cosmological content [15, 19]. The full amplitude is a truncation of a Feynman “sum over geometries” in the bulk [11].
3. They are finite. This is the key result. A theorem states that the vertex amplitude with cosmological constant is finite [15, 19]. The Planck length and the cosmological constant provide physical ultraviolet and infrared cut off, respectively.<sup>17</sup>

Notice the following structure of the theory:



The amplitudes on a given truncation approximate the truncated dynamics of classical general relativity on a given triangulation. These approximate continuous general relativity when the truncation is refined. As in QCD and in QED, the truncation is expected to offer already a good approximation to the full dynamics, in appropriate regimes.

Although a number of important technical issues, which must not be underestimated, remain open (see a discussion in [33]) equations (17–18) give a definition of a quantum theory of gravity which is finite and yields classical general relativity in an appropriate limit. The construction provides a good ground for discussing the physical question of the nature of quantum spacetime.

## 4 Quantum spacetime

I can now summarize the conceptual structure that has emerged, for a quantum theory of spacetime.

Measurements directly involving the gravitational field are measurements of geometrical length, areas or volumes. For instance, any measurement of a cross-section is the measurement of an area. Gravitational wave detectors measure (the variation of) a length. Outcome of these measurements are

described by the spin-network Hilbert space. No measurement measures an infinite number of quantities: we always have access only to a finite number of outcomes. Hence a truncation of the degrees of freedom is sufficient to describe the outcome of any measurement.

The theory predicts that measurements of some geometrical quantities yield discrete values [30, 39]. According to the theory, for instance, a physical cross section cannot be smaller than the Planck scale, it can only take values which are in the spectrum of the area operator.<sup>18</sup>

Dynamics is given by associating an amplitude to each process. A process is determined its boundary state, namely the outcome of a measurement (or a generic interaction) on its boundary. The relative probability of distinct processes can be computed from these amplitudes. The formalism does not require to go at infinite distance from an interaction to have well-defined physical amplitudes.

Spacial and temporal specifications make sense only on the boundary of a process, in the context of an interaction. In other words, space and time themselves are reduced to quantum entities like the position of a quantum particle, which is determined only at interaction time, otherwise is fluctuating.

## 5 Empirical evidence

For a long time the possibility of observing effects of quantum gravitational phenomena appeared completely out of our reach. We expect quantum gravity phenomena to become relevant at the Planck scale, and this scale is fifteen orders of magnitude beyond the scales probed at LHC. The smallness of the Planck length appeared to represent an absolute barrier. Things have now changed (see for instance [2]), astrophysical measurements at this scale have already been performed, and the possibility of measuring quantum gravity effects does not appear remote anymore.

There are also laboratory experiments that have probed

scales not far from the Planck scale. For instance, the SU(5) grand unified theory has been falsified by proton- decay experiments, which effectively probe the unification scale, only a few orders of magnitudes below the Planck scale.

In astrophysics, the MAGIC observations [1] and the observations on the Crab Nebula have put quantitative limit to the possibility of Lorentz invariance violations up to scales beyond the Planck scale, already ruling out quantum-gravity theories that had been considered by theoreticians. Early MAGIC observations found hints of a nontrivial energy dependence on the time of arrival from distant events [1]. The result raised hopes for a possible direct observations of quantum gravity phenomena, because such dependence was predicted by some quantum gravity theories. These hopes have now faded for two reasons: on the one hand, the theory suggests that Lorentz invariance is not violated [36, 37]; on the other, the MAGIC data appear to have different explanations [22]. Thus there is consistency, although is the somewhat disappointing context of a null result. Still, these developments have shown that if a Lorentz-violating quantum-gravity effect existed, we would have the technology to observe it. The Planck scale is not out of reach.

BICEP2 has recently claimed to have observed of the long awaited B-modes in the CMB, likely generated by primordial gravitational waves. According to current models, these are in turn generated by primordial quantum gravitational fluctuations. If confirmed, these measurements can be interpreted also as indirect evidence for quantisation of gravity. According to current cosmological model, indeed, primordial Bmodes in the CMB are generated by gravitational waves which, in turn, are generated by quantum fluctuations of the gravitational field. If confirmed, BICEP2 result suggests that the onset of inflation is around  $10^{16}$  GeV, which is just three orders of magnitude below the Planck scale (12 orders of magnitude beyond LHC).

The Planck mission has presented its first set of results [28] and is expected to release more data soon. More data are

expected from Polarbear and the South Pole telescope. High-precision CMB data may soon be used to test quantum gravity scenarios. When the power spectrum of gravity waves is measured, going beyond the detection itself, new perspective for testing quantum gravity will open immediately. Planck scale measurements are within reach in cosmology.

Other ideas for observational windows on quantum gravity have been recently put forward. Loop quantum gravity appears to predict that a strong pressure, with its origin in quantum geometry, develops when matter reaches the Planck density. This indicates that the size of the core of the black holes where quantum effects prevent the collapse be much larger than Planckian. Taking this into account, a primordial black hole ending its life today could emit radiation in the wavelength [38]

$$\lambda = \sqrt[3]{\frac{t_H}{t_{Pl}}} l_{Pl} \sim 10^{-14} \text{ cm}$$

(19)

( $t_H$  is the Hubble time,  $t_{Pl}$  and  $l_{Pl}$  the Planck time and Planck length.) Note that  $\lambda$  is in the GeV range of cosmic rays, explored for instance by the FERMI satellite, and may be detectable. At higher energies the HESS telescope provides data up the several tens of TeV. Such measurements have already been successfully used to derive stringent limits on possible Lorentz invariance violation and therefore on possible Planck scale effects implying a modified dispersion relation. Again, they do probe the Planck scale.

Roger Penrose and collaborators has observed that events in an eventual pre-Big-Bang phase, which different theories including Loop quantum cosmology are currently exploring, could have left observable large-angle traces in the CMB [18, 24, 25]. The observational status of these is controversial, but the suggestion is intriguing.

There are other channels for possible Planck scale observations. Neutrinos astrophysics is improving rapidly and energies up to 400 TeV have been detected for atmospheric

neutrinos in experiments such as ICECube opening a new promising channel for testing Lorentz and CPT invariance at high energies. At even higher energies, the Auger observatory measures cosmic-rays beyond  $10^{20}$  eV with a remarkable statistics albeit substantial uncertainties are still present on the exact nature of these ultra high energy particles. This latter fact is so far preventing conclusive statements about the available constraints. However, preliminary studies (based both on ultra high energy cosmic rays as well as on their byproducts such as cosmogenic neutrinos and ultra high energy gamma rays) have proven able to cast effective constraints on dispersion relations modified by  $(\text{Energy}/\text{Planckmass})^2$  suppressed terms (corresponding to mass-dimension five and six operators in an effective field theory framework) [21]. Specific Planck-scale effects can now be tested, and ruled out.

To be sure, no direct quantum gravitational effect has been observed so far, but quantum-gravity, long a purely theoretical effort, is now beginning to confront experiments.

**Acknowledgment:** Parts of this article reproduce a text by the author appeared in the Springer Handbook on Spacetime.

## Bibliography

- [1] J. Albert et al., *Probing Quantum Gravity using Photons from a Mkn 501 Flare Observed by MAGIC*, 2007.
- [2] Amelino-Camelia, Gravity in quantum mechanics, *Nature-Physics* **10** (2014), 254.
- [3] A. Ashtekar, New Variables for Classical and Quantum Gravity, *Phys. Rev. Lett.* **57** (1986), 2244–2247.
- [4] A. Ashtekar and J. Lewandowski, Quantum theory of geometry. I: Area operators, *Class. Quant. Grav.* **14** (1997), A55–A82.
- [5] B. Bahr, F. Hellmann, W. Kaminski, M. Kisielowski and J. Lewandowski, Operator Spin Foam Models, *Class.*

*Quant. Grav.* **28** (2011), 105 003.

- [6] J. W. Barrett, R. J. Dowdall, W. J. Fairbairn, F. Hellmann and R. Pereira, Asymptotic analysis of Lorentzian spin foam models, *PoS QGQGS2011* (2011), 9.
- [7] E. Bianchi, *Talk at the 2012 Marcel Grossmann meeting*, July 2012.
- [8] E. Bianchi, P. Donà and S. Speziale, Polyhedra in loop quantum gravity, *Phys. Rev. D* **83** (2011), 44 035.
- [9] M. P. Bronstein, Kvantovanie gravitatsionnykh voln (Quantization of Gravitational Waves), *Zh. Eksp. Tear. Fiz.* **6** (1936), 195.
- [10] M. P. Bronstein, Quantentheorie schwacher Gravitationsfelder, *Phys. Z. Sowjetunion* **9** (1936), 140–157.
- [11] F. Conrady and L. Freidel, Path integral representation of spin foam models of 4d gravity, *Class. Quant. Grav.* **25** (2008), 245 010.
- [12] M. Dupuis, J. P. Ryan and S. Speziale, Discrete gravity models and Loop Quantum Gravity: a short review, (2012).
- [13] J. Engle, E. Livine, R. Pereira and C. Rovelli, LQG vertex with finite Immirzi parameter, *Nucl. Phys. B* **799** (2008), 136–149.
- [14] J. Engle, R. Pereira and C. Rovelli, The loop-quantum-gravity vertex-amplitude, *Phys. Rev. Lett.* **99** (2007), 161301.
- [15] W. J. Fairbairn and C. Meusburger, Quantum deformation of two four-dimensional spin foam models, *J. Math. Phys.* **53** (2012), 22501.
- [16] L. Freidel and K. Krasnov, A New Spin Foam Model for 4d Gravity, *Class. Quant. Grav.* **25** (2008), 125018.
- [17] L. Freidel and S. Speziale, Twisted geometries: A geometric parametrisation of SU(2) phase space, *Phys. Rev. D* **82** (2010), 84040.
- [18] V. G. Gurzadyan and R. Penrose, CCC-predicted low-variance circles in CMB sky and LCDM, (2011).
- [19] M. Han, 4-dimensional Spinfoam Model with

Quantum Lorentz Group, *J. Math. Phys.* **52** (2011), 72501.

- [20] M. Han, Cosmological Constant in LQG Vertex Amplitude, *Phys. Rev. D* **84** (2011), 64010.
- [21] S. Liberati and L. Maccione, Lorentz Violation: Motivation and new constraints, *Ann. Rev. Nucl. Part. Sci.* **59** (2009), 245–267.
- [22] S. Liberati and L. Maccione, Quantum Gravity phenomenology: achievements and challenges, (2011).
- [23] E. R. Livine and S. Speziale, A new spinfoam vertex for quantum gravity, *Phys. Rev. D* **76** (2007), 84028.
- [24] W. Nelson and E. Wilson-Ewing, Pre-big-bang cosmology and circles in the cosmic microwave background, *Physical Review D* **84** (2011), 043508.
- [25] R. Penrose, Before the big bang: An outrageous new perspective and its implications for particle physics, *Conf. Proc. C060626* (2006), 2759–2767.
- [26] R. Pereira, Lorentzian LQG vertex amplitude, *Class. Quant. Grav.* **25** (2008), 85013.
- [27] A. Perez, The Spin Foam Approach to Quantum Gravity, *Living Rev.Rel.* **16** (2012), 3.
- [28] Planck Collaboration, Planck 2013 results. I. Overview of products and scientific results, (2013).
- [29] C. Rovelli and F. Vidotto, *Introduction to covariant loop quantum gravity*, Cambridge University Press, to appear., 2015.
- [30] C. Rovelli, A Generally covariant quantum field theory and a prediction on quantum measurements of geometry, *Nucl. Phys. B* **405** (1993), 797.
- [31] C. Rovelli, Relational Quantum Mechanics, *Int. J. Theor. Phys.* **35** (1996), 1637.
- [32] C. Rovelli, *Comment on ‘Are the spectra of geometrical operators in Loop Quantum Gravity really discrete?’ by B. Dittrich and T. Thiemann*, 2007.
- [33] C. Rovelli, Zakopane lectures on loop gravity, *PoS QGQGS2011* (2011), 3.
- [34] C. Rovelli and L. Smolin, Discreteness of area and

- volume in quantum gravity, *Nucl. Phys. B* **442** (1995), 593–622.
- [35] C. Rovelli and L. Smolin, Spin Networks and Quantum Gravity, *arXiv.org gr-qc* (1995).
  - [36] C. Rovelli and S. Speziale, Reconcile Planck-scale discreteness and the Lorentz–Fitzgerald contraction, *Phys. Rev. D* **67** (2003), 64019.
  - [37] C. Rovelli and S. Speziale, Lorentz covariance of loop quantum gravity, *Phys. Rev. D* **83** (2011), 104029.
  - [38] C. Rovelli and F. Vidotto, Planck stars, (2014), *arXiv* : 1401.6562.
  - [39] L. Smolin, Finite diffeomorphism invariant observables in quantum gravity, *Phys. Rev. D* **49** (1994), 4028.
  - [40] W. Wieland, Complex Ashtekar variables and reality conditions for Holst’s action, *Annales Henri Poincaré* **13** (2012), 425.

# Index

$\Omega_\Lambda$  at different redshifts [1](#)

## A

ABCD matrix [1](#)

accretion [1](#)

accretion disc [1](#), [2](#), [3](#), [4](#), [5](#), [6](#)

adiabatic (reversible) change [1](#)

AdS/CFT correspondence [1](#)

advance of perihelion [1](#)

AGN feedback [1](#), [2](#), [3](#)

albedo [1](#)

Anaximander [1](#)

anomalies of the Earth gravity field [1](#)

antigravity [1](#)

Ashtekar's formulation of general relativity [1](#)

atmospheric drag [1](#), [2](#)

## B

Beckenstein [1](#)

BEIDOU [1](#)

bending of light rays by Sun [1](#)

black hole [1](#)

black hole entropy [1](#)

black hole evaporation [1](#)

Blandford-Znajek [1](#)

Blandford-Znajek effect [1](#)  
Blandford-Znajek mechanism [1](#)  
blazar [1](#)  
boundary data [1](#)  
boundary state [1](#)  
boundary transition amplitudes [1](#)  
Boyer & Lindquist [1](#)  
Boyer-Lindquist coordinates [1](#)  
Brill-Cohen metric [1](#)  
broad iron line method [1](#)  
bulk [1](#)

## C

calendar, Gregorian [1](#)  
calendar, Julian [1](#)  
calendars [1](#)  
cellular decomposition [1](#)  
CenA/M83 group zero gravity radius [1](#)  
centrifugal force [1](#)  
Chankillo [1](#)  
Chern-Simons gravity [1, 2, 3](#)  
Chern-Simons gravity and String theories [1](#)  
circular orbit [1](#)  
circular photon orbit [1](#)  
clock [1, 2](#)  
- coordinate [1](#)  
clocks and dragging of inertial frames [1-2, 3](#)  
cluster of galaxies [1](#)  
Coma cluster  
- dark energy effects [1](#)  
- mass estimates [1](#)  
- predicted outflow [1](#)  
- profiles [1](#)  
Coma masses from matter profiles [1](#)

Compton hump [1](#)  
coordinate time [1](#), [2](#)  
- barycentric [1](#)  
coordinate, isotropic [1](#)  
coordinates, emission [1](#)  
Copernicus [1](#), [2](#), [3](#)  
cosmological constant [1](#), [2](#)  
Cotton-York tensor [1](#)  
criterion for stability [1](#)  
Curtis [1](#)  
curvature invariants and gravitomagnetism [1](#), [2](#)  
Cyg X-1 [1](#), [2](#)

## D

dark energy [1](#)  
dark energy constancy [1](#)  
dark energy from Local Group motions [1](#)  
Darwin [1](#)  
de Sitter [1](#), [2](#)  
de Sitter universe [1](#)  
Dirac [1](#), [2](#), [3](#)  
discrete  
- area [1](#)  
- volume [1](#)  
discreteness [1](#)  
disk  
- thickness [1](#)  
- viscosity [1](#)  
drag-free satellite [1](#)  
dragging of inertial frames, or frame-dragging [1-2](#)  
- and error analyses [1-2](#)  
- inside a hollow sphere [1](#)  
- on clocks and photons [1-2](#)  
- on orbital period [1](#)

- time-delay and gravitational lensing [1-2](#)  
dynamical time, barycentric [1](#)  
Dyson [1](#)

## E

Earth-centered inertial frame [1](#)  
Earth-Centered-Earth-Fixed frame (ECEF) [1](#)  
Earth's lowest degree even zonal harmonics [1](#)  
Earth's spherical harmonics [1-2](#), [3-4](#)  
Earth-Yarkovsky effect [1](#)  
eccentricity [1](#), [2](#)  
- correction [1](#), [2](#)  
Eddington [1](#)  
Eddington limit [1](#)  
edge [1](#)  
effective gravitating density [1](#)  
effective potential [1](#), [2](#)  
efficiency [1](#)  
Ehrenfest [1](#)  
EIGEN-6C3stat Earth gravity model [1](#)  
Einstein [1](#), [2](#)  
Einstein equations [1](#)  
Einstein tensor [1](#)  
Einstein-Straus vacuole [1](#)  
energy momentum tensor [1](#)  
entropy [1](#)  
Ephemeris time [1](#)  
equivalence, principle of [1](#), [2](#), [3](#)  
ergosphere [1](#)  
European Space Agency (ESA) [1](#)  
even zonal harmonics [1](#), [2](#), [3](#)  
event horizon area [1](#)  
experimental triumph of General Relativity [1](#)  
extended configuration space [1](#)

extremal black hole [1](#)

## F

f(R) gravitational theories [1](#)

face [1](#)

factory frequency offset [1](#)

field equations and gravitomagnetism [1](#), [2](#)

field equations in Chern-Simons gravity [1](#)

finiteness [1](#)

First law of black hole thermodynamics [1](#)

flattening [1](#)

force [1](#)

fountain, cesium [1](#)

frame-dragging see dragging of inertial frames

free fall, universality [1](#)

Friedmann [1](#)

- universe [1](#)

## G

Galactic Centre [1](#)

GALILEO [1](#)

Galileo Galilei [1](#), [2](#), [3](#)

Gamma-Ray Bursts [1](#)

Geocentric Coordinate Time [1](#)

geodesic equation in gravitomagnetism [1](#), [2](#)-[3](#)

geodesic equation in weak field and slow motion [1](#)

geodesic motion [1](#)-[2](#)

geodesic motion and the LARES satellite [1](#)-[2](#)

geodesics [1](#)

geoid [1](#)

German Research Centre for Geosciences (GFZ) [1](#)

Giacconi [1](#)  
Global Navigation Satellite Systems and General Relativity [1](#)  
Global Positioning System (GPS) [1](#), [2](#)  
GLONASS [1](#)  
graph [1](#)  
Gravitational lensing  
- and dragging of inertial frames [1-2](#)  
- time-delay and frame-dragging [1-2](#)  
gravitational lensing [1](#)  
gravitational redshift [1](#), [2](#)  
gravitational wave [1](#)  
gravitational waves [1](#)  
gravitomagnetic  
-analogy [1-2](#)  
- field [1](#)  
- field and Chern-Simons gravity [1](#), [2](#)  
- potential [1](#)  
gravitomagnetism and curvature invariants [1](#), [2](#)  
gravitomagnetism and magnetism [1](#), [2](#)  
gravity probe B [1](#)  
Gravity Recovery and Climate Experiment (GRACE) [1](#), [2](#), [3](#), [4](#)  
gyroscope [1](#)  
gyroscopes precession [1](#), [2](#)

## H

Hamilton function [1](#)  
Hamilton function of General Relativity [1](#)  
hamiltonian constraint [1](#)  
Hamilton-Jacobi equation [1](#)  
Hawking radiation [1](#)  
Hawking temperature [1](#)  
Hazard [1](#)  
Hernquist cluster profile [1](#)  
high energy jet [1](#)

horizon [1](#)

horizon angular velocity [1](#)

horizon area [1](#)

Hubble diagram, local [1](#)

Hubble redshift distance relation [1](#)

## I

inertia [1](#)

inertial frames [1](#)

Innermost Stable Circular Orbit (ISCO) [1](#)

Intermediate Mass Black Hole (IMBH) [1](#)

International Astronomic Union (IAU) [1](#)

International Celestial Reference Frame (ICRF) [1](#)

International Earth Rotation Service (IERS) [1](#)

International Laser Ranging Service (ILRS) [1](#)

International Terrestrial Reference Frame (ITRF) [1](#)

iron-line emission [1](#)

Italian Space Agency (ASI) [1](#)

## J

J2000 [1](#)

Järnefelt [1](#)

JUNO mission [1](#)

## K

Kahn-Woltjer method with dark energy [1](#)

Kant [1](#)

Kepler [1](#)

Kerr [1](#), [2](#)

Kerr black hole [1](#), [2](#), [3](#), [4](#), [5](#)

Kerr metric [1](#), [2](#), [3](#)

Kerr parameter [1](#)

Killing vector [1](#)

Killing vector field [1](#)

King Lear [1](#)

Kuhn [1](#)

## L

LAGEOS (LAser GEodynamics Satellite) [1](#)

LAGEOS 2 [1](#)

LAGEOS satellites [1-2](#)

- and dragging of inertial frames [1](#), [2](#)

- and limits on Chern-Simons gravity [1](#)

- and limits on heterotic String theories [1](#)

Lambda [1](#)

LARES (LAser RElativity Satellite) [1](#), [2](#)

- and limits on Chern-Simons gravity [1](#)

- and limits on heterotic String theories [1](#)

- frame-dragging and error analyses [1-2](#)

- space experiment to test frame-dragging [1-2](#)

LARES and geodesic motion [1](#)

LARES, LAGEOS and LAGEOS 2 [1](#)

laser-ranged satellites [1](#)

lattice Yang-Mills theory [1](#)

laws of black hole thermodynamics [1](#)

Le Verrier [1](#)

leading order spin-orbit equations [1](#)

Lemaître [1](#), [2](#)

Lense-Thirring effect [1](#), [2-3](#)

Lense-Thirring effect observation [1](#)

light bending [1](#), [2](#), [3](#), [4](#)

light bending effect [1](#)

limits on Chern-Simons gravity by LAGEOS and LARES [1](#)  
limits on heterotic String theories by LAGEOS and LARES [1](#)  
line element [1](#)  
link [1](#)  
local dark energy  
- density estimate [1](#)  
- estimate methods [1](#)  
- estimating [1](#)  
- limits [1](#)  
local group zero gravity radius [1](#)  
local Hubble flows and  $\Omega_\Lambda$  [1](#)  
Loop quantum gravity [1](#)  
Lorentz contraction [1](#)  
Lorentz equation for gravitomagnetism [1](#)  
Lunar Laser Ranging and general relativity [1](#)  
Lynden-Bell [1](#)

## M

M81/M82 group zero gravity radius [1](#)  
Mach [1, 2](#)  
Mach's principle [1](#)  
magnetic field [1](#)  
Magorrian relation [1](#)  
Maxwell-Ampère equation in Chern-Simons gravity [1](#)  
McVittie [1](#)  
measurement [1](#)  
metric inside a hollow sphere [1](#)  
metric of spherically symmetric object [1](#)  
metric tensor [1, 2](#)  
metric theories of Gravitation [1](#)  
Michell [1](#)  
microlensing [1](#)  
Minkowski metric [1](#)  
Modified Julian Date (MJD) [1](#)

Monte Carlo simulations of the LARES experiment [1](#), [2](#)  
motion under gravity and dark energy [1](#)  
Mozart [1](#)

## N

Narayan [1](#)  
NASA [1](#)  
nature of quantum spacetime [1](#)  
navigation [1](#), [2](#)  
Newton [1](#), [2](#), [3](#)  
NFW cluster matter profile [1](#)  
no hair theorem [1](#)  
node [1](#)  
non-metric theories of Gravitation [1](#)  
nutation [1](#)

## O

obliquity [1](#)  
Odyssey [1](#)  
OJ287  
- Bardeen-Petterson effect [1](#)  
- binary black hole system [1](#)  
- flare processes [1](#)  
- flares and jet [1](#)  
- jet position angle [1](#)  
- light curve [1](#)  
- model parameter and outbursts [1](#)  
- past light curve [1](#)  
- past outburst times [1](#)  
- predicted outbursts [1](#)  
- strong field  $q$  [1](#)

- system parameters [1](#), [2](#)
- test of no-hair theorem [1](#)
- orbital determination programs [1](#), [2](#)
- orbital frequency [1](#)
- orbital precession [1](#)
- outbursts, OJ287 predictions [1](#)
- Outer event horizon [1](#)
- outflow in groups [1](#)
- outflows with/without dark energy [1](#)

## P

- Parametrized Post-Newtonian (PPN) approximation [1](#)
- particle energy [1](#)
- particle motion [1](#)
- Penrose process [1](#), [2](#), [3](#), [4](#)
- perihelion precession [1](#)
- perihelion precession of Mercury due to frame-dragging [1](#)
- Perlmutter [1](#)
- photon orbit [1](#)
- Planck scale [1](#)
- polar motion [1](#)
- Pontryagin pseudo-invariant [1](#)
- post Newtonian binary models [1](#)
- post Newtonian corrections of General Relativity [1](#)
- post Newtonian force law [1](#)
- post Newtonian modeling of OJ287 [1](#)
- potential, quadrupole [1](#)
- potential, tidal [1](#)
- precession [1](#), [2](#)
- precession of black hole binary 2.5 PN [1](#)
- precession of Mercury's orbit [1](#)
- primordial black hole [1](#)
- problem of time [1](#)
- process [1](#), [2](#)

proper time [1](#), [2](#)

PSR 1913+16 [1](#)

## Q

quadrupole moment

-  $J_2$  [1](#)

- of the Earth [1](#), [2](#)

-  $Q$  [1](#)

quadrupole parameter  $q_2$  [1](#)

quadrupole-monopole interaction [1](#)

quasar [1](#), [2](#), [3](#), [4](#), [5](#)

quasi-periodic oscillation [1](#), [2](#)

Quasi-Zenith Satellite System (QZSS) [1](#)

quintessence [1](#)

## R

radiatively inefficient flows (RIAFS) [1](#)

redshift [1](#)

reflection spectrum [1](#), [2](#)

Regge action [1](#)

Reissner-Nordstrom [1](#), [2](#), [3](#)

Reissner-Nordstrom-de Sitter [1](#)

relative evolution [1](#)

relativistic jet [1](#)

relativistic perihelion shift [1](#)

reverberation [1](#)

Ricci scalar [1](#), [2](#)

Ricci tensor [1](#)

Riemann tensor [1](#), [2](#)

Riess [1](#)

rotating black hole [1](#)

rotating black hole Kerr metric [1](#)

## S

Sagnac effect [1](#)

Salpeter [1](#)

Satellite Laser Ranging [1](#), [2](#), [3](#), [4](#)

scalar invariant [1](#)

Schmidt [1](#)

Schwarzschild black hole [1](#), [2](#), [3](#), [4](#)

Schwarzschild metric [1](#)

Schwarzschild radius [1](#)

Schwarzschild-de Sitter [1](#), [2](#)

second, definition of [1](#)

Seyfert [1](#)

Sgr A\* [1](#), [2](#), [3](#), [4](#), [5-6](#)

Sgr a\* no-hair theorem test [1](#)

shadow [1](#)

shadow radius [1](#)

singularities [1](#)

Sistine Chapel [1](#)

SOFA software [1](#)

solar and terrestrial radiation pressure [1](#)

solar radiation pressure [1](#)

spacetime geodesic [1](#)

speed of light [1](#), [2](#)

spikes [1](#)

spin [1](#)

spin network [1](#), [2](#)

spin network basis [1](#)

spin orbit binary solutions [1](#)

spin orbit coupling, relativistic [1](#)

spin orbit OJ287 orbit precession [1](#)

spin orbit OJ287 spin-vector precession [1](#)

stability [1](#)

stability of circular orbits [1](#)  
Starobinski [1](#)  
Stationary limit surface [1, 2](#)  
statistical states [1](#)  
Stonehenge [1](#)  
String and Brane-World theories and solar system tests [1](#)  
String theories and Chern-Simons gravity [1, 2](#)  
strong field [1](#)  
strong limit test, O287 [1](#)  
sum over geometries [1](#)  
Sun's Schwarzschild radius [1](#)  
supermassive black hole (SMBH) [1](#)  
supermassive black holes [1](#)  
superradiance [1](#)  
surface gravity [1, 2, 3](#)  
synchronization [1, 2](#)

## T

TAI [1](#)  
temporal variations of the Earth's gravitational field [1](#)  
terrestrial tides [1](#)  
tesseral tide [1](#)  
test of frame-dragging with the LAGEOS satellites [1-2](#)  
test of frame-dragging with the LARES satellites [1-2](#)  
test particle [1, 2](#)  
tests of frame-dragging [1](#)  
thermal thrust [1, 2](#)  
Thirring metric inside a hollow sphere [1](#)  
Thorne no-hair theorem test [1](#)  
tidal disruption event [1](#)  
time [1](#)  
time delay [1](#)  
time dilation [1](#)  
time, ephemeris [1](#)

time, GPS [1](#)  
time, TAI [1](#)  
time-delay, frame-dragging and gravitational lensing [1-2](#)  
timelike geodesic [1](#)  
transition amplitudes [1](#)  
triangulation [1](#)  
two-complex [1](#)

## U

UltraLuminous X-ray source (ULX) [1](#)  
universal time [1](#)  
Unruh [1](#)

## V

vacuole [1](#)  
vacuum density [1](#)  
vacuum fluctuation [1](#)  
van der Klis [1](#)  
VEGA [1](#)  
VEGA launcher [1](#)  
vertex [1](#)  
vertex amplitude [1](#)  
Very Long Baseline Interferometry and General Relativity [1](#)  
virial theorem with dark energy [1](#)  
virtual pair [1](#)

## W

weak field limit [1](#)

Weinberg [1](#)

Weyl tensor [1](#)

WGS84 [1](#)

Wheeler-deWitt equation [1](#)

## X

X-ray astronomy [1](#)

X-ray Background [1](#), [2](#)

X-ray reverberation [1](#)

## Y

Yarkovsky effect [1](#)

## Z

Zeldovich [1](#)

zero gravity

- distance [1](#)

- surface [1](#)

zonal harmonic of degree four [1](#)

# List of contributors

## **Prof. Dr. Andrew Fabian**

University of Cambridge  
Institute of Astronomy  
Madingley Road, Cambridge  
CB3 0HA, UK  
**acf@ast.cam.ac.uk**

## **Prof. Dr. Anthony Lasenby**

University of Cambridge  
Department of Physics, Cavendish Laboratory,  
JJ Thomson Avenue, Cambridge  
CB3 0HE, UK  
**a.n.lasenby@mrao.cam.ac.uk**

## **Prof. Dr. Gene Gilbert Byrd**

University of Alabama  
Department of Physics and Astronomy  
206 Gallalee Hall, 514 University Blvd.  
Tuscaloosa, AL  
35487-0324, USA  
**byrd@ua.edu**

## **Prof. Dr. Arthur D. Chernin**

Moscow State University  
Sternberg Astronomical Institute  
Universitetskii pr. 13, Moscow  
119992, Russia  
**chernin@sai.msu.ru**

## **Dr. Pekka Teerikorpi**

University of Turku  
Department of Physics and Astronomy  
Väisäläntie 20, Piikkiö  
21500, Finland  
**pekkatee@utu.fi**

**Prof. Dr. Mauri Valtonen**  
University of Turku  
Department of Physics and Astronomy  
Väisäläntie 20, Piikkiö  
21500, Finland  
**mauri.valtonen@utu.fi**

**Prof. Dr. Ignazio Ciufolini**  
University of Salento  
Department of Innovation Engineering  
Via per Monteroni, Lecce  
73100, Italy  
and  
Centro Fermi  
Piazza del Viminale 1, Roma  
00184, Italy  
**ignazio.ciufolini@unisalento.it**

**Prof. Dr. Neil Ashby**  
University of Colorado  
Department of Physics  
390 UCB, Boulder CO  
80309-0390, USA  
**ashby@boulder.nist.gov**

**Prof. Dr. Carlo Rovelli**  
Aix-Marseille University  
Centre de Physique Théorique de Luminy  
Case 907, Luminy

13288 Marseille, France  
**rovelli@cpt.univ-mrs.fr**

# **De Gruyter Studies in Mathematical Physics**

## **Volume 27**

Ivan A. Lukovsky, Peter V. Malyshев

Nonlinear Dynamics, 2015

ISBN 978-3-11-031655-1, e-ISBN (PDF) 978-3-11-031657-5,  
e-ISBN (EPUB) 978-3-11-038973-9, Set-ISBN 978-3-11-031658-2

## **Volume 26**

Vincent Laude

Phononic Crystals, 2015

ISBN 978-3-11-030265-3, e-ISBN (PDF) 978-3-11-030266-0,  
e-ISBN (EPUB) 978-3-11-038791-9, Set-ISBN 978-3-11-030267-7

## **Volume 25**

Sergey D. Algazin, Igor A. Kijko

Aeroelastic Vibrations and Stability of Plates and Shells, 2015

ISBN 978-3-11-033836-2, e-ISBN (PDF) 978-3-11-033837-9,  
e-ISBN (EPUB) 978-3-11-038945-6, Set-ISBN 978-3-11-040491-3

## **Volume 24**

Igor O. Cherednikov, Tom Mertens, Frederik F. van der Veken

Wilson Lines in Quantum Field Theory, 2014

ISBN 978-3-11-030910-2, e-ISBN (PDF) 978-3-11-030921-8,  
e-ISBN (EPUB) 978-3-11-038293-8, Set-ISBN 978-3-11-030922-5

## **Volume 23**

Boris A. Arbuzov

Non-perturbative Effective Interactions in the Standard Model,  
2014

ISBN 978-3-11-030292-9, e-ISBN (PDF) 978-3-11-030521-0,  
e-ISBN (EPUB) 978-3-11-038805-3, Set-ISBN 978-3-11-030522-7

[www.degruyter.com](http://www.degruyter.com)

1

There are other examples of this: Copernicus, for instance.

2

Cartesian physics had nothing of the sort: extension was a property of things and “space” was just a relation between things, so “empty space” was a logical impossibility.

3

The metaphor is Einstein’s.

4

Contemporary fundamental physics, perhaps because of over-reading Kuhn has a certain difficulty in digesting this fact; I suspect this to be big part of its recent sterility.

5

Einstein’s letter to Ehrenfest, 5 January 1916

6

Even less they are a priori conditions for our experience. Recall that Kant got so wrong as to state that Newtonian mechanics is true a priori. We certainly make large use of forms of understanding that do not follow from experience, but as the consequence of experience these evolve.

7

He wrote a paper claiming that the Schwarzschild solution could not describe reality.

8

3C273 is not just a pointlike object but has an associated linear structure which we now know to be a relativistic jet, but the above argument is good for the main central source.

9

Newman et al [48] found the solutions to the Einstein-Maxwell field equations in 1965 and the resulting Kerr-Newman geometry describes spinning charged black holes.

10

This means that the mass of any galactic disc is ignored.

11

If the mass accretion rate can be measured (from observations of the outer disc, say) and the total luminosity is known, then in principle spin can be deduced from the radiative efficiency using the relation in [Figure 15](#).

12

No clock can be decoupled from the gravitational field.

13

The general relativistic time coordinate does not perform the same role as the Schrödinger equation time, because the corresponding generator is a constraint, and vanishes. The notion of a physical time measured by a clock reappears in general relativity under other forms, such as the proper length of a timelike worldline.

14

Or “measurement”, in standard parlance. “Measurement” has a misleading connotation: it sounds as to refer to the presence of a human being, or a recording device. Nothing of this is required to make sense of quantum theory [31].

15

The boundary does not need to be considered split a priori into past and future. If we do consider this split, then the boundary Hilbert space splits into the tensor product of an “in” and an “out” Hilbert spaces. Tensor product states correspond to pure states, while generic states include statistical states.[7]

16

For simplicity, I consider here only two-complexes that are dual to a 4d triangulation [5]. **6** For a recent clarifying discussion on the hamiltonian structure of the theory, see [40].

17

Without cosmological constant, they are still ultraviolet finite, but there are diverging radiative corrections describing large “spikes” of the geometry. These are cut off by the cosmological constant.

18

See [32] for a discussion on the “diff invariance” of these prediction.