Lecture Notes in Physics

The Editorial Policy for Edited Volumes

The series *Lecture Notes in Physics* (LNP), founded in 1969, reports new developments in physics research and teaching - quickly, informally but with a high degree of quality. Manuscripts to be considered for publication are topical volumes consisting of a limited number of contributions, carefully edited and closely related to each other. Each contribution should contain at least partly original and previously unpublished material, be written in a clear, pedagogical style and aimed at a broader readership, especially graduate students and nonspecialist researchers wishing to familiarize themselves with the topic concerned. For this reason, traditional proceedings cannot be considered for this series though volumes to appear in this series are often based on material presented at conferences, workshops and schools.

Acceptance

A project can only be accepted tentatively for publication, by both the editorial board and the publisher, following thorough examination of the material submitted. The book proposal sent to the publisher should consist at least of a preliminary table of contents outlining the structure of the book together with abstracts of all contributions to be included. Final acceptance is issued by the series editor in charge, in consultation with the publisher, only after receiving the complete manuscript. Final acceptance, possibly requiring minor corrections, usually follows the tentative acceptance unless the final manuscript differs significantly from expectations (project outline). In particular, the series editors are entitled to reject individual contributions if they do not meet the high quality standards of this series. The final manuscript must be ready to print, and should include both an informative introduction and a sufficiently detailed subject index.

Contractual Aspects

Publication in LNP is free of charge. There is no formal contract, no royalties are paid, and no bulk orders are required, although special discounts are offered in this case. The volume editors receive jointly 30 free copies for their personal use and are entitled, as are the contributing authors, to purchase Springer books at a reduced rate. The publisher secures the copyright for each volume. As a rule, no reprints of individual contributions can be supplied.

Manuscript Submission

The manuscript in its final and approved version must be submitted in ready to print form. The corresponding electronic source files are also required for the production process, in particular the online version. Technical assistance in compiling the final manuscript can be provided by the publisher‘s production editor(s), especially with regard to the publisher’s own LaTeX macro package which has been specially designed for this series.

LNP Homepage (springerlink.com)

On the LNP homepage you will find:
−The LNP online archive. It contains the full texts (PDF) of all volumes published since 2000. Abstracts, table of contents and prefaces are accessible free of charge to everyone. Information about the availability of printed volumes can be obtained.
−The subscription information. The online archive is free of charge to all subscribers of the printed volumes.
−The editorial contacts, with respect to both scientific and technical matters.
−The author’s / editor’s instructions.

S.G. Karshenboim E. Peik (Eds.)

Astrophysics, Clocks and Fundamental Constants

Springer

Editors

Savely G. Karshenboim
D.I. Mendeleev Institute for Metrology
198005 St. Petersburg
Russia
and
Max-Planck-Institut für Quantenoptik
85748 Garching
Germany

Ekkehard Peik
Physikalisch-Technische Bundesanstalt
Bundesallee 100
38116 Braunschweig
Germany

S. G. Karshenboim E. Peik (Eds.), *Astrophysics, Clocks and Fundamental Constants*, Lect. Notes Phys. **648** (Springer, Berlin Heidelberg 2004), DOI 10.1007/b13178

Library of Congress Control Number: 2004107591

ISSN 0075-8450
ISBN 3-540-21967-6 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by the authors/editor
Data conversion: PTP-Berlin Protago-TeX-Production GmbH
Cover design: *design & production*, Heidelberg

Printed on acid-free paper
54/3141/ts - 5 4 3 2 1 0

Lecture Notes in Physics

For information about Vols. 1–604
please contact your bookseller or Springer-Verlag
LNP Online archive: springerlink.com

Vol.605: G. Ciccotti, M. Mareschal, P. Nielaba (Eds.), Bridging Time Scales: Molecular Simulations for the Next Decade.

Vol.606: J.-U. Sommer, G. Reiter (Eds.), Polymer Crystallization. Obervations, Concepts and Interpretations.

Vol.607: R. Guzzi (Ed.), Exploring the Atmosphere by Remote Sensing Techniques.

Vol.608: F. Courbin, D. Minniti (Eds.), Gravitational Lensing:An Astrophysical Tool.

Vol.609: T. Henning (Ed.), Astromineralogy.

Vol.610: M. Ristig, K. Gernoth (Eds.), Particle Scattering, X-Ray Diffraction, and Microstructure of Solids and Liquids.

Vol.611: A. Buchleitner, K. Hornberger (Eds.), Coherent Evolution in Noisy Environments.

Vol.612: L. Klein, (Ed.), Energy Conversion and Particle Acceleration in the Solar Corona.

Vol.613: K. Porsezian, V.C. Kuriakose (Eds.), Optical Solitons. Theoretical and Experimental Challenges.

Vol.614: E. Falgarone, T. Passot (Eds.), Turbulence and Magnetic Fields in Astrophysics.

Vol.615: J. Büchner, C.T. Dum, M. Scholer (Eds.), Space Plasma Simulation.

Vol.616: J. Trampetic, J. Wess (Eds.), Particle Physics in the New Millenium.

Vol.617: L. Fernández-Jambrina, L. M. González-Romero (Eds.), Current Trends in Relativistic Astrophysics, Theoretical, Numerical, Observational

Vol.618: M.D. Esposti, S. Graffi (Eds.), The Mathematical Aspects of Quantum Maps

Vol.619: H.M. Antia, A. Bhatnagar, P. Ulmschneider (Eds.), Lectures on Solar Physics

Vol.620: C. Fiolhais, F. Nogueira, M. Marques (Eds.), A Primer in Density Functional Theory

Vol.621: G. Rangarajan, M. Ding (Eds.), Processes with Long-Range Correlations

Vol.622: F. Benatti, R. Floreanini (Eds.), Irreversible Quantum Dynamics

Vol.623: M. Falcke, D. Malchow (Eds.), Understanding Calcium Dynamics, Experiments and Theory

Vol.624: T. Pöschel (Ed.), Granular Gas Dynamics

Vol.625: R. Pastor-Satorras, M. Rubi, A. Diaz-Guilera (Eds.), Statistical Mechanics of Complex Networks

Vol.626: G. Contopoulos, N. Voglis (Eds.), Galaxies and Chaos

Vol.627: S.G. Karshenboim, V.B. Smirnov (Eds.), Precision Physics of Simple Atomic Systems

Vol.628: R. Narayanan, D. Schwabe (Eds.), Interfacial Fluid Dynamics and Transport Processes

Vol.629: U.-G. Meißner, W. Plessas (Eds.), Lectures on Flavor Physics

Vol.630: T. Brandes, S. Kettemann (Eds.), Anderson Localization and Its Ramifications

Vol.631: D. J. W. Giulini, C. Kiefer, C. Lämmerzahl (Eds.), Quantum Gravity, From Theory to Experimental Search

Vol.632: A. M. Greco (Ed.), Direct and Inverse Methods in Nonlinear Evolution Equations

Vol.633: H.-T. Elze (Ed.), Decoherence and Entropy in Complex Systems, Based on Selected Lectures from DICE 2002

Vol.634: R. Haberlandt, D. Michel, A. Pöppl, R. Stannarius (Eds.), Molecules in Interaction with Surfaces and Interfaces

Vol.635: D. Alloin, W. Gieren (Eds.), Stellar Candles for the Extragalactic Distance Scale

Vol.636: R. Livi, A. Vulpiani (Eds.), The Kolmogorov Legacy in Physics, A Century of Turbulence and Complexity

Vol.637: I. Müller, P. Strehlow, Rubber and Rubber Balloons, Paradigms of Thermodynamics

Vol.638: Y. Kosmann-Schwarzbach, B. Grammaticos, K.M. Tamizhmani (Eds.), Integrability of Nonlinear Systems

Vol.639: G. Ripka, Dual Superconductor Models of Color Confinement

Vol.640: M. Karttunen, I. Vattulainen, A. Lukkarinen (Eds.), Novel Methods in Soft Matter Simulations

Vol.641: A. Lalazissis, P. Ring, D. Vretenar (Eds.), Extended Density Functionals in Nuclear Structure Physics

Vol.642: W. Hergert, A. Ernst, M. Däne (Eds.), Computational Materials Science

Vol.643: F. Strocchi, Symmetry Breaking

Vol.644: B. Grammaticos, Y. Kosmann-Schwarzbach, T. Tamizhmani (Eds.) Discrete Integrable Systems

Vol.645: U. Schollwöck, J. Richter, D.J.J. Farnell, R.F. Bishop (Eds.), Quantum Magnetism

Vol.646: N. Bretón, J. L. Cervantes-Cota, M. Salgado (Eds.), The Early Universe and Observational Cosmology

Vol.647: D. Blaschke, M. A. Ivanov, T. Mannel (Eds.), Heavy Quark Physics

Vol.648: S. G. Karshenboim, E. Peik (Eds.), Astrophysics, Clocks and Fundamental Constants

Vol.649: M. Paris, J. Rehacek (Eds.), Quantum State Estimation

Foreword

We know that we live in an expanding and changing universe, but nevertheless we prefer some solid foundations upon which to base our reasoning. This idea is reflected in the introduction of quantities that we call the "fundamental physical constants". These constants do, as John D. Barrow said, "capture at once our greatest knowledge and our greatest ignorance about the universe". The interrelations between them give us hints where to look for underlying common foundations of phenomena that are observed in different branches of physics. At the same time, we also call these quantities "fundamental" because we have not so far been able to deduce their numerical values from a deeper level of understanding.

As supposedly universal and invariable quantities, the fundamental constants constitute an essential pillar for the definition and representation of the SI units. The replacement of man-made artefacts by quantum standards for the representation of the units has led to enormous progress in precision measurements. Provided a sufficient theoretical understanding of the quantum standard is available, its output may be linked to the fundamental constants and that is why many metrology institutes, like the PTB, are actively involved in the high precision determinations of these values.

While for practical purposes we would like to trust in the universality and invariability of the constants, as physicists we may be urged to challenge these postulates. Possible variations in time or space would have far reaching consequences and may point the way to a unified theory of the fundamental interactions and the establishment of a quantum theory of space-time. Discrepancies with respect to the presently believed interrelations (e.g. between the value of the Planck constant h from electrical measurements and from the Avogadro project with the silicon single-crystal method) may lead to new insights as well.

The study of the fundamental constants and their possible variation is an interdisciplinary and "global" task, i.e. involving experts from very different fields of physics all over the world. It is therefore essential to bring scientists from these different fields together in symposia like "Astrophysics, Clocks and Fundamental Constants". Fortunately, the speakers at that conference agreed to make their contributions available to a wider audience and contributed to this book. Astrophysics provides information about the state of the universe in the remote past and about regions of space far away from us, geophysics teaches us the history of our planet, laboratory experiments (especially with atomic clocks, as frequency and time are the most precisely measurable physical quantities)

can furnish a precise snapshot of present temporal derivatives, whereas satellite missions may be used to probe the solar system with precision instruments. This book presents a collection of excellent reviews on these topics.

Physikalisch-Technische Bundesanstalt *Ernst O. Göbel*
Braunschweig
May 2004

Preface

Fundamental physical constants play an important role in modern physics. The "old fashioned" celestial mechanics knew only one such constant, the Newtonian constant of gravitation G. Now the list of basic universal constants is much longer. The speed of light c is a basic constant of relativistic physics and its value is a part of the international system of SI units. It is not even possible to imagine contemporary physics without the fine structure constant α, the Planck constant $\hbar$, the electron mass m_e, the Rydberg constant Ry, etc. A frequent appearance of the same constant in different branches of physics demonstrates its universality and significance for fundamental physics and numerous applications.

Most of the fundamental constants entered physics in the late nineteenth or early twentieth century. The early twentieth century was not only a time when a substantial part of the constants were introduced into modern physics with the establishment of quantum physics but that was also the time when the question of the constancy of the "constants" was raised by P.A.M. Dirac. He considered the problem in the context of a comparison of the electromagnetic and gravitational interactions. Dirac's original idea is now rather out of consideration; however, the link between the grand unification of all basic interactions and a variability of their coupling constants has survived and still seems to be important.

A search for variations in the values of fundamental physical constants gives us a rare but excellent example of the strong interplay between new fundamental physics and the development of applied studies such as frequency metrology and space navigation.

Our book is formed of invited reviews presented at an international conference "Astrophysics, Clocks and Fundamental Constants" organized as the 302nd WE-Heraeus-Seminar which took place in June, 2003, in Bad Honnef in Germany. We tried to combine in our meeting the contributions on

- *Astrophysics*, a science which among other branches of physics is most involved in the study of the time evolution of our world; it also allows us to probe objects located far away from us in time and space;
- *Fundamental Constants*, which involves a broad variety of questions ranging from deeply fundamental properties of Nature to the development of new standards;
- *Clocks*, the modern frequency standards, which offer the most accurate measurements with an accuracy superseding any other measurements by several

orders of magnitude and therefore provide the most favourable opportunity for a search of a possible time variation of fundamental constants in the laboratory.

The field of variations of fundamental constants involves various pieces of theory and experiments from so many parts of physics that there may be no single person in the world who is a real expert in the field. We had the hope, however, that the conference participants all together form a kind of collective expert. We think now that we were right and that this book has been written by such an expert covering most aspects of the field related to astrophysics, cosmology, geochemistry, molecular, atomic, nuclear and particle physics, quantum field theory, space science and metrology. The contributions to the book have been prepared with the strongly interdisciplinary character of the conference in mind. It was our intention to encourage an exchange between the various specialized subfields so that the book addresses a wide audience of physicists from any of these fields as well as students looking for an introduction into this exciting and topical area of research.

We live in an expanding universe at a time when even the number of dimensions of our world is not a completely clear problem and in experiencing the changing Nature we may wonder why the fundamental constants should be left unchanged. Should they?

We are grateful to the WE-Heraeus-Stiftung for their support and substantial help in organizing the conference.

Braunschweig
May 2004

Savely Karshenboim,
Ekkehard Peik

Contents

Part VIII Space Missions and General Relativity

List of Contributors

Sebastien Bize,
BNM-SYRTE,
Observatoire de Paris,
61 avenue de l'Observatoire,
75014 Paris, France,
sebastien.bize@obspm.fr

Gerhard Börner,
Max-Planck-Institut für Astrophysik,
Karl-Schwarzschild-Str. 1,
85748 Garching, Germany,
gboerner@mpa-garching.mpg.de

Marc Fischer,
Max-Planck-Institut
für Quantenoptik,
85748 Garching, Germany,
mcf@mpq.mpg.de

Victor V. Flambaum,
School of Physics,
University of New South Wales,
Sydney N.S.W. 2052, Australia,
v.flambaum@unsw.edu.au

Jeff Flowers,
National Physical Laboratory,
Teddington, UK,
jeff.flowers@npl.co.uk

Harald Fritzsch,
Ludwig-Maximilians-University
Munich,
Sektion Physik,
Theresienstraße 37,
80333 München, Germany,
fritzsch@mppmu.mpg.de

Yasunori Fujii,
Advanced Research Institute
for Science and Engineering,
Waseda University,
Tokyo 169-8555, Japan,
fujii@e07.itscom.net

Savely G. Karshenboim,
D. I. Mendeleev Institute
for Metrology,
198005 St. Petersburg, Russia,
and
Max-Planck-Institut
für Quantenoptik,
85748 Garching, Germany,
sek@mpq.mpg.de

Claus Kiefer,
Institut für Theoretische Physik,
Universität zu Köln,
Zülpicher Str. 77,
50937 Köln, Germany,
kiefer@thp.uni-koeln.de

Michael Kramer,
University of Manchester,
Jodrell Bank Observatory,
Cheshire SK11 9DL, UK,
mkramer@jb.man.ac.uk

Sergei A. Levshakov,
Department
of Theoretical Astrophysics,
Ioffe Physico-Technical Institute,
194021 St. Petersburg, Russia,
lev@astro.ioffe.rssi.ru

Lute Maleki,
Jet Propulsion Laboratory,
4800 Oak Grove Drive,
Pasadena, CA 91109 USA,
lute.maleki@jpl.nasa.gov

William J. Marciano,
Brookhaven National Laboratory,
Upton, NY 11973, USA,
marciano@bnl.gov

Michael T. Murphy,
Institute of Astronomy,
University of Cambridge,
Madingley Road,
Cambridge CB3 0HA, UK,
mim@ast.cam.ac.uk

Lev B. Okun,
ITEP, Moscow,
117218, Russia,
okun@heron.itep.ru

Ekkehard Peik,
Physikalisch-Technische
Bundesanstalt,
Bundesallee 100,
38116 Braunschweig, Germany,
ekkehard.peik@ptb.de

Fritz Riehle,
Physikalisch-Technische
Bundesanstalt,
Bundesallee 100,
38116 Braunschweig, Germany,
fritz.riehle@ptb.de

Stephan Schiller,
Universität Düsseldorf,
Institut für Experimentalphysik,
Universitätsstr. 1,
40225 Düsseldorf, Germany,
step.schiller@uni-duesseldorf.de

Christian Tamm,
Physikalisch-Technische
Bundesanstalt,
Bundesallee 100,
38116 Braunschweig, Germany,
christian.tamm@ptb.de

Slava G. Turyshev,
Jet Propulsion Laboratory,
4800 Oak Grove Drive,
Pasadena, CA 91109 USA
turyshev@jpl.nasa.gov

Joachim von Zanthier,
MPI für Quantenoptik
and
Sektion Physik
der LMU München,
Hans-Kopfermann-Str.1,
85748 Garching, Germany,
jvz@mpq.mpg.de

Jun Ye,
JILA, National Institute
of Standards and Technology
and
University of Colorado,
Boulder, Colorado 80309-0440, USA,
ye@jila.colorado.edu

An Introduction to Varying Fundamental Constants

Savely G. Karshenboim[1,2] and Ekkehard Peik[3]

[1] D. I. Mendeleev Institute for Metrology (VNIIM), 198005 St. Petersburg, Russia
[2] Max-Planck-Institut für Quantenoptik, 85748 Garching, Germany
[3] Physikalisch-Technische Bundesanstalt, 38116 Braunschweig, Germany

1 Introduction

The way physics explains Nature is to speak in terms of the consequences of a few very basic equations. The electromagnetic phenomena were explained within classical physics by Maxwell's equations. Newtonian gravity and the rules of mechanics completed the classical picture. Assuming an atomic nature of substance, statistical physics allowed thermodynamics to be established from mechanics. Relativity discovered deep connections between the symmetry of space-time and the laws of Nature. Later, relativity was generalized to incorporate gravitational effects. Quantum physics in its turn introduced new mechanics and new theories of electromagnetic, weak and strong interactions. It explained most of the then-known quantum phenomena and predicted a number of new ones. It also successfully explained the atomic nature of substance and classical physics as a limit of the quantum world. Efforts to combine quantum physics with gravity are on their way nowadays.

Reducing all problems of our world to such a few basic identities one has to acknowledge the importance of their parameters which play a prominent role in physics. These parameters are the fundamental constants.

For a while classical mechanics recognized two basic constants: Newton's gravitational constant G and the free-fall acceleration g. Today, the former is one of the three most fundamental constants that determine the Planck length scale where we suspect that space (and time) possess new and unexpected properties. Meanwhile, the latter is clearly seen not to be a constant at all. Historically, these two "constants" seemed to be universal. Perhaps g was even more universal because it was applied to various problems, while G was really crucial only for a very few. Obviously G was not important e.g. for the description of planetary motion in the first approximation because it always appeared together with the mass of the Sun. Looking at more complicated systems (like Sun-Earth-Moon) one can approach the universal gravitational constant but only by estimating the masses of the Earth, Sun or Moon can one deal with G directly. Now we know that g is not constant, but possesses small space and time gradients (being affected by Sun, Moon and even by trains and cars, etc.). Despite that, it is still a very important quantity for practical measurements, since weighing an appropriate mass is still the best practical approach for realizing a determined value of force.

S.G. Karshenboim, E. Peik, An Introduction to Varying Fundamental Constants, Lect. Notes Phys. **648**, 1–18 (2004)
http://www.springerlink.com/

So our experience tells us that a quantity which looks like a universal constant may happen to be the result of an interaction between some fundamental effects and the environment on a 'large' scale. In the case of cosmological evolution of the universe the environment is the universe as a whole. We have learned that it is obviously a "creative" environment which can really affect the fundamental constants and in particular the basic properties of particles. For instance, if the so-called inflation of the universe took place in the remote past (and that is a very likely scenario – see e.g. the contribution by Börner [1] and [2]), our world has experienced one or several phase transitions which strongly affected the electron mass and the fine structure constant. The universe is also quite a puzzling environment which we still hardly understand. Only a small portion (nearly 5%) of it corresponds to known forms of matter, while the rest is so-called 'dark matter' and 'dark energy' [1], which interacts very weakly with 'normal matter', but still it does interact and in particular determines the expansion of the universe.

In the early days of quantum mechanics an understanding of new physics went along with the development of new technology. These two directions (fundamental and technological) were nearly identical because new technology was often designed either as a probe for new physics or as a result of its immediate realization. In present days, fundamental physics and applied physics – searching for new technologies – these two basic directions in physics are strongly separated. Indeed, advanced technology offers us better experimental data and thus helps to develop fundamental physics. However, this is not the same as it was in the time of the 'young' quantum physics. The work for conceptually new fundamental physics is concentrated today in the field of high energy physics (maybe even ultra-high energy physics which can only be studied theoretically) and in a few problems of cosmology such as the problem of dark matter and dark energy. Meanwhile, new technology mainly comes from 'low energy physics' such as atomic and molecular physics, (quantum) electronics or space missions.

The search for a variation of the fundamental constants gives us a very rare example of a strong overlap between a look into the deepest fundamental problems of grand unification, quantum gravity, cosmology and highly advanced applied technology such as is found in frequency metrology and space navigation. Metrological work for the development of better frequency standards has turned out to be a probe for new physics. In return, the study of grand unification and the even more 'unpractical' quantum gravity may help us with a better understanding of frequency standards. And the problem of the variability of the fundamental constants is a key issue of such an intriguing interaction of fundamental and applied physics. Meanwhile, precision space navigation is strongly related to high precision tests of general relativity with a search for a possible variation of the gravitational constant G and for possible corrections of Einstein's theory.

The topics related to the fundamental constants and their possible variability involve pieces from very different branches of fundamental and applied physics. No single scientist can be an expert in such an interdisciplinary problem. That was a reason for organizing a conference in order to bring together experts from different subfields. The conference *Astrophysics, Clocks and Fundamental Constants* (ACFC) took place in Bad Honnef (Germany) in June 2003 and most in-

vited speakers have kindly agreed to contribute to this book. The book presents review contributions of the topics discussed at the meeting. Very different topics are covered here and our introductory paper is meant to give a brief overview of the subject and of the book.

2 Fundamental Constants

In physics we prefer not to use formal definitions like in mathematics but we should still fix notations and thus we need first of all to decide what are the fundamental constants and what is their role in physics and metrology. And we can see immediately that this is not a simple question.

Classical physics offers us the Newtonian gravitational constant G, the Boltzmann constant k, speed of light c, the magnetic constant μ_0 and in part the Avogadro constant N_A. We stress that the latter is a classical constant only in part since the atomic nature of substance is a consequence of quantum physics. In principle, classical physics can even provide us with (semi)classical measurements of particle properties like mass, charge, magnetic moment, etc. However, those are clearly of quantum origin. Quantum physics first explained the identity of objects and provided powerful tools to verify this via interference and in particular the Pauli principle and Bose condensation. It has thus allowed us to speak about the electron charge as a quantity which describes each individual electron, and not just an average electron (like e.g. temperature).

What can we say about the classical constants? The gravitational constant G and the speed of light c are fundamental properties of space and time (see the contribution by Okun [3]), while the other three (k, μ_0 and N_A) are in part a result of convention. Still they are important because we clearly recognize that measurements of temperature and of atomic kinetic energy are related to two quite different experiments and thus for practical reasons these two closely connected quantities are to be measured in different units. The same is true with other constants. Their origin is mainly due to the requirements of measurements, which are considered in the contribution by Flowers and Petley [4].

Quantum physics introduced more fundamental constants. Indeed, the most important is the Planck constant h which is related to properties of space, while the other quantum constants either are properties of particles (electrons, protons), or describe macroscopic quantum effects. The latter, realized with devices of macroscopical 'classical' size – like a Josephson junction for example – are of a special importance in metrology.

There are two very different kinds of constants. Some carry dimensions and thus their numerical values are results of conventions. Okun suggested naming them 'fundamental units' [3]. The others are dimensionless and their numerical values are objective. The most well-known example of these is the fine structure constant

$$\alpha = \frac{e^2}{4\pi\epsilon_0\hbar c} \,, \tag{1}$$

where ϵ_0 is the electric constant.

These two kinds of constants (dimensional and dimensionless) are important in different ways. Some 'fundamental units' are directly or indirectly involved in the definition, realization or maintenance of practical units. These have a long history and were realized by different methods. Even now, the international system of units (SI) involves such different units as the *kilogram* (defined by an artefact chosen by a convention), the *second* (defined by an atomic transition chosen by a convention), the *metre* (defined via the *second* and a fixed value of the speed of light c) and the *ampere* (defined via mechanical units and the magnetic constant μ_0).

Indeed, from the point of view of fundamental physics we would like to define and realize all our practical units in terms of quantum effects (which provide us with time-independent identical objects) and fundamental constants. The present practical situation is considered in [4].

The importance of dimensionless constants is in part different. However, only in part, because we can still consider them as conversion factors between different units. For example, a natural unit of resistance related to the SI system (i.e. a unit of classical physics) is

$$R_{\rm cl} = \mu_0 c = \sqrt{\frac{\mu_0}{\epsilon_0}}\ . \tag{2}$$

Since SI suggests some numerical values of the dimensional constants c and μ_0, not equal to unity, the *ohm* of SI is determined via $R_{\rm cl}$, but is not equal to $R_{\rm cl}$:

$$1\ \text{ohm} = \left(376.730\,313\,451...\right)^{-1} R_{\rm cl}\ . \tag{3}$$

However, there is also a natural quantum unit which comes from the quantum Hall effect, also called the von Klitzing effect. The quantum unit of resistance is the von Klitzing constant

$$R_K = \frac{h}{e^2}\ . \tag{4}$$

This unit is very important for the practical application in pure electrical measurements [5]; however, one has to follow the SI units in order to provide consistency between electromagnetic and mechanical measurements. One can easily find the conversion factor between these two natural units

$$R_K = \frac{1}{2\alpha}\, R_{\rm cl}\ , \tag{5}$$

which involves the fine structure constant.

This example demonstrates that the practical units are often determined by certain natural units, but are not equal to them, and that there may be several natural units for the same quantity. The natural units based on fundamental constants are often related to special ranges of phenomena and can be very useful for their description. A ratio of two such units can bridge two different areas of measurements (e.g. atomic and nuclear magnetic moments) and be quite helpful in the establishment of official or inofficial units for special cases (e.g. atomic mass unit and nuclear magneton). The determination of the 'best' values of the 'practical' constants which play a fundamental role in the realization of the SI and non-SI units is done regularly by CODATA [6].

3 Variability of Fundamental Constants

A physical measurement is always a comparison. Some measurements are realized as relative, but some pretend to be 'absolute'. They are in fact quite different. The absolute measurement is a comparison of a given quantity to a certain special quantity. The latter, for some measurements, is a result of convention (like the SI second), while for others it is a result of measurements with quantities of different dimensions and a certain dimensional constant fixed by a convention (e.g. SI ampere). Still, to realize the conventional units one has to build a device. The unit itself may be fixed without any uncertainty, whereas the realization of the unit always involves uncertainty. This uncertainty of the realization may be a source of variability of the numerical values of dimensional constants. Another source could be the SI convention itself. Firstly, it involves an artefact (the prototype of the SI kilogram). Secondly, it fixes the conditions of the realization officially or inofficially. While the frequency of the hyperfine interval of the cesium atom may well be a constant, improper understanding of the conditions to realize the SI definition of the second may in practice create certain troubles. And indeed, the value of the second can be affected by a 'real' variability of the fundamental constants.

Variability of the fundamental constants should affect not only units and standards based on some accepted or recommended values of the fundamental constants, but also different properties of quantum objects, such as transition frequencies. Even for the hydrogen atom such a frequency cannot be calculated with an accuracy compatible with the one of contemporary frequency standards. However, as long as we consider variations of the fundamental constants we have to look for a possible variation of properties of objects, which are far from fundamental. In the context of present accuracies and constraints a question related to the variability of the constants is the stability of the involved standards, mainly those for frequency, i.e. for atomic clocks.

The dimensionless fundamental constants play an important role in the question of a possible time-variation of fundamental constants. Indeed, time- and space-variations are possible both for dimensional and dimensionless values; however, the experiments performed to look for them are quite different (see e.g. [7]). The direct study of possible time- and space-gradients of dimensional quantities is much less sensitive than a comparison of two results that are widely separated in time or in space. In this case, since the units used in the comparison may also vary, we have to deal with dimensionless quantities only.

In discussing the dimensionless quantities which are in principle determined by the fundamental constants we need to understand the origin of such quantities and in particular to respond to several questions such as

- If the measurable quantities can be expressed in a model-independent way in terms of certain basic constants;
- If these basic constants are *fundamental*;
- If these basic constants are *constants*;
- If (and how) these basic constants are correlated.

From a practical point of view, the question of variability of the constants is a question about the stability of natural clocks. A variety of stable periodic phenomena is available in:

- Atomic clocks;
- Molecular vibrations and rotations;
- Nuclear transitions (see [8,9]);
- Solid oscillators (see e.g. [10]);
- Planetary motion;
- Rotation of pulsars.

For the dating of events from the remote past of the Earth and of meteorites one can use different radioactive clocks.

The characteristic time interval (or frequency) related to different clocks is also a subject of consideration in terms of fundamental constants. Because of the large variety of clocks we can receive controversial data and for this reason it is of practical importance to be able to compare results from different measurements in the same terms. From a practical point of view, the basic constants used for cross-comparisons may not necessarily be the fundamental constants, but very likely they will, because the fundamental constants appear as the parameters which universally enter equations from different branches of physics.

3.1 'Fundamentality' of the Fundamental Constants

One part of the answer to the question on the 'fundamentality' of the constants is the attempt to classify the various properties under question ([7]). We do not pretend to have found the best classification scheme, but it suits our purposes (cf. e.g. Table 2 of [4]). One could classify:

- Constants of space-time, the scene where particles exist; those constants are $c, \hbar, G$ (cf. [3]);
- Coupling constants of fundamental interactions; an example is the fine structure constant α;
- Properties of fundamental particles such as their masses (in particular, the Rydberg constant Ry, which is the basic spectroscopic constant, is a product of a few constants involving the electron mass);
- Secondary quantum constants such as frequencies of atomic and molecular transitions (which, with a certain accuracy, could be connected in terms of more fundamental parameters);
- Secondary quantum constants such as the position of the samarium resonance (which plays an important role in geochemical estimations (see Sect. 4 and [11,12] for more details)) and other nuclear properties (which cannot be connected with fundamental constants in a model-independent way);
- Properties of classical macroscopic objects (such as the Sun or a pulsar); those can vary not only because of variations in the parameters of particles and atoms but also because of a variation in the number of atoms which form them or because of effects of their internal structure.

We would like to point out a few questions raised by this classification:

- What are *fundamental* interactions? Is the interaction with the Higgs sector fundamental? In contrast to quantum electrodynamics and similar theories of weak and strong interactions it involves a large number of different coupling constants. And this interaction determines such observable constants as the electron mass.
- Which particles are *fundamental*? And in particular: is the proton (neutron) a fundamental particle? Or are quarks the fundamental particles? Then we would need to be able to derive proton properties from a more fundamental theory of quarks (and that is not easy).

We leave these questions for a moment, noting that they are closely related to intercorrelations between different constants and the hierarchy of their variation rates in certain scenarios (see below).

3.2 'Constancy' of the Fundamental Constants

Before we discuss any details of their possible variations we need to question the 'constancy' of constants. Different reasons for variations of fundamental constants and different scenarios have already been discussed in the literature. Here, we would like to mention a cosmological trace of this story. The contemporary accepted model of the evolution of our universe assumes that in the very early stage our world went through certain phase transitions and that during one of them the electron and quarks received their mass. Quantum electrodynamics with massless charged fermions, which existed before the transition, is quite different from the one we observe now where all charged particles possess masses. The fine structure constant defined in a standard way was then infinite because of renormalization effects. Details of the cosmological evolution of the universe can be found in the contribution by Börner [1] in this book. As it is discussed in [7], inflation may (under certain conditions) also produce some space variations of fundamental constants. Another good reason for a variability may be extra dimensions (see e.g. the papers by Marciano [13,14]).

3.3 Intercorrelations Between the Fundamental Constants

Early estimations of variation of fundamental constants were often based on the idea that *one* of them is varying (e.g. the fine structure constant) while the others remain unchanged. This is obviously inconsistent. However, we cannot simply say that all constants should vary at approximately the same rate because some hierarchy may be in place. The enhancement or suppression factor may be up to 100 in different scenarios (see e.g. [13–15]). To understand the problem we need to discuss possible correlations between the fundamental constants and for this purpose we need to return to the question of their 'fundamentality' which is strongly related to their origin.

One of the problems can be explained with a simple example: Should we consider the fine structure constant as a 'truly' fundamental constant? From the

point of view of atomic physics or electromagnetic macroscopic quantum effects (such as the Josephson effect or the quantum Hall effect), we have to acknowledge its fundamental character. However, a crucial problem is renormalization [7,14]. We tend to believe that the 'truly' fundamental theory of everything (TOE) is determined on the Planck scale in simple terms (cf. [3,14–16]) and that the constants may even be calculable from first principles. However, the fine structure constant α_0 once defined there will be affected by renormalization

$$\alpha_0 \to \alpha = \alpha_0 + \delta\alpha \ .$$

Thus 'true' fundamentality is to be assigned to an unobservable value of the unperturbed fine structure constant α_0 determined at the Planck scale $M_{\rm Pl}$. Another discrepancy between 'truly' fundamental and observable constants appears with the electron and proton masses. From a spectroscopic point of view (and that is related to all available precision data) a proper constant is m_e/m_p. However, the origin of the electron and proton masses is so different (see e.g. [14,15]) that the theoretical fundamental constants are $m_e/M_{\rm Pl}$ and $m_p/M_{\rm Pl}$, which should be studied separately.

Several models of hierarchy and certain scenarios for variability of the fundamental constants can be found in [7,12,14,15,17].

3.4 Variability of Fundamental Constants and Equivalence Principle

In different scenarios there is a connection between variability of the fundamental constants and a certain violation of Einstein's equivalence principle (see e.g. [17,12]). There is also a practical connection between these two matters:

- Precision tests of general relativity via advanced celestial mechanics [18] question both the validity of Einstein's theory and the invariability of the Newtonian gravitational constant G.
- The same dilemma may appear in the interpretation of some pulsar spectra (see the contribution by Kramer [19]).
- Frequency measurements are the most accurate measurements and one may try to use them to look for both a variability of the constants and a violation of Lorentz invariance (see e.g. the contribution by Bize et al. [10]).

In particular Einstein's equivalence principle assumes as one of its parts the local position invariance (LPI). The latter predicts that the results of measurements should not depend on their location in space and time. However, we prefer to avoid setting an equality mark between LPI and 'constancy' of the constants. A problem is that the LPI, as an invariance of special relativity, takes place in a free flat space. If the gravitational effects are weak we may introduce proper corrections and still speak about LPI. We indeed pretend that locally we deal with a flat space. However, we cannot claim this for the universe on the cosmological scale.

Studies of LPI and local Lorentz invariance (another important part of Einstein's equivalence principle) have to remove any possible effects due to the

evolution of the universe, due to their topology (which may be locally nearly flat, but non-trivial globally) and even the very existence of the universe. All such effects are to be treated as an environment on the same ground as the influence of the motion of Sun or Moon, etc. A situation with variability of the constants is different. If there is a variation driven by the existence or evolution of matter at a cosmological scale (i.e. of the universe) we would say that the constants vary.

4 Astrophysical and Geophysical Search for a Variability of Constants

A number of contributions in this book are devoted to the practical search for a possible time variation of the fundamental constants. Probably the strongest limits come from studies of nuclear properties. One option is to consider the 93 meV resonance

$$^{149}\mathrm{Sm} + n \rightarrow {}^{150}\mathrm{Sm} + \gamma \; . \tag{6}$$

Such a low-lying resonance may be extremely sensitive to a minor variation of the fundamental constants. The reaction is strongly enhanced for thermal neutrons. Their flux may be studied in a very remote past (about 2 billion years ago) due to *fossil nuclear reactors* [11]. Details can be found in the contribution by Fujii in this book [12].

Another promising approach is the study of some radioactive elements and in particular of the decay rate of $^{187}_{75}\mathrm{Re}$ [20,21] found in meteorites. However, a weak point of all such studies is the proper understanding of effects of the strong interaction. There is no model-independent way to translate the results in terms of the fundamental constants. Since a number of parameters involved may be subject to variation and since all their variations are to be correlated, the lack of a reliable model-independent interpretation should significantly weaken the result.

A reliable interpretation may be easily achieved in the study of atomic and molecular transition frequencies. Their behavior can be expressed in terms of fundamental constants in the leading non-relativistic approximation. We summarize the result in Table 1. Relativistic corrections are not negligible if heavy nuclei are involved. Their importance for the hyperfine structure was first emphasized in [22], while for other atomic transitions it was discussed in [23].

Application of non-relativistic and relativistic calculations to astrophysical data provides a probe for variability of the fundamental constants in quasar absorption spectra related to the remote past (about 10 billion years ago). The astrophysical results can be found in the contributions by Murphy, Flambaum et al. [26] and by Levshakov [27] in this book and in [28]. One of these investigations has so far provided the only positive evidence for a variation of a fundamental constant: The present situation is characterized by a signal of variability of the fine structure constant on the level of $\Delta\alpha \simeq 0.5(1) \times 10^{-5}$ [26].

Table 1. Functional dependence on the fundamental constants for transition frequencies related to different kinds of energy intervals. The coefficient of proportionality is left out. *Ry* stands for the Rydberg constant, μ is the nuclear magnetic moment. The nuclear mass is approximated by $A\,m_p$. Higher vibrational levels involve a certain correction factor which depends on m_e/m_p which is not included. The references are related to papers where the use of this scaling behavior in a search for variations was pointed out.

Transition		Energy scaling	Refs.
Atomic	Gross structure	Ry	[24]
	Fine structure	α^2Ry	[24]
	Hyperfine structure	$\alpha^2(\mu/\mu_B)$Ry	[24]
Molecular	Electronic structure	Ry	[25]
	Vibrational structure	$(m_e/m_p)^{1/2}$Ry	[25]
	Rotational structure	(m_e/m_p)Ry	[25]
Relativistic corrections		Function of α^2	[22,23]

5 New Frequency Standards and Constraints on Variation of Fundamental Constants

In the astrophysical data [26–28] the nuclear magnetic moment used to be related to the hyperfine structure of atomic hydrogen and thus to the proton magnetic moment. In laboratory experiments, the hyperfine structure of other atoms is available. The most important results from precision experiments were obtained with: cesium, rubidium [10], ytterbium [30], and mercury [31]. In particular, a comparison of the rubidium and cesium hyperfine intervals offers the most stringent limit on the relative variation of two atomic frequencies.

We note that the involvement of several nuclei leads to complications because of their different magnetic moments. One can try to describe all magnetic moments with the help of the Schmidt model; however this model is far from being perfect [32]. The model suggests that in any nucleus with an odd number of nucleons all nucleons but one are coupled (with antiparallel spins) and do not contribute to the nuclear magnetic moment. The latter is a result of spin and orbital contributions of the remaining uncoupled nucleon. For example, for cesium-133 the model predicts

$$\mu(^{133}\mathrm{Cs}) = \frac{7}{18}\left(10 - g_p\right)\,\mu_\mathrm{N} \simeq\ 1.72\,\mu_\mathrm{N}\ , \tag{7}$$

where $g_p = 5.585$ is the g factor of the proton and μ_N is the nuclear magneton.

Table 2. Results of the most accurate optical measurements obtained until 2001 and recent results reproduced until the present days (2003).

Atom	Frequency (2001) [Hz]	Frequency (2003) [Hz]
H	2 466 061 413 187 103(46), [34]	2 466 061 413 187 074(34), [35]
Ca	455 986 240 494 158(26), [39]	455 986 240 494 150(8), [36]
Yb^+	688 358 979 309 312(6), [37]	688 358 979 309 310(6), [38]
Hg^+	1 064 721 609 899 143(10), [39]	1 064 721 609 899 144(11), [40]

However the real value ($\mu(^{133}\mathrm{Cs}) = 2.58\ \mu_\mathrm{N}$) is approximately 50% larger, while for most other atoms of interest for precision spectroscopy it is by 20-25% smaller than the Schmidt model prediction (see [32,7] for details). The model may be quite reasonable if the nuclear shells are completed and it is hard to excite the nucleus. If one of the shells is not filled the excitation energy is smaller and there is a relatively big admixture of excited states with different nuclear magnetic moments. In the case of a cancelation between spin and orbital contributions (like in cesium-133) the effect of the admixtures may be more important because the leading contribution from the ground state is reduced.

As we can see, the Schmidt model is oversimplified for many nuclei and one may introduce some more advanced models that involve additional parameters describing the nuclear magnetic moment. Since we need to express all data in terms of very few parameters we are constrained to using an *ab initio* model and there is no way for any improvement of the Schmidt model in this way.

This makes comparisons between optical transitions more attractive (see e.g. [33]). It is now relatively easy to compare an optical frequency to the one related to the cesium hyperfine interval [9,34–40]. This is because the cesium standard as the basis of the realization of the SI second is the best understood frequency standard and is available in a number of laboratories and also because of the dramatic progress in the creation of the so-called frequency comb [35,41,42].

We started the preparation for the ACFC meeting in 2001 inspired in part by the impressive progress in frequency metrology. The progress in frequency standards can be illustrated with the examples of ytterbium [9,37] and mercury [39,40] frequency measurements with a single trapped ion. The ion can be kept in a trap for months, while the measurement typically takes one hour. That means that a result for the transition frequency with an uncertainty of one part in 10^{14} can be repeatedly obtained from just one ion. In 2001 most optical frequencies that are investigated in candidate frequency standards were measured accurately only once. To detect or limit a possible time variation one needs to perform at least two measurements. Recently some of these frequencies were remeasured and we can derive some new constraints on the variability of the fundamental constants. The laboratory results are summarized in Table 2.

Table 3. Laboratory limits for a variation of the fundamental constants in 2001 (see [7,32,33] for details) and 2003. All results but the limit for g_p and m_e/m_p are model independent. The latter are based on the Schmidt model [7,32].

Fundamental constant	Characteristic limits for variation rate	
	2001	2003
α	1×10^{-13} yr^{-1}	2×10^{-15} yr^{-1}
μ_p/μ_e	2×10^{-13} yr^{-1}	5×10^{-15} yr^{-1}
μ_n/μ_e	4×10^{-13} yr^{-1}	5×10^{-14} yr^{-1}
g_n/g_p	3×10^{-13} yr^{-1}	5×10^{-14} yr^{-1}
m_e/m_p	2×10^{-13} yr^{-1}	4×10^{-15} yr^{-1}
g_p	2×10^{-15} yr^{-1}	4×10^{-16} yr^{-1}

We have performed an evaluation of characteristic limits on the variability of fundamental constants from atomic spectroscopy and the results are summarized in Table 3. To simplify the notation we use approximate one-digit numbers (like 3 instead of 3.2) and use the Schmidt model for rubidium assuming it to be correct. We note that this is just for simplification of notation: this allows us to use μ_p; otherwise we would have to use $\mu_{\rm Rb}$. The same holds for neutron and ytterbium magnetic moments and g factors. If one really likes to consider the limits for μ_p and μ_n the results become slightly model dependent because the Schmidt model for rubidium and ytterbium predicts reasonably good results to within 20–25%.

Thus our evaluation leads to results which are

- Model independent (for α);
- Slightly model dependent (for $\mu_p, \mu_n, g_p/g_n$);
- Strongly model dependent (for $g_p, m_e/m_p$).

To obtain the model-independent results for the α-variation we present all optical frequencies in the form (cf. [7,33])

$$f = \text{const.} \cdot c\,\text{Ry} \cdot F(\alpha)\ . \tag{8}$$

Thus for the time dependence we find

$$\frac{\partial \ln f}{\partial t} = A + B \cdot \frac{\partial \ln F}{\partial \ln \alpha}\ , \tag{9}$$

where

$$A = \frac{\partial \ln \text{Ry}}{\partial t} \tag{10}$$

and

$$B = \frac{\partial \ln \alpha}{\partial t}\ . \tag{11}$$

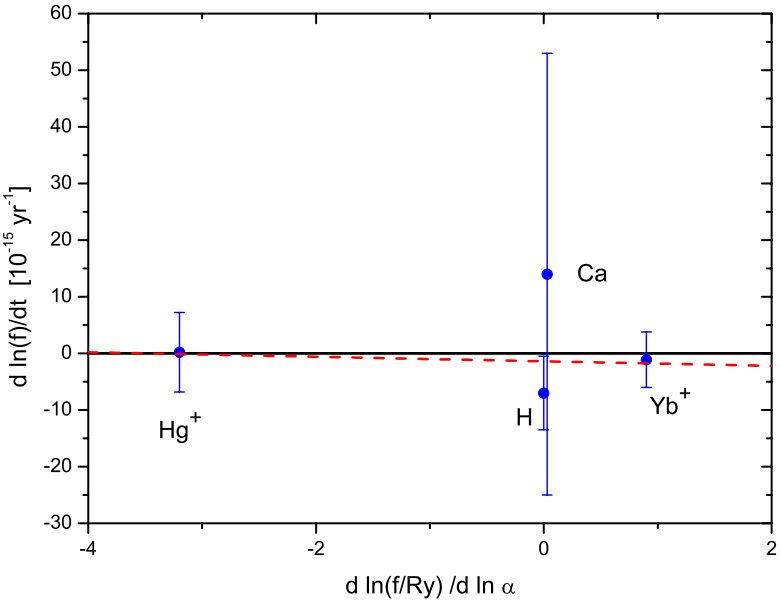


Fig. 1. Data of optical frequency measurements and a search for time variation of the fine structure constant α. The straight line is the result of a mean squares fit (see (9)–(13), the uncertainty of the fit is not shown)

The values of $\partial \ln F/\partial \ln(\alpha)$ were calculated in [23] and the most accurate frequency data are taken from Table 2. The results for the evaluation over these three points is presented in Fig. 1. The result of the mean squares fit are

$$\frac{\partial \ln \alpha}{\partial t} \simeq -0.3(2.0) \times 10^{-15}\ \mathrm{yr}^{-1}\ , \tag{12}$$

$$\frac{\partial \ln \mathrm{Ry}}{\partial t} \simeq -1.5(3.2) \times 10^{-15}\ \mathrm{yr}^{-1}\ . \tag{13}$$

We have to emphasize that the interpretation of the variation of the Rydberg constant in (13) is not a simple matter. One has to distinguish between a constant and its numerical value. The Rydberg constant is a fundamental constant being a combination of fundamental values

$$\mathrm{Ry} = \frac{\alpha^2 m_e c}{2h}\ . \tag{14}$$

However, its numerical value in SI units is a product of three factors

$$\{\mathrm{Ry}\} = \{\nu(\mathrm{Cs})\} \left\{\frac{1}{c}\right\} \left\{\frac{c\mathrm{Ry}}{\nu(\mathrm{Cs})}\right\}\ . \tag{15}$$

The first of them is the numerical value of the hyperfine interval in cesium,

while the second is the numerical value of the speed of light. Both are fixed by definition in the SI system

$$\begin{aligned} \big\{\nu(\mathrm{Cs})\big\} &= 9\ 192\ 631\ 770\ , \\ \big\{c\big\} &= 299\ 792\ 458\ . \end{aligned}$$

The third is a dimensionless value which does not depend on units and it characterises the cesium hyperfine interval measured in natural (atomic) units. Only the third factor may vary and it involves the magnetic moment of the cesium-133 nucleus.

Rigorously speaking, we have done our evaluation not for a variation of frequencies f and the Rydberg constant Ry but for their numerical values $\{f\}$ and $\{\mathrm{Ry}\}$ in SI units and the result in (13) should actually be interpreted as $\partial\{\mathrm{Ry}\}/\partial t$. A non-zero value of this quantity would mean that the optical clocks are inconsistent with the cesium clock and the SI definition of the *second*. The value of A in (9) is not suitable for a model-independent interpretation. This is because the cesium magnetic moment is not well explained by the Schmidt model. For our evaluation we take advantage of an accurate comparison between hyperfine intervals in cesium and rubidium [10,29] and for further analyses we translate the comparisons of optical transitions with cesium to their comparisons with rubidium. This allows us to obtain the *slightly model-dependent* data. Still, the comparison of rubidium to cesium can give us a limit on the variation of the proton g factor and we have to use the Schmidt model for that. The result is indeed strongly dependent on the accuracy of the application of the Schmidt model. The limit for a variability of the electron-to-proton mass ratio comes from limits on proton-to-electron magnetic moment and on g_p. In principle, molecular spectroscopy could deliver such a limit in a model-independent way; however the present accuracy (see references in [43]) is quite low. We can however hope that accurate experiments will be realized in the future [42,43].

Resuming a discussion about results from new frequency standards we have to note that there are two kinds of clocks. One kind is potential primary clocks similar to the cesium clocks. Their frequency is directly linked to the atomic or molecular transition frequency. To work with such a clock is the same as measuring the transition frequency. If one of them will supersede the cesium clock some time in future we may like to change the definition of the second. The other kind of clocks are secondary standards. Their frequency is determined not only by the atomic transition but also by some properties of the apparatus or the environment. A well-known example is the hydrogen maser whose frequency is different from the atomic frequency because of the so-called wall shift (see e.g. [44]). Such clocks are built for the purpose of giving a highly stable reference value; however, a drift of the output frequency is possible (like in the H-maser) and it can have an origin different than a variation of the atomic frequency. We state that comparisons involving clocks of this kind may serve for a search of a violation of LPI or Lorentz invariance [10,45] (because one looks for a signal at a definite period and phase) but not for determining a slow variation of fundamental constants [32].

6 Summary: Results and Open Questions

The purpose of the ACFC conference has been to consider a number of topics in detail and to get an answer to the question: *Where do we stand now?* The answer is not completely clear. A number of basic questions still remain to be solved. Here we briefly summarize them.

- How should the laboratory results for the cesium-to-rubidium comparison [10,29] (negative) with the Oklo constraints [12] (negative) and the astrophysical data [26] (positive) be compared?
- How should the Oklo result and other nuclear data be treated in a model-independent way?
- What is the hierarchy of variations of fundamental constants?
- What is the dependence of the proton mass and magnetic moment on the quark masses?
- How should the nuclear magnetic moments be treated in a model-independent way?
- Do we have time variations or space variations or both?
- Do we have a slow drift or oscillations?
- What is the most important cause of a variation of the fundamental constants?

Some possible answers to these questions are presented in this book. However, the opinions of the editors are not always in unison with the ones of the contributors. We cannot pretend to possess the truth ourselves and thus did not attempt to 'adjust' the contributions to a single point of view. But we still doubt if we have clear anwers to any of the questions above. And without that we cannot answer the main question: "Do we see the variations now?" But we do see strong progress in different fields related to the problem and we hope that we will achieve a clarification in the near future.

Acknowledgments

We are grateful to the WE-Heraeus-Stiftung and in particular to Jutta Lang and Ernst Dreisigacker for their tremendous efforts in organizing the ACFC conference. We are very grateful to the speakers at ACFC for their contributions to this book and acknowledge the many stimulating discussions with participants of the conference and our colleagues at Physikalisch-Technische Bundesanstalt, D. I. Mendeleev Institute for Metrology (VNIIM) and Max-Planck-Institut für Quantenoptik. We would like to thank PTB and DFG SFB-407 for support in the organization of the conference and the preparation of the book.

References

1. G. Börner, Time and the Universe, Lect. Notes Phys. **648**, 21–32 (2004)
2. A. Linde, In *Three Hundred Years of Gravitation* (Ed. by S. W. Hawking and W. Israel, Cambridge University Press, Cambridge, 1987), p. 604;
   S. K. Blau and A. H. Guth, *ibid.*, p. 524.
3. L.B. Okun, Fundamental Units: Physics and Metrology, Lect. Notes Phys. **648**, 57–74 (2004)
4. J. Flowers, B. Petley, Constants, Units and Standards, Lect. Notes Phys. **648**, 75–93 (2004)
5. B. N. Taylor and T. J. Witt, Metrologia **26**, 47 (1989).
6. P. J. Mohr and B. N. Taylor, Rev. Mod. Phys. **72**, 351 (2000). The updated version of 'Recommended values of fundamental constants' will hopefully be published in 2004.
7. S. G. Karshenboim, Eprints physics/0306180 and physics/0311080, to be published.
8. E. Peik and Chr. Tamm, Europhys. Lett. **61**, 181 (2003).
9. C. Tamm, T. Schneider, E. Peik, Trapped Ion Optical Frequency Standards for Laboratory Tests of Alpha-Variability, Lect. Notes Phys. **648** 247–261 (2004)
10. S. Bize et al., Cold Atom Clocks, Precision Oscillators and Fundamental Tests, Lect. Notes Phys. **648**, 189–207 (2004)
11. A. I. Shlyakhter, Nature (London) **264**, 340 (1976);
    T. Damour and F. Dyson, Nucl. Phys. B**480**, 596 (1994).
12. Y. Fujii, Oklo Constraint on the Time-Variability of the Fine-Structure Constant, Lect. Notes Phys. **648**, 167–185 (2004);
    Y. Fujii, A. Iwamoto, T. Fukahori, T. Ohnuki, M. Nakagawa, H. Hidaka, Y. Oura, and P. Moller, Nucl. Phys. B**573**, 377 (2000).
13. W. J. Marciano, Phys. Rev. Lett. **52**, 489 (1984).
14. W.J. Marciano, Time Varying Fundamental Constants, Extra Dimensions and the Renormalization Group, Lect. Notes Phys. **648**, 97–105 (2004)
15. H. Fritzsch, Fundamental Constants and Their Possible Time Dependence, Lect. Notes Phys. **648**, 107–113 (2004);
    X. Calmet and H. Fritzsch, Eur. Phys. J. C**24**, 639 (2002);
    H. Fritzsch, eprint hep-ph/0212186.
16. C. Kiefer, Quantum Gravity and Fundamental Constants, Lect. Notes Phys. **648**, 115–127 (2004)
17. J.-P. Uzan, Rev. Mod. Phys. **75**, 403 (2003).
18. S.G. Turyshev et al., 35 Years of Testing Relativistic Gravity: Where Do We Go from Here?, Lect. Notes Phys. **648**, 311–330 (2004)
19. M. Kramer, Millisecond Pulsars as Tools of Fundamental Physics, Lect. Notes Phys. **648**, 33–54 (2004)
20. M. Lindner et al., Nature **320**, 246 (1986).
21. K. Olive et al., eprint hep-ph/0205269.
22. J. D. Prestage, R. L. Tjoelker, and L. Maleki, Phys. Rev. Lett. **74**, 3511 (1995).
23. V. A. Dzuba, V. V. Flambaum, and J. K. Webb, Phys. Rev. A **59**, 230 (1999);
    V. A. Dzuba and V. V. Flambaum, Phys. Rev. A **61**, 034502 (2000);
    V. A. Dzuba, V. V. Flambaum, and M. V. Marchenko, Phys. Rev. A **68**, 022506 (2003).
24. M. P. Savedoff, Nature **178**, 688 (1956).
25. R. I. Thompson, Astrophys. Lett. **16**, 3 (1975).

26. M.T. Murphy, V.V. Flambaum et al., Constraining Variations in the Fine-Structure Constant, Quark Masses and the Strong Interaction, Lect. Notes Phys. **648**, 131–150 (2004);
J. K. Webb, M. T. Murphy, V. V. Flambaum, S. J. Curran, Astrophys. Space Sci. **283**, 565 (2003).
27. S.A. Levshakov, Astrophysical Constraints on Hypothetical Variability of Fundamental Constants, Lect. Notes Phys. **648**, 151–166 (2004)
28. D. A. Varshalovich, A. V. Ivanchik, A. V. Orlov, A. Y. Potekhin and P. Petitjean. In: S. G. Karshenboim and V. B. Smirnov (Eds). *Precision Physics of Simple Atomic Systems* (Springer-Verlag, Berlin, Heidelberg, 2003), pp. 199–209;
A. Ivanchik, E. Rodriguez, P. Petitjean, and D. Varshalovich, Astron. Lett. **28**, 423 (2002); A. Ivanchik, P. Petitjean, E. Rodriguez, and D. Varshalovich, Astrophys. Space Sci. **283**, 583 (2003).
29. H. Marion, F. Pereira Dos Santos, M. Abgrall, S. Zhang, Y. Sortais, S. Bize, I. Maksimovic, D. Calonico, J. Gruenert, C. Mandache, P. Lemonde, G. Santarelli, Ph. Laurent, A. Clairon, and C. Salomon, Phys. Rev. Lett. **90**, 150801 (2003).
30. R. B. Warrington, P. T. H. Fisk, M. J. Wouters, and M. A. Lawn, In *Proceedings of the 6th Symposium Frequency Standards and Metrology*, ed. by P. Gill (World Sci., 2002), p. 297.
31. L. Maleki, J. Prestage, Search for New Physics with Atomic Clocks, Lect. Notes Phys. **648**, 331–341 (2004)
32. S. G. Karshenboim, Can. J. Phys. **78**, 639 (2000).
33. S. G. Karshenboim. In, *Laser Physics at the Limits*, ed. by H. Figger, D. Meschede and C. Zimmermann (Springer-Verlag, Berlin, Heidelberg, 2001), p. 165.
34. M. Niering, R. Holzwarth, J. Reichert, P. Pokasov, Th. Udem, M. Weitz, T. W. Hänsch, P. Lemonde, G. Santarelli, M. Abgrall, P. Laurent, C. Salomon, and A. Clairon, Phys. Rev. Lett. **84**, 5496 (2000).
35. M. Fischer et al., Precision Spectroscopy of Atomic Hydrogen and Variations of Fundamental Constants, Lect. Notes Phys. **648**, 209–227 (2004)
36. F. Riehle et al., An Optical Frequency Standard with Cold and Ultra-cold Calcium Atoms, Lect. Notes Phys. **648**, 229–244 (2004); J. Helmcke, G. Wilpers, Th. Binnewies, C. Degenhardt, U. Sterr, H. Schnatz and F. Riehle, IEEE Transactions IM-**52**, 250 (2003). See also G. Wilpers, T. Binnewies, C. Degenhardt, U. Sterr, J. Helmcke, and F. Riehle, Phys. Rev. Lett. **89**, 230801 (2002).
37. J. Stenger, C. Tamm, N. Haverkamp, S. Weyers, and H. R. Telle, Opt. Lett. **26**, 1589 (2001).
38. E. Peik, B. Lipphardt, H. Schnatz, T. Schneider, Chr. Tamm, eprint physics/0402132. See also [9] for details.
39. T. Udem, S. A. Diddams, K. R. Vogel, C. W. Oates, E. A. Curtis, W. D. Lee, W. M. Itano, R. E. Drullinger, J. C. Bergquist, and L. Hollberg, Phys. Rev. Lett. **86**, 4996 (2001).
40. S. Bize, S. A. Diddams, U. Tanaka, C. E. Tanner, W. H. Oskay, R. E. Drullinger, T. E. Parker, T. P. Heavner, S. R. Jefferts, L. Hollberg, W. M. Itano, D. J. Wineland, and J. C. Bergquist, Phys. Rev. Lett. **90**, 150802 (2003).
41. T. Udem, J. Reichert, R. Holzwarth, S. Diddams, D. Jones, J. Ye, S. Cundiff, T. Hänsch, and J. Hall, In *Hydrogen atom, Precision physics of simple atomic systems*, ed by S. G. Karshenboim et al., (Springer, Berlin, Heidelberg, 2001) p. 125;
J. Reichert, M. Niering, R. Holzwarth, M. Weitz, Th. Udem, and T. W. Hänsch, Phys. Rev. Lett. **84**, 3232 (2000);
R. Holzwarth, Th. Udem, T. W. Hänsch, J. C. Knight, W. J. Wadsworth, and P.

St. J. Russell, Phys. Rev. Lett. **85**, 2264 (2000);
S. A. Diddams, D. J. Jones, J. Ye, S. T. Cundiff, J. L. Hall, J. K. Ranka, R. S. Windeler, R. Holzwarth, Th. Udem, and T. W. Hänsch, Phys. Rev. Lett. **84**, 5102 (2000).
42. J. Ye et al., Applications of Femtosecond Laser Comb to Nonlinear Molecular Spectroscopy, Lect. Notes Phys. **648**, 275–295 (2004)
43. U. Fröhlich et al., Ultracold Trapped Molecules: Novel Systems for Tests of the Time-Independence of the Electron-to-Proton Mass Ratio, Lect. Notes Phys. **648**, 297–307 (2004)
44. P. W. Zitzewitz and N. F. Ramsey, Phys. Rev. A **3**, 51 (1971);
J. Vanier and C. Audoin, *Quantum Physics of Atomic Frequency Stardards*. Vol. 2. (A. Hilger, Bristol, Philadelpia, 1989).
45. A. Bauch, S. Weyers, Phys. Rev. D **65**, 081101(R) (2002).

Time and the Universe

Gerhard Börner

Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany

Abstract. The cosmological parameters are well defined now, after the satellite WMAP has measured anisotropies in the cosmic microwave background with high accuracy. Adding to this the information from the Hubble diagram of type Ia supernovae, and from the analysis of the large scale galaxy distribution, we find that we live in a Universe that undergoes accelerated expansion. The energy densities reach the critical value, and relative to that baryons are 0.04, dark matter 0.26, and a mysterious dark energy balances everything with its contribution of 0.7. The measurements agree well with the predictions of the inflationary model. The inflation models can also nicely accommodate speculations on varying constants of nature. A toy model involving dark energy for a varying fine structure constant α is briefly mentioned.

1 Introduction

The last century saw dramatic changes of our world view. On one side Quantum Mechanics led to a new understanding of the fundamental constituents of matter and of the forces between them. On the other side the American astronomer Edwin P. Hubble demonstrated that all distant galaxies move away from us. His discovery suggested a changing and evolving universe. The idea of a static uniform distribution of stars which constitutes the world – a picture that initially seemd attractive even to Einstein – had to be given up.

Another big step forward in the exploration of the cosmos was the discovery of a radiation field in the microwave range with a temperature of 2.7 Kelvin. Arno Penzias and Robert Wilson detected the first signal of this radiation, when they were working on antennae for the Bell Telephone Company in 1964. The cosmic origin of this radiation was soon established, and it has been called the Cosmic Microwave Background (CMB). Penzias and Wilson later received the Nobel Prize in physics for their discovery.

These two observational milestones lead almost straightforwardly to an interesting interpretation of the history of the universe: if the galaxies move away from each other today, then earlier they must have been closer together. The radiation too must have been compressed and hotter in the past. The inevitable conclusions seems to be that the evolution of the universe began in a hot early phase, whose heat did not tolerate the existence of galaxies and stars, but everything was dissolved in a dense and hot mixture of radiation and matter.

This "hot big bang model" has become the standard model of cosmology. It finds its mathematical expression in simple solutions of Einstein's theory of

G. Börner, Time and the Universe, Lect. Notes Phys. **648**, 21–32 (2004)
http://www.springerlink.com/

gravitation, the theory of General Relativity. These solutions can accommodate the many, partly very precise observations surprisingly well [1].

2 The Cosmological Models

Alexander Friedman and George Lemâitre were the first to derive solutions from the field equations of General Relativity which could serve as cosmological models. In these Friedmann-Lemâitre (FL) models the expansion is idealized as the flow of a uniform fluid which is characterized by its homogeneous density $\rho(t)$ and pressure $p(t)$. The fluid particles which may be pictured as representatives of the galaxies have fixed positions, but their distance changes, because the space-time between them expands. The distances get bigger in proportion to a function of time, the so-called expansion factor $R(t)$.

The dynamics of the system is determined by ordinary differential equations for $R(t)$:

$$\frac{\dot{R}^2}{R^2} = \frac{8\pi G}{3}\rho - \frac{K}{R^2} + \frac{\Lambda}{3}\ ; \tag{1}$$

$$\frac{d}{dt}(\rho R^3) + p\frac{d}{dt}(R^3) = 0\ ; \tag{2}$$

$$\ddot{R} = -\frac{4\pi}{3}\left(\rho + 3p\right)GR + \frac{1}{3}\Lambda R; \tag{3}$$

K/R^2 is the Gaussian curvature of space-time. By an appropriate choice of standard units, K can be set equal to ± 1, or 0. Λ is the cosmological constant introduced by Einstein with the aim to find a static solution, where the repulsive (for $\Lambda > 0$) cosmological constant balances the attractive gravitational force. Such a solution is, however, unstable. Einstein later called the introduction of a Λ-term "meine größte Eselei", because he had missed to predict the expansion of the universe from his equations. Meanwhile, as we shall see, the cosmological constant has reappeared as a physical quantity in its own right.

Whenever $\dot{R} \neq 0$, any two of the equations (1), (2), and (3) imply the third.

A photon emitted at an earlier time t arrives in the detector of an astronomer at the present time t_0 with a redshift z, where

$$1 + z = \frac{R(t_0)}{R(t)}\ . \tag{4}$$

The wavelength of the photon is stretched out proportional to the expansion factor $R(t)$. We say, for example, that a redshift $z = 1$ means that these photons were emitted at a time, when the universe was just half its present size.

Thus, within the context of the cosmological model, redshift and time are intimately connected. For small redshifts, and times close to t_0, the linear approximation to the redshift definition, is just Hubble's law:

$$cz = H_0 \cdot d\ , \tag{5}$$

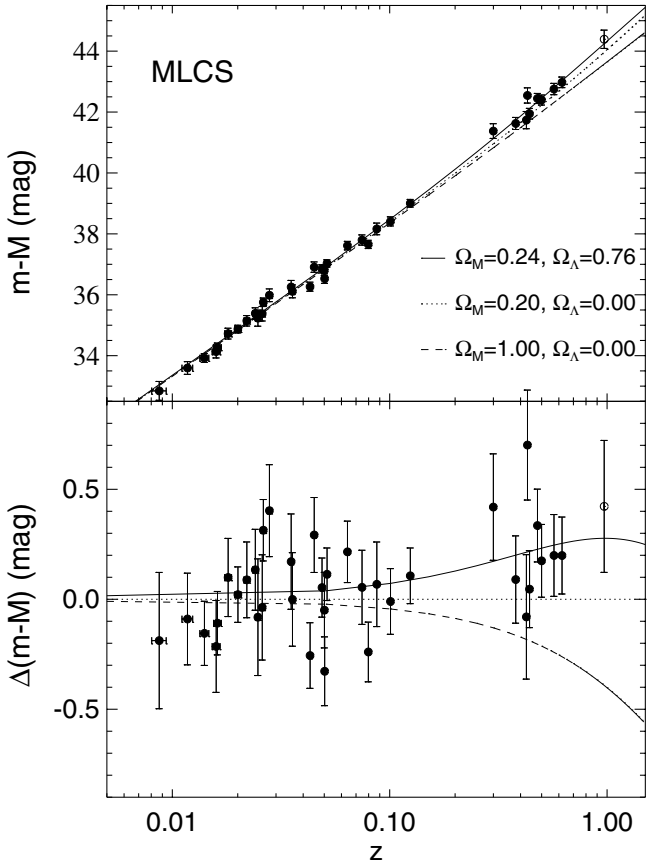


Fig. 1. The plot of Type Ia Supernovae redshifts vs. the distance modulus $(m - M)$, a logarithmic distance scale. The upper panel shows the perfect fit to a linear Hubble law, the lower panel displays the data in different FL models indicating a preference for a model with $\Omega_\Lambda > 0$.

where d is the distance, c the velocity of light, and

$$H_0 = \frac{1}{R}\frac{dR}{dt}(t = t_0)$$

is the expansion rate at the present time. H_0 is known as the "Hubble constant". Measurements of the Hubble constant have reached an acceptable level of precision only quite recently – during the last five years – when it was realized that a certain class of stellar explosions, so-called Supernovae of type Ia, were good standard candles, and could be used to measure distances. These type Ia supernovae are probably exploding white dwarfs. The maximum luminosity and the decay of the supernova lightcurve are closely correlated. In Fig. 1 the Hubble diagram obtained from such supernovae is shown. The measurements

determine H_0 as

$$H_0 = \left(70 \pm 7\right) \,\mathrm{km\,s^{-1}\,Mpc^{-1}} \,.$$

(This astronomical system of units means that every 3.26 million lightyears the expansion velocity increases by 70 km s^{-1}.)

The reciprocal, H_0^{-1}, is a characteristic time of the expansion equal to 14 $\cdot 10^9$ years with an uncertainty of about 10% [1,2].

The curves in Fig. 1 show a dependence on the cosmic density, expressed as usual as the ratio Ω of the present mean density to a critical density ρ_c. For matter, e.g.

$$\Omega_M \equiv \frac{\rho_M}{\rho_c} \,.$$

The critical density ρ_c is

$$\rho_c = \frac{3}{8\pi}\,\frac{H_0^2}{G} \,.$$

Not only the masses, but according to relativity theory all other forms of energy as well, contribute to the total density. All the different components can be added up to a total density Ω_{tot}, where each component is written as a fraction Ω of the critical density. Thus we know that besides the normal baryonic matter (Ω_B) which makes up the stars, and ourselves, there must be a significant nonbaryonic component. A so-called "Dark Matter" ($\Omega_{\mathrm{CDM}} - CDM$ for "Cold Dark Matter", where "cold" indicates that the dark matter particles move with nonrelativistic velocities) is found by the effects of its gravity, whereas an "energy-like" component which is also around, is quite mysterious. It seems to correspond to a cosmological constant at present, and therefore we name it Ω_Λ.

In Fig. 1 you can see that the supernovae measurements at large redshift seem to favour a cosmological model with a positive value of Ω_Λ. From equation (3) you can see that $\ddot{R}$ is positive, if Ω_Λ dominates, i.e. the universe undergoes an accelerated expansion.

In Fig. 2 various types of solutions for $R(t)$ have been plotted schematically, and the curve suggested by the supernovae data has been labelled "$\Omega_\Lambda > 0$".

At the present epoch t_0, we find the following connection between the different cosmological quantities:

$$t_0 H_0 = \int_0^1 \frac{dy}{\left(1 - \Omega_M - \Omega_\Lambda + \Omega_M y + \Omega_\Lambda y^2\right)^{1/2}} \,. \tag{6}$$

This relation provides a stringent test for the simple FL models, since all 4 quantities t_0, H_0, Ω_M, and Ω_Λ can be measured.

A reasonable estimate for t_0 is the age of the oldest stars in globular clusters which turns out to be

$$t_0 = (12 \pm 1.2) \cdot 10^9 \text{ years} \,.$$

It is gratifying that this agrees with the "expansion age" H_0^{-1}. We shall see later that Ω_M and Ω_Λ are estimated at values which satisfy nicely relation (6).

At a fixed time t the cosmological models (FL models) describe a three-dimensional space with constant curvature K/R^2, with 3 different types of space:

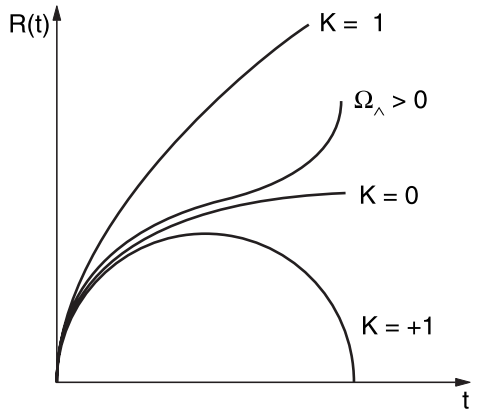


Fig. 2. $R(t)$ in different Friedmann-Lemâitre (FL) models.

$K = \pm 1$, or $K = 0$. At the present epoch

$$\frac{K}{R_0^2} = H_0^2 \, (\Omega_\Lambda + \Omega_M - 1) \; . \tag{7}$$

If $\Omega_M + \Omega_\Lambda = 1$, i.e. if $\Omega_{\rm tot}$ reaches the critical density, the curvature of space is zero. The geometry of space is Euclidean in this case.

3 The History of the Universe

The cosmic expansion changes the energy density of radiation and matter in proportion to R^{-4} and R^{-3} respectively. The temperature of radiation T is proportional to R^{-1}. In terms of the redshift this can be written as

$$\begin{aligned} T &= T_0 \, (1+z) \, ; \\ \rho_\gamma &= \rho_{\gamma 0} \, (1+z)^4 \quad \text{for radiation} \, ; \\ \rho_M &= \rho_{M0} \, (1+z)^3 \quad \text{for matter} \, . \end{aligned}$$

At the present epoch the energy density of radiation is much smaller than that of matter, but for z greater than 10^4 radiation becomes the dominant energy density.

Starting from the present state the history of the universe can be reconstructed within the context of the cosmological model chosen. The cosmological quantities change only with respect to time due to the assumed overall homogeneity and isotropy. This is an approximation which must be modified, when e.g. the formation of structure is discussed. The idea here is that the models are slightly perturbed by small fluctuations in the matter and energy density. The matter fluctuations then grow by the action of gravity, and collapse eventually to form dense clumps which evolve into stars and galaxies. The seeds of these structures are also imprinted on the CMB, since in the early phase the photons of the CMB were tightly coupled to the matter by Compton scattering on free electrons.

4 The Cosmic Microwave Background

In Fig. 3 the spectrum of the CMB is displayed. The plot shown here summarizes 9 minutes of observation by the NASA satellite COBE (Cosmic Background Explorer), lanched in 1989. The fit to a black-body spectrum with temperature T = 2.728 K is excellent [1]. Besides the spectrum, COBE also produced evidence for temperature fluctuations across the sky at a level of 30 μK. This was the first glimpse of the seed fluctuations for galaxy formation. The angular resolution of the instruments aboard COBE was only 7°, and therefore the structures seen by COBE in the CMB are just an averaged picture. Recently the satellite WMAP (Wilkinson Microwave Anisotropy Probe) has observed the CMB with a much better angular resolution of 15′ from its position at the outer Lagrange point of the system Earth–Sun, about 1.5 million kilometers away from the Earth. The sky maps constructed from its first year of taking data show the expected fine-grained pattern of hot and cold spots (see Fig. 4) [3]. Expanding the auto-covariance function of the temperature fluctuations $\Delta T(\mathbf{n})$ in spherical harmonics

$$c(\theta) \equiv \langle \Delta T(\mathbf{n}) \Delta T(\mathbf{n}') \rangle \ , \tag{8}$$

with

$$\cos\theta = (\mathbf{n} \cdot \mathbf{n}') \ ,$$

leads to the power spectrum of Fig. 5 [3].

What is the physics behind this sequence of maxima? Well, it is a bit complicated, but very interesting. First of all we must emphasize the fact that the amplitudes of the fluctuations in the CMB are very small: $\Delta T/T \simeq 10^{-5}$. The amplitudes of the matter fluctuations are then less than $\Delta\rho/\rho = 3 \cdot 10^{-5}$. Until

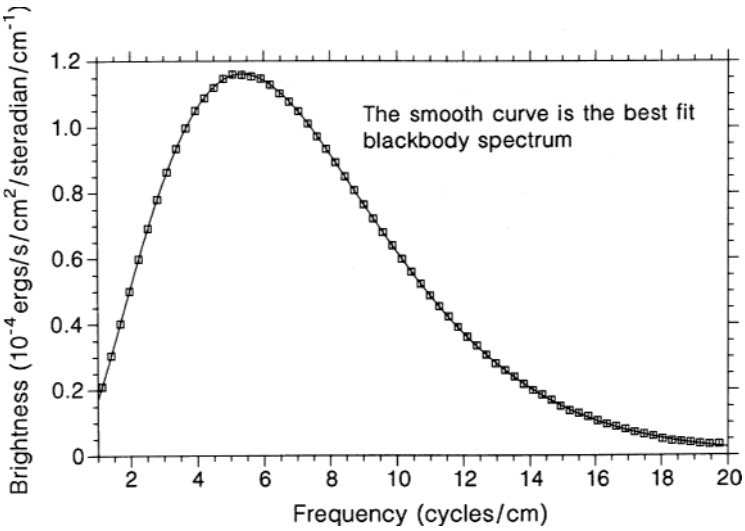


Fig. 3. The spectrum of the CMB. The best fit Planck curve at $T = 2.726$ K is the solid line.

Fig. 4. A sky map of the CMB as obtained by WMAP. Courtesy of the WMAP team [3]; – http://lambda.gsfc.nasa.gov.

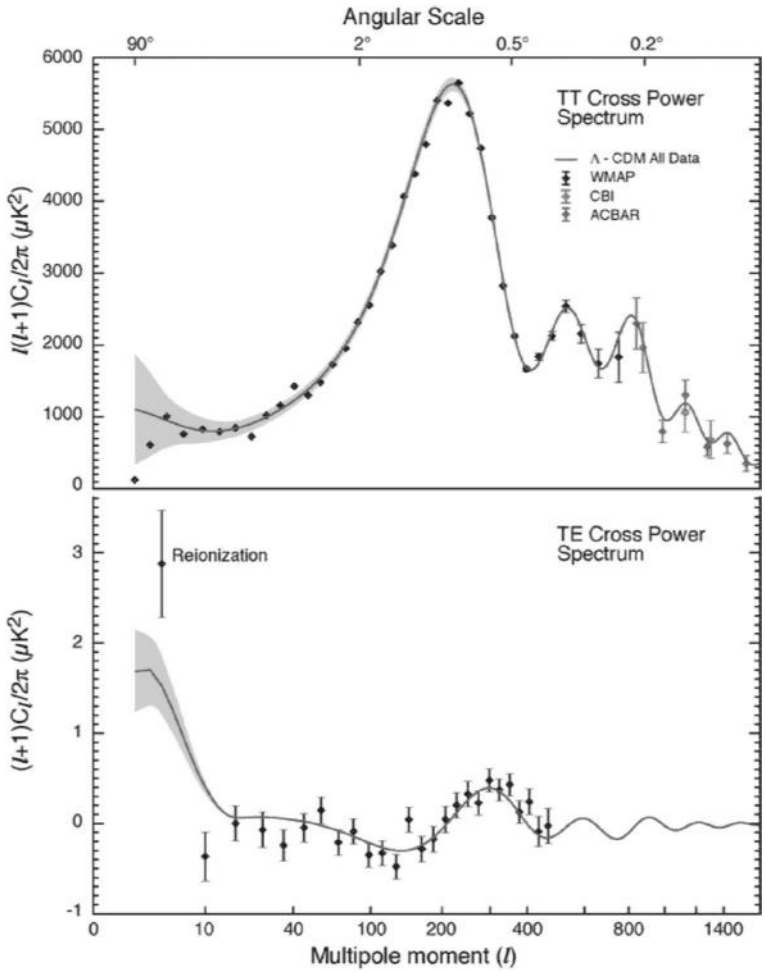


Fig. 5. The power spectrum of WMAP. Upper panel: Multipole Component of $(\Delta T/T)^2$ vs. multiple index. Lower panel: A cross correlation of temperature and polarization variations. Courtesy of the WMAP team; – http://lambda.gsfc.nasa.gov.

the radiation has cooled down to 3000 K the cosmic material acts like a photon-baryon fluid, because of the tight coupling between radiation and matter. Below 3000 K there are no longer enough photons with energies larger than 13.6 eV, and the hydrogen atoms cannot be kept ionised. At this epoch, 400 000 years after the big bang, hydrogen atoms form, photons and the matter particles decouple, and the radiation propagates freely. When we observe the CMB, we see the radiation field at this epoch of "recombination", only redshifted by a factor $(1+z_R) \simeq 1100$. Then $10^{-5} \cdot (1 + z_R) \simeq 10^{-2}$ is the maximum growth density fluctuations can have up to the present epoch, and this means we would still find an almost uniform density field – no galaxies, and no stars! The way out is, of course, non-

baryonic dark matter which interacts with photons only by gravity, and thus dark matter fluctuations can gain the additional growth factor needed of about 100. Dark matter fluctuations provide the gravitational potential wells, where the baryons flow in to form galaxies, stars, planets, and eventually astronomers and physicists.

Now, just before recombination there already existed traces of mass concentrations of baryon-photon clouds within the gravitational potentials of the dark matter.

The baryons pulled those clouds together by the gravitational force acting between them, the photons exerted a pressure which worked on driving the clouds apart. The interplay of forces led to oscillations of these clouds – in complete analogy to sound waves. The clouds all oscillate in phase, synchronized by the big bang. During contraction the photon gas got heated up, during rarefaction it cooled. Quite generally the temperature differences can be calculated in linear approximation as

$$\frac{\Delta T}{T} = \frac{1}{3}\psi \cos\left(k r_s\right) - \left(1 + R_s\right)\psi \ . \tag{9}$$

Here ψ is the relativistic analogon of the Newtonian gravitational potential, k is the wave number, $r_s = c_s \eta$ is the sound horizon, η is conformal time $R\, d\eta = dt$; the sound speed c_s is given by

$$c_s^2 = \frac{1}{3}\,\frac{c^2}{1 + R_s} \ ,$$

and R_s is determined by the ratio of baryonic ρ_B to radiation energy density

$$R_s = \frac{3}{4}\,\frac{\rho_B}{\rho_\gamma} \ .$$

At the time of recombination η_R the photons left the plasma clouds with temperatures corresponding to the phase of the oscillation. Then we interpret $\cos(k r_s)$ not as a function of η, but we fix $\eta = \eta_R$, hence r_s, and consider $\Delta T/T$ as functions of k. Extremal values of the temperature fluctuations should occur at

$$k = \frac{\pi m}{r_s} \text{ for } m = 1, 2, 3, ... $$

When we analyze the $(\Delta T)^2$ variations of the CMB averaged over different angles (or better expanded in terms of spherical harmonics to find the multiple structure) we expect to see the traces of the acoustic oscillations of the plasma clouds just before recombination. That is indeed what we see, and what Fig. 5 shows.

The position in angle of the peaks and their relative amplitudes (as displayed in Fig. 5) depend on all the different cosmic parameters. An analysis of the power spectrum of Fig. 5 determines the cosmic parameters with unprecedented accuracy. For instance, the position of the first peak determines the total energy density. $l = 200$ means $\Omega_{\text{tot}} = 1$. Precisely $\Omega_{\text{tot}} = 1.02 \pm 02$ as given by the WMAP data.

The density of matter (mainly cold dark matter (CDM)) is also determined. With the preferred value of the Hubble constant of $H_0 = (70 \pm 7)\,\mathrm{Mpc}^{-1}$, one finds

$$\Omega_{\mathrm{CDM}} = 0.29 \pm 0.01\,,$$

and for the baryon density

$$\Omega_B = 0.048 \pm 0.008\,.$$

The value for Ω_B agrees nicely with the values derived from big bang nucleosynthesis, i.e. from the theory of the production of the light elements He, Li, and D during the first 3 minutes of the big bang universe. The comparison of the theoretical predictions with observations leads to limits in the present-day baryonic density. These estimates coincide with the CMB results.

But $\Omega_{\mathrm{CDM}} + \Omega_B \simeq 0.3$ only (contributions by known non-barionic matter, such as neutrionos or photons, are certainly much smaller than Ω_B, and thus do not matter in the overall balance), they do not reach the value $\Omega_{\mathrm{tot}} = 1$. To balance the energy density of the cosmos one has to postulate another component, Ω_Λ, with

$$\Omega_\Lambda = 0.67 \pm 0.04\,.$$

This additional component corresponds to a cosmological constant which at the present epoch leads to an accelerated expansion of the universe. This has been detected from the supernova observations. But at the moment this "dark energy" as it has been called is totally mysterious.

5 The Inflationary Model

One aspect of the very early universe deserves a more detailed discussion. It is the speculation known under the name of "inflation" [1].

Imagine a state very close to the big bang, say $10^{-35}s$ later, where the universe is filled with scalar fields φ, whose energy can be described by an effective potential $V(\varphi)$. The energy density of such a scalar field is

$$\rho_\varphi = \dot{\varphi}^2 + V(\varphi) \tag{10}$$

and the pressure

$$p_\varphi = \dot{\varphi}^2 - V(\varphi)\,.$$

Whenever $\dot{\varphi}^2$ is small compared to $V(\varphi)$, then $\rho_\varphi \sim p_\varphi$. If in addition the field energy dominates, it changes the expansion of the universe to an exponential acceleration

$$R(t) \propto \exp\left(H_\varphi t\right)\,,$$

where

$$H_\varphi^2 = \frac{8\pi G}{3} V(\varphi)\,. \tag{11}$$

If the potential is chosen appropriately this exponential inflation can lead to an immense increase of the expansion factor $R(t)$. If inflation lasts for about 10^{-33} s,

then $R(t)$ increases by huge factors, while in the normal radiation filled universe $R(t)$ would increase only by a factor 10. An inflationary epoch smoothes out any inhomogeneities, the curvature: K/R^2 goes to zero, and therefore at the end of inflation the total energy density $\Omega_{\rm tot}$ should be equal to one. This prediction is nicely confirmed by the CMB measurements.

Of course, to arrive at this value the universe must undergo a transition at the end of inflation, in which the scalar field energy is transformed into radiation. The models for inflation usually do not give a convincing explanation for that transition. Within field theory it is the problem of the coupling of the "inflaton" – as the scalar field responsible for inflation is usually called – to relativistic particles. This coupling must be small enough to allow a sufficiently long epoch of inflation, but large enough to bring about the transformation of the potential energy of the inflation, and the exit from the exponential expansion.

Another point in favour of the inflationary model is the prediction of a spectrum of fluctuations $P(k) \simeq k^n$, with $n \equiv 1$. Again, this prediction has been confirmed by the CMB observations.

Another interesting aspect of inflation models is the fact that only a small part of a much larger space-time structure must undergo inflation to produce our observable universe. One can imagine that other parts evolve in quite different ways with different laws and constants of nature [4]. The values of the constants of nature that we find are accidental – we just happen to be in an inflationary bubble with parameters suitable for our existence.

This philosophical attitude leaves, of course, room for a lot of speculation on how the constants of nature acquire their present values. A change with cosmic time seems almost natural in such a context.

6 Variation of the Fine Structure Constant

Here I just want to draw your attention to an amusing idea of John Barrow ([5]), who considers a cosmological model with a time-varying fine-structure constant. In that model alpha is varying, because the electric charge e is connected with cosmic time by a scalar field. This field enters the cosmological equations in such a way that a time variation $\alpha \propto \log t$ occurs for a radiation or matter dominated epoch, but when the cosmological constant dominates the time variation ceases.

7 Conclusions

We have seen that modern cosmology enters a phase of precision measurements. The determination of the cosmic parameters, such as the Hubble constant and the mean density, which were the big questions in cosmology 25 years ago, has been successfully done. Fortunately there are new and interesting questions remaining. What is the nature of dark matter, of dark energy? How did galaxies form? You need little imagination to see that these will be the hot topics of cosmological research in the near future. In addition there is the very early universe,

the realm of fascinating speculations, and perhaps eventually some testable result. We have seen that already the presently discussed theories – like inflation – leave room to accommodate spatial or temporal variations of the laws and constants of nature, should such features be required by experiments. We can look forward to an interesting decade of cosmological research.

References

1. G. Börner. *The Early Universe – Facts and Fiction.* (Springer-Verlag, Berlin, Heidelberg, 2003) (this text-book contains all relevant cosmological results up to 2002; only a few additional references are added).
2. A.G. Riess et al., Ap. J. **504**, 1998, 935;
   S. Perlmutter et al., Nature **391**, 1998, 51.
3. C.L. Bennett et al., Ap. J. Suppl. **148**, 2003, 1;
   a web-site for WMAP is http://lambda.gsfc.nasa.gov.
4. A.D. Linde. *Particle Physics and Inflationary Cosmology* (Harvard Acad. 1990).
5. J. Barrow, Ap. Space Sci. **283**, 2003, 645.

Millisecond Pulsars as Tools of Fundamental Physics

Michael Kramer

University of Manchester, Jodrell Bank Observatory, Cheshire SK11 9DL, UK

Abstract. A new era in fundamental physics began when pulsars were discovered in 1967. Soon it became clear that pulsars were useful tools for a wide variety of physical and astrophysical problems. Further applications became possible with the discovery of the first binary pulsar in 1974 and the discovery of millisecond pulsars in 1982. Ever since pulsars have been used as precise cosmic clocks, taking us beyond the weak-field limit of the solar-system in the study of theories of gravity. Their contribution is crucial as no test can be considered to be complete without probing the strong-field realm of gravitational physics by finding and timing pulsars. This is particularly highlighted by the discovery of the first double pulsar system in 2003. In this review, I will explain some of the most important applications of millisecond pulsar clocks in the study of gravity and fundamental constants.

1 Introduction

The title of this volume, "Astrophysics, Clocks, and Fundamental Constants", would also be a suitable title for this contribution describing the use of radio pulsars in the study of fundamental physics. Indeed, pulsar astronomy is an extraordinary discipline which removes the distinction between physics and astrophysics that is often made. Such a distinction may be justified by the fact that in a terrestrial laboratory we can modify the experimental set-up and control the environment. In contrast, in astrophysical experiments we remain an observer, deriving all our information simply from observing photons and their properties. Thereby, terrestrial experiments are typically more precise and, most importantly, can be reproduced – at least in principle – in any other laboratory on Earth. However, when probing the limits of our understanding of fundamental physics, we often have to study conditions that are too extreme to be encountered on Earth. One may take the experiment into space, like "LISA", "STEP" or "Gravity Probe-B", but even then we are limited, particularly if we want to study gravity. While solar system tests provide a number of very stringent tests for general relativity, none of the experiments made or proposed for the future will ever be able to test the strong field limit. For such studies, pulsars are and will remain the only way to test and enhance our understanding. Additionally, pulsars not only provide us with the only means to perform strong-field experiments, but these experiments are also amazingly precise. It is this unique aspect that I will review in the following. The interested reader may also consult the excellent reviews by Will [1] and Turyshev [2] on PPN formalism, by Wex [3] and Stairs [4] on strong gravity tests, and by Lorimer [5] on pulsars in general.

M. Kramer, Millisecond Pulsars as Tools of Fundamental Physics, Lect. Notes Phys. **648**, 33–54 (2004)
http://www.springerlink.com/

2 Pulsars

Pulsars are highly magnetized, rotating neutron stars which emit a narrow radio beam along the magnetic dipole axis. As the magnetic axis is inclined to the rotation axis, the pulsar acts like a cosmic light-house emitting a radio pulse that can be detected once per rotation period when the beam is directed towards Earth (Fig. 1). For some very fast rotating pulsars, the so-called millisecond pulsars, the stability of the pulse period is similar to that achieved by the best terrestrial atomic clocks. This is not surprising if we consider that they have large rotational energies of $E = 10^{43-45}$ J and low energy loss rates. Using these astrophysical clocks by accurately measuring the arrival times of their pulses, a wide range of experiments is possible, some of which are presented here. While it is not of utmost importance for the remainder of this review *how* the radio pulses are actually created, we will consider some of the basic pulsar properties below.

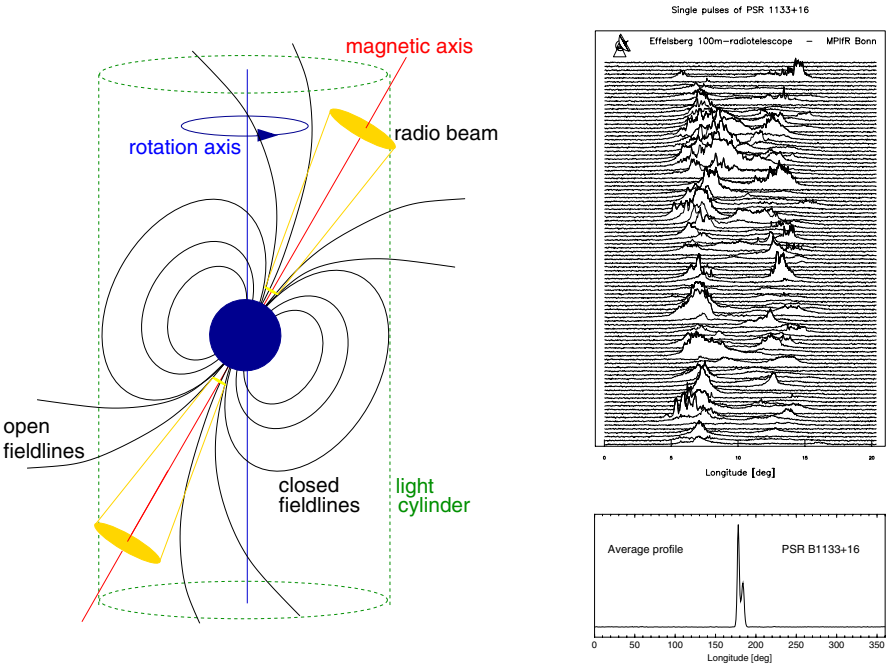


Fig. 1. (left) A pulsar is a rotating, highly magnetised neutron star. A radio beam centred on the magnetic axis is created at some height above the surface. The tilt between the rotation and magnetic axes makes the pulsar in effect a cosmic lighthouse when the beam sweeps around in space. (right) Individual pulses vary in shapes and strength (top) average profiles are stable (bottom). The typical pulse width is only ∼4% of the period.

2.1 Pulsars as Neutron Stars

Pulsars are born in supernova explosions of massive stars. Created in the collapse of the stars' core, neutron stars are the most compact objects next to black holes. From timing measurements of binary pulsars (see Sect. 5.2), we determine the masses of pulsars to be within a narrow range of $(1.35 \pm 0.04)\, M_\odot$[6]. Modern calculations for different equations of state produce results for the size of a neutron star quite similar to the very first calculations by Oppenheimer & Volkov [7], i.e. about 20 km in diameter. Such sizes are consistent with independent estimates derived from modelling light-curves and luminosities of pulsars observed in X-rays (e.g. [8]).

As rotating magnets, pulsars emit magnetic dipole radiation as the dominant effect for an increase in rotation period, P, described by $\dot{P}$. Equating the corresponding energy output of the dipole to the loss rate in rotational energy, we obtain an estimate for the magnetic field strength at the pulsar surface from

$$B_S = 3.2 \times 10^{19} \sqrt{P\dot{P}}\ \text{Gauss}, \tag{1}$$

with P measured in s and $\dot{P}$ in s s^{-1}. Sometimes twice the value is quoted to reflect the field at the poles. Typical values are of order 10^{12} G, although field strengths up to 10^{14} have been observed [9]. Millisecond pulsars have lower field strengths of the order of 10^8 to 10^{10} Gauss which appear to be a result of their evolutionary history (see Sect. 3). These magnetic fields are consistent with values derived from X-ray spectra of neutron stars where we observe cyclotron lines [10].

2.2 Pulsars as Radio Sources

The radio signal of a pulsar is usually weak, both because the pulsar is distant and the size of the actual emission region is small. Estimates range down to a few metres, resulting in brightness temperatures of up to 10^{37} K [11]. Such values require a coherent emission mechanism which, despite 35 years of intensive research, is still unidentified. However, we seem to have some basic understanding, in which the magnetized rotating neutron star induces an electric quadrupole field which is strong enough to pull out charges from the stellar surface (the electrical force exceeds the gravitational force by a factor of $\sim 10^{12}$!). The magnetic field forces the resulting dense plasma to co-rotate with the pulsar. This *magnetosphere* can only extend up to a distance where the co-rotation velocity reaches the speed of light[1]. This distance defines the so-called light cylinder which separates the magnetic field lines into two distinct groups, i.e. *open and closed field lines*. The plasma on the closed field lines is trapped and co-rotates with the pulsar forever. In contrast, plasma on the open field lines can reach highly relativistic velocities and can leave the magnetosphere, creating the observed radio beam at a distance of a few tens to hundreds of km above the pulsar surface (e.g. [12], see Fig. 1).

[1] Strictly speaking, the Alfvén velocity will determine the co-rotational properties of the magnetosphere.

Most pulsars are not strong enough for us to allow studies of their individual radio pulses. Then, only an integrated pulse shape, the "pulse profile", can be observed. Individual pulses reflect the instantaneous plasma processes in the pulsar magnetosphere, resulting in often seemingly random pulses (see Fig. 1). In contrast, the average pulse profile reflects the global constraints mostly given by a conal beam structure and geometrical factors and is thereby stable. It is this profile stability which allows us to time pulsar to high precision.

3 A Pulsar's Life

The evolution in pulsar period, P, and slow-down, $\dot{P}$, can be used to describe the life of a pulsar. This is usually done in a (logarithmic) P-$\dot{P}$-diagram as shown in Fig. 2 where we can draw lines of constant magnetic field (see (1)) and constant "characteristic age" estimated from

$$\tau = \frac{P}{2\dot{P}} = -\frac{\nu}{2\dot{\nu}}, \tag{2}$$

using either period, P, or the spin frequency, ν, and their derivatives in standard units. This quantity is a valid estimate for the true age under the assumption that the initial spin period is much smaller than the present period and that the

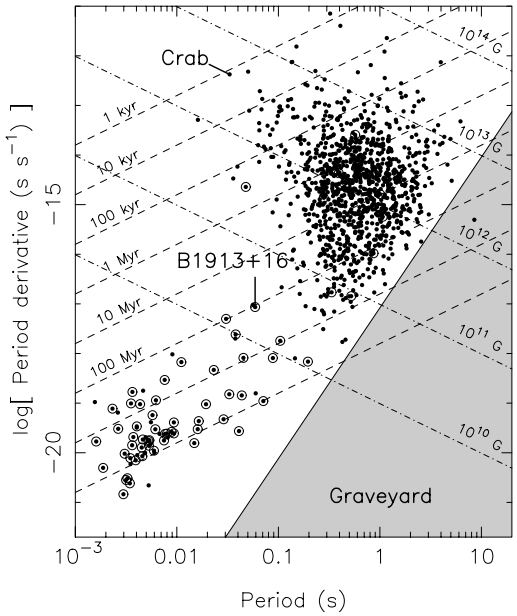


Fig. 2. The $P - \dot{P}$–diagram for the known pulsar population. Lines of constant characteristic age and surface magnetic field are shown. Binary pulsars are marked by a circle. The solid line represents the pulsar "death line" enclosing the "pulsar graveyard" where pulsars are expected to switch off radio emission.

spin-down is fully determined by magnetic dipole braking. While it had been assumed that pulsars are born with birth periods similar to that estimated for the Crab pulsar, $P_0 = 19$ ms [13], recent estimates for a growing number of pulsars suggest a wide range of initial spin periods from 14 ms up to 140 ms [14]. Pulsars are therefore born in the upper left area of Fig. 2 and move into the central part where they spent most of their lifetime.

3.1 Normal Pulsars

Most known pulsars have spin periods between 0.1 and 1.0 sec with period derivatives of typically $\dot{P} = 10^{-15}$ s s^{-1}. Selection effects are only partly responsible for the limited number of pulsars known with very long periods, the longest known period being 8.5 s [15]. The dominant effect is due to the "death" of pulsars when their slow-down has reached a critical state. This state seems to depend on a combination of P and $\dot{P}$ which can be represented in the $P - \dot{P}$-diagram as a "pulsar death-line". To the right and below this line (see Fig. 2) the electric potential above the polar cap may not be sufficient to produce the particle plasma that is responsible for the observed radio emission. While this model can indeed explain the lack of pulsars beyond the death-line, the truth may be more complicated as the position of the 8.5-sec pulsar deep in the "pulsar graveyard" indicates. Nevertheless, it is clear that the normal life of radio pulsars is limited and that they die eventually after tens to a hundred million years.

3.2 Millisecond Pulsars

Inspecting the approx. 1600 sources shown in the $P - \dot{P}$-diagram, it is obvious that the position of a sub-set of about 100 pulsars located in the lower left part of the diagram cannot be explained by the above picture of normal pulsar life. Instead, these pulsars simultaneously have small periods (of the order of milliseconds) and small period derivatives, $\dot{P} \lesssim 10^{-18}$ s s^{-1}. They appear much older than ordinary pulsars (see (2)) and, indeed, these so-called "millisecond pulsars" represent the oldest population of pulsars with ages up to $\sim 10^{10}$ yr. A model for their evolutionary history was proposed soon after the discovery of PSR B1937+21 by Backer et al. in 1982 [16]. This first millisecond pulsar has a period of only 1.56 ms and remains the pulsar with the shortest period known.

It is suggested that millisecond periods are obtained when mass and thereby angular momentum is transferred from an evolving binary companion while it overflows its Roche lobe [17]. In this model, millisecond pulsars are recycled from a dead binary pulsar via an X-ray binary phase. This model implies a number of observational consequences: a) most normal pulsars do not develop into a millisecond pulsar as they have long lost a possible companion during their violent birth event; b) for surviving binary systems, X-ray binary pulsars represent the progenitor systems for millisecond pulsars; c) the final spin period of recycled pulsars depends on the mass of the binary companion. A more massive companion evolves faster, limiting the duration of the accretion process; d) the majority of millisecond pulsars have low-mass white-dwarf companions as the

remnant of the binary star. These systems evolve from low-mass X-ray binary systems (LMXBs); e) high-mass X-ray binary systems (HMXBs) represent the progenitors for double neutron star systems (DNSs). DNSs are rare since these systems need to survive a second supernova explosion. The resulting millisecond pulsar is only mildly recycled with a period of tens of millisecond.

The properties of millisecond pulsars and X-ray binaries are consistent with the described picture. For instance, it is striking that $\sim 80\%$ of all millisecond pulsars are in a binary orbit while this is true for only less than 1% of the non-recycled population. For millisecond pulsars with a low-mass white dwarf companion the orbit is nearly circular. In case of double neutron star systems, the orbit is affected by the unpredictable nature of the kick imparted onto the newly born neutron star in the asymmetric supernova explosion of the companion. If the system survives, the result is typically an eccentric orbit with an orbital period of a few hours. As we will see, both types of system can be used to test different aspects of gravitational theories.

4 Pulsars as Clocks

By measuring the arrival time of the received ticks of the pulsar clock very precisely, we can study effects that determine the propagation of the pulses in four-dimensional space-time. Millisecond pulsars are the most useful objects for these investigations: their pulse arrival times can be measured much more accurately than for normal pulsars (the measurement precision scales essentially with spin frequency) and their rotation is much smoother, making them intrinsically better clocks. Specifically, they do not exhibit rotational instabilities known for normal pulsars, namely "timing noise" and "glitches". Glitches are associated with young pulsars and they represent a sudden increase in rotation frequency that is probably caused by an abrupt change in the internal structure of the neutron star. The origin of timing noise is much less understood. It manifests itself in a quasi-random walk in one or more of the rotational parameters on timescales of months to years. Again, it appears mostly for young pulsars and scales with some power of the period derivative, $\dot{P}$. Hence, millisecond pulsars generally do not show timing noise, although it has been detected for few sources such as PSR B1937+21 [18] albeit on a much smaller amplitude scale than for normal pulsars.

4.1 Time Transfer

In order to study effects that change the pulse travel time, we first have to find an expression that describes the pulsar rotation in a reference frame co-moving with the pulsar. We start by expressing the spin frequency of the pulsar in a Taylor expansion,

$$\nu(t) = \nu_0 + \dot{\nu}_0(t - t_0) + \frac{1}{2}\ddot{\nu}_0(t - t_0)^2 + ..., \tag{3}$$

where ν_0 is the spin frequency at reference time t_0, i.e. $\nu_0 = \nu(t_0) = 1/P_0$ with P_0 being the corresponding pulse period. While ν_0 and its derivatives refer to values measured at a certain epoch, $\dot{\nu}$, $\ddot{\nu}$ are determined by the physical process responsible for the pulsar slow-down and should, in principle, be constant for most time-spans considered. We expect a relation

$$\dot{\nu} = -\text{const.}\ \nu^n, \tag{4}$$

and hence $\ddot{\nu} = -\text{const.} \times n \times \nu^{n-1}\ \dot{\nu}$ where the "braking index", n, has a value of $n = 3$ for magnetic dipole braking, relating to (2). Measuring $\ddot{\nu}$ can yield the braking index via $n = \nu\ddot{\nu}/\dot{\nu}^2$ so that the assumption of dipole braking can be tested. Timing noise can mimic a significant but time-varying value of $\ddot{\nu}$ that reflects timing noise rather than regular spin-down. In these cases, derived braking indices are meaningless in terms of global spin-down. For most millisecond pulsars $\ddot{\nu}$ is too small to be of significance although some sources show a non-zero $\ddot{\nu}$ due to timing noise.

Relating the spin frequency to the pulse number N, we find

$$N = N_0 + \nu_0(t - t_0) + \frac{1}{2}\dot{\nu}_0(t - t_0)^2 + \frac{1}{6}\ddot{\nu}_0(t - t_0)^3 + \cdots \tag{5}$$

where N_0 is the pulse number at the reference epoch t_0. If t_0 coincides with the arrival of a pulse and the pulsar spin-down is accurately known, the pulses should therefore appear at integer values of N when observed in an inertial reference frame.

Our observing frame is not inertial, as we are using telescopes that are located on a rotating Earth orbiting the Sun. Before analysing corresponding TOAs measured with the observatory clock ("topocentric arrival times"), we need to transfer them to the centre of mass of the solar system as the best approximation to an inertial frame available. By using such "barycentric arrival times", we can easily combine transferred topocentric TOAs measured at different observatories at different times. The transformation of a topocentric TOA to a barycentric arrival time, t_{SSB}, is given by

$$\begin{aligned} t_{\mathrm{SSB}} = \quad & t_{\mathrm{topo}} - t_0 + t_{\mathrm{corr}} - D/f^2\ , && \text{(6a)} \\ & + \Delta_{\mathrm{Roemer},\odot} + \Delta_{\mathrm{Shapiro},\odot} + \Delta_{\mathrm{Einstein},\odot}\ , && \text{(6b)} \\ & + \Delta_{\mathrm{Roemer,Bin}} + \Delta_{\mathrm{Shapiro,Bin}} + \Delta_{\mathrm{Einstein,Bin}}. && \text{(6c)} \end{aligned}$$

We have split the transformation into three lines. The first two lines apply to every pulsar whilst the third line is only applicable to binary pulsars. We discuss each term in detail.

Clock and Frequency Corrections. The observatory time is typically maintained by local H-maser clocks that are compared to UTC (NIST) by the Global Positioning System (GPS). Offsets are monitored and retroactively applied as clock corrections, t_{corr}, in the off-line analysis which often uses UTC of BIPM as time standard. Further corrections take into account that the Earth is not

rotating uniformly, so that leap seconds are occasionally inserted into UTC to keep it close mean solar time. All leap second are removed from the used UTC time standard to produce a TOA measured in International Atomic Time (TAI). The TAI is maintained as an average of a large number of selected atomic clocks by the *Bureau International des Poids et Mesures* (BIPM), which also publishes a retroactive uniform atomic time standard known as Terrestrial Time, TT (formerly known as Terrestrial Dynamical Time, TDT). The unit of TT is the SI second and may be regarded as the time that would be kept by an ideal atomic clock on the geoid with TT=TAI+32.184 seconds, where the offset of about 32 s stems from historic reasons. This time scale should be used in the final analysis by correcting the initially measured TOAs to TT(BIPM).

As the pulses are delayed due to dispersion in the interstellar medium, the arrival time depends on the observing frequency, f. The TOA is therefore corrected for a pulse arrival at an infinitely high frequency, thereby removing dispersion from the data (last term in 6a). The corresponding Dispersion Measure (DM) is determined during the discovery of the pulsar and can be measured accurately by observations at multiple frequencies. For some pulsars, the dispersion measure is observed to change with time. In order to avoid time-varying drifts introduced into the TOAs in such cases, the above term needs to be modified to include time-derivatives of DM, i.e. $\dot{\mathrm{DM}}$, $\ddot{\mathrm{DM}}$ and so on. These can be determined if monitoring observations at two or more frequencies are available when they provide an estimate for the change in electron density along the line-of-sight as a result of 'interstellar weather'. For high-precision timing of millisecond pulsars such multi-frequency observations are essential.

Barycentric Corrections. The terms in (6b) describe the corrections necessary to transfer topocentric to barycentric TOAs.

The *Roemer delay*, $\Delta_{\mathrm{Roemer},\odot}$, is the classical light-travel time between the phase centre of the telescope and the solar system barycentre (SSB). Given a unit vector, $\vec{\hat{s}}$, pointing from the SSB to the position of the pulsar and the vector connecting the SSB to the observatory, $\vec{r}$, we find:

$$\Delta_{\mathrm{Roemer},\odot} = -\frac{1}{c}\,\vec{r}\cdot\vec{\hat{s}} = -\frac{1}{c}\,(\vec{r}_{\mathrm{SSB}} + \vec{r}_{\mathrm{EO}})\cdot\vec{\hat{s}}. \tag{7}$$

Here c is the speed of light and we have split $\vec{r}$ into two parts. The vector, $\vec{r}_{\mathrm{SSB}}$, points from the SSB to centre of the Earth (geocentre). Computation of this vector requires accurate knowledge of the locations of all major bodies in the Solar system and uses *solar system ephemerides* such as the 'DE200' or 'DE405' published by *Jet Propulsion Laboratory* (JPL) [19]. The second vector $\vec{r}_{\mathrm{EO}}$, connects the geocentre with the phase centre of the telescope. In order to compute this vector accurately, the non-uniform rotation of the Earth has to be taken into account, so that the correct relative position of the observatory is derived. This is achieved using appropriate UT1 corrections published by International Earth Rotation Service (IERS).

The *Shapiro delay*, $\Delta_{\mathrm{Shapiro},\odot}$, is a relativistic correction that corrects for extra delays due to the curvature of space-time caused by the presence of masses

in the solar system [20]. The delays are largest for a signal passing the Sun's limb (~ 120 μs) while Jupiter can contribute as much as 200 ns. In principle one has to sum over all bodies in the solar system, yielding

$$\Delta_{\mathrm{Shapiro},\odot} = (1+\hat{\gamma}) \sum_i \frac{GM_i}{c^3} \ln \left[\frac{\vec{\hat{s}} \cdot \vec{r}_i^E + r_i^E}{\vec{\hat{s}} \cdot \vec{r}_i^P + r_i^P} \right], \tag{8}$$

where M_i is the mass of body i, $\vec{r}_i^P$ is the pulsar position relative to it, and $\vec{r}_i^E$ is the telescope position relative to that body at the time of closest approach of the photon (see [21]). The parameter $\hat{\gamma}$ is one of the Parameterised-Post-Newtonian (PPN) parameters that will be discussed in Sect. 5.1. It describes how much space-curvature is produced by unit rest mass and takes the value $\hat{\gamma} = 1$ in general relativity [2] In practice, $\hat{\gamma}$ is adopted as unity and only the Sun, and in some cases Jupiter, are accounted for in this calculation.

The last term in (6b), $\Delta_{\mathrm{Einstein},\odot}$, is called *Einstein delay* and it describes the combined effect of time dilation due to the motion of the Earth and gravitational redshift caused by the other bodies in the Solar system. This time varying effect takes into account the variation of an atomic clock on Earth in the changing gravitational potential as the Earth follows its elliptical orbit around the Sun. The delay amounts to an integral of the expression [21]

$$\frac{d\Delta_{\mathrm{E}instein,\odot}}{dt} = \sum_i \frac{GM_i}{c^2 r_i^E} + \frac{v_E^2}{2c^2} - \text{constant}, \tag{9}$$

where the sum is again over all bodies in the solar system but this time excluding Earth. The distance r_i^E is again the distance between Earth and body i, while v_E is the velocity of the Earth relative to the Sun.

Relative Motion. Equation 6a-6c is sufficient to measure the clock rate as produced by the pulsar if no further motion or acceleration between pulsar and SSB occurs. If the pulsar is moving relative to the SSB, only the transverse component of the velocity, v_t, can be observed from timing. A radial motion is not measurable practically (though theoretically possible), leaving resulting Doppler corrections to observed periods, masses etc. undetermined. The situation changes if the pulsar has an optically detectable companion such as a white dwarf for which Doppler shifts can be measured from optical spectra. In contrast, a transverse motion will change the vector $\vec{\hat{s}}$ in (7), adding a linear time-dependent term to our transfer equation, and can therefore be measured as proper motion, μ.

Another effect arising from a transverse motion is the *Shklovskii effect*, also known in classical astronomy as "secular acceleration". With the pulsar motion, the projected distance of the pulsar to the SSB is increasing, leading to a

[2] Usually, this parameter is denoted without the "hat" simply as γ (see [2]), but it is common practice in the study of binary pulsars to use the symbol γ to describe the amount of time dilation and gravitational redshift caused by a pulsar companion.

correction that is quadratic in time [21],

$$\Delta t_S = \frac{v_t^2}{2dc}\, t^2. \tag{10}$$

Since this delay scales with inverse of the distance d to the pulsar, the correction is usually too small to be considered. However, it has the effect that any observed change in a periodicity (i.e. change in pulse or orbital period) is increased over the intrinsic value by

$$\frac{\dot{P}}{P} = \frac{1}{c}\,\frac{v_t^2}{d}. \tag{11}$$

For millisecond pulsars where $\dot{P}$ is small, a significant fraction of the observed change in period can be due to the Shklovskii effect. This effect also needs to be considered, when studying the decay of an orbital period due to gravitational wave emission, where the observed value is increased by the Shklovskii term.

Similarly, any line-of-sight acceleration a of the pulsar due to an external gravitational field changes the observed period derivative by $\dot{P}/P = a/c$. This effect is commonly observed for pulsars in globular clusters where the acceleration through the cluster's gravitational field towards our line of sight can often be large enough to reverse the sign of $\dot{P}$. As a result, the pulsars appear to be spinning up rather than down! Such pulsars place useful constraints on the cluster mass distributions and the intracluster medium [22].

Finally, a related term that needs to be considered for nearby pulsars describes an annual parallax given by [21]

$$\Delta t_\pi = -\frac{1}{2cd}\left(\vec{r}\times\vec{\hat{s}}\right)^2 = \frac{1}{2cd}\left((\vec{r}\cdot\vec{\hat{s}})^2 - |\vec{r}|^2\right). \tag{12}$$

In comparison to the more familiar positional parallax, this *timing parallax* corresponds to measuring the time delay due to the curvature of the emitted wavefronts at different positions of the Earth orbit. This effect imposes a signal with an amplitude of $l^2\cos\beta/(2cd)$ where l is the Earth-Sun distance and β is the ecliptic latitude of the pulsar. For a pulsar at $d = 1$ kpc, this delay amounts to only $\lesssim 1.2\mu$s, and hence it is only measurable for a few millisecond pulsars where it provides a precise distance estimate. Similarly difficult to measure is the *annual-orbital parallax* for binary pulsars which manifests itself as a periodic change in the observed projected semi-major axis of the pulsar's orbit due to viewing the system from slightly different directions during the Earth's orbit. In contrast, a secular change of the semi-major axis due to a proper motion of the system on the sky has been measured for a number of binary millisecond pulsars (see [4]).

4.2 Pulsar Timing

The transfer equation (6a)–(6c) contains a number of parameters which are not known *a priori* (or only with limited precision after the discovery of a pulsar) and

need to be determined precisely in a least-squares fit analysis of the measured TOAs. These parameters can be categorised into three groups:

Astrometric parameters: The astrometric parameters include the position of the pulsar, and its proper motion and parallax. While the position is only known within a telescope beam after the discovery, the precision can be greatly improved by timing the pulsar for about a year (full Earth orbit). Proper motion and parallax only become evident after a longer time-span.

Spin parameters: These include the rotation frequency of the pulsar, ν, and its derivatives (3).

Binary parameters: For pulsars in a binary orbit, the initial observations will typically show a periodic variation in observed pulse period. Five Keplerian parameters then need to be determined: orbital period, P_b; the projected semi-major axis of the orbit, $x \equiv a \sin i$ where i is the (usually unknown) inclination angle of the orbit; the orbital eccentricity, e; the longitude of periastron, ω; and the time of periastron passage, T_0. For a number of binary systems this Newtonian description of the orbit is not sufficient and relativistic corrections need to be applied, e.g. ω is replaced by $\omega + \dot{\omega}t$. The measurement of the *Post-Keplerian* (PK) *parameters* such as $\dot{\omega}$ allows a comparison with values expected in the framework of specific theories of gravity. We discuss these aspects further below.

Given a minimal set of starting parameters, a least-squares fit is needed to match the measured arrival times to pulse numbers according to (5). The aim is to obtain a phase-coherent solution that accounts for every single rotation of the pulsar between two observations. One starts off with a small set of TOAs that were obtained so closely in time, that the accumulated uncertainties in the starting parameters do not exceed one pulse period. Gradually, the data set is expanded, maintaining coherence in phase. When successful, post-fit residuals expressed in pulse phase show a random distribution around zero (see Fig 3). After starting with fits for only period and pulse reference phase over some hours and days, longer time spans slowly require fits for parameters like spin-frequency derivative(s) and position. Incorrect or incomplete timing models cause systematic structures in the post-fit residuals identifying the parameter that needs to be included or adjusted (see Fig. 3). The precision of the parameters improves with length of the data span and the frequency of observation, but also with orbital coverage in the case of binary pulsars. Sufficient data sets then enable measurements with amazing precision, e.g. the period determined for PSR B1937+21 is known to a relative precision of 10^{-15}.

Obviously, the above description shows that the process of pulsar timing is elaborate. Fortunately, sophisticated software packages have been developed that combine time transfer and the least-squares fit of the timing model. The three major packages are TEMPO (ATNF/Princeton University) [23], PSRTIME (Jodrell Bank Observatory) [24], and TIMAPR (Pushchino Observatory/MPIfR) [25]. The most widely used package is TEMPO.

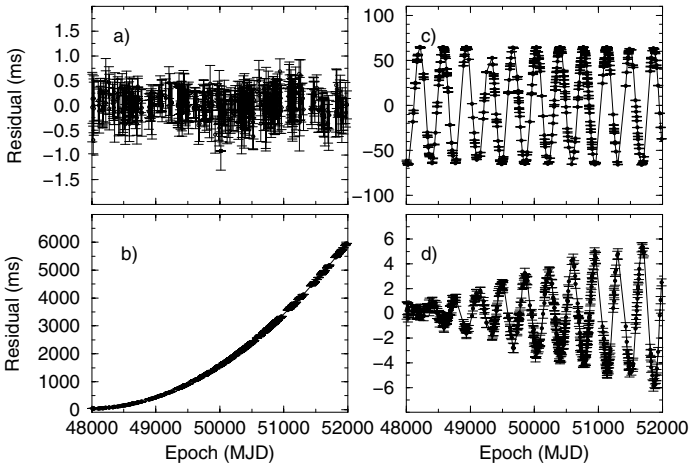


Fig. 3. Timing residuals for the 1.19-s pulsar B1133+16. A fit of a perfect timing model should result in randomly distributed residuals, shown in a). A parabolic increase in the residuals in part b) is obtained if $\dot{P}$ is underestimated, here by 4%. An offset in position produces sinusoid residuals shown in part c) where the declination has an error of 1 arcmin. Part d) demonstrates the effect of neglected proper motion, here of $\mu = 380$ mas/yr. Note the different scales on the y-axes.

5 Applications of Pulsars

Pulsars are unique and versatile objects which can be used to study an extremely wide range of physical and astrophysical problems. Beside testing theories of gravity one can study the Galaxy and the interstellar medium, stars, binary systems and their evolution, solid state physics and the interior of neutron stars. Investigating the radio emission of pulsars provides insight into plasma physics under extreme conditions. In the following we will concentrate on the application of pulsars as clocks, paying in particular attention to tests of theories of gravity. Some of these tests involve studies of *PPN parameters* and possible related time variation in the Gravitational Constant, G.

5.1 PPN Parameters

Metric theories of gravity assume *(i)* the existence of a symmetric metric, *(ii)* that all test bodies follow geodesics of the metric and *(iii)* that in local Lorentz frames the non-gravitational laws of physics are those of special relativity. Under these conditions we can study metric theories with the *Parameterised Post-Newtonian* (PPN) formalism by describing deviations from simple Newtonian physics in the slow-motion and weak-field limit. This is possible in a theory-independent fashion, such that the only differences in these theories occur in the numerical coefficients that appear in the metric, characterised by a set of 10 real-valued PPN-Parameters [26]. Each of the parameters can be associated

with a specific physical effect, like the violation of conservation of momentum or equivalence principles, and certain values are assigned to them in a given theory. Thereby, comparing measured PPN parameters to their theoretical values can single out wrong theories in a purely experimental way. A more complete description of the PPN formalism and the physical meaning of PPN parameters is presented by Turyshev in this volume [2]. A recent review was given by Will [1]. Here we summarize the studies of those PPN parameters that can be constrained by pulsars. A more detailed account of related pulsar tests is given by Wex [27,3] and Stairs [4].

Violations of the Strong-Equivalence-Principle (SEP). The *Strong Equivalence Principle* (SEP) is completely embodied into general relativity, while alternative theories of gravity predict a violation of some or all aspects of SEP. The SEP is, according to its name, stronger than both the *Weak Equivalence Principle* (WEP) and the *Einstein Equivalence Principle* (EEP). The WEP states that all test bodies in an external gravitational field experience the same acceleration regardless of the mass and composition. While the WEP is included in all metric theories of gravity, the EEP goes one step further and also postulates *Lorentz-invariance* and positional invariance. Lorentz-invariance means that no preferred frame exists, so the outcome of a local non-gravitational experiment is independent from the velocity of the apparatus, while positional invariance renders it unimportant where this experiment is being performed. The SEP includes both the WEP and the EEP, but postulates them also for gravitational experiments. As a consequence, both Lorentz- and positional invariance should be independent of the gravitational self-energy of the bodies in the experiment. Obviously, all bodies involved in terrestrial lab experiments possess only a negligible fraction of gravitational self-energy, so that tests of SEP require the involvement of astronomical objects.

A violation of SEP means that there is a difference between gravitational mass, M_g, and inertial mass, M_i. The difference can be written as

$$\frac{M_g}{M_i} \equiv 1 + \delta(\epsilon) = 1 + \eta\epsilon + \mathcal{O}(\epsilon^2), \tag{13}$$

where ϵ is the gravitation self-energy in units of mc^2 and η is a parameter characterising the violation of SEP. The latter parameter was introduced by Nordvedt (1968) who suggested studying the Earth-Moon system to test for violations of SEP. Due to their different self-energy (Earth: $\epsilon \sim -4.6 \times 10^{-10}$, Moon: $\epsilon \sim -0.2 \times 10^{-10}$), Earth and Moon would fall differently in the external gravitational field of the Sun, leading to a polarization of the Earth-Moon orbit ("Nordvedt-effect"). Lunar-laser-ranging experiments can be used to put tight limits on η which is a linear combination of PPN parameters representing effects due to preferred locations, preferred frames and the violation of the conservation of momentum. However, even in the Earth-Moon case, or the Solar system in general, the self-energies involved are still small and do not test the SEP in strong-field regimes where deviations to higher order terms of ϵ could be

present. It is at that point where the circular pulsar-white dwarf systems become important.

For neutron stars, $\epsilon \sim 0.15$, which is large, in particular considering $\epsilon = 0.5$ for a black hole, and much larger than the self-energy of a white dwarf, $\epsilon \sim 10^{-4}$. Therefore, the pulsar and white dwarf companion of a binary system should feel a different acceleration due to the external Galactic gravitational field if SEP is violated. Similar to the Nordvedt effect, this should lead to a polarisation of the pulsar-white dwarf orbit ("gravitational Stark effect" [28]). The eccentricity vector of such a binary system should therefore have two components, one constant component due to the external acceleration, and another one that evolves in time following relativistic periastron advance. Since the present direction of the evolving eccentricity vector is unknown, a careful analysis of all relevant systems in a statistical manner is needed. Significant contributions to the results are made by long orbital-period and small-eccentricity systems, i.e. where P_b/e^2 is large [27].

Preferred-Frames and Conservation Laws. Some metric theories of gravity violate SEP specifically by predicting preferred-frame and preferred-location effects. A preferred universal rest frame, presumably equivalent with that of the Cosmic Microwave Background (CMB), may exist if gravity is mediated (in part) by a long-range vector field. This violation of local Lorentz-invariance is described by the two PPN parameters α_1 and α_2. While both parameters can be tightly constrained in the weak-field limit of the solar system, α_1 can also be studied in the strong-field regime by analysing the same low-eccentricity pulsar-white dwarf systems with a figure-of-merit given by $P_b^{1/3}/e$: If α_1 were different from zero, a binary system moving with respect to a preferred universal rest frame would again suffer a long-term change in its orbital eccentricity. In a statistical analysis similar to that for the study of the gravitational Stark-effect, the binary 5.3-ms pulsar PSR J1012+5307 is particularly valuable. It not only has an extremely small orbital eccentricity, for which only an upper limit of $e < 8 \times 10^{-7}$ (68% C.L.) was found from Jodrell Bank and Effelsberg observations [29], but its optically detected white dwarf companion also provides full 3-d velocity information relative to the CMB. Using this and other systems, Wex [27,3] derives $|\alpha_1| < 1.2 \times 10^{-4}$ (95% C.L.) which is slightly better than the solar system limit (see [2]).

In cases where theories both violate the Lorentz-invariance and the conservation of momentum, the equation of motion for a rotating body in the post-Newtonian limit contains so-called self-acceleration terms. This self-acceleration of the body's centre depends on the internal structure of the rotating body and results from the breakdown in conservation of total momentum. Another term in the self-acceleration involves the body's motion relative to a universal rest frame. Both contributions relate to the PPN parameter α_3 that can be tested using pulsars as isolated rotating objects [30,31], or as bodies in binary systems where both pulsar and companion suffer self-acceleration, leading to polarized orbits [32]. The limits derived by this second method using circular pulsar-white

dwarf systems are much tighter than studying the spin periods of isolated pulsars, resulting in $|\alpha_3| < 1.5 \times 10^{-19}$ (95% C.L.) [31]. This result for α_3 and a limit set on $(\alpha_3 + \zeta_2) < 4 \times 10^{-5}$ [33] constrains the PPN parameter ζ_2, which describes the non-conservation of momentum. Derived from a limit on the second period derivative, $\ddot{P}$, of PSR B1913+16 and, hence, its acceleration, the interpretation of this limit may be complicated as non-zero $\ddot{P}$ could arise from a number of sources such as timing noise [4].

Gravitational Dipole Radiation. Essentially any metric theory of gravity that embodies Lorentz-invariance in its field equations predicts gravitational radiation. However, the details of these predictions may differ in the speed of the gravity waves, the polarization of the waves and/or the multi-polarities of the radiation. If a theory satisfies SEP, like general relativity, gravitational dipole radiation is not expected, and the quadrupole emission should be the lowest multipole term. This arises because the dipole moment (centre of mass) of isolated systems is uniform in time due to the conservation of momentum and because the inertial mass that determines the dipole moment is the same as the mass that generates gravitational waves. In alternative theories, while the *inertial* dipole moment may remain uniform, the *gravity wave* dipole moment may not, since in a violation of SEP the mass generating gravitational waves depends differently on the internal gravitational binding energy of each body than does the inertial mass [1]. If dipole radiation is predicted, the magnitude of this effect depends on the difference in gravitational binding energies, expressed by the difference in coupling constants to a scalar gravitational field, $(\hat{\alpha}_p - \hat{\alpha}_c)$. For a white dwarf companion $|\alpha_c| \ll |\alpha_p|$, so that the strongest emission should occur for short-orbital period pulsar-white dwarf systems. Again, the binary pulsar J1012+5307 becomes extremely useful. Given its vanishing eccentricity, the change in orbital period due to dipole radiation becomes

$$\dot{P}_b^{dipole} \simeq \frac{4\pi^2 G_*}{c^3 P_b} \frac{M_p M_c}{M_p + M_c} \hat{\alpha_p}^2 + \mathcal{O}\left(\frac{v^5}{c^5}\right), \tag{14}$$

where G_* is the "bare" gravitational constant. With the optically detected companion, the measured radial velocity can be use to correct for Doppler effects. For this system [29], Wex [3] derives a limit of $|\hat{\alpha}_p|^2 < 4 \times 10^{-4}$ (95% C.L.).

Time Variability of the Gravitational Constant. Three different pulsar tests are available to test the time variability of the gravitational constant, G, and to derive an upper limit for $\dot{G}/G$. A time variability is only allowed if SEP is violated due to preferred locations in space and time. In the case of pulsars, a changing G would change the gravitational binding energy of a neutron star and thereby possibly also its moment of inertia, which would cause a change in the spin-down behaviour, namely a contribution to $\dot{P}$. A comparison with observed values leads to limits of the order of $\dot{G}/G < 10^{-11}$ yr^{-1} [4]. A slightly more stringent limit can be derived from the effects that a varying G would

have on orbital periods [34,35]. Both limits are still about an order of magnitude above the limits set by solar system tests. Moreover, they depend to some degree on the compactness of the neutron star and its equation of state, so that they are not truly theory independent [1]. An interesting alternative test uses the mass determination for neutron stars [36], utilising that the Chandrasekhar mass depends directly on G. Studying the mass of millisecond pulsars as function of pulsar age, a strong limit of $\dot{G}/G < (-0.6 \pm 4.2)10^{-12}$ yr^{-1} (95% C.L.) is derived. However, while the mass of neutron stars can be determined quite accurately in relativistic binaries (see Sect. 5.2), an age estimation relying on (2) can contain considerable uncertainty.

5.2 Tests Using Double Neutron Stars

Even though in all metric theories matter and non-gravitational fields respond only to the space-time metric, it is possible that scalar or vector fields exist in addition to the metric. Damour & Esposito-Farèse developed a framework to study theories at a second post-Newtonian (2PN) level where gravity is mediated by a tensor field and one or more scalar field [38]. These theories are interesting since scalar partners to gravitons arise naturally in quantum gravity and unified theories. Damour & Esposito-Farèse show that it is possible to construct corresponding theories where deviations from general relativity are not visible in the weak field but only manifest themselves in a "spontaneous scalarization" if the strong field limit is approached. They conclude that current solar system tests and also upcoming satellite missions will not be able to replace the strong field tests provided by radio pulsars. Indeed, they use the DNSs, PSR B1534+12 and PSR B1913+16, together with PSR B0655+64 and solar system tests, to significantly constrain the parameters describing the coupling of matter to the scalar field [39]. More stringent limits have been presented recently using also results for the pulsar-white dwarf system PSR J1141−4565 [40]. A more classical approach using DNSs for tests of theories of gravity is made with the measurement of post-Keplerian (PK) parameters as observables.

Because of the strong gravitational fields, we expect DNSs to suffer large relativistic effects. In this case, we cannot necessarily assume that we understand the underlying physics, even though general relativity appears to describe the physics in the solar system to high precision. Therefore, a purely theory-independent approach like the PPN approximation is difficult to realize. Instead, one can only use an existing theory of gravity and check if the observations are consistently described by the measured Keplerian and PK parameters. In each theory, for point masses with negligible spin contributions, the PK parameters should only be functions of the a priori unknown pulsar and companion mass, M_p and M_c, and the easily measurable Keplerian parameters. With the two masses as the only free parameters, an observation of two PK parameters will already determine the masses uniquely in the framework of the given theory. The measurement of a third or more PK parameters then provides a consistency check. In general relativity, the five most important PK parameters are given to

lowest Post-Newtonian order by (e.g. [3]):

$$\dot{\omega} = 3T_{\odot}^{2/3} \left(\frac{P_b}{2\pi}\right)^{-5/3} \frac{1}{1-e^2} (M_p + M_c)^{2/3}, \tag{15}$$

$$\gamma = T_{\odot}^{2/3} \left(\frac{P_b}{2\pi}\right)^{1/3} e \frac{M_c(M_p + 2M_c)}{(M_p + M_c)^{4/3}}, \tag{16}$$

$$\dot{P}_b = -\frac{192\pi}{5} T_{\odot}^{5/3} \left(\frac{P_b}{2\pi}\right)^{-5/3} \frac{\left(1 + \frac{73}{24}e^2 + \frac{37}{96}e^4\right)}{(1-e^2)^{7/2}} \frac{M_p M_c}{(M_p + M_c)^{1/3}}, \tag{17}$$

$$r = T_{\odot} M_c, \tag{18}$$

$$s = T_{\odot}^{-1/3} \left(\frac{P_b}{2\pi}\right)^{-2/3} x \frac{(M_p + M_c)^{2/3}}{M_c}, \tag{19}$$

where P_b is the period and e the eccentricity of the binary orbit. The masses M_p and M_c of pulsar and companion, respectively, are expressed in solar masses ($M_{\odot}$). We define the constant $T_{\odot} = GM_{\odot}/c^3 = 4.925490947\mu$s where G denotes the Newtonian constant of gravity and c the speed of light. The first PK parameter, $\dot{\omega}$, is the easiest to measure and describes the relativistic advance of periastron. According to (15) it provides an immediate measurement of the total mass of the system, $(M_p + M_c)$. The parameter γ denotes the amplitude of delays in arrival times caused by the varying effects of the gravitational redshift and time dilation (second order Doppler) as the pulsar moves in its elliptical orbit at varying distances from the companion and with varying speeds. The decay of the orbit due to gravitational wave damping is expressed by the change in orbital period, $\dot{P}_b$. The other two parameters, r and s, are related to the Shapiro delay caused by the gravitational field of the companion. These parameters are only measurable, depending on timing precision, if the orbit is seen nearly edge-on.

Until very recently, only two DNSs had more than two PK parameters determined, the 59-ms pulsar B1913+16 and the 38-ms PSR B1534+12. For PSR B1913+16 with an eccentric ($e = 0.61$) 7.8-hr orbit, the PK parameters $\dot{\omega}$, γ and $\dot{P}_b$ are measured very precisely. Correcting the observed $\dot{P}_b$ value for effects of relative motion (see Sect. 4.1), the measured value is in excellent agreement with the prediction of general relativity for quadrupole emission (see Fig. 4). This result demonstrates impressively that general relativity provides a self-consistent and accurate description of the system which can be described as orbiting point masses, i.e. the structure of the neutron stars does not influence their orbital motion as expected from SEP. The precision of this test is limited by our knowledge of the Galactic gravitational potential and the corresponding correction to $\dot{P}_b$. The timing results for PSR B1913+16 provide us with the most precise measurements of neutron star masses so far, i.e. $M_p = (1.4408 \pm 0.0003)M_{\odot}$ and $M_c = (1.3873 \pm 0.0003)M_{\odot}$ [37]. It is worth pointing out that these values include the unknown Doppler factor.

The 10-hr orbit of the second DNS PSR B1534+14 ($e = 0.27$) is observed under fortunate circumstances, it is seen nearly edge-on. Thereby, in addition to the three PK parameters observed for PSR B1913+16, the Shapiro-delay parameters r and s can be measured, enabling non-radiative aspects of gravitational

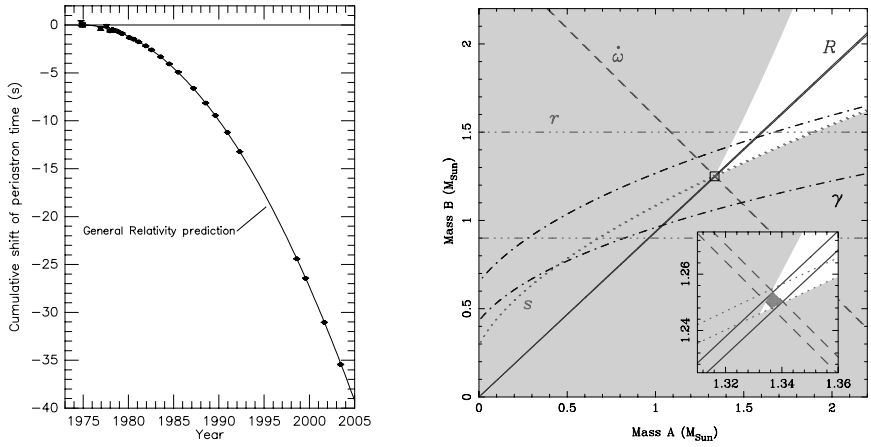


Fig. 4. (left) Shift in the periastron passage of the DNS PSR B1913+16 plotted as a function of time, resulting from orbital energy loss due to the emission of gravitational radiation. The agreement between the data, now spanning almost 30 yr, and the predicted curve due to gravitational quadrupole wave emission is now better than 0.5%. Figure provided by Joel Weisberg and Joe Taylor. (right) "Mass-mass" diagram showing the observational constraints on the masses of the neutron stars in the double-pulsar system J0737–3039. The shaded regions are those which are excluded by the Keplerian mass functions of the two pulsars. Further constraints are shown as pairs of lines enclosing permitted regions as predicted by general relativity: (a) the measurement of $\dot{\omega}$ gives the total system mass $m_A + m_B = 2.59\ \mathrm{M}_\odot$; (b) the measurement of the mass ratio $R = m_A/m_B = 1.07$; (c) the measurement of the gravitational redshift/time dilation parameter γ; (d) the measurement of the two Shapiro delay parameters r and s. Inset is an enlarged view of the small square encompassing the intersection of the three tightest constraints, representing the area allowed by general relativity and the present measurements.

theories to be tested, as $\dot{P}_b$ is not necessarily needed. In fact, the observed value of $\dot{P}_b$ seems to be heavily influenced by Shklovskii-terms, so that the corresponding line fails to meet the others in a M_p-M_c diagram. However, assuming that general relativity is the correct theory of gravitation, the deviation from the predicted value and the measured proper motion, μ, can be used to compute the necessary correction and hence the distance to the pulsar, $d = 1.02 \pm 0.05$ kpc [41].

5.3 Tests Using Profile Structure Data

In addition to the use of pulsars as clocks, strong gravity effects can also be tested using pulse structure data, namely the effects of "geodetic precession" in the DNSs PSR B1913+16 and PSR B1534+14. In both cases, the pulsar spin axis appears to be misaligned with the orbital angular momentum vector. In such a case, general relativity predicts a relativistic spin-orbit coupling, analogous to spin-orbit coupling in atomic physics. The pulsar spin precesses about the

total angular momentum, changing the relative orientation of the pulsar towards Earth. As a result, the angle between the pulsar spin axis and our line-of-sight changes with time, so that different portions of the emission beam are observed [42]. Consequently, changes in the measured pulse profile and its polarization are expected. In extreme cases, the precession may even move the beam out of our line-of-sight and the pulsar may disappear as predicted for PSR B1913+16 for the year 2025 [43]. See the review by Kramer [44] for a detailed description of this effect and its observations.

5.4 Recent Discoveries

The tremendous success of recent surveys, in particular those using the Parkes telescope (e.g. [45]), has lead not only to the discovery of more than 700 new pulsars, but also to some very exciting new binary systems. Until recently, only five DNSs were known. This situation has changed and now eight systems can be studied. The most recent addition is a DNS discovered in the Parkes Multibeam (PM) Survey, PSR J1756−2251 (Faulkner et al., in prep.). This system shows similarities with PSR B1913+16 as its orbital period is somewhat less than 8 hours although its eccentricity is smaller ($e = 0.18$).

The recent discoveries benefit from larger available computing power which enables so-called "acceleration searches" for fast orbiting binary pulsars. Such techniques try to correct for the usually made assumption that the pulse period remains constant during the observations. For compact binary systems, this assumption is violated due to large Doppler shifts in period, resulting in much reduced sensitivity in standard Fourier searches. The employment of acceleration codes has therefore lead to a number of new binary pulsars with short orbital periods. Another example is PSR J1744−3922 (Faulkner et el., in prep.). This 172-ms pulsar is in an almost circular 4.6-hr orbit and hence only the second long-period pulsar in such a short orbit. The other such pulsar is PSR J1141−6545 which is a 393-ms PM pulsar in an eccentric 4.5-hr orbit [46]. Both pulsars appear to have a white dwarf companion, but while PSR J1141−6545's companion is heavy ($M_c \sim 1 M_\odot$), the new pulsar's companion is probably much lighter ($M_c \geq 0.08 M_\odot$). Whilst these are indeed exciting discoveries, the most stunning success is clearly the recent discovery of the first double-pulsar system, J0737−3039.

6 The Double-Pulsar

The 22.8-ms pulsar J0737-3039 was discovered in April 2003 [47]. It was soon found to be a member of the most extreme relativistic binary system ever discovered: its short orbital period ($P_b = 2.4$ hrs) is combined with a remarkable high value of periastron advance ($\dot{\omega} = 16.88 \pm 0.09\,\mathrm{deg/yr}$, i.e. four times larger than for PSR B1913+16!) and a short coalescing time (~ 85 Myr). The latter time-scale boosts the hopes for detecting a merger of two neutron stars with first-generation ground-based gravitational wave detectors by about an order of

magnitude compared to previous estimates based on only the DNSs B1534+12 and B1913+16 [47]. Consequently, during the lecture I had presented this pulsar already as the most beautiful laboratory for testing general relativity found so far, pointing also out that with a geodetic precession period of only 70 yr future studies should reveal interesting and exciting results. But little did we know then which surprise was still waiting for us.

In October 2003, our team detected radio pulses from the second neutron star when data sets covering the full orbital period were analysed [48]. The reason why signals from the 2.8-s pulsar companion (now called PSR J0737−3039B, hereafter "B") to the millisecond pulsar (now called PSR J0737−3039A, hereafter "A") had not been found earlier, became clear when it was realized that B was only visible clearly for two short parts of the orbits. For the remainder of the orbit, the pulsar B is extremely weak and only detectable with the most sensitive equipment. The detection of a young pulsar-companion B clearly confirmed the evolution scenario presented in Sect. 3 and made this already exciting system sensational, providing a truly unique testbed for relativistic gravity.

Indeed, we have now measured A's $\dot{\omega}$ and γ and we have also detected the Shapiro delay in the pulse arrival times of A due to the gravitational field of B, providing a precise measurement of the orbital inclination of $\sin i = 0.9995(^{+4}_{-32})$. Obviously, as another strike of luck, we are observing the system almost completely edge-on which allows us to also probe pulsar magnetospheres for the very first time by a background beacon. The measurements already provide four measured PK parameters, resulting in a $m_A - m_B$ plot shown in Fig. 4. The orbital decay due to gravitational wave emission is already visible in the data, but the uncertainties are yet too large to provide a useful constraint. However, in addition to tests with these PK parameters, the detection of B as a pulsar opens up opportunities that go well beyond what has been possible so far. With a measurement of the projected semi-major axes of the orbits of both A and B, we obtain a precise measurement of the mass ratio, $R(m_\mathrm{A}, m_\mathrm{B}) \equiv m_\mathrm{A}/m_\mathrm{B} = x_\mathrm{B}/x_\mathrm{A}$, providing a further constraint displayed in Fig. 4. For every realistic theory of gravity, we can expect the mass ratio, R, to follow this simple relation [49]. Most importantly, the R-line is not only theory-independent, but also independent of strong-field (self-field) effects which is not the case for PK-parameters. This provides a stringent and new constraint for tests of gravitational theories as any intersection of the PK-parameters *must* be located on the R-line. At the same time, it provides us already with very accurate mass measurements for the neutron stars, $M_A = (1.337 \pm 0.005) M_\odot$ and $M_B = (1.250 \pm 0.005) M_\odot$, respectively, making B the least-massive neutron star ever observed.

The equations for the PK parameters given in Sect. 5.2 are all given to lowest Post-Newtonian order. However, higher-order corrections may become important if relativistic effects are large and timing precision is sufficiently high. Whilst this has not been the case in the past, the double pulsar system may allow measurements of these effects in the future [48]. One such effect involves the prediction by general relativity that, in contrast to Newtonian physics, the neutron stars' spins affect their orbital motion via spin-orbit coupling. This effect would be visible clearest as a contribution to the observed $\dot{\omega}$ in a secular [50] and

periodic fashion [51]. For the J0737−3039 system, the expected contribution is about an order of magnitude larger than for PSR B1913+16, i.e. 2×10^{-4} deg yr^{-1} (for A, assuming a geometry as determined for PSR B1913+16 [43]). As the exact value depends on the pulsars' moment of inertia, a potential measurement of this effect allows the moment of inertia of a neutron star to be determined for the first time [52]. Obviously, the double pulsar system offers improved but also new tests of general relativity. The current data already indicate an agreement of the observed with the expected Shapiro parameter of $s_{\mathrm{obs}}/s_{\mathrm{exp}} = 1.00007 \pm 0.00220$ (Kramer et al. in prep.) where the uncertainties are likely to decrease.

7 Conclusions and Outlook

Millisecond pulsars find a wide range of applications, in particular for precise tests of theories of gravity. After the discovery of pulsars thereby marked the beginning of a new era in fundamental physics, pulsars discovered and observed with the future Square-Kilometer-Array (SKA) will further transform our understanding of gravitational physics. The SKA's sensitivity will discover the majority of pulsars in the Galaxy, almost certainly providing the discovery of the first pulsar-black hole system. For tests of general relativity such a system would have a discriminating power that surpasses all its present and foreseeable competitors [39]. In particular, we could directly test black hole properties as predicted by general relativity, such as the Cosmic Censorship Conjecture or the "no-hair" theorem. Moreover, the pulsars discovered with the SKA would act as arms of a huge gravitational wave detector enabling the study of a possible gravitational wave background in a frequency range that is inaccessible to LIGO or even LISA. Clearly, the SKA will provide yet another leap in our understanding and application of pulsars.

References

1. C. M. Will, Living Rev. Relativity **4**, 4. [Online article]: cited on 1 Oct 2003, http://www.livingreviews.org/lrr-2001-4 (2001)
2. S.G. Turyshev et al., 35 Years of Testing Relativistic Gravity: Where Do We Go from Here?, Lect. Notes Phys. **648**, 311–330 (2004)
3. N. Wex, In: *Gyros, Clocks, Interferometers...: Testing Relativistic Gravity in Space*, eds C. Lämmerzahl, C. W. F. Everitt, & F. W. Hehl, (Springer, 2001)
4. I. H. Stairs, Living Rev. Relativity **6**, 5. [Online article]: cited on 1 Oct 2003, http://www.livingreviews.org/lrr-2003-5 (2003)
5. D. R. Lorimer, Living Rev. Relativity **4**, 5. [Online article]: cited on 1 Oct 2003, http://www.livingreviews.org/lrr-2001-5 (2001)
6. S. E. Thorsett & D. Chakrabarty, Ap. J. **512**, 288 (1999)
7. J. R. Oppenheimer & G. Volkoff, Phys. Rev. **55**, 374 (1939)
8. V. E. Zavlin and G. G. Pavlov, A&A **329**, 583 (1998)
9. M. A. McLaughlin, I. H. Stairs, V. M. Kaspi et al., Ap. J. **591**, L135 (2003)
10. G. F. Bignami, P. A. Caraveo, A. D. Luca & S. Mereghetti, Nature **423**, 725 (2003)
11. T. H. Hankins, J. S. Kern, J. C. Weatherall & J. A. Eilek, Nature **422**, 141 (2003)

12. M. Kramer, K. M. Xilouris, A. Jessner, et al., A&A, **322**, 846 (1997)
13. A. G. Lyne, R. S. Pritchard & F. G. Smith, MNRAS **265**, 1003 (1993)
14. M. Kramer, A. G. Lyne, G. Hobbs, et al., Ap.J. **593**, L31 (2003)
15. M. D. Young, R. N. Manchester & S. Johnston, Nature **400**, 848 (1999)
16. D. C. Backer, S. R. Kulkarni, C. Heiles, M. M. Davis & W. M.Goss, Nature **300**, 615 (1982)
17. M. A. Alpar, A. F. Cheng, M. A. Ruderman & J. Shaham, Nature **300**, 728 (1982)
18. V. M. Kaspi, J. H. Taylor & M. Ryba, Ap. J. **428**, 713 (1994)
19. E. M. Standish, A&A **114**, 297 (1982)
20. I. I. Shapiro, Phys. Rev. Lett. **13**, 789 (1964)
21. D. C. Backer & R. W. Hellings, Ann. Rev. Astr. Ap. **24**, 537 (1986)
22. E. S. Phinney, Philos. Trans. Roy. Soc. London A**341**, 39 (1992)
23. http://pulsar.princeton.edu/tempo/
24. http://www.jb.man.ac.uk/research/pulsar/observing/progs/progs.html
25. http://www.mpifr-bonn.mpg.de/div/pulsar/former/olegd/soft.html
26. K. Nordtvedt, Phys. Rev. **170**, 1186 (1968)
27. N. Wex, In: *Pulsar Astronomy - 2000 and Beyond, IAU Colloquium 177*, eds M. Kramer, N. Wex & R. Wielebinski, R., ASP Conf. Series Vol. 202 (PASP, San Francisco 2000), p. 113
28. T. Damour & G. Schäfer, Phys. Rev. Lett. **66**, 2549 (1991)
29. C. Lange, F. Camilo, N. Wex et al., MNRAS **326**, 274 (2001)
30. C. M. Will, *Theory and Experiment in Gravitational Physics*, (Cambridge University Press, Cambridge 1993)
31. J. F. Bell, Ap. J. **462**, 287 (1996)
32. Bell, J. F. & Damour, T., Class. Quantum Grav., **13**, 3121 (1996)
33. C. M. Will, Ap. J. **393**, L59 (1992)
34. T. Damour, G. W. Gibbons & J. H. Taylor, Phys. Rev. Lett. **61**, 1151 (1988)
35. Z. Arzoumanian, *PhD thesis*, Princeton University (1995)
36. S. E. Thorsett, Phys. Rev. Lett. **77**, 1432 (1996)
37. J. M. Weisberg & J. H. Taylor, In: *Radio Pulsars*, eds M. Bailes, D.J. Nice & S.E. Thorsett, ASP Conf. Series Vol. 302 (PASP, San Francisco 2003), p. 93
38. T. Damour & G. Esposito-Farèse, Phys. Rev. D**53**, 5541 (1996)
39. T. Damour & G. Esposito-Farèse, Phys. Rev. D**58**, 042001 (1998)
40. Esposito-Farèse, G. contribution to 10th Marcel Grossmann meeting, gr-qc/0402007 (2004)
41. I. H. Stairs, S. E. Thorsett, J. H. Taylor & A. Wolszczan, Ap. J. **581**, 501 (2002)
42. T. Damour & R. Ruffini, Academie des Sciences Paris Comptes Rendus Ser. Scie. Math. **279**, 971 (1974)
43. M. Kramer, Ap. J. **509**, 856 (1998)
44. M. Kramer, In: *The Ninth Marcel Grossmann Meeting*, eds V.G. Gurzadyan, R.T. Jantzen & R. Ruffini (World Scientific, Singapore 2002) p. 219
45. R. N. Manchester, A. G. Lyne, F. Camilo, et al., MNRAS **328**, 17 (2001)
46. V. M. Kaspi, A. G. Lyne, R. N. Manchester, et al., Ap. J. **543**, 321 (2000)
47. Burgay, M., D'Amico, N., Possenti, et al., Nature, **426**, 531 (2003)
48. Lyne, A. G., Burgay, M., Kramer, M., et al., Science, **303**, 1153 (2004)
49. Damour, T. & Taylor, J. H., Phys. Rev. D, **45**, 1840 (1992)
50. Barker, B. M. & O'Connell, R. F., Phys. Rev. D, **12**, 329 (1975)
51. Wex, N., 1995, *Class. Quantum Grav.*, **12**, 983 (1995)
52. Damour, T. & Schäfer, G., Nuovo Cim., **101**, 127 (1988)

Fundamental Units: Physics and Metrology

Lev B. Okun

ITEP, Moscow, 117218, Russia

Abstract. The problem of fundamental units is discussed in the context of achievements of both theoretical physics and modern metrology. On one hand, due to the fascinating accuracy of atomic clocks, the traditional macroscopic standards of metrology (second, metre, kilogram) are giving way to standards based on fundamental units of nature: velocity of light c and quantum of action h. On the other hand, the poor precision of the gravitational constant G, which is widely believed to define the "cube of theories" and the units of the future "theory of everything", does not allow to use G as a fundamental dimensional constant in metrology. The electromagnetic units in SI are actually based on concepts of prerelativistic classical electrodynamics such as ether, electric permittivity and magnetic permeability of vacuum. Concluding remarks are devoted to terminological confusion which accompanies the progress in basic physics and metrology.

1 Introduction

The problem of fundamental units has many facets, three of which seem to be most important: theoretical, experimental and technological. At present they are inseparable. Theory, the so called Standard Model, formulates basic physical laws and mathematical methods of their application. Theoretical laws were established and continue to be established and tested on the basis of ingenious experiments and astronomical observations of higher and higher accuracy for an expanding space of parameters.

Precision experiments and observations, in their turn, are unthinkable without modern high technologies, including lasers and computers. These technologies are indissolubly connected with metrology – creation, perfection and unification of standards of physical units, while metrology is widely using the results of such theories as quantum mechanics, relativity theory, electrodynamics, condensed matter theory etc. Thus the circle is closed.

The situation is additionally complicated by the fact that the Standard Model is not a complete theory. It has many unsolved problems. Perhaps the most burning is the problem of the existence of fundamental scalar particles (higgses), responsible for the masses of all fundamental particles (from the heaviest one – the t-quark – to the lightest of the three neutrinos). We still do not understand the role of the three families of leptons and quarks. It would be too naive to think that their only justification is CP-violation. We still lack a successful theory unifying electroweak and strong interactions. The hypothesis of the existence of moderately violated supersymmetry – symmetry between fundamental bosons

L.B. Okun, Fundamental Units: Physics and Metrology, Lect. Notes Phys. **648**, 57–74 (2004)
http://www.springerlink.com/

and fermions – is still not confirmed by experiments. Mathematical constructs of the type of superstrings and M-theory which in the beginning were considered as attempts to unify quantum gravity with electroweak and strong interaction, as time goes by, withdraw into a separate field of mathematics, whose practioners do not promise any applications to the real physical world. The situation might become drastically different if manifestations of extra space dimensions are discovered; in particular if laws of gravity turn out to change at TeV scale.

Sections 2–6 are devoted to the history of units based on c, h, G. Sections 7–9 deal with the units based on c, h, e and precision frequency measurements. Section 10 compares the Gaussian units and SI units. It is argued that while the latter are more convenient for practical purposes, the former allow to better understand the basic notions of modern physics. Therefore the use of both systems of units should be allowed in physics textbooks. Section 11 contains concluding remarks.

2 Fundamental Parameters and Units

The essence of theoretical physics is expressed by dimensionless equations for dimensionless quantities. However one cannot do experimental physics (and teach physics) without dimensional quantities and dimensional units.

In what follows we shall refer to dimensionless fundamental constants such as $e^2/\hbar c$, or m_e/m_p as fundamental parameters (here e is the electron charge, $\hbar$ is the reduced quantum of action ($\hbar = h/2\pi$) and of angular momentum, c is the velocity of light in vacuum, m_e and m_p are the electron mass and the proton mass, respectively).

In the absence of established terminology we shall refer to dimensional fundamental constants as fundamental units. Examples of units: c (for velocity), $\hbar$ (for action and angular momentum). According to our definition G is also a fundamental unit (though indirectly).

3 Planck Units

When in 1899-1900 Planck discovered h [1], he used this discovery to introduce universal units, which at present are written in the form

$$l_P = \hbar/m_P c\ ,\ \ t_P = \hbar/m_P c^2\ ,\ \ m_P = (\hbar c/G)^{1/2}\ \ , \tag{1}$$

where G is Newton's gravitational constant.

Planck derived his units by using dimensional order of magnitude relations:

$$c = \frac{l_P}{t_P}\ ,\ \ \frac{Gm_P^2}{l_P} = m_P c^2\ ,\ \ \frac{Gm_P^2}{l_P} t_P = \hbar\ \ . \tag{2}$$

He was inspired by the idea that his units are universal (contrary to "handcrafted" earthbound ordinary units – meter, second, gram): they are the same at any far away corner of the universe.

Planck also considered as universal the Planck temperature $T_P = m_P c^2/k$. But Boltzmann's k is not a universal unit, it is a conversion factor: $k = 8.6 \cdot 10^{-5}$ eV/K (hint: $\hbar\omega/kT$).

4 *c, h, G* – Units

From the point of view of the future "theory of everything" it is natural to use c, h, G as fundamental dimensional constants.

In 1928 Gamov, Ivanenko and Landau [2] considered the theory "of the world as a whole" in terms of dimensional fundamental constant c, h, G. In 1928 Landau was 20 years old, Gamow and Ivanenko twentyfour (see Figs. 1,2). They had written the paper "World Constants and Limiting Transitions" (see Fig. 3) as a humorous birthday present to their friend, a young lady. None of them ever referred to this paper in their subsequent publications. But the ideas of the paper were fundamental. In 1936 Bronstein [3] worked on a theory in which all three constants are finite. It was one of the first papers on relativistic quantum gravity. In 1967 ideas of [2,3] were presented in the form of a cube (see Figs. 4,5) by Zelmanov [4]. Later on it was further developed by others [5,6].

The vertices of this cube represent nonrelativistic mechanics (NM), nonrelativistic gravity (NG), nonrelativistic quantum mechanics (QM), special relativity (SR), quantum field theory (QFT), general relativity (GR) and finally relativistic quantum gravity (QGR) or theory of everything (TOE).

The cube, made of units, is "endowed" with dimensionless parameters like α, α_s, mixing angles, mass ratios, etc. Their values are expected to follow from TOE. Similar to the cube is "dimensional pyramid" (Kuchar [7], Sanchez [8]) with 4 vertices and 4 planes, Fig. 6.

Note that Einstein tried to build a unified theory of electricity and gravity ("TOE") in the left-hand vertical plane of the cube, without Quantum Mechanics, without $\hbar$.

Planck units allow one to deal in the equations of TOE only with dimensionless functions of dimensionless variables and dimensionless parameters of the type $\alpha = e^2/\hbar c$ or m_e/m_p. Conceptually Planck units are excellent, but practically they have serious shortcomings, caused by G, the same G which allows to bring gravitation and cosmology into the realm of quantum phenomena. Thus the source of strength at the same time turns out to be a source of weakness.

5 Planck Units Are Impractical

The obvious shortcoming of Planck units is that they differ by many orders of magnitude from atomic units commonly used in physics. Their values are natural for the early universe and TOE, but not for mundane physics:

$$l_P = 10^{-35} \text{ m} , \quad t_P = 10^{-43} \text{ s} , \quad E_P = m_P c^2 = 10^{19} \text{ GeV} \quad .$$

The energy which corresponds to the Planck mass is unattainable by accelerators even of the remotest future. (Note, however, that it is only a few orders

Fig. 1. Meeting in Kharkov, 1928, attended by Gamow, Ivanenko, and Landau

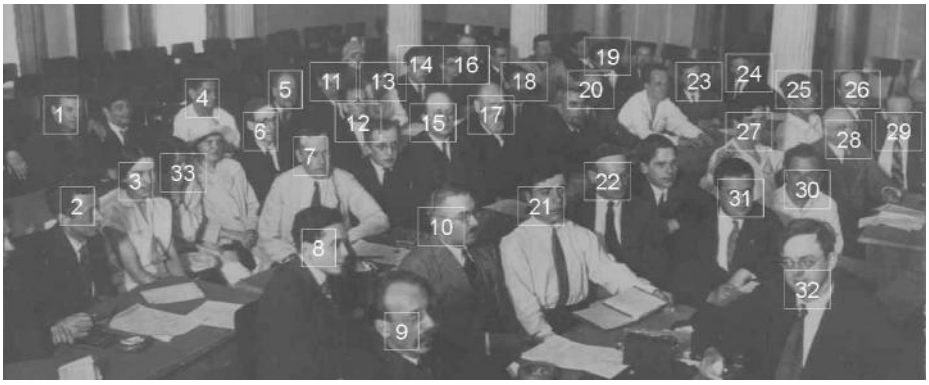
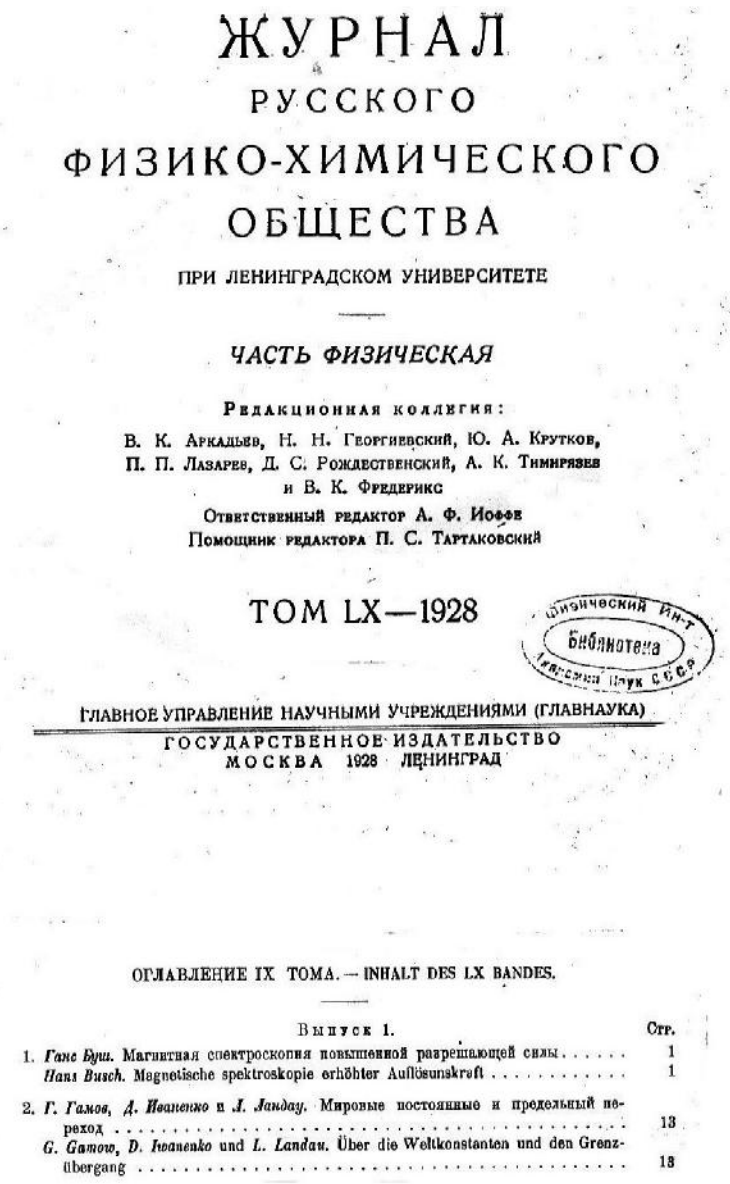


Fig. 2. Who is who in Fig.1.

1. Dunaev	8. Ivanenko	15. Efimovich	22. Mandel	29. Frenkel
2. Heitler	9. Obreimov	16. Ogievetsky	23. Vereschagin	30. Rosenkevich
3. Arsenyeva	10. Fock	17. Grommer	24. Slutsky	31. Finkelshtein
4. Davydov	11. Leipunsky	18. Muskelischvili	25. Gamov	32. Jordan
5. Todorovich (?)	12. Kopp	19. Korsunsky	26. Shubnikov	33. Timoreva
6. Frish	13. Kotsarova	20. Gorwitz	27. Landau	
7. Bursian	14. Khalfin	21. Ambartsumian	28. Shtrumm	



ЖУРНАЛ
РУССКОГО
ФИЗИКО-ХИМИЧЕСКОГО
ОБЩЕСТВА

ПРИ ЛЕНИНГРАДСКОМ УНИВЕРСИТЕТЕ

ЧАСТЬ ФИЗИЧЕСКАЯ

Редакционная коллегия:
В. К. Аркадьев, Н. Н. Георгиевский, Ю. А. Крутков,
П. П. Лазарев, Д. С. Рождественский, А. К. Тимирязев
и В. К. Фредерикс
Ответственный редактор А. Ф. Иоффе
Помощник редактора П. С. Тартаковский

ТОМ LX—1928

ГЛАВНОЕ УПРАВЛЕНИЕ НАУЧНЫМИ УЧРЕЖДЕНИЯМИ (ГЛАВНАУКА)
ГОСУДАРСТВЕННОЕ ИЗДАТЕЛЬСТВО
МОСКВА 1928 ЛЕНИНГРАД

ОГЛАВЛЕНИЕ IX ТОМА. — INHALT DES LX BANDES.

Fig. 3. The title page and a part of the contents of the journal where the paper by G. Gamow, D. Ivanenko, and L. Landau was published.

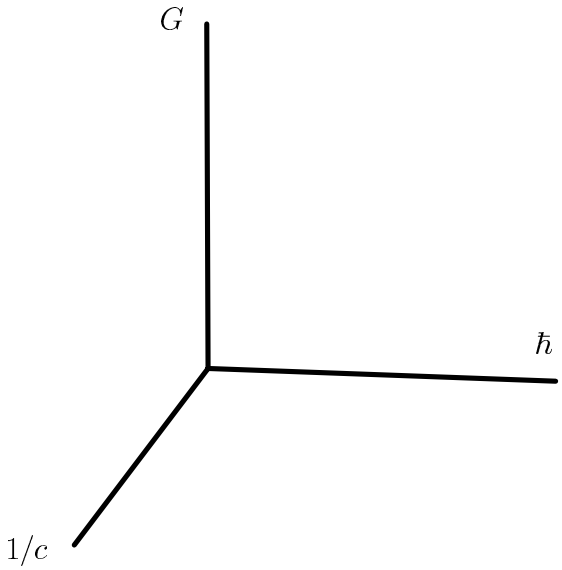


Fig. 4. Orthogonal axes. At the origin there is no gravity, no maximal velocity, no quantum effects

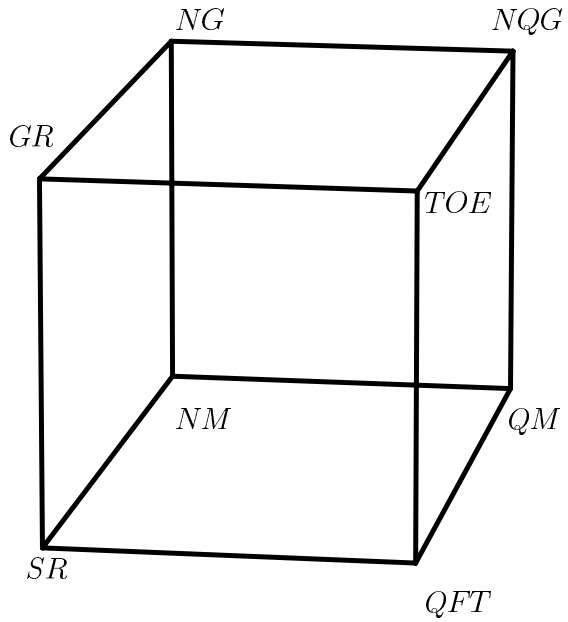


Fig. 5. Cube of theories

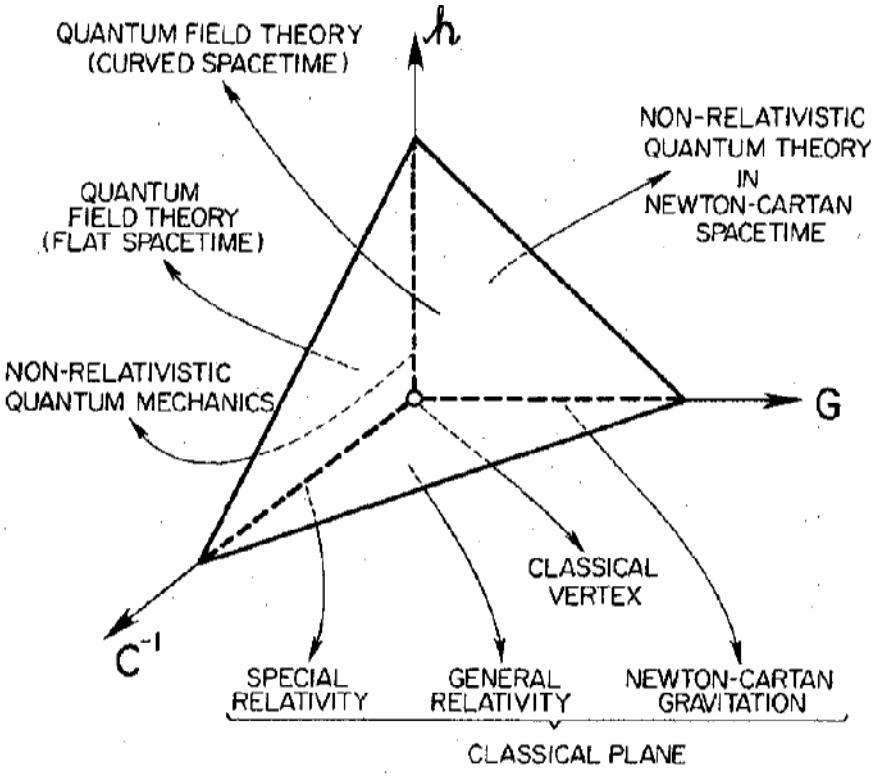


Fig. 6. Dimensional pyramid

of magnitude larger than the grand unification scale of electroweak and strong interactions.) The Planck units of length and time are vanishingly small compared with atomic units. Of course the huge powers of ten are not frightful by themselves. As is well known, atomic units also differ by many orders of magnitude from SI units, which does not prevent atomic standards to be the base of modern metrology.

Much more essential is another shortcoming of Planck units, which stems from the fact that G is known with rather poor accuracy [9] (by several orders worse than those of c and h and by approximately ten orders worse than the precision of atomic clocks). Thus it is impossible to use the Planck units as standards in modern precision physics and technology.

6 Units of Stoney

Planck's use of G as a basis for defining the unit of mass was caused by the absence at the beginning of the 20th century of another natural, not "hand-crafted", candidate for the unit of mass. In that respect Planck's universal units resemble the universal units suggested 30 years earlier by the Irish physicist

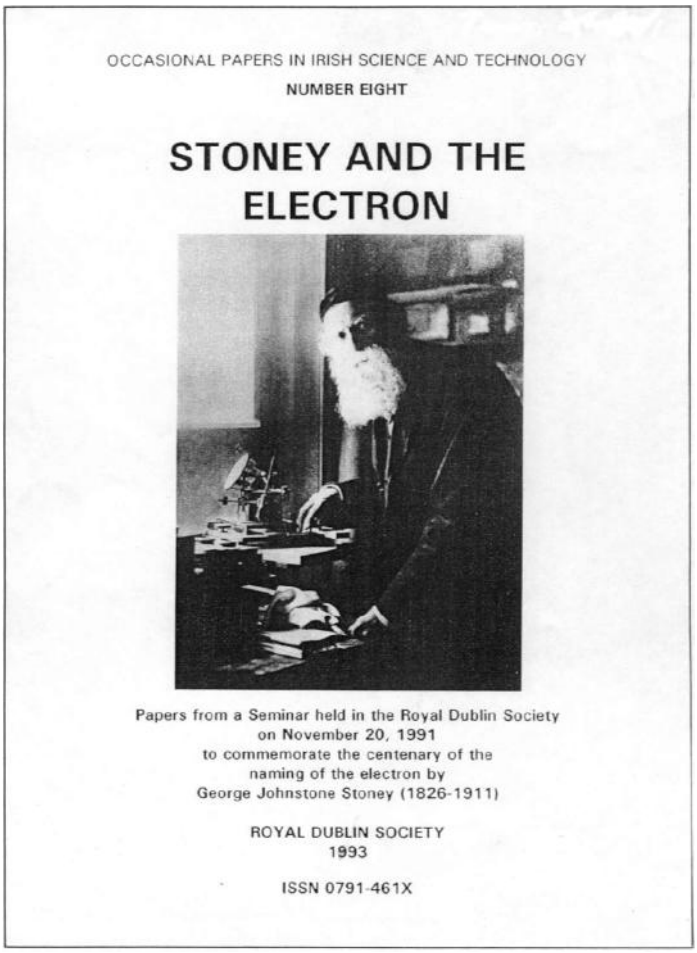


Fig. 7. Cover of the book dedicated to G.J. Stoney (with permission of the Royal Dublin Society)

Stoney (1826 - 1910), secretary of the Royal Dublin Society (see Figs. 7,8). By studying electrolysis, he was the first who measured the value of elementary charge e and introduced into physics the term "electron" for the carrier of this charge (in modern terminology he was referring to ions). From e, c, G Stoney [10] constructed in 1870 - 1880 universal units with dimensions of length, time and mass:

$$l_{\rm S} = e\sqrt{G}/c^2 \ , \ \ t_{\rm S} = e\sqrt{G}/c^3 \ , \ \ m_{\rm S} = e/\sqrt{G} \ , \tag{3}$$

which he derived from dimensional equations:

$$c = l_{\rm S}/t_{\rm S} \ , \ \ e^2 = Gm_{\rm S}^2 \ , \ \ e^2/l_{\rm S} = m_{\rm S}c^2 \ .$$

Let us note that the units of Stoney are only by a factor $\sqrt{\alpha}$ smaller than those of Planck.

THE

LONDON, EDINBURGH, AND DUBLIN

PHILOSOPHICAL MAGAZINE

AND

JOURNAL OF SCIENCE.

VOL. XI.—FIFTH SERIES.

JANUARY—JUNE 1881.

LII. *On the Physical Units of Nature.* *By* G. JOHNSTONE STONEY, *D.Sc., F.R.S., Vice-President of the Royal Dublin Society**.

1. WHEN mathematicians apply the sciences of measurement to the investigation of Nature, they find it convenient to select such units of the several kinds of quantity with which they have to deal as will get rid of any coefficients in their equations which it is possible in this way to avoid. Every advance in our knowledge of Nature enables us to see more distinctly that it would contribute to our further progress if we could effect this simplification, not only with reference to certain classes of phenomena, but throughout the whole domain of Nature.

* From the 'Scientific Proceedings' of the Royal Dublin Society of February 16, 1881, being a paper which had been read before Section A of the British Association, at the Belfast Meeting in 1874. Communicated by the Author.

Fig. 8. The title page of "Philosophical Magazine" and the beginning of article [10] by G.J. Stoney

Stoney's units look "tailored" for Einstein's unified theory. Constants e, c, G contain the gist of classical electrodynamics and gravity. There is no $\hbar$ in them. Comparison with c, $\hbar$, G shows that $\hbar$ is brought into Stoney's set of constants "through the backdoor of α". Therefore e, c, G do not form a cube of theories with its limiting transitions considered by Gamov, Ivanenko and Landau [2].

7 Atomic Clocks and c

During the 20th century the situation with standards of mass (time, length) has changed drastically. The fundamental identity of elementary particles and hence of atoms produced many candidates for a standard of mass, known with much, much better precision than G. Thus, from the point of view of dimensions, the necessity to use G disappeared. However from the point of view of unifying physics the Planck units became even more attractive.

Let us now look at two other fundamental constants: c and h.

Let us start from c and the frequencies of light and radio waves. In the second half of the 20th century physicists learned how to measure them in a digital way by counting the number of crests. This raised the accuracy of atomic (Cesium-133) clocks (first suggested by I. Rabi in 1945) to the level of 1 second in 300 years (NBS, 1955). (Now this has become 1 second in $20 \cdot 10^6$ years: LPTF, NIST, PTB.) But even the first figure was sufficient for the introduction into SI of an atomic unit of a second (in 1967):

"1 s = 9 192 631 770 periods of radiation in the transition between levels of hyperfine splitting of the atomic ground state of Cs-133".

This, together with the independence of the velocity of light on its frequency, impelled Bay et al. [11] to suggest, instead of the unit of length (meter), to use as the basic unit the unit of velocity, namely the velocity of light c. In 1983 the definition

$$c = 299\ 792\ 458 \text{ m/s} \tag{4}$$

was introduced into the SI. The traditional standard of length gave way to the new standard based on the value of the velocity c. This velocity is defined as a number without uncertainty. Further improvements of experiments which measured c would mean further improvement of the realization of the meter. An international report "Practical realization of the definition of the metre, including recommended radiations of other optical frequency standards" (2001) was published by T. Quinn in 2003 [12]. (Note that both spellings "metre" and "meter" are used in the literature, the former in metrology, while the latter one in physics.)

Further progress in the accuracy of atomic clocks is connected with passing from microwave to optical frequencies [13,14].

8 Towards a Kilogram Based on h

Thus metrology made two momentous steps in the direction of fundamental physics: the place of macroscopic clocks and ruler (the famous rod at BIPM, in Sèvres, near Paris) became occupied by the velocity of light and by atoms of Cs-133. There remains now only one macroscopic standard – the kilogram at Sèvres. The prospect of expressing it through the quantum of action h is connected with precision measurements in atomic and condensed matter physics. There are many promising quantities which are good candidates for such measurements. I shall touch upon only one project which is connected with two outstanding discoveries

in condensed matter physics: the Josephson effect [15] (Nobel Prize 1973) and the von Klitzing effect [16] (Nobel Prize 1985).

Josephson theoretically predicted the existence of a supercurrent and its remarkable properties. A supercurrent is a current of Cooper pairs tunneling through an insulator separating two superconductors. A supercurrent can exist without external voltage. An external voltage V creates an alternating supercurrent of frequency ν. The steps in V are given by the relation:

$$V(n) = \nu n / K_J \ , \tag{5}$$

where n is an integer, while the coefficient K_J is universal and is called the Josephson constant. It is reproduced in various experiments with unprecedented accuracy and is determined only by the ratio of fundamental constants:

$$K_J = 2e/h \ . \tag{6}$$

The effect, discovered by von Klitzing, is called the quantum Hall effect. This effect shows that there exists in Nature a universal electric resistance, one which can be expressed in terms of fundamental constants.

As is well known, the ordinary Hall effect occurs in a solid conductor (or semiconductor) with density of current **j** in a magnetic field **H** which produce an electric field **E** (with voltage V_H) orthogonal both to **j** and **H**.

The quantum Hall effect was discovered in a two-dimensional electron system separating two parts of a silicon field transistor at very low temperature (< 4 K) and very strong magnetic field ($\sim$ 14 Tesla). It was established that the Hall resistance

$$R_H = V_H / I \ , \tag{7}$$

where I is the total current, has quantum jumps:

$$R_H(n) = \frac{R_K}{n} \ , \tag{8}$$

where n is an integer, while R_K is the von Klitzing constant:

$$R_K = h/e^2 \ . \tag{9}$$

It is obvious that

$$h = 4K_J^{-2} R_K^{-1} \ . \tag{10}$$

This permits the measurement of h using a macroscopic apparatus. A special two-story-high watt balance compared electrical and mechanical forces:

$$VI/v = mg \ , \tag{11}$$

where m is the measured mass of a body, g the local gravitational acceleration, V the voltage in a coil moving with a vertical velocity v in a magnetic field, while I is the current in the same coil, this time fixed in the same magnetic field. By calibrating V and V/I through the Josephson and von Klitzing effects Williams

et al. [17] succeeded in connecting h and the kilogram within an uncertainty of $8.7 \cdot 10^{-8}$.

It is hoped that in the not too distant future this accuracy might be improved by an order of magnitude, which would allow to use the watt balance for gauging the standards of mass and thus get rid of the Sèvres kilogram and to define the value of h. As a result the value of h would have no uncertainties in the same way as it occurred with c. Thus fundamental units of nature c and h would become fundamental SI units of metrology.

9 Kilogram as Frequency ν_K

Another definition of the kilogram has been suggested [18] on the basis of equations

$$E = h\nu \ , \tag{12}$$

$$E = mc^2 \ : \tag{13}$$

"The kilogram is the mass of a body at rest whose equivalent energy equals the energy of collection of photons whose frequencies sum to $13.5639274 \times 10^{49}$ hertz".

This definition should be taken with a grain of salt. The combined use of (12) and (13) implies that a photon of frequency ν has mass $h\nu/c^2$. This implication persists in spite of the words "equivalent energy". The words "the mass of the body at rest" imply that mass is not Lorentz invariant, but depends on the velocity of a reference frame. It would be proper to replace (13) by

$$E_0 = mc^2 \ , \tag{14}$$

where E_0 is the rest energy (see e.g. [19]). But then it would take some additional considerations in order to define the frequency ν_K corresponding to one kilogram. In particular, massive atoms emitting and absorbing photons should be taken into account. From a practical point of view the measurement of "frequencies sum" of order 10^{50} hertz is by eight orders of magnitude more difficult than that of the Planck frequency $\nu_P = 1/t_P$.

10 Electromagnetism and Relativity

Electromagnetism – the kinship of electricity and magnetism – discovered in 1820 by Oersted, rather soon became the foundation of Ampére's electrodynamics. The development of the latter by Faraday and other outstanding physicists culminated in 1873 in the Treatise of Maxwell [20] who linked electric currents with electric and magnetic fields and with the properties of light. None of these great physicists knew the genuine nature of the phenomena. Maxwell considered a vacuum filled with ether; the carriers of charges were unknown to him. The electromagnetic field was described by four vector quantities: electric field **E**, electric induction (or displacement) **D**, magnetic field **H**, and magnetic induction (or flux density) **B**.

On the basis of these notions practical units (such as volt, ampere, coulomb, joule) were introduced by International Electrical Congresses in the 1880s. The electric permittivity ε_0 and magnetic permeability μ_0 ascribed by Maxwell to the ether were accepted by the community of engineers and physicists: $\mathbf{D} = \varepsilon_0 \mathbf{E}$, $\mathbf{B} = \mu_0 \mathbf{H}$. In the middle of the 20th century these practical units became the basis of the Système International d'Unités (SI).

The end of the 19th and beginning of the 20th century were marked by great successes in understanding and applying classical electrodynamics. On the practical side it was the use of electric currents in industry, transport and radio communications. On the theoretical side it was the unification of electrodynamics, optics and mechanics in the framework of special relativity [21].

According to special relativity, the position four-vector is $x^i = (ct, \mathbf{r})$ $(i = 0, 1, 2, 3)$, the momentum four-vector is $p^i = (E/c, \mathbf{p})$, the four-potential of electromagnetic field $A^i = (\varphi, \mathbf{A})$, the density of the four-current $j^i = (c\rho, \mathbf{j})$, where $\mathbf{j} = \rho \mathbf{v}$, and $\rho = e\delta(\mathbf{r} - \mathbf{r}_a)$, e is the electric charge. (The current j^i is consistent with the definitions of p^i and A^i, due to an appropriate coefficient c in front of ρ. The source of the field, the charge, is pointlike. Otherwise there appears a problem of the field inside the finite-size cloud of charge.) The upper index i of a four-vector indicates a contravariant four-vector; a lower index i indicates a covariant four-vector, its space components have a minus sign. Raising or lowering of indices is done with the diagonal metric tensors g^{ik} or g_{ik} respectively.

The three-vectors $\mathbf{E}$ and $\mathbf{H}$ are components of the four-tensor of the electromagnetic field

$$F_{ik} = \frac{\partial A_k}{\partial x^i} - \frac{\partial A_i}{\partial x^k} \ . \tag{15}$$

The tensors F_{ik} and F^{ik} can be represented by matrices:

$$F_{ik} = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & -H_3 & H_2 \\ -E_2 & H_3 & 0 & -H_1 \\ -E_3 & -H_2 & H_1 & 0 \end{pmatrix} \ , \tag{16}$$

and

$$F^{ik} = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & -H_3 & H_2 \\ E_2 & H_3 & 0 & -H_1 \\ E_3 & -H_2 & H_1 & 0 \end{pmatrix} \ , \tag{17}$$

respectively, or in a condensed form:

$$F_{ik} = (\mathbf{E}, \mathbf{H}) \ , \tag{18}$$

$$F^{ik} = (-\mathbf{E}, \mathbf{H}) \ . \tag{19}$$

This four-tensor is obviously antisymmetric. From the definition of F_{ik} it follows that the dimensions of $\mathbf{E}$ and $\mathbf{H}$ are the same: $[\mathbf{E}] = [\mathbf{H}]$.

The field equations have the form in Gaussian units:

$$\frac{\tilde{F}^{ik}}{\partial x^k} = 0 \ , \tag{20}$$

$$\frac{\partial F^{ik}}{\partial x^k} = -\frac{4\pi}{c} j^i \quad . \tag{21}$$

Here

$$\tilde{F}^{ik} = \frac{1}{2}\varepsilon^{iklm} F_{lm} \quad , \tag{22}$$

where ε^{iklm} is fully antisymmetric tensor ($\varepsilon^{0123} = +1$).

The equation describing the motion of a charge in the electromagnetic field is given by

$$\frac{d\mathbf{p}}{dt} = e\mathbf{E} + \frac{e}{c}[\mathbf{vH}] \quad , \tag{23}$$

where

$$\mathbf{v} = \frac{\mathbf{p}c^2}{E} \quad . \tag{24}$$

Note that according to special relativity there is no ether, $\varepsilon_0 \equiv \mu_0 \equiv 1$, and the strength of magnetic field in vacuum $\mathbf{H}$ has the same dimension as that of $\mathbf{E}$; the identity of $\varepsilon_0 \equiv \mu_0 \equiv 1$ immediately follows from the fact that the same e determines the action of the charge on the field and of the field on the charge. (See expression for the action in [22], (27.6).) Thus, there is no need to consider $\mathbf{B}$ and $\mathbf{D}$ in the case of vacuum. In classical electrodynamics they appear only in the continuous media due to polarization of the latter [23].

In a number of classical monographs and textbooks on classical electrodynamics $\mathbf{E}$ and $\mathbf{H}$ are consistently used for the description of electric and magnetic fields in vacuum with $\varepsilon_0 \equiv \mu_0 \equiv 1$ ([21,22,24–26]). Their authors use Gaussian or Heaviside-Lorentz (with $1/4\pi$ in the Coulomb law) units.

Many other authors use $\mathbf{B}$ instead of $\mathbf{H}$, sometimes calling $\mathbf{B}$ magnetic field and sometimes magnetic induction in vacuum [27]. Most of them use the SI units, according to which ε_0 and μ_0 are dimensional: $\mu_0 = 4\pi \cdot 10^{-7}\mathrm{HA}^{-2}$, $\varepsilon_0\mu_0 = c^2$, where H is henry, while A is the ampere. The classical electromagnetic fields in vacuum are described by four physical quantities $\mathbf{D}$, $\mathbf{H}$ and $\mathbf{E}$, $\mathbf{B}$, all four of them having different dimensions at variance with the spirit of special relativity.[1] In that respect, the vacuum is similar to a material body. The SI units are very convenient for engineers, but not for theorists in particle physics.

In fact, theorists are not less responsible than metrologists for the gap between the deductive basis of modern physics and the mainly prerelativistic inductive basis of modern metrology. A good example is the 1935 article [28] by A. Sommerfeld and his book "Electrodynamics" based on lectures given in 1933-34 [29].

His argument against an absolute system (that is based on units of time, length and mass) was the presence of fractional exponents (for instance from the Coulomb law the dimension of charge is $\mathrm{g}^{1/2}\ \mathrm{cm}^{3/2}\mathrm{s}^{-1}$). This argument was not very compelling in the 1930s and is even less so today. His argument against the

[1] Sometimes one can hear that the identity $\varepsilon_0 \equiv \mu_0 \equiv 1$ is similar to putting $c = 1$, when using c as a unit of velocity. However this similarity is superficial. In the framework of special relativity one can use any unit for velocity (for instance, m/s). But the dimensions and values of ε_0 and μ_0 are fixed in SI.

Gaussian or Heaviside-Lorentz system was based on an inductive, prerelativistic view on electromagnetism. Though he was not quite happy[2] with the new clumsy expression for the fine structure constant α introduced by him during World War I, he kept insisting on MKSA units and against Gaussian units. His authority was not the least in the decision to legally enforce after World War II the SI as the obligatory system of units for all textbooks in physics.

Coming back to classical electrodynamics, let us note that it is not a perfect theory: it has serious problems at short distances. To a large part these problems are solved by quantum electrodynamics (QED). Therefore the latter should be used as a foundation of a system of electromagnetic units. By the way, QED is used to extract the most accurate value of α from the precision measurements of the magnetic moment of the electron.

In the framework of QED, α is not a constant but a function of momentum transfer due to the polarization of vacuum. Let us stress that this polarization has nothing to do with purely classical non-unit values of ε_0 and μ_0.

11 Concluding Remarks

The mutually fruitful "crossing" of fundamental physics and metrology gives numerous practical applications. One of them should be specially mentioned: the use of general relativity in global positioning systems [30,31].

Remarkable achievements of metrology are not always accompanied by elaboration of adequate terminology. Here we will mention only a few of wide-spread delusions.

The choice of c as a unit of velocity leads many authors to the false conclusion that c should be excluded from the set of fundamental units. They insist that $c = 1$, because c in units of c is equal to 1. (The same refers to h in units of h.) But the number 1 is not a unit of measurement, because such units are always dimensional. Equations $c, h = 1$ are simply wide spread jargon. Some authors go even further by identifying space and time. (A detailed discussion can be found in [32].)

The number of physical units is not limited. When solving a given problem the choice of units is determined by considerations of convenience. However, from the point of view of "the world as a whole" c, h and G (or instead of G some other quantity representing gravity) are definitely singled out as fundamental dimensional constants. Of course they must be accompanied by a number of dimensionless parameters. But the number of fundamental units could not be less than three [32].

The inclusion of the candela into the set of base units (see Fig. 9) seems to be unconvincing from the point of view of physics. Of course, practically it is convenient to use it when discussing the brightness of light. But it does not look logical to put it on the same footing as units of length, time and mass.

[2] "What is especially painful for me is that the fine structure constant is no more $e^2/\hbar c$, but $e^2/4\pi\varepsilon_0\hbar c$". Z. Phys. **36** (1935) 818.

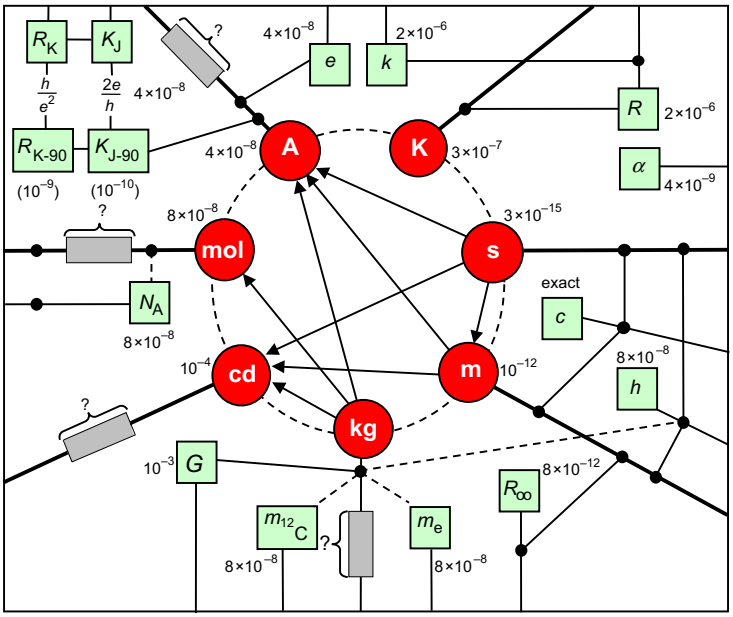


Fig. 9. The base units of the SI, with their present uncertainties of realization, and some of their links to atomic and fundamental constants with their present uncertainties in terms of the SI. The absence of a useful quantitative estimate of the long-term stability of the kilogram, indicated by "?", is reflected in three of the other base units. The dashed lines to the kilogram indicate possible routes to a new definition. (See the article by T.J. Quinn "Base Units of the Système International d'Unités, their Accuracy, Dissemination and International Traceability", Metrologia **31** (1994/95) 515-527.)

As SI is imposed on the physics literature by governmental laws, the obligatory usage in textbooks of such notions as permittivity ε_0 and permeability μ_0 of vacuum, makes it difficult to appreciate the beauty of the modern electrodynamics and field theory. It corresponds to the prerelativistic stage of physics.

This list can be extended, but it seems that the above remarks are sufficient for a serious discussion. The metrological institutes and SI are of great importance for science and technology. Therefore the metrological legal documents should be to a greater degree based on modern physical concepts. Especially they should give more freedom to the usage of Gaussian and Heaviside-Lorentz systems of units in the textbooks.

Acknowledgements

I would like to thank A. Clairon, H. Fritzsch, J. Jackson, S. Karshenboim, N. Koshelyaevsky, H. Leutwyler, E. Peik, N. Sanchez, M. Tatarenko, V. Telegdi, Th. Udem, L. Vitushkin and H. Wagner for fruitful discussions. The work has been partly supported by the Russian Federal Special Scientific-Technological

Program of Nuclear Physics Fund 40.052.1.1.1112 and by A. von Humboldt award.

References

1. M. Planck: Sitz.-Ber. Preuss. Akad. Wiss. (1899) 449;
   M. Planck: Ann. d. Phys. **4**, 1, 69 (1900). Reprinted in Max Planck. Physikalische Abhandlungen und Vorträge, Band I, Fried. Vieweg. 1958, pp. 500-600, 614-667.
2. G. Gamov, D. Ivanenko, L. Landau: Journal of Russian Physicochemical Society, Ser. Phys. **LX**, 13 (1928) (in Russian), reprinted in Yad. Fiz. **65**, 1406 (2002), translated into English in Physics of Atomic Nuclei **65**, 1373 (2002).
3. M. Bronstein: Phys. Zeitschrift der Sowjetunion **9**, 140 (1936).
4. A. Zelmanov: 'Kosmologia'. In *Razvitie Astronomii v SSSR* (Nauka, Moskva, 1967) p. 323 (in Russian).
5. G. Gorelik: *Razmernost' prostranstva* (MGU, Moskva, 1983) Chapter 5 (in Russian).
6. L. Okun: Sov. Phys. Usp. **34**, 818 (1991).
7. K. Kuchar: Phys. Rev. D **22**, 1285 (1980).
8. N. Sanchez: 'Recent Progress in String Cosmology'. *International school of astrophysics "D. Chalonge", 5th Course "Current topics in astrofundamental physics", Erice, September 1996.*
9. J. Flowers and B. Petley: In: *Astrophysics, Clocks and Fundamental Constants*, ed. by S. G. Karshenboim and E. Peik, Lecture Notes in Physics Vol. 648 (Springer, Berlin, Heidelberg 2004), pp. 75–93.
10. G.J. Stoney: The Philosophical Magazine and Journal of Science **11**, 381 (1881).
11. Z. Bay, G.G. Luther, J.A. White: Phys. Rev. Lett. **29**, 189 (1972);
    Z. Bay, J.A. White: Phys. Rev. D **5**, 796 (1972).
12. T.J. Quinn: Metrologia **40**, 103 (2003).
13. Th. Udem, R. Holzwarth, T.W. Hänsch: Nature **416**, 233 (14 March 2002).
14. S.G. Karshenboim, F.S. Pavone, F. Bassani, M. Inguscio, T.W. Hänsch (Eds.): *The Hydrogen Atom. Precision Physics of Simple Atomic Systems* (Springer, 2001).
15. B.D. Josephson: Phys. Lett. **1**, 251 (1962).
16. K.V. Klitzing, G. Dorda, M. Pepper: Phys. Rev. Lett. **45**, 494 (1980).
17. E.R. Williams, R.L. Steiner, D.B. Newell, P.T. Olsen: Phys. Rev. Lett. **81**, 2404 (1998).
18. R.N. Taylor, P.J. Mohr: Metrologia **36**, 63 (1999).
19. L.B. Okun: Physics Today, June 1989, 31; May 1990, 13, 15, 115-117; Uspekhi Fiz. Nauk **158**, 511 (1989); Sov. Phys. Usp. **32**, 629 (1989).
20. J.C. Maxwell: *A Treatise on Electricity and Magnetism.* (Oxford Univ. Press, 1873) Part IV, Chapters IX, X, XIX. 3rd edn. (1891); reprint Dover, New York (1954)
21. A. Einstein: *The meaning of relativity*, 4th edn. (Princeton, 1953)
22. L.D. Landau and E.M. Lifshitz: *The Classical Theory of Fields*, 4th edn. (Pergamon Press, Oxford, 1987)
23. L.D. Landau and E.M. Lifshitz: *Electrodynamics of continuous Media*, 2nd edn. (Addison-Wesley, Reading, MA, 1984)
24. C.Möller: *The Theory of Relativity*, 2nd edn. (Clarendon Press, Oxford, 1974)
25. P.G. Bergmann: *Introduction to the Theory of Relativity* (Dover reprint, 1976)
26. W. Pauli: *Theory of Relativity* (Pergamon Press, New York, 1958)
27. J.D. Jackson: *Classical Electrodynamics*, 3rd edn. (John Wiley and sons, Inc. 1998)

28. A. Sommerfeld: Phys. Z. **36**, 814 (1935)
29. A. Sommerfeld: *Electrodynamik* (Leipzig, 1949. Academic Press, New York, 1952)
30. N. Ashby, D. Allen: Radio Science **14**, 649 (1979).
31. B. Guinot: Metrologia **34**, 261 (1997).
32. M.J. Duff, L.B. Okun, G. Veneziano: JHEP **03**, 023 (2002), arXiv, physics/0110060.

Constants, Units and Standards

Jeff Flowers and Brian Petley

National Physical Laboratory, Teddington, UK

Abstract. Our knowledge of the Universe depends on our ability to make measurements. Measurement is essentially a comparison process. In order to make comparisons valid for different positions in space and time a system of invariant units is required. The fundamental constants are a natural system of units in physics and have increasingly been used in practice to provide what is assumed to be an invariant, practical system of standards and units that is sufficiently constant for present-day technology.

Measurement of a fundamental constant has often been limited by our ability to realise the appropriate unit. The situation then has been inverted and the relevant fundamental constant has been used in practice to define or maintain the unit.

We consider measurement as a comparison, the fundamental constants as units, whether the constants are constant, the practical realizations of the SI units, their interaction with the fundamental constants and present accuracy.[1]

1 Introduction

Although we can learn much about the nature of the universe from contemplation we soon find, as did the ancient Greeks, that we can learn even more by recourse to measurement. Measurement is essentially a comparison process whereby a quantity Q_1 is compared with a like quantity Q_2 to yield a numerical result n so that

$$Q_1 = nQ_2. \tag{1}$$

We often make our measurements in terms of internationally agreed standard quantities, which are known as units. Thus, the value of a quantity is the product of a number and an internationally agreed unit.

$$\text{Quantity} = \{\text{numerical value}\} \cdot [\text{unit}] \tag{2}$$

so that

$$Q_1 = \{Q_1\} \cdot [Q_1] \text{ and } Q_2 = \{Q_2\} \cdot [Q_2]. \tag{3}$$

A whole series of national comparison and international comparison measurements are often required before the quantities can be expressed in identical units to ensure that

$$[Q_1] = [Q_2]\,, \tag{4}$$

[1]

J. Flowers, B. Petley, Constants, Units and Standards, Lect. Notes Phys. **648**, 75–93 (2004)
http://www.springerlink.com/

and hence that the ratio of the two quantities may be measured as the ratio of their two numerical values, as

$$\frac{Q_1}{Q_2} = \frac{\{Q_1\}}{\{Q_2\}} = n. \tag{5}$$

In some cases there is a further topic involved and that is whether physics at the two locations, characterised by P and P' respectively, is the same. That is we really are testing whether

$$\left(\frac{Q_1}{Q_2}\right)_P = \left(\frac{Q_1}{Q_2}\right)_{P'}. \tag{6}$$

One cannot make two measurements at the same point in spacetime and so it must be established that the physics and the quantities are both invariant at the level of precision required.

1.1 Early Measurements

For most of the earlier measurements it was reasonable to assume that, by suitable control of the measurement environment, one could achieve the equality of physics and quantities; where necessary by applying suitable corrections to ensure that this was so.

The increased accuracy of modern measurements has made it necessary to take more sophisticated phenomena into account than hitherto to achieve this identity. Thus the modern accuracy achieved is such that measurements made with atomic clocks, that are at different altitudes and temperatures, must take account both of gravitational red-shifts and also of black body radiation shifts in their frequencies [1].

Where one is comparing the physics at one epoch with that of the present-day terrestrial environment, or radiation from a distant astronomical source, one must similarly be careful to transfer the measurements from the environment at the source to those pertaining in empty space of that epoch – because even the properties of the local vacuum at the source may be modified in some way by the presence of the source.

For terrestrial measurements, with the passage of time and increased sophistication of measurements, the comparison of Q and Q' has been made successively through comparisons with standard representations of these quantities that are maintained at a local, national, and international level. At the same time the greater attention has been given to achieving ever more accurate definitions and realizations of the basic minimum set of these standard quantities in terms of base and derived 'units'.

1.2 The Fundamental Physical Constants

Measurements in science and technology are usually based on an internationally agreed set of units involving combinations of a basic set of standard quantities.

Table 1. The group conventionally known as the fundamental physical constants

Fundamental Quantity	Symbol
Speed of electromagnetic radiation in free space	c
Elementary charge (of proton)	e
Mass of electron at rest	m_e
Mass of proton at rest	m_p
Planck constant	h
Avogadro constant	N_{A}
Newtonian constant of gravitation	G
Boltzmann constant	k
Magnetic and electric constants	μ_0, ε_0

The number of such base standard quantities required depends on the internationally agreed system of units. Our present international system, the SI, is based on seven base quantities. The original concept of the centimetre, gram, second (CGS) system, was of a three base unit system, the SI expanded the number of base units to cover electrical, thermodynamic chemical and optical measurement. The number of base units required is a matter of convention and practical need.

Theoretical physics is expressed in terms of quantity equations rather than units. Some quantities that are essentially invariant occur so frequently and in so many different contexts in such equations, that they are termed fundamental physical constants. The conventional core group of quantities known as the fundamental physical constants comprises those given in Table 1. They have been recognised as such for about the last one hundred years, some, such as the speed of light and Newtonian constant of gravitation for a considerably longer period.

As soon as one uses the term fundamental physical constant one is immediately faced with the question: are they constant? Although we return to this later, there is one immediate sense in which they are not constant, and that is their role in physics, which often evolves from one classification to another as a result of discovery. The broad classifications of the fundamental constants due to Lévy-Leblond [2], are summarised in Table 2. Thus, the speed of light has evolved from applying to light (Type A), through to being a constant of electromagnetism (Type B) and, with the advent of special and general relativity to a universal constant (Type C). This process continues still, for example there are ongoing measurements concerning whether the mass of the neutrino is a Type E constant or a Type A constant.

In addition to the fundamental physical constants there are also certain important combinations of these. Two very important combinations are the fine–structure constant α, given by

$$\alpha = \frac{\mu_0 c e^2}{2h} \,, \tag{7}$$

Table 2. The Lévy-Leblond classification of the constants

Class of constant	Characteristics of Class
Type A	Properties of particular physical objects, e.g. masses and moments of fundamental particles
Type B	Characteristics of classes of physical phenomena, e.g. coupling constants
Type C	Universal constants, e.g. speed of light, Planck constant
Type D	Invisible constants, e.g. isotropy of space, equivalence of inertial and gravitational mass
Type E	Constants indistinguishable from zero, e.g. mass of photon, neutrality of matter

and the Rydberg constant R_∞, given by

$$R_\infty = \frac{m_e c \alpha^2}{2h} \,. \tag{8}$$

There are many other important combinations that occur sufficiently frequently to be designated by separate symbols and given in tables of values of the fundamental constants, for example, the Compton wavelengths of fundamental particles, the Bohr and nuclear magnetons.

2 Units and Standards

For most of the history of measurement these standard representations have been based on physical artifacts, such as the prototype metre bar, or the prototype kilogram. Gradually these are being replaced by atomic or quantum standards which are thought to be really invariant, or at least, are sufficiently invariant for the present-day measurement accuracies. Unfortunately, as Einstein and others recognised, all clocks, rulers, *etc.* that we may employ are of necessity inside the universe and hence are likely to be affected by it. This must be taken more carefully into account as we move to greater levels of measurement accuracy.

It is important to recognise that, although we are moving towards a system of units based on atomic or quantum phenomena, it has long been clear that there is unlikely to be a unique set of such units. Thus, physicists and cosmologists frequently make use of Planck units or Hartree (atomic) units. Such natural units were pioneered by Stoney [3] and Planck [4]. Units must be suitable for use throughout the whole range of scientific and technological measurement.

Although the above units suffice for atomic phenomena they are less suitable for nuclear phenomena or those of an astronomical scale. We must recognise that the 'ultimate' set of natural units, the set suitable for all purposes, may yet be discovered and agreed. Meanwhile we have different systems for different purposes and metrologists are primarily concerned about serving the scientific

and technological community with a set of reference quantities that are of adequate for present day scientific and technological needs. Achieving practicality transcends the aesthetics of achieving the ultimate system.

2.1 Use of the Fundamental Constants to Form Systems of Units

As far as theoretical physics is concerned, some constants occur so frequently that they are omitted from equations by setting their values at unity, thereby giving them the function of units. There are at least four schemes in common use in different parts of physics, as shown in Table 3.

Table 3. Some common unit systems based on fundamental constants

Scheme	System defining equation
Stoney units	$\{c\} = \{e\} = \{G\} = \{4\pi\varepsilon_0\} = 1$
Hartree units	$\{\hbar\} = \{e\} = \{m_\mathrm{e}\} = \{4\pi\varepsilon_0\} = 1$
Plank units	$\{\hbar\} = \{c\} = \{G\} = 1\,, \{4\pi\varepsilon_0\} = 1$ or $\mu_0 = 1$
Schrödinger units	$\{\hbar\} = \{e\} = \{G\} = 1\,, \{4\pi\varepsilon_0\} = 1$ or $\mu_0 = 1$

2.2 Are the Constants Really Constant?

Any physical measurement apparatus is inside the universe and hence is affected by it. It is therefore difficult to find something that is more constant to compare the fundamental constants against. It is also difficult to avoid obtaining a null result from such comparisons. Thus, in the present-day SI a fixed value of the speed of light is built into the present definition of the metre – so one can no longer meaningfully measure the speed of light in SI units. The general conclusion from this is that only dimensionless constants can be measured independently of the unit system, but this is a necessary, but not sufficient condition. Examples of such dimensionless quantities are: $\alpha\,,\ m_\mathrm{p}/m_\mathrm{e}\,,\ g_\mathrm{p}$ and $\alpha_\mathrm{G}(= Gm_\mathrm{p}^2/\hbar c)$.

In the system of units such that

$$\{\hbar\} = \{e\} = \{m_\mathrm{e}\} = \{4\pi\varepsilon_0\} = 1\ ,$$

we find that the unit of length is the Bohr radius and that $\{c\} = \{1/\alpha\}$. Consequently, if α varies with time c will also vary in a similar way in this system of units. Evidently, by using this system of units, one is assuming that h, e, m_e, and ε_0 are more likely to be constant than are either c or α.

2.3 The CODATA Evaluations

CODATA is an organisation set up by the Council of International Scientific Unions to co-ordinate the data for science and technology. One key aspect of its

work lies in the evaluations of the best values to adopt for the fundamental physical constants by the CODATA Task Group on the Fundamental Constants of Physics and Chemistry. Although the evaluations are authored by two individuals, presently Peter Mohr and Barry Taylor, the Task Group comprises experts in the field who can advise them on the integrity of the variety of different theories and experimental measurements involved [5].

This work succeeded the pioneering evaluations by Raymond T Birge[2], Jessie DuMond, Richard Cohen and others between 1929 and 1973 when the CODATA Task Group took over.

2.4 Changing Accuracy

There is a sense in which the measured values change with time and that is in the uncertainty in their values. This is steadily decreasing with time as a result of more sophisticated methods and improved apparatus. The uncertainties in the values of some of the 'constants' with successive evaluations is shown in Fig. 1.

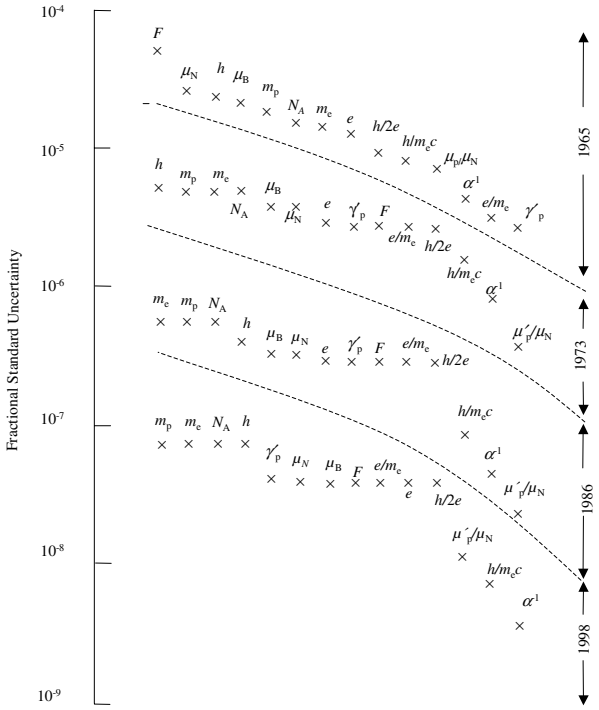


Fig. 1. Showing how the fractional accuracy of a group of fundamental constants improved with each evaluation between 1965 and 1998

[2] The first evaluation by R T Birge was appropriately the first article of the first volume of the Physical Review Supplement that became known as the journal: 'Reviews of Modern Physics'.

One consequence of this improvement is that the associated value changes. On the average the change has been about two standard deviations. In principle such an average change is rather improbable, but it is a reflection that the amount of unknown physics is of comparable size to the measurement accuracy.

If we look at two particular examples, the fine structure constant and the elementary charge, we see that the improved accuracy has followed a reasonably straight line on a logarithmic scale giving about tenfold improvement every fifteen years or so as shown in Fig. 2.

If we look at the improvement in the accuracy of the realization of the ohm and ampere respectively we see that these have been overtaken by a higher rate of improvement in the corresponding measurements of α and e. In the same way the accuracy of the speed of light measurements caught up with the realization of the metre in around 1972 and this led to a redefinition of the metre to the present-day one. The latter incorporates a fixed value for the speed of light. The question of what to do about the ohm and ampere will be discussed in Sect. 6.1.

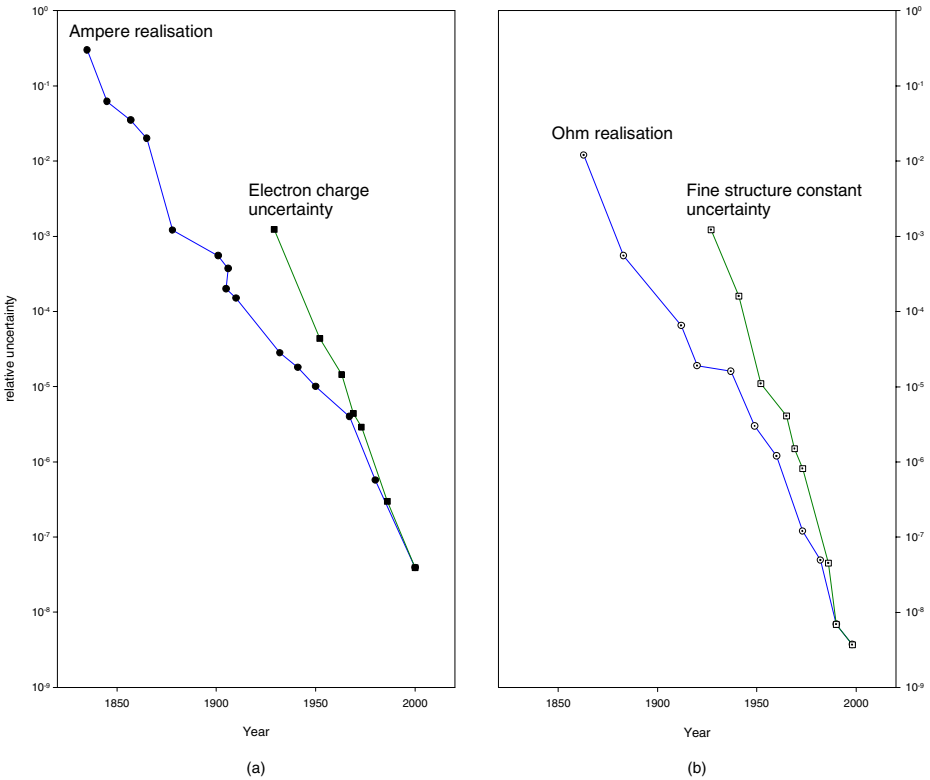


Fig. 2. Illustrating how the accuracy of (a) the ampere realization and (b) the ohm realization have been caught up by the elementary charge and fine structure constant respectively since the evaluation by R T Birge in 1929

Table 4. Present-day accuracy of realization of the SI base units

Quantity	Unit	Approximate relative accuracy	Associated fundamental constant
Length	metre	1×10^{-12}	c, R_∞
Mass	kilogram	1×10^{-8}	m_u, eV/c^2
Time	second	1×10^{-15}	$\Delta\nu(^{133}\mathrm{Cs})$
Electric current	ampere	1×10^{-8}	μ_0, K_J
Thermodynamic temperature	kelvin	1×10^{-7}	k, R
Luminous intensity	candela	1×10^{-4}	k
Amount of substance	mole	1×10^{-8}	N_A

3 Accuracy of Realization of the SI Units

The present accuracy of realization of the SI base units is shown in Table 4. It is evident that there is a considerable variation in accuracy spanning some ten orders of magnitude. This variation is a reflection of the variation of the measurement capability in different areas of technology. Time and frequency measurements may be made with outstanding superior accuracy. This provides a role for the fundamental constants for, in an increasing number of cases, they are involved explicitly in the relationship between frequency and the quantity that is required. Thus if we know the frequency of a laser we know its wavelength (c, or R_∞), if we measure a voltage via the Josephson effect we involve a frequency (h/e), or a flux density by nuclear magnetic resonance, we involve ($2\mu_\mathrm{p}/h$), and so on.

4 Practical Realizations of the SI Units and the Involvement of Fundamental Constants

We can usually maintain a unit more stably than we can realize it absolutely. This is a desirable aspect of metrology since we must otherwise make frequent absolute realizations of the unit. Artifact standards usually drift slowly with time and so any method of maintaining a unit using a combination of fundamental constants is intrinsically more stable. Aside from time and frequency measurements, even if the fundamental constants drift slowly on a cosmological timescale the variation is negligible, or if they do vary such variation is likely to be sufficiently predictable for present-day applications. There are, of course, already many examples of manifestations of the effects of quantum fluctuations and Johnson noise that must be allowed for.

A more stable standard is an almost essential precursor to an improved method of realizing a unit. Voltage measurements are almost as essential to

modern measurement methods as frequency measurements and so we begin by looking at the Josephson effects that have so revolutionised the accuracy with which a stable voltage unit may be maintained in standards laboratories throughout the world.

Measurements of many of the fundamental constants are essential to modern metrology and we illustrate some of these below.

4.1 The Josephson Effect Voltage Standard

A superconducting tunnelling current between two close (∼1 nm) superconductors irradiated by a microwave frequency, ν, produces a potential difference V_{J} between them, where

$$2eV_{\mathrm{J}} = nh\nu \; . \tag{9}$$

The quantity $2e/h$ has been assigned the symbol K_{J} and is known as the Josephson constant.

The agreement between Josephson effect voltage standards constructed in different laboratories is about 3 parts in 10^9. A null test between two junctions subject to the same frequency has been set at 2 parts in 10^{16}.

4.2 The Quantised Hall Resistance

In 1980 Klaus von Klitzing demonstrated [6] that the Hall effect resistance of a silicon MOSFET device showed a magnetic field independent plateau when at cryogenic temperatures. The plateau are given by

$$R = \frac{h}{ie^2} = \frac{R_{\mathrm{K}}}{i} \; , \tag{10}$$

where i is an integer. This quantum Hall effect is an independent route to experimentally determine the fine structure constant. The quantity h/e^2 has been assigned the symbol R_{K}, known as the von Klitzing constant, and

$$R_{\mathrm{K}} = \frac{h}{e^2} = \frac{\mu_0 c}{2\alpha}. \tag{11}$$

It is usually necessary to extrapolate the measurements to 0 K for this identity to apply. The equality between two such MOSFET devices has been tested to a few parts in 10^9.

4.3 The Calculable Capacitor

The realizations of the ohm represent the involvement of a second fundamental constant, the impedance of space $\mu_0 c$. The related condition

$$\varepsilon_0 \mu_0 = c^2 \tag{12}$$

must also be also satisfied.

Both of these conditions are satisfied in the SI by making them intrinsic to the definition of the ampere. The theorem in electrostatics that is utilised

to realize the ohm necessarily incorporates this. This theorem embodied in the calculable capacitor was discovered by Thompson and Lampard in 1956 [7]. They found that the cross-capacitance per unit length between four infinite parallel conducting cylinders was given by

$$\frac{\mathrm{d}C}{\mathrm{d}z} = \left(\frac{\varepsilon_0}{\pi}\right) \ln 2 = 1.953\ldots \,\mathrm{pF\,m^{-1}} \tag{13}$$

independent of cross section of the cylinders. In a practical device a small pip is located at the end of the moveable guard electrode to compensate for end effects due to the need to use finite cylinders. The guard can only be moved by a metre or so in practical devices but, despite the small capacitance per metre the ohm can be obtained to a few parts in 10^8.

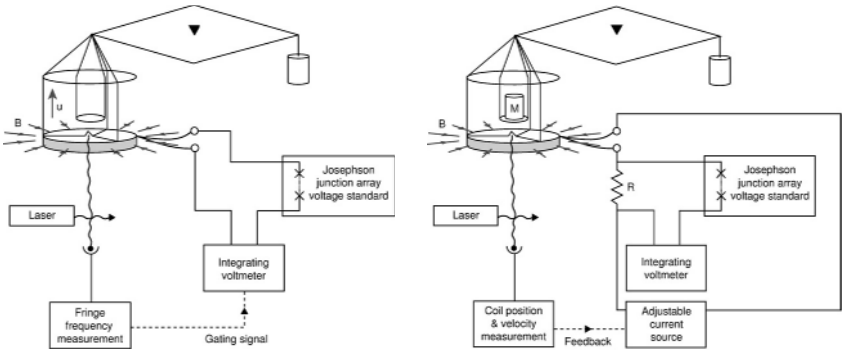
4.4 The Moving Coil Watt Realization of Kibble

This method was devised by Kibble [8] and the apparatus is shown is Fig. 3 with its vacuum container removed. This has permitted a major advance in the methods used to realize the ampere or watt. This has surpassed the accuracy achieved previously which necessitated tedious detailed measurements of precision wound solenoids.

In the dynamic mode we have

$$E = Bk_{\mathrm{c}}\frac{dz}{dt}\ , \tag{14}$$

where E is the induced emf, B the magnetic flux density, k_c the coil constant and $\mathrm{d}z/\mathrm{d}t$ the vertical velocity of the coil.



(a) (b)

Fig. 3. The scheme of the measurement (a) the force on the coil is balanced by a mass on the other arm of the balance (b) the coil is moved with a constant vertical velocity an the induced emf measured. (Figures courtesy of Ian Robinson, NPL)

In the static mode we have

$$mg = Bk_{\mathrm{c}}I \ , \tag{15}$$

where m is the mass required to balance the electrical force on the coil when a current I flows through it.

Very careful attention to the details of the design and construction of the apparatus is required to ensure that the product Bk_c is the same in both cases. The two equations may be combined to eliminate the product Bk_c in which case we obtain an electrical power as

$$EI = mg\frac{dz}{dt} \ . \tag{16}$$

Thus the apparatus expresses an electrical watt in terms of a mechanical watt and so may be used to calibrate the maintained electrical standards of power.

Beyond this, if the power is measured in terms of a Josephson voltage standard as k_1/K_J , and a resistance standard k_2R_K then we have

$$EI = \frac{k_1^2}{\left(K_J\right)^2 k_2 R_K} = \left(\frac{k_1^2}{k_2}\right) \cdot \left(\frac{1}{\left(2e/h\right)^2 \left(h/e^2\right)}\right) = \left(\frac{k_1^2}{k_2}\right) \cdot \frac{h}{4} \tag{17}$$

and finally

$$h = \frac{4mgk_2}{k_1^2} \cdot \frac{dz}{dt} \ . \tag{18}$$

We see therefore that the measurement provides an electro-mechanical measurement of the Planck constant as well as providing a realization of the watt.

4.5 The Kilogram

The SI kilogram still relies on the prototype kilogram of more than a hundred years ago. This is an artifact standard made from a platinum-iridium alloy. It is not thought to provide a constant realizable mass standard with an accuracy of much better than a part or so in 10^8. A possible replacement based on a fundamental constant is looming on the horizon. A very promising one is suggested by the above watt realization for we may rearrange the equation to obtain:

$$m = h \Big/ \left(\frac{4gk_2}{k_1^2} \cdot \frac{dz}{dt}\right) \ . \tag{19}$$

4.6 The Anomalous *g*-Factor of the Electron

Formidable measurements have been made of the g-factor of the electron at the University of Washington [9] by measuring the difference between the cyclotron and spin precession frequencies of a single electron in a Penning trap in a very stable magnetic field produced by a compensated superconducting magnet whose persistent current coils were wound to maximise the screening from extraneous

magnetic fields. Ground-state electrons have been trapped and measured for several months. The measurements yield a value for the g-factor anomaly a_e given by

$$a_{\mathrm{e}} = \frac{|g_{\mathrm{e}}| - 2}{2} = \frac{|\mu_{\mathrm{e}}|}{\mu_{\mathrm{B}}} - 1 \,. \tag{20}$$

The g-factor anomaly may be expressed in terms of a power series in α using

$$a_{\mathrm{e}}\,(\mathrm{theory}) = a_{\mathrm{e}}\,(\mathrm{QED}) + a_{\mathrm{e}}\,(\mathrm{weak}) + a_{\mathrm{e}}\,(\mathrm{hadron})\,. \tag{21}$$

The weak and hadronic contributions are not significant at the present level of accuracy leaving

$$a_{\mathrm{e}}\,(\mathrm{QED}) = C_{\mathrm{e}}^{(2)} \left(\frac{\alpha}{\pi}\right) + C_{\mathrm{e}}^{(4)} \left(\frac{\alpha}{\pi}\right)^2 + C_{\mathrm{e}}^{(6)} \left(\frac{\alpha}{\pi}\right)^3 + \ldots \tag{22}$$

Terms up to $C_{\mathrm{e}}^{(8)}$ have been evaluated by Kinoshita and his colleagues, and a correction due to a miscalculation has been found recently [10]. The $C_{\mathrm{e}}^{(10)}$ will be required if a value of α accurate to a part in 10^{10} is ever to be achieved by this method. As far as is presently known, the contribution of the higher order terms is negligible. It should be noted that the calculation of the $C_{\mathrm{e}}^{(8)}$ term requires the accurate computation of some 891 Feynman diagrams. Evidently, theoretical work on this aspect of QED requires lengthy calculations on powerful computers by most dedicated theoreticians.

The comparison of the results of theory (including knowledge of α obtained by other methods) and experiment combine to provide one of the most stringent tests of QED theory (and if one accepts the theory), currently the most accurate method of obtaining a value for the fine structure constant.

4.7 The Rydberg Constant

The Rydberg constant is derived from the optical spectroscopy of atomic hydrogen, the accuracy is such that QED and other corrections are needed to the theory. Two transitions are needed to obtain the Rydberg constant, the metastable 1S–2S transition and one of the 2S–nD transitions. Biraben *et al.* measured n=8,12 [11,12]. The Max-Planck-Institute measurements of Hänsch have yielded a measurement of the 1S–2S transition frequency by two-photon spectroscopy [13]. Accurate frequency measurements were facilitated by direct multiplication between the caesium-133 microwave frequency used to define the second and the optical region. Their value of 2 466 061 413 187 103(46) Hz has been incorporated as one of the transitions specified in the *mise en pratique* for the realization of the SI definition of the metre.

The accuracy puts stringent requirements on the theory of the energy levels of atomic hydrogen. One of the limiting factors currently being our experimental knowledge of the electric charge radius of the proton, and two loop corrections to the Lamb shifts.

The measurements of the Rydberg constant provide the most accurately known input values for the CODATA evaluation of the best values fundamental constants.

4.8 The Newtonian Constant of Gravitation

At present the limits on any time variation of G, established by planetary ranging, are less than 10^{-12} per year [14, Table IV p.448]. The measurements of G are much less accurate and it is difficult to achieve part in 10^5 accuracy. Laboratory measurements of G are very difficult because the gravitational attractive force between laboratory masses is some 10^{11} times smaller than the gravitational attraction of the Earth. If G is to be measured to parts in 10^5 it is apparent that the attractive force due to the Earth must be eliminated to parts in 10^{13}.

The small size of a fundamental physical constant may not permit it to be measured directly, but it does not prevent its being measured experimentally in combination with another constant. Recent attempts to improve the accuracy have led to measurements that differ by more than the assigned uncertainties, suggesting that one or more of them have undiscovered effects shifting their values. As a result of this the CODATA recommended value of G has an expanded uncertainty, as shown in Fig. 4.

The related gravitational fine structure constant α_{G} is one such combination that is of considerable importance in cosmology. Unfortunately the very small magnitude of this constant makes it very unlikely that any terrestrial method of measuring it will be devised,

$$\alpha_{\mathrm{G}} = \frac{G m_{\mathrm{p}}^2}{\hbar c} \sim 5 \times 10^{-39} \, . \tag{23}$$

As so often happens in this difficult measurement topic area, the most accurate of the current measurements of G disagree by more than their assigned uncertainties. The usual solution, pioneered by R T Birge [15] is to arrive at a best value with an expanded uncertainty. With hindsight we can see that it would have been safer to double the uncertainty arrived at in this way and, indeed that may be the prudent way of dealing with the present discrepancies with the G measurements.

5 Underpinning of the SI by the Fundamental Physical Constants

Figure 5, due to T.J. Quinn, BIPM [16], illustrates very well how the accuracies with which we can currently realize the fundamental physical constants helps reinforce our knowledge of how well we can realize or maintain the SI units. It is already an impressive list. In the past we have generally achieved a ten-fold improvement in measurement accuracy on a roughly fifteen years timescale. In doing so we have often moved to fresh ways of realizing the SI units and different ways of measuring the fundamental constants. In many cases this illustrates the underlying role of the fine structure constant.

5.1 The Importance of the Fine Structure Constant in Metrology

Although the fine structure constant is dimensionless it has a profound underlying effect on metrology. The comparison of different methods of measuring α is

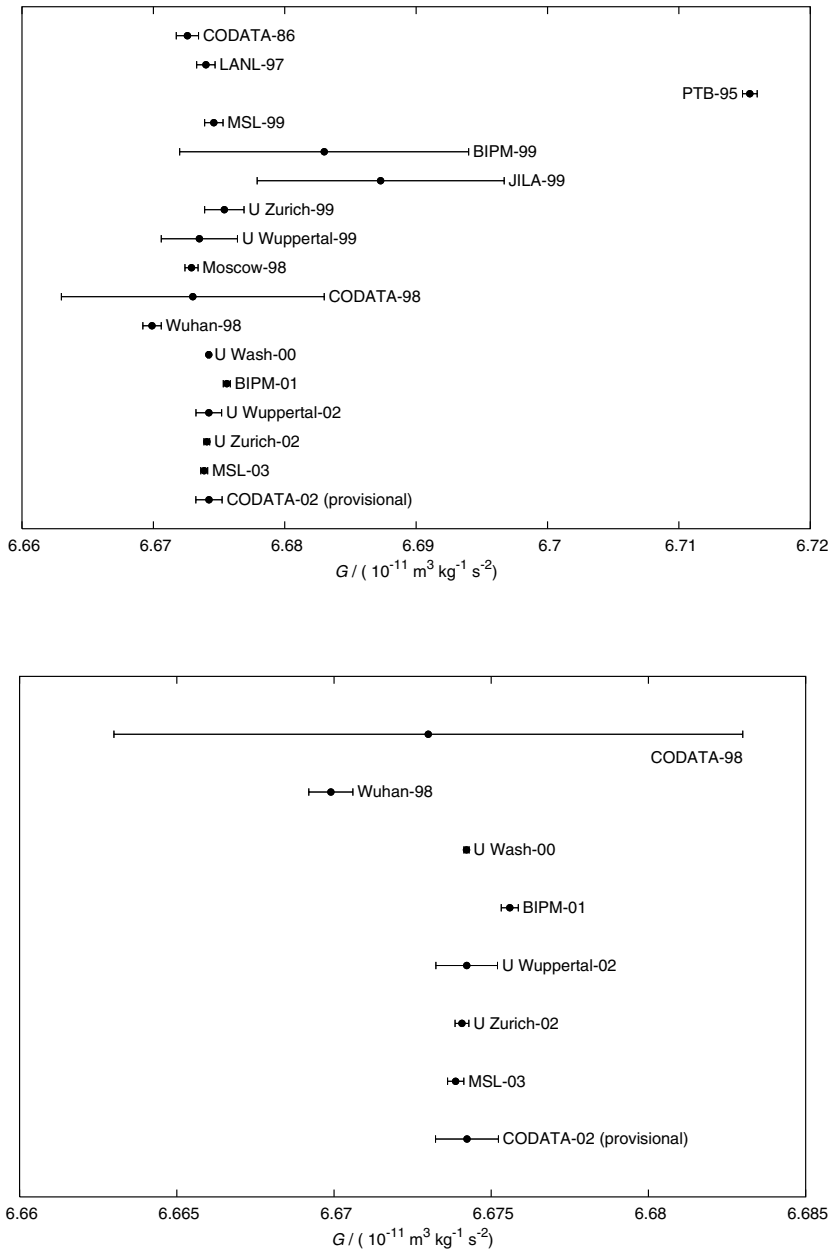


Fig. 4. The present most accurately determined experimental values for G showing a lack of consistency. The lower graph is an expanded view of the post 1998 points, note the change of scale.

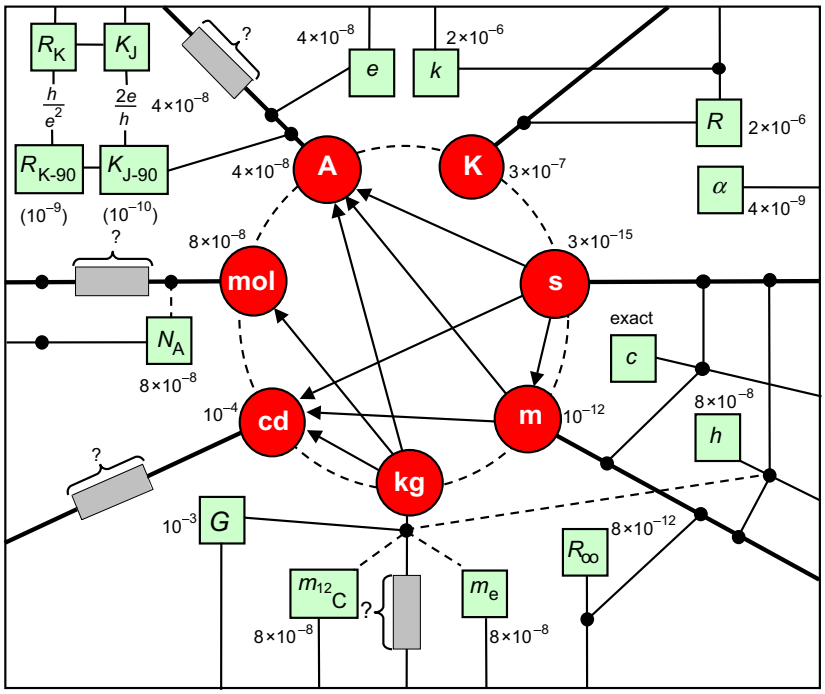


Fig. 5. Illustrating our knowledge of the SI units, and the fractional uncertainties of the underpinning fundamental physical constants. The solid boundary represents total certainty. Also shown are the uncertainties in the quantities $R_{\mathrm{K}-90}$ and $K_{\mathrm{J}-90}$ used to maintain representations of the ohm and volt respectively. The dependence of the other base units on the accuracy of realization of the second is also readily seen (diagram by Quinn, BIPM)

free from questions of the absolute accuracy of SI units and therefore provides a valuable check on how well we understand the physics involved in measurements by very methods. The principle measurements contributing to our present knowledge of α are shown in Fig. 6.

The two most accurate measurements are the $g_e - 2$ [17] and caesium recoil [18] and these are in good agreement. The 1998 CODATA recommended value for α was slightly higher than the 1986 value and the 2002 value may be slightly higher than the 1998 one.

At this point it is worth pointing out that different realizations of the units have so far involved α to different powers and that if we find the caesium-133 definition of the second no longer adequate we may move to an optical or higher frequency transition involving a change in α dependence by a factor of α^2. If α were shown to be varying then this fact would have to be incorporated into the definition of the second.

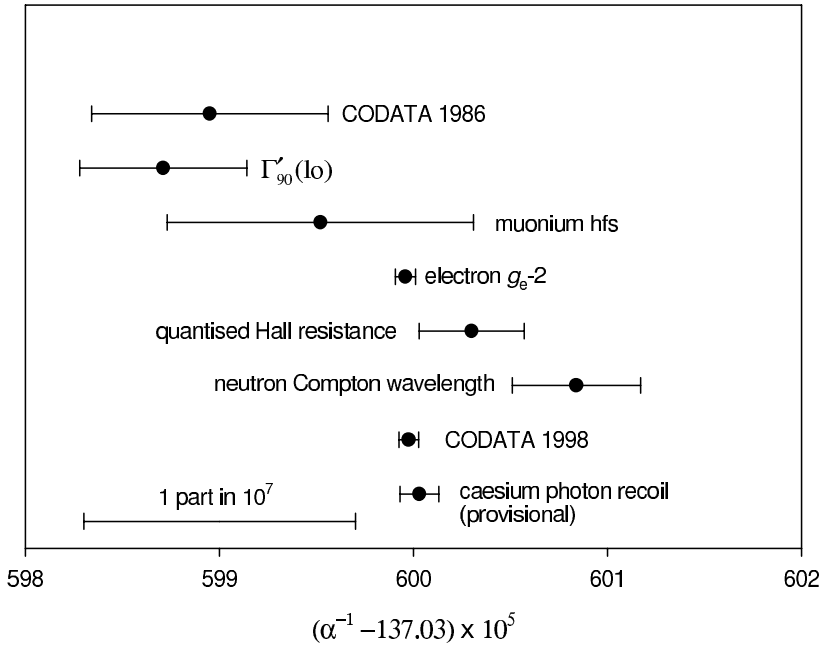


Fig. 6. Showing the principle most accurate determinations from which a value for α may be deduced. $\Gamma'_{\mathrm{p}-90}(\mathrm{lo})$ is the NIST low field measurement of the gyromagnetic ratio of the proton.

6 Conclusion

Metrology fundamentally seeks to achieve a set of units that are suitable and slightly more adequate than is required by present-day science and technology. It therefore does not attempt to achieve ultimate units, although, obviously, every effort is made to adopt a definition having a longer duration than the krypton-86 definition of the metre. It may not be possible to find an explicit expression in terms of fundamental constants. Thus, the expression for the hydrogen 1S–2S frequency in terms of fundamental constants is [19]

$$\nu\,(1\mathrm{S}-2\mathrm{S}) = \frac{3R_\infty c}{4}\left[1 - \frac{m_\mathrm{e}}{m_\mathrm{p}} + \frac{11}{48}\alpha^2 - \frac{28}{9}\frac{\alpha^3}{\pi}\ln\alpha^{-2} - \frac{14}{9}\left(\frac{2\pi\alpha R_\mathrm{p}}{\lambda_\mathrm{C}}\right)^2 + \ldots\right],$$

where λ_C is the Compton wavelength of the electron.

The root mean square electric charge radius of the proton (R_p) is one of the factors limiting our 7.6×10^{-12} fractional accuracy of this expression whose value compares well with the experimental value, which can be measured to 1.8×10^{-14}. In the case of R_p, this illustrates well how accurate measurements can test our knowledge acquired from different parts of physics (spectroscopy and high energy nuclear physics).

We can also see that the underlying role of the fundamental constants in a particular SI unit may not be calculable with sufficient accuracy for the requirements of modern science and technology.

6.1 Future

It seems very unlikely that we have discovered all of the simple quantum and other phenomena that might be used to define or realize the fundamental constants in terms of the SI units. Much of the improved accuracy has come from our ability to make very accurate measurements at much higher frequencies than hitherto. Of late, attention has moved from the microwave to visible frequencies, but may well ultimately continue on through the ultraviolet region to the gamma ray region. Indeed, the wavelength of the hydrogen 1S–2S transition is already one of the *mise en pratiques* for the realization of the metre with $\sim 10^{-12}$ fractional accuracy.

Some countries prefer the definition of a unit to suggest a method for its practical realization, whereas others prefer it to be more abstract [20]. One advantage of the latter solution is that a particular definition may thereby endure for much longer. For example, the krypton-86 definition was difficult to sustain once more stable and accurately realizable laser sources became available, and only endured for 23 years. One possibility might have been to specify a laser standard that was based on a particular transition. However, such a definition would have necessitated a new definition of the metre each time that a 'better' laser was developed. The more abstract definition based on the speed of light has no such constraints, except via the definition of the second. Particular ways of realizing the metre are specified in the associated *mise en pratique* for the metre. It is therefore similarly desirable that any redefinition of the second does not rely on specifying the periodicity of a particular transition in the longer term, although it may be necessary to do so in the interim. We might, for example, one day define the second in terms of the Planck time, or the period of rotation of an electron, or the time for an electron to complete a single orbit in the ground-state of a hydrogen atom.

A further feature of modern metrology is our ability to make measurements in terms of time frequency and fundamental constants are involved in relating the quantity required to frequency as in length measurement involves the speed of light, and the Josephson measurement of voltage via frequency involves $h/2e$.

Today we have the situation that most of the SI base units, and many of the derived units, may be realized in terms of fundamental constants. This has made it much easier to compare the ratio Q_1/Q_2 than hitherto – often simply by exchanging frequency values rather than risking changes to artifacts in transit to a calibration laboratory.

We must be careful in our use of the fundamental physical constants to define an SI unit. Thus, the present definition of the ampere implicitly defines the value of the ohm in terms of $\mu_0 c$. If we were to define the ohm in terms of the von Klitzing constant (h/e^2), the two definitions would also fix the value of α in the SI system!

7 Summary

Fundamental constants are woven intimately into the SI system of units of today. Our knowledge of their values improves as science and technology progress – in many cases by more than tenfold every fifteen years or so. The national standards laboratories make strong efforts to provide a unit that is sufficiently invariant for most purposes. Where it is not for a particular phenomenon, that phenomenon may be harnessed to improve their accuracy. Although all units rely implicitly on the fundamental constants we can expect to harness more of them to play an explicit role in sustaining the SI units, possibly leading to new definitions or even alternative SI base units – for example, energy instead of mass by fixing h. Finally, measurements from very different parts of physics can be intercompared as an essential part of the evaluation of their best values. Such comparisons show where theory and experiment are inadequate for the required accuracy. This also leads to progress in our understanding of the universe.

References

1. S. R. Jefferts, J. Shirley, T. E. Parker, T. P. Heavner, D. M. Meekhof, C. Nelson, F. Levi, G. Costanzo, A. De Marchi, R. Drullinger, L. Hollberg, W. D. Lee, and F. L. Walls, Metrologia **39**, 321, 2002.
2. J.-M. Lévy-Leblond. *The importance of being (a) constant.* In: G. Toraldo di Francia, editor, *Problems in the Foundations of Physics; Proceedings of the International School of Physics "Enrico Fermi" Course LXXII*, pages 237–263. North-Holland, Amsterdam, 1979.
3. G. J. Stoney. *On the physical units of nature.* Philosophical Magazine **11**, 381, 1881.
4. M. Planck. *The theory of heat radiation.* Dover, New York, 1959. See p. 218 for Planck units.
5. P. J. Mohr and B. N. Taylor, *CODATA recommended values of the fundamental physical constants 1998.* Reviews of Modern Physics **72**, 351, 2000.
6. K. von Klitzing, G. Dorda, and M. Pepper, Physical Review Letters **45**, 494, 1980.
7. A. M. Thompson and D. G. Lampard, Nature **177**, 888, 1956.
8. B. P. Kibble. *A measurement of the gyromagnetic ratio of the proton by the strong field method.* In: J. H. Sanders and A. H. Wapstra, editors, *Atomic Masses and Fundamental Constants 5*, pages 545–551. Plenum Press, London and New York, 1975.
9. R. S. Van Dyck, Jr., P. B. Schwinberg, and H. G. Dehmelt, Physical Review D **34**, 722, 1986.
10. T. Kinoshita and M. Nio, Physical Review Letters **90**, 021803, 2003.
11. B. de Beauvoir, F. Nez, L. Julien, B. Cagnac, F. Biraben, D. Touahri, L. Hilico, A. Acef, A. Clarion, and J. J. Zondy, Physical Review Letters **78**, 440, 1997.
12. C. Schwob, L. Jozefowski, B. De Beauvoir, L. Hilico, F. Nez, L. Julien, F. Biraben, O. Acef, J.-J. Zondy, and A. Clarion, Physical Review Letters **82**, 4960, 1999.
13. M. Niering, R. Holzwarth, J. Reichert, P. Pokasov, Th Udem, M. Weitz, T. W. Hänsch, P. Lemonde, G. Santarelli, M. Abgrall, P. Laurent, C. Salomon, and A. Clairon, Physical Review Letters **84**, 5496, 2000.
14. J.-P. Uzan, Reviews of Modern Physics **75**, 403, 2003.

15. R. T. Birge. *Probable values of the general physical constants (as of January 1, 1929)*. Physical Review Supplement **1**, 1 1929.
16. `http://www.bipm.fr/enus/3_SI/si_fig.html`
17. R. S. Van Dyck, Jr., P. B. Schwinberg, and H. G. Dehmelt, Physical Review Letters **59**, 26, 1987.
18. A. Wicht, J. M. Hensley, E. Sarajlic, and S. Chu. *A preliminary measurement of $h/m(Cs)$ with atom interferometry*. In: P. Gill, editor, *Proceedings of the 6th Symposium on Frequency Standards and Metrology*, pages 193–212. World Scientific, Singapore, 2002.
19. P.J. Mohr, B.N. Taylor, Fundamental constants and the hydrogen atom, Lect. Notes Phys. **627**, 145–156 (2001)
20. V. Kose, B. R. L. Siebert, and W. Wöger, Metrologia **40**, 146, 2003.

Time Varying Fundamental Constants, Extra Dimensions and the Renormalization Group

William J. Marciano

Brookhaven National Laboratory, Upton, NY 11973, USA

Abstract. Time varying fundamental constants find a natural setting in the cosmology of extra space-time dimensions. Changes in the sizes of additional compact spatial dimensions manifest themselves as modifications of coupling strengths and masses in our effective low energy four-dimensional world. Those variations and their interconnection are best studied using the renormalization group equations of the Standard Model or its supersymmetric and grand unified extensions. A generic analysis predicts that time changes in the fine structure constant, α, are relatively suppressed by vacuum polarization effects while the QCD coupling variation is relatively enhanced due to asymptotic freedom. The latter effect suggests that time changes in the proton to electron mass ratio, m_p/m_e may be a particularly sensitive probe of the physics of extra dimensions, with the sign $(+/-)$ of the variation relative to other changing quantities differentiating possible cosmologies.

1 Dirac Revisited – The Hierarchy Problem

In 1937, Dirac put forward his "Big Numbers Hypothesis" [1]. He argued that extremely large dimensionless ratios were not likely to be calculable in a single theory. So, for example, any attempt to combine ordinary electromagnetism of protons and electrons with gravity would have to explain the very large ratio (for $\hbar = c = 1$)

$$\frac{\alpha}{G_N m_p m_e} \simeq 10^{39} \, , \tag{1}$$

where G_N is Newton's constant, $\alpha \simeq 1/137$ is the fine structure constant and m_p, m_e represents proton and electron masses. He noted that a similar numerical ratio resulted in a comparison of the universe lifetime, $\tau_u \simeq 1.5 \times 10^9$yr with atomic time, $1/m_e$,

$$\tau_u m_e \simeq 10^{39} \, . \tag{2}$$

That "coincidence" led Dirac to conjecture a simple time dependence for $G_N \sim 1/t$. Could the strength of gravity be decreasing with time? Stated another way, the Planck scale

$$m_{\rm Pl} = 1/\sqrt{G_N} \simeq 1.2 \times 10^{19} \text{ GeV} \tag{3}$$

might be growing in time

$$m_{\rm Pl} \sim t^{1/2} \, . \tag{4}$$

W.J. Marciano, Time Varying Fundamental Constants, Extra Dimensions and the Renormalization Group, Lect. Notes Phys. **648**, 97–105 (2004)
http://www.springerlink.com/

Efforts to confirm Dirac's provocative conjecture have proven negative, leading to the rather stringent bound on $\dot{G}_N/G_N$ ($\dot{G}_N = dG_N/dt$)

$$\dot{G}_N/G_N \lesssim 10^{-12}\ \mathrm{yr}^{-1} \tag{5}$$

"assuming" a simple continuous $1/t$ time dependence.

Emulating Dirac, other notable theorists later speculated that the fine structure constant α may vary with time [2]. Efforts to observe variations in α have also generally proven negative particularly when combined with assumptions of a simple linear dependence on time, leading to stringent constraints [3]. However, recent astrophysical absorption line spectroscopy studies seem to indicate possible variations in α during the Universe's early history [4] followed by a long period of relative constancy. Although such an interpretation of the data is controversial, it has stimulated renewed interest in Dirac's conjecture that fundamental constants may indeed vary with time.

It is interesting to note that Dirac's idea of a time varying G_N can be viewed as the first proposed solution to what we today call the "Hierarchy Problem", i.e. Why are ordinary elementary particle masses so much smaller than the Planck scale, m_{Pl}, and can the hierarchy of masses be preserved in quantum field theory? Supersymmetry is often touted as a way to preserve the hierarchy, but it does little to explain its basic origin. Purely dynamical theories (without fundamental scalars) do better, since mass ratios only appear in logarithms. It is much easier to accommodate

$$\ln\left(\frac{m_{\mathrm{Pl}}}{m_e}\right) \simeq 51\ . \tag{6}$$

Than the enormous mass ratio $m_{\mathrm{Pl}}/m_e \simeq 2 \times 10^{22}$. However, fully dynamical mass generating models have so far fallen short of their goal to provide a viable alternative to the Standard Model and its fundamental Higgs scalar mechanism.

Dirac was clearly ahead of his time. His time varying constants (a true oxymoron) conjecture did not have much of a following although from time to time new limits would be published for $\dot{G}_N/G_N$, $\dot{\alpha}/\alpha$, $\dot{m}_p/m_p$ etc. Renewed interest in such a possibility resulted from a number of confluent developments:

1) Astronomical indications from absorption line spectroscopy that α may have varied in the distant past [4].
2) New cosmological mysteries such as dark energy and indications of a cosmological constant have inspired far out speculations.
3) The possible existence of compact extra dimensions, introduced for elementary particles model building, provides a natural framework for discussing time variations of fundamental parameters by connecting them with the cosmology of compact extra dimensions [5].

Turning that point around, observation of time varying fundamental constants could be taken as possible evidence for extra dimensions and a window to their basic role in nature. This chapter is based on that perspective. Its main contents were published about 20 years ago [5] when the possible reality of extra dimensions and the mathematics of superstrings were finding their way back into contemporary physics.

2 Fundamental Constants from a Modern Perspective

Any discussion of time varying fundamental "constants" must stipulate what is fundamental, what can vary and how such variations may be interconnected. From a particle physics perspective, it must be framed in the context of relativistic quantum field theory and the Standard $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ Model or its various extensions. For that reason, it is best to assume time invariant natural units

$$\hbar = c = 1 \, . \tag{7}$$

To do otherwise risks violation of the basic tenets of quantum mechanics and special relativity, an area I would not want to tread into. One can, in principle fix an additional scale quantity such as m_e and specify all 3 units. That can have potential drawbacks in discussing some aspects of early time astrophysical spectroscopy. Nevertheless, to simplify my discussion I will subsequently assume $\dot{m}_e = 0$, i.e. m_e will effectively be used as a fixed mass unit. In astrophysics, variations of m_e will be difficult to disentangle from redshift effects and the possible time variation of m_p/m_e. That issue deserves a thorough discussion, but it will not be given here.

The Standard Model has many independent fundamental parameters. They include 3 gauge couplings $g_i, i = 1, 2, 3$, two Higgs parameters (m_H and the vacuum expectation value v) and 36 complex Yakawa couplings of left and right handed fermions to the Higgs scalars. Those Yukawa couplings lead to fermion masses and mixing. In a supersymmetric (SUSY) or grand unified theory (GUT) there can be many more independent parameters. To allow all of these parameters to vary independently with time would lead to complete chrono pandemonium; so, I will limit my considerations to a manageable subset and link them by symmetries. Note, however, that variation in the Higgs vacuum expectation value, which will not be discussed here, can have very interesting implications for Dark Energy considerations.

In modern quantum field theory, couplings and effective masses depend on the energy scale μ or distance $1/\mu$ being probed. One employs running quantities $\alpha_i(\mu) \equiv g_i^2(\mu)/4\pi$ or effective masses $m(\mu)$ parametrized by the sliding mass scale μ. So, for example, the $\alpha_i(\mu), i = 1, 2, 3$ of the Standard Model are often approximately specified at a scale $\mu = m_Z \simeq 91$ GeV to be (in the $\overline{\mathrm{MS}}$, modified minimal subtraction scheme)

$$\alpha_3(m_Z)_{\overline{\mathrm{MS}}} \simeq 0.118 \, , \tag{8}$$

$$\alpha_2(m_Z)_{\overline{\mathrm{MS}}} \simeq 0.034 \, , \tag{9}$$

$$\alpha_1(m_Z)_{\overline{\mathrm{MS}}} \simeq 0.017 \, . \tag{10}$$

The short-distance $\alpha(\mu)$ of QED is related to those quantities via [6]

$$\alpha^{-1}(\mu)_{\overline{\mathrm{MS}}} = \frac{5}{3}\alpha_1^{-1}(\mu)_{\overline{\mathrm{MS}}} + \alpha_2^{-1}(\mu)_{\overline{\mathrm{MS}}} \, , \tag{11}$$

which implies [7]

$$\alpha(m_Z)_{\overline{\mathrm{MS}}} \simeq 1/128 \tag{12}$$

(I subsequently drop the $\overline{\rm MS}$ subscript.) rather than the usual fine structure constant (at $\mu = 0$)

$$\alpha = 1/137.036 \tag{13}$$

we all know and love.

What is more fundamental $\alpha = 1/137$ or $\alpha(m_Z) \simeq 1/128$? The latter by today's standards. That hopefully dispels the mystical beliefs often associated with "137". In fact, the shorter the distance examined (larger μ), the more fundamental. The difference between 1/128 and 1/137 stems from fermion vacuum polarization effects that screen the short-distance couplings. In the case of alpha

$$\alpha^{-1}(\mu) = \alpha^{-1} - \frac{2}{3\pi}\sum_f Q_f^2 \ln\left(\frac{\mu}{m_f}\right)\theta(\mu - m_f) + \frac{7}{2\pi}\ln\left(\frac{\mu}{m_W}\right)\theta(\mu - m_W) + \text{ higher orders}\,, \tag{14}$$

where the sum is over all fermions (quarks and leptons) with electrical charge Q_f.

The relationship in (14) illustrates an important point regarding time variation. All physical quantities are connected via quantum loops (at some level). There may be scenarios where all vary in a way that makes the relative variations unobservable. However, it is unlikely (impossible) that only a single parameter such as α varies in time. One must consider the relationships among parameters induced by quantum loops in discussing time varying constants and their implications. This is, unfortunately, a principle almost never adhered to in the literature where $\dot\alpha$ is often discussed in isolation and little distinction between long and short-distance is made. If there is an underlying source of time variation, it will be most simply manifested in the very short-distance parameters which are inherently more fundamental. Those effects will be dressed by quantum loop corrections in α, m_p etc. Any interpretation of an observed $\dot\alpha \neq 0$ would have to unfold those quantum corrections.

What is the relationship between the time variation in long and short distance parameters? To properly address that issue really requires a complete theory. However, some generic features can be gleamed simply from the running properties of the couplings. For example, $\alpha(\mu)$ increases, albeit slowly, as μ increases. The value $\alpha \simeq 1/137$ effectively corresponds to $\mu \simeq 0$. The running in (14) suggests (ignoring fermion mass variations)

$$\frac{\dot\alpha(\mu)}{\alpha^2(\mu)} \simeq \frac{\dot\alpha}{\alpha^2} \tag{15}$$

or

$$\frac{\dot\alpha}{\alpha} \simeq \frac{\alpha}{\alpha(\mu)}\,\frac{\dot\alpha(\mu)}{\alpha(\mu)}\,. \tag{16}$$

The factor $\alpha/\alpha(\mu)$ represents a suppression factor. In the case of QCD, a similar analysis suggests

$$\frac{\dot\alpha_3(\mu)}{\alpha_3(\mu)} \simeq \frac{\alpha_3(\mu)}{\alpha_3(\mu_0)}\,\frac{\dot\alpha_3(\mu_0)}{\alpha_3(\mu_0)}\,. \tag{17}$$

For $\mu \ll \mu_0$, the factor $\alpha_3(\mu)/\alpha_3(\mu_0)$ represents a potentially large enhancement factor due to asymptotic freedom. It can be quite large for small μ where $\alpha_3(\mu)$ blows up. That simple relationship suggest quite generically that low energy QCD quantities such as the proton mass, m_p, may be among the best places to search for time variation [5]. In fact, the ratio m_p/m_e which is present in hyperfine spectroscopy, provides a very sensitive probe of time dependence.

3 Extra Dimensions

The speculation that we live in a space-time of dimension > 4 has a long and fascinating history. It stimulated great interest when Kaluza (in 1921) and Klein (in 1926) showed [8] that a 5 dimensional metric $g_{\mu\nu}(x)$, $\mu, \nu = 0, 1, 2, 3, 4$ in which the 5th dimension corresponded to a closed circle of radius $R_{\rm KK} \simeq 10^{-32}$cm could simultaneously describe gravity and electromagnetism, an amazing observation. In such a theory, $g_{\mu\nu}$, $\mu, \nu = 0, 1, 2, 3$ corresponds to the usual metric of general relativity, $g_{\mu 4}(x) = A_\mu(x)$ the electromagnetic gauge field and $g_{44}(x)$ is a Brans-Dicke scalar field. The electromagnetic coupling α and Newton's constant, G_N are related via

$$\alpha = \frac{e^2}{4\pi} = \frac{4G_N}{R_{\rm KK}^2} \ . \tag{18}$$

The idea of Kaluza-Klein unification can be extended to higher spatial dimensions. With 11 space-time dimensions [9] (4 ordinary and 7 compact) the entire $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ gauge structure of the Standard Model can emerge from the metric. Superstrings, which are 10 or 11 dimensional share many of the Kaluza-Klein features, but offer the attractive possibility of a finite quantum theory of gravity. In all such higher dimensional theories, one finds the generic relations

$$\alpha_i = K_i \frac{G_N}{R_{\rm KK}^2} \ , \tag{19}$$

where the K_i are pure numbers determined by the topology of the compact spatial dimensions. At what scale are the α_i couplings in (19) defined? The natural answer is $\mu \simeq m_{\rm KK} = 1/R_{\rm KK} \simeq 10^{18}$ GeV. That is also roughly the unification scale of some grand unified theories (GUTS). The short-distance nature of those couplings is consistent with my statement that short-distance parameters are more fundamental and more closely connected with the basic underlying physics. Of course, the long distance couplings can be obtained from (19) by including quantum loop effects, provided we know the full theory content at scales $< m_{\rm KK}$.

Are extra dimensions a physical reality or merely a model building tool? If they are real, can we find evidence for their existence? Assuming they are real, forces one to generalize our space-time cosmology and provides a potentially rich new perspective regarding the origin of our universe. Are the extra dimensions contracting, expanding or even oscillating as a function of time? If (generically) $\dot{R}_{\rm KK} = dR_{\rm KK}/dt \neq 0$, it will give rise to time varying fundamental constants at low energies and such effects could be our window to the physics of those extra

dimensions. My original interest in this subject stems from the early studies by Chodos and Detweiler [10] along with Freund [11] that incorporating expansion of our ordinary 4 dimensional space-time in Kaluza-Klein theories naturally leads to cosmologies with $\dot{R}_{\rm KK} \neq 0$. Today, such speculations are even more interesting because of indications that our universe may have accelerated or decelerated in the past and may be dominated by Dark Energy. Small non-static extra spatial dimensions not only provide a framework for discussing time varying fundamental constants, but may also hold insights regarding the origin of Dark Energy and the unusual evolutionary aspects exhibited by our universe.

4 Renormalization Group Connections

Simple connections among fundamental parameters are provided by renormalization group equations which describe their common dependence on the energy scale μ. Here, I will focus on gauge couplings, but a similar analysis could be carried out for particle masses.

The following relations are assumed to hold

$$\alpha_i(m_{\rm KK}) = K_i G_N / R^2_{\rm KK} = K_i G_N m^2_{\rm KK} \ , \qquad i = 1,2,3 \ , \tag{20}$$

and

$$\alpha_i^{-1}(\mu) \simeq \alpha_i^{-1}(m_{\rm KK}) + \sum_j b_i \left[\ln \frac{m_{\rm KK}}{m_j} + \theta(\mu - m_j) \ln \frac{m_j}{\mu} \right] + \cdots \ , \tag{21}$$

where the b_i are given by (in the Standard Model)

$$b_i = \begin{pmatrix} b_1 \\ b_2 \\ b_2 \end{pmatrix} = -\frac{1}{2\pi} \begin{pmatrix} -\frac{4}{3} n_g - \frac{1}{10} N_H \\ \frac{22}{3} - \frac{4}{3} n_g - \frac{1}{6} N_H \\ 11 - \frac{4}{3} n_g \end{pmatrix} \ , \tag{22}$$

$$\mu \frac{\partial}{\partial \mu} \alpha_i(\mu) = b_i \alpha_i^2 + \cdots ,$$

n_g = number of fermion generations (3 are known) and N_H = number of Higgs doublets. In a supersymmetric extension of the Standard Model

$$b_i = -\frac{1}{2\pi} \begin{pmatrix} -2n_g - \frac{3}{10} N_H \\ 6 - 2n_g - \frac{1}{2} N_H \\ 9 - 2n_g \end{pmatrix} \ . \tag{23}$$

This theory has the advantage of being embedable in a GUT such as SU(5) or SO(10) with coupling unification.

$$\alpha_1^{-1}(m_{\rm GUT}) = \alpha_2^{-1}(m_{\rm GUT}) - \frac{1}{6\pi} = \alpha_3^{-1}(m_{\rm GUT}) - \frac{1}{4\pi} \tag{24}$$

at $m_{\rm GUT} \simeq 10^{16}$ GeV if $m_{SUSY} \simeq 1$ TeV. In such theories the evolution of the $\alpha_i(\mu)$ are correlated by the unification condition in (24). Another interesting SUSY GUT example based on E_6 with **27** plet generations has [12]

$$b_i = -\frac{1}{2\pi}\begin{pmatrix} -3n_g - \frac{3}{10}N_H \\ 6 - 3n_g - \frac{1}{2}N_H \\ 9 - 3n_g \end{pmatrix} . \tag{25}$$

In that model, unification occurs at $M_{\rm GUT} \simeq 10^{18}$ GeV where $\alpha_i(m_{\rm GUT}) \sim \mathcal{O}(1)$. Interestingly, the QCD coupling $\alpha_3(\mu)$ has $b_3 = 0$ in leading order for that model when $n_g = 3$.

For purposes of illustration, I will assume unification $K_1 = K_2 = K_3$. Also, all light fermion masses and Higgs parameters are assumed static $\dot{m}_e = \dot{m}_f = \dot{m}_W = \dot{m}_Z = 0$. That assumption is not likely completely correct if the couplings change with time; however, I will assume for illustration that such effects are not leading. Taking the time derivative of (20) and (21), one finds (with $\dot{K}_i = 0$)

$$\frac{\dot{\alpha}_i(m_{\rm KK})}{\alpha_i(m_{\rm KK})} = \frac{\dot{G}_N}{G_N} + 2\frac{\dot{m}_{\rm KK}}{m_{\rm KK}} , \tag{26}$$

$$\frac{\dot{\alpha}_i(\mu)}{\alpha_i^2(\mu)} = \frac{\dot{\alpha}_i(m_{\rm KK})}{\alpha_i^2(m_{\rm KK})} - b_i\frac{\dot{m}_{\rm KK}}{m_{\rm KK}} . \tag{27}$$

These equations generically describe the interconnection among time varying parameters. For further discussion, GUT unification $\alpha_1(m_{\rm KK}) \simeq \alpha_2(m_{\rm KK}) \simeq \alpha_3(m_{\rm KK})$ is assumed and

$$\frac{\dot{m}_p}{m_p} \simeq \frac{\dot{\Lambda}_{QCD}}{\Lambda_{QCD}} \simeq \frac{\dot{\alpha}_3(\mu)}{\alpha_3(\mu)} \text{ at } \mu < 1 \text{ GeV} \tag{28}$$

will be employed.

5 Examples

Let us consider the following higher dimensional cosmologies. The first due to Chodos and Detweiler [10] has

$$\dot{m}_{\rm KK} \neq 0 \text{ but } \dot{\alpha}_i(m_{\rm KK}) = 0 . \tag{29}$$

That leads to [5]

$$\begin{aligned} \frac{\dot{G}_N}{G_N} &= -2\frac{\dot{m}_{\rm KK}}{m_{\rm KK}} = 2\frac{\dot{R}_{\rm KK}}{R_{\rm KK}} , \\ \frac{\dot{\alpha}_i(\mu)}{\alpha_i^2(\mu)} &= \frac{1}{2}b_i\frac{\dot{G}_N}{G_N} , \\ \frac{\dot{\alpha}}{\alpha^2} &= \frac{1}{2}\left(\frac{5}{3}b_1 + b_2\right)\frac{\dot{G}_N}{G_N} , \end{aligned} \tag{30}$$

where the relationship in (11) has been employed. Assuming the b_i in (23) (i.e. a SUSY GUT) leads to

$$\frac{\dot{\alpha}}{\alpha} \simeq 3\frac{\alpha}{\pi}\frac{\dot{G}_N}{G_N} , \tag{31}$$

$$\frac{\dot{m}_p}{m_p} \simeq -50\frac{\dot{\alpha}}{\alpha} \simeq -\frac{3}{8}\frac{\dot{G}_N}{G_N} . \tag{32}$$

Notice that $\dot{\alpha}/\alpha$ is suppressed relative to $\dot{G}_N/G_N$ while $\dot{m}_p/m_p$ is roughly similar. Also, the relative signs in (32) are not the same.

In a second cosmology considered by Freund [11]

$$\dot{m}_{\rm KK} = 0 , \qquad \dot{\alpha}_i(m_{\rm KK}) \neq 0 , \tag{33}$$

and one finds [5]

$$\frac{\dot{\alpha}_i(\mu)}{\alpha_i^2(\mu)} = \frac{1}{\alpha_i(m_{\rm KK})}\frac{\dot{G}_N}{G_N} = \frac{\dot{\alpha}_i(m_{\rm KK})}{\alpha_i^2(m_{\rm KK})} ,$$

$$\frac{\dot{m}_p}{m_p} \simeq 40\frac{\dot{\alpha}}{\alpha} \simeq \frac{\alpha_3(\mu)}{\alpha_3(m_{\rm KK})}\frac{\dot{G}_N}{G_N} . \tag{34}$$

Note that the relative sign has changed in the relationship between $\dot{m}_p/m_p$ and $\dot{\alpha}/\alpha$ compared with the first example; but again $\dot{m}_p/m_p$ is enhanced in magnitude relative to $\dot{\alpha}/\alpha$. (It is also enhanced relative to $\dot{G}_N/G_N$.)

The above results are rather generic. They suggest that m_p/m_e is particularly sensitive to time variation. Also if a variation in that quantity is observed the relative sign $(+/-)$ of the effect compared with $\dot{\alpha}/\alpha$ or $\dot{G}_N/G_N$ can differentiate underlying cosmologies [5,13].

6 Discussions

Dirac's conjecture regarding time varying constants is realized if $m_{\rm KK} \sim t^{1/2}$ which leads to $\dot{G}_N/G_N = -1/t$. However, stringent bounds on $\dot{G}_N/G_N$, $\dot{\alpha}/\alpha$ and $\dot{m}_p/m_p$ (via m_p/m_e) make such a simple time dependence unlikely. More realistic are scenarios where fundamental constants undergo time variation in the early universe but such effects are damped out with time. It would be interesting to correlate possible time evolutions in α with hints of acceleration in early universe expansion.

Many possibilities can be imagined. Indeed, with the advent of Dark Energy anything goes in cosmology. Higher dimension theories provide a rich natural setting for such studies. The extra dimensions may be contracting, expanding, oscillating etc. Experiments of $\dot{G}_N/G_N$, $\dot{\alpha}/\alpha$ and $\dot{m}_p/m_p$ should be pushed as far as possible with m_p/m_e providing the best bet because of relative enhancements stemming from asymptotic freedom in QCD.

Acknowledgements

This work is supported by the U.S. Department of Energy under contract number DE-AC02-98-CH-10886.

References

1. P.A.M. Dirac, Nature **139**, 323 (1937) and Proc. Roy. Soc. London Ser. A **165**, 199 (1938).
2. E. Teller, Phys. Rev. **73**, 801 (1948);
   P.J. Peebles and R.H. Dicke, Phys. Rev. **128**, 2006 (1962);
   G. Gamow, Phys. Rev. Lett. **19**, 759 (1967).
3. A. Shlyakhter, Nature **264**, 340 (1976).
4. J.K. Webb, V. Flambaum, C. Churchill, M. Drinkwater and J. Barrow, Phys. Rev. Lett. **82**, 884 (1999);
   V. Flambaum, physics/0309107.
5. W. Marciano, Phys. Rev. Lett. **52**, 489 (1984).
6. H. Georgi, H. Quinn and S. Weinberg, Phys. Rev. Lett. **33**, 451 (1974);
   W. Marciano and A. Sirlin in Proceedings of the Second Workshop on Grand Unification, Ann Arbor 1981, edited by J. Leveille, L. Sulak and D. Unger (Birkhäuser, Boston 1981).
7. W. Marciano, Phys. Rev. **D20**, 274 (1979).
8. Th. Kaluza, Sitzungsber. Preuss. Akad. Wiss. 966 (1921);
   O. Klein, Z. Phys. **37**, 895 (1926).
9. E. Witten, Nucl Phys. **B186**, 412 (1981).
10. A. Chodos and S. Detweiler, Phys. Rev. **D21**, 2167 (1980).
11. P. Freund, Nucl. Phys. **B209**, 146 (1982); Phys. Lett. **B120**, 335 (1983).
12. W. Marciano and M. Goldhaber in "Paul Adrien Maurice Dirac", Cambridge Univ. Press, Eds. B. Kursunoglu and E. Wigner (1987) p. 163.
13. The potential enhancement of $\dot{m}_p/m_p$ pointed out in [5] was more recently observed for GUT theories in: P. Langacker, G. Segre and M. Strassler, Phys. Lett. **B528**, 121 (2002); X. Calmet and H. Fritzsch, Phys. Lett. **B540**, 173 (2002). As noted in this chapter, that feature is generically related to asymptotic freedom in QCD.

Fundamental Constants and Their Possible Time Dependence

Harald Fritzsch

Ludwig-Maximilians-University Munich, Sektion Physik, Theresienstraße 37, 80333 München, Germany

Abstract. Astrophysical indications that the fine structure constant has undergone a small time variation during the cosmological evolution are discussed within the framework of the standard model of the electroweak and strong interactions and of grand unification. A variation of the electromagnetic coupling constant could either be generated by a corresponding time variation of the unified coupling constant or by a time variation of the unification scale, or by both. The various possibilities, differing substantially in their implications for the variation of low energy physics parameters like the nuclear mass scale, are discussed. The case in which the variation is caused by a time variation of the unification scale is of special interest.

1 Introduction

Usually in physics, especially in particle physics, we deal with the local laws of nature, say the field equations of gauge theory. But when it comes to the fundamental constants, like the fine structure constant α, we must keep in mind that questions about the boundary conditions of the universe come up. We do not know, where these constants, like α or α_s or the lepton and quark masses, come from, but it could well be that at least a few of them are products of the Big Bang, and if the Big Bang would be repeated, these constants would take different values. But if things would go this way, it is clear that the constants could never be calculated.

So in connection to the fundamental constants the question comes up, whether they are really cosmic accidents, or whether they are determined by the dynamics, whether they are changing in time or in space, or whether they are indeed calculable in a hypothetical theory going far above the present Standard Model. Also considerations related to the Anthropic Principle should be made. Life in our universe can exist only if the values of the fundamental constants take on certain values. In a universe in which, for example, the u–quark is heavier than the d-quark, the proton would decay into a neutron, and life would not exist, at least not in a form known to us.

One can also take the attitude, taken e. g. by Einstein, who believed that the constants of nature are all calculable, i. e. fixed by the dynamics. In a letter written in 1950 to Ilse Rosenthal–Schneider he wrote: "Dimensionless constants in the laws of nature, which from the purely logical point of view can just as well have different values, shall not exist. To me, with my "trust in God" this appears to be evident, but there will be few who are of the same opinion."

H. Fritzsch, Fundamental Constants and Their Possible Time Dependence, Lect. Notes Phys. **648**, 107–113 (2004)
http://www.springerlink.com/

One aspect which needs to be stressed is the fact that fundamental constants do not exist a priori, but are depending on the specific theory. Only the underlying theory decides, which parameters are fundamental.

An example is the fine structure constant α, introduced by Sommerfeld in 1916. The inverse of α is, of course, close to the integer number 137. The number excited many physicists. Heisenberg speculated about it. W. Pauli, who was very sick, was able to fix the number of the room in the Zürich clinic, in which he died, to be 137.

On one of the occasions I had lunch with R. Feynman, he told me that every theoretician should have written on the blackboard in his office: 137 – how little we know. When we came back from lunch, I checked Feynman's office, and nothing was written on his blackboard. So I took a piece of chalk and wrote in big letters: 137 – how little we know. One hour later Feynman came to me and thanked me for the remark.

In superstring–theory people might find a way to fix or calculate the fine structure constant. In this connection I like to draw attention again to a formula which was derived in the seventies by Wyler,

$$\alpha = \frac{9}{8\pi^4} \left(\frac{\pi^5}{2^4 5!} \right)^{1/4} . \tag{1}$$

This formula works very well and was derived by considering ratios of certain topological spaces – perhaps in string theory something similar can happen.

Of course, today α is just the interaction constant, describing e. g. electron–scattering at low energies,

$$\alpha^{-1} = 137.035\,999\,76(50) . \tag{2}$$

But it is remarkable. Based on this number, one can calculate all effects in QED to an accuracy of about 1 : 10.000.000, e. g. the magnetic moment of the electron. Of course, QED is only a part of the Standard Model of today, based on a superposition of QCD and the $SU(2) \times U(1)$ – electroweak theory, and α is just one of at least 18 parameters that enter the Standard Model.

One of the fundamental quantities is the proton mass. I should like to stress that the proton mass is a rather complicated object in the Standard Model. The coupling constant of QCD follows in leading order the equation

$$\begin{aligned} \alpha_s \left(Q^2\right) &= \frac{2\pi}{b_0} \left(\frac{Q}{\Lambda} \right) , \\ b_0 \qquad &= 11 - \frac{2}{3} n_f . \end{aligned} \tag{3}$$

Here the scale parameter Λ enters, which has been determined to be

$$\Lambda = 214^{+38}_{-35} \ \text{MeV} . \tag{4}$$

Λ is a free parameter of QCD, and all numbers of QCD scale with Λ, at least in the limit where the masses of the quarks are set to zero. But Λ can be expressed

in terms of MeV, i. e. it is given in reference to the electron mass, which is outside QCD. The physical parameters like the proton mass are simply proportional to Λ. The scale of confinement of the quarks is inversely proportional to Λ. The proportionality constants have been calculated in lattice QCD and seem to be in agreement with the observations.

Let me mention one remarkable success of lattice QCD. The pion decay constant has been calculated:

$$\frac{F_\pi}{\Lambda} = 0.56 \pm 0.05\,. \tag{5}$$

The experimental value is 0.62 ± 0.10, i. e. in good agreement with the theoretical value.

The experimental value of the nucleon mass is 938.272 MeV. However, in the theory there is a contribution, given by the proton expectation value of $m_u \bar{u}u + m_d \bar{d}d$, as the σ–term, which is about 45 MeV, but it is not known to an accuracy of better than about 10%.

Also s–quarks contribute. If we say, in accordance with experiment, that the $\bar{s}s$–pairs should contribute about 10% as much as the $\bar{u}u$ and $\bar{d}d$–pairs, the proton expectation value for $m_s \bar{s}s$ is about 40 MeV, i. e. about as large as the $\bar{u}u/\bar{d}d$–contribution. This implies that the nucleon in a world without s–quarks would have a mass of about 900 MeV. The same is true for charmed quarks. There are, of course, much less charmed quarks in the nucleon than strange quarks, but when it comes to the expectation value of $m_c \bar{c}c$, one finds

$$\langle p | m_c \bar{c}c | p \rangle \sim 30 \text{ MeV}\,,$$

with a large uncertainty.

We can write down the nucleon mass

$$M = c_0 \Lambda + \Delta M\,, \tag{6}$$

where ΔM is defined in such a way that is vanishes when the quarks masses are set to zero and electromagnetic effects are neglected.

If we say, the physical proton mass is 100%, then [1] about 79% are due to Λ, about 4.6% due to the u– and d–quarks, about 10% due to the s–quarks, 6% due to the c–quarks, and about 0.2% due to electromagnetic effects. Something similar can be said about the magnetic moments of the nucleon. It will be dominated by a term given by Λ, while the quark mass terms contribute not much.

I should also remind you that Grand Unification imposes that the parameters α_s, α and α_w are not independent, but are related to each other, and related to the unified coupling constant, describing the interaction at the unification scale Λ_{un}. It is also known that the group $SU(5)$ does not describe the observations, since the three coupling constants do not converge precisely. If supersymmetric particles are added at an energy scale of about 1 TeV, a convergence takes place, however [2]. In $SO(10)$ [3] the situation is different, since in this group the unification is a two–step process, where another mass scale, the mass scale

for the right-handed W–boson, enters. If this mass scale is chosen in the right way, the unification can be achieved without supersymmetry.

After these preparations let me come to the question of time dependence. A group of physicists from Australia, England and the US has recently published their evidence that the fine structure constant had a different value billions of years ago [4]. They were investigating the light from about 150 quasars, being on its way for up to 11 billion years, using the so–called "many multiplet method". They were looking at the fine–structure of atomic lines, originating from elements like Fe, Ni, Mg, Sn, Ag etc.

One particular aspect is that the fine–structure is a rather complex phenomenon, fluctuating in particular also in the sign of the effect. These sign changes have been observed and used in fixing the experimental values of α. The result is

$$\frac{\Delta\alpha}{\alpha} = (-0.72 \pm 0.18) \cdot 10^{-5} \,. \tag{7}$$

Thus α was slightly smaller in the past. If one takes a linear approximation and uses a cosmic lifetime of 14 billion years, the effect is $\dot{\alpha}/\alpha \approx 1.2 \cdot 10^{-15}$ per year.

I should mention that considerations related to a time–dependence of fundamental parameters have a long history. In the 1930s Dirac [5] considered a time–variation of G, and independently also Milne [6]. P. Jordan [7] looked at the time–dependence of other parameters, e. g. nuclear parameters. L. Landau [8] speculated in the 1950s about a time–variation of α.

If α depends on time, the question arises how this time–variation is generated. Since $\alpha = e^2/\hbar c$, a time variation could come from a time variation of $\hbar$ or c. Both cases are, I think, not very likely. If c depends on time, it would mean, that we have a serious problem with relativity. If $\hbar$ would depend on time, atomic physics runs into a problem. So I think that a time dependence of α simply means that e is time–dependent.

Let me also mention that according to the results of Dyson and Damour [9] there is a rather strong constraint on a time–variation of α. If no other parameters change as well, the relative change $(\dot{\alpha}/\alpha)$ per year cannot be more than 10^{-17}, i. e. there is a problem with the astrophysical measurements, unless the rate of change for α has become less during the last 2 billion years. The constraint is derived by looking at the position of a nuclear resonance in Samarium, which cannot have changed much during the last 2 billion years. However, I tend not to take this constraint very seriously. According to the Grand Unification, α_s and Λ should have changed as well, and the two effects (change of α and of Λ) might partially cancel each other.

2 Variation of Fundamental Constants and Grand Unification

The idea of Grand Unification implies that the gauge group $SU(3)$ of the strong interactions and the gauge group $SU(2) \times U(1)$ of the electroweak sector are subgroups of a simple group, which causes the unification.

Both the groups $SU(5)$ and $SO(10)$ are considered in this way. I like to emphasize that the group $SO(10)$ has the nice property that all leptons and quarks of one generation are described by one representation, the 16–representation. It includes a right-handed neutrino, which does not contribute to the normal weak interaction, but it is essential for the appearance of a mass of the neutrino, which is expected in the $SO(10)$–Theory. In $SU(5)$ two representations of the group are needed to describe the leptons and quarks of one generation, a 10– and a $(\bar{5})$–representation.

I should also like to emphasize that the gauge couplings α_s, α_w and α meet in the $SU(5)$–theory only, if one assumes that above about 1 TeV supersymmetry is realized. In the $SO(10)$–theory this is not needed. A convergence of the coupling constants can be achieved, since at high energies another energy scale enters, which has to be chosen in a suitable manner.

A change in time of α can be obtained in two different ways. Either the coupling constant $\alpha_{\rm un}$ stays invariant or the unification scale changes. I consider both effects in the $SU(5)$–model with supersymmetry. In this model the relative changes are related:

$$\frac{1}{\alpha}\frac{\dot{\alpha}}{\alpha} = \frac{8}{3}\frac{1}{\alpha_s} - \frac{10}{\pi}\frac{\dot{\Lambda}_{\rm un}}{\Lambda_{\rm un}} \tag{8}$$

One may consider the following scenarios:

1) Λ_G invariant, $\alpha_u = \alpha_u(t)$. This is the case considered in [10] (see also [11]), and one finds

$$\frac{1}{\alpha}\frac{\dot{\alpha}}{\alpha} = \frac{8}{3}\frac{1}{\alpha_s}\frac{\dot{\alpha}_s}{\alpha_s} \tag{9}$$

and

$$\frac{\dot{\Lambda}}{\Lambda} = -\frac{3}{8}\frac{2\pi}{b_3^{SM}}\frac{1}{\alpha}\frac{\dot{\alpha}}{\alpha}. \tag{10}$$

2) α_u invariant, $\Lambda_G = \Lambda_G(t)$. One finds

$$\frac{1}{\alpha}\frac{\dot{\alpha}}{\alpha} = -\frac{1}{2\pi}\left(b_2^S + \frac{5}{3}b_1^S\right)\frac{\dot{\Lambda}_G}{\Lambda_G}, \tag{11}$$

$$\frac{\dot{\Lambda}}{\Lambda} = \left(\frac{b_3^S}{b_3^{SM}}\frac{1}{\alpha}\frac{\dot{\alpha}}{\alpha}\right) \approx -30.8\frac{\dot{\alpha}}{\alpha}. \tag{12}$$

3) $\alpha_u = \alpha_u(t)$ and $\Lambda_G = \Lambda_G(t)$. One has

$$\begin{aligned}\frac{\dot{\Lambda}}{\Lambda} &= -\frac{2\pi}{b_3^{SM}}\frac{1}{\alpha_u}\frac{\dot{\alpha_u}}{\alpha_u} + \frac{b_3^S}{b_3^{SM}}\frac{\dot{\Lambda}_G}{\Lambda_G} \\ &= -\frac{3}{8}\frac{2\pi}{b_3^{SM}}\frac{1}{\alpha}\frac{\dot{\alpha}}{\alpha} - \frac{3}{8}\frac{1}{b_3^{SM}}\left(b_2^S + \frac{5}{3}b_1^S - \frac{8}{3}b_3^S\right)\frac{\dot{\Lambda}_G}{\Lambda_G} \\ &= 46\frac{\dot{\alpha}}{\alpha} + 1.07\frac{\dot{\Lambda}_G}{\Lambda_G}\end{aligned} \tag{13}$$

where theoretical uncertainties in the factor $R = (\dot{\Lambda}/\Lambda)/(\dot{\alpha}/\alpha) = 46$ have been discussed in [10]. The actual value of this factor is sensitive to the inclusion of the quark masses and the associated thresholds, just like in the determination of Λ. Furthermore higher order terms in the QCD evolution of α_s will play a role. In [10] it was estimated: $R = 38 \pm 6$.

The case in which the time variation of α is not related to a time variation of the unified coupling constant, but rather to a time variation of the unification scale, is of particular interest. Unified theories, in which the Standard Model arises as a low energy approximation, might well provide a numerical value for the unified coupling constant, but allow for a smooth time variation of the unification scale, related in specific models to vacuum expectation values of scalar fields. Since the universe expands, one might expect a decrease of the unification scale due to a dilution of the scalar field. A lowering of Λ_G implies according to (11)

$$\frac{\dot{\alpha}}{\alpha} = -\frac{1}{2\pi}\alpha\left(b_2^S + \frac{5}{3}b_1^S\right)\frac{\dot{\Lambda}_G}{\Lambda_G} = -0.014\frac{\dot{\Lambda}_G}{\Lambda_G}. \tag{14}$$

If $\dot{\Lambda}_G/\Lambda_G$ is negative, α increases in time, consistent with the experimental observation. Taking $\Delta\alpha/\alpha = -0.72 \times 10^{-5}$, we would conclude $\Delta\Lambda_G/\Lambda_G = 5.1 \times 10^{-4}$, i.e. the scale of grand unification about 8 billion years ago was about 8.3×10^{12} GeV higher than today. If the rate of change is extrapolated linearly, Λ_G is decreasing at a rate $\dot{\Lambda}_G/\Lambda_G = -7 \times 10^{-14}$/yr.

According to (12) the relative changes of Λ and α are opposite in sign. While α is increasing with a rate of 1.0×10^{-15}/yr, Λ and the nucleon mass are decreasing, e.g. with a rate of 1.9×10^{-14}/yr. The magnetic moments of the proton μ_p as well of nuclei would increase according to

$$\frac{\dot{\mu}_p}{\mu_p} = 30.8\frac{\dot{\alpha}}{\alpha} \approx 3.1 \times 10^{-14}/\text{yr}. \tag{15}$$

The time variations of the ratio M_p/m_e and α discussed here are such that they could by discovered by precise measurements in quantum optics. The wave length of the light emitted in hyperfine transitions, e.g. the ones used in the cesium clocks being proportional to $\alpha^4 m_e/\Lambda$ will vary in time like

$$\frac{\dot{\lambda}_{hf}}{\lambda_{hf}} = 4\frac{\dot{\alpha}}{\alpha} - \frac{\dot{\Lambda}}{\Lambda} \approx 3.5 \times 10^{-14}/\text{yr} \tag{16}$$

taking $\dot{\alpha}/\alpha \approx 1.0 \times 10^{-15}$/yr. The wavelength of the light emitted in atomic transitions varies like α^{-2},

$$\frac{\dot{\lambda}_{at}}{\lambda_{at}} = -2\frac{\dot{\alpha}}{\alpha}. \tag{17}$$

One has $\dot{\lambda}_{at}/\lambda_{at} \approx -2.0 \times 10^{-15}$/yr. A comparison gives

$$\frac{\dot{\lambda}_{hf}/\lambda_{hf}}{\dot{\lambda}_{at}/\lambda_{at}} = -\frac{4\dot{\alpha}/\alpha - \dot{\Lambda}/\Lambda}{2\dot{\alpha}/\alpha} \approx -17.4. \tag{18}$$

At present the time unit second is defined as the duration of 9.192.631.770 cycles of microwave radiation emitted or absorbed by the hyperfine transmission of cesium-133 atoms. If Λ indeed changes, as described above, it would imply that the time flow measured by the cesium clocks does not fully correspond with the time flow defined by atomic transitions.

Recently a high precision experiment was done at the MPQ in Munich, using the precise cesium clock from Paris. The preliminary result is consistent with no change of the frequencies – one measures for the transition in hydrogen a frequency of 2 466 061 413 187 127 (18) Hz. This gives a change of $2.8(5.7)\cdot 10^{-15}$ per year.

According to (20) the effect should be about ten times larger. Although this result is still preliminary, one is supposed to think what might be the reason for the small effect.

One possibility is, of course, that the astrophysical measurements of the change of α are not correct. Another interesting possibility, however, needs to be studied. It might be that both α_{uu} and Λ_G change such that the result of λ_{hf} is essentially zero – both effects cancel each other in leading order. Nevertheless on the level of 10^{-15} an effect should be seen. More refined experiments are needed to search for a time dependence of Λ.

Acknowledgements

We should like to thank A. Albrecht, J. D. Bjorken, E. Bloom, G. Börner, S. Brodsky, P. Chen, S. Drell, G. Goldhaber, T. Hänsch, M. Jacob, P. Minkowski, A. Odian, H. Walther and A. Wolfe for useful discussions.

References

1. J. Gasser, H. Leutwyler and M. E. Sainio, Phys. Lett. B **253**, 252 (1991).
2. U. Amaldi, W. de Boer and H. Furstenau, Phys. Lett. **260**, 447 (1991).
3. H. Fritzsch and P. Minkowski, Annals of Physics **93**, 193 (1975).
4. J. K. Webb *et al.*, Phys. Rev. Lett. **87** 091301 (2001).
5. P. M. Dirac, Nature **192**, 235 (1937).
6. E. A. Milne. *Relativity, Gravitation and World Structure.* (Clarendon Press, Oxford, (1935).
7. P. Jordan. *Naturwiss., 25*, **513** (1937); Z. Physik, **113**, 660 (1939).
8. L. D. Landau. *Niels Bohr and the Development of Physics.* Ed. by W. Pauli (McGraw-Hill, NY, 1955) 52.
9. T. Damour and F. Dyson, Nucl. Phys. B **480**, 37 (1996).
10. X. Calmet and H. Fritzsch, Eur. Phys. J. C **24**, 639 (2002).
11. P. Langacker, G. Segre and M. J. Strassler, Phys. Lett. B **528**, 121 (2002).
12. T. W. Hänsch, "Laser Frequency Combs and Ultraprecise Spectroscopy," In: *Laser Spectroscopy XVI, Proceedings of ICOLS 2003*, Ed. by P. Hannaford, A. Sidorov, H. Bachor and K. Baldwin (World Scientific, Singapore, 2004).

Quantum Gravity and Fundamental Constants

Claus Kiefer

Institut für Theoretische Physik, Universität zu Köln, Zülpicher Str. 77, 50937 Köln, Germany

Abstract. Two popular approaches towards a quantum theory of gravity are quantum general relativity and string theory ('M-theory'). I review their essential ideas with special emphasis on the role of fundamental constants. A special section is devoted to Kaluza–Klein theories. Most of the 'constants' may have their origin in quantum fields of the fundamental theory and may be space- and time-dependent in their classical limit.

1 Introduction

Quantum theory is a general framework for physical theories and seems to be of a universal nature. It describes successfully all non-gravitational interactions. This holds in particular for the standard model of strong and electroweak interactions, which is a quantum field theory that has so far passed all experimental tests. A final quantum framework does, however, not yet exist for the gravitational field. Gravity is successfully described by a classical theory—Einstein's theory of general relativity (GR), more appropriately called geometrodynamics because gravity is there understood as a manifestation of the geometry of spacetime. It is given by the 'Einstein–Hilbert action', which apart from a surface term reads

$$S_{\rm EH} = \frac{c^4}{16\pi G}\int_{\mathcal{M}} {\rm d}^4x\, \sqrt{-g}\,(R-2\Lambda)\ , \tag{1}$$

where G and Λ denote the gravitational constant and the cosmological constant, respectively. Here, $\mathcal{M}$ denotes a region of the spacetime manifold. In addition one has a 'matter' action $S_{\rm m}$ describing non-gravitational fields coupled to gravity. It gives rise to an energy-momentum tensor

$$T_{\mu\nu} = \frac{2}{\sqrt{-g}}\frac{\delta S_{\rm m}}{\delta g^{\mu\nu}(x)}\ , \tag{2}$$

which acts as a 'source' for the gravitational field. The sum of both actions gives rise to the Einstein field equations.

In spite of its success, there are many reasons to believe that GR must be superseded at the most fundamental level by a *quantum* theory of gravity. Since no experimental hints are available up to now, the reasons are therefore of a pure theoretical nature. The main motivations for quantum gravity are the following, see [1] and [2] for more details and references:

C. Kiefer, Quantum Gravity and Fundamental Constants, Lect. Notes Phys. **648**, 115–127 (2004)
http://www.springerlink.com/

- **Unification.** The history of science shows that a reductionist viewpoint has been very fruitful in physics. The standard model of particle physics is a *quantum* field theory which has united in a certain sense all non-gravitational interactions. The universal coupling of gravity to all forms of energy would make it plausible that gravity has to be implemented in a quantum framework, too. Moreover, attempts to construct an exact semiclassical theory, where gravity stays classical but all other fields are quantum, have failed up to now. This demonstrates in particular that classical and quantum *concepts* (phase space versus Hilbert space, etc.) are most likely incompatible.
- **Cosmology and black holes.** As the *singularity theorems* and the ensuing breakdown of general relativity demonstrate, a fundamental understanding of the early universe—in particular its initial conditions near the 'big bang'—and of the final stages of black-hole evolution requires a more fundamental theory. From the historical analogue of quantum mechanics (which due to its stationary states has rescued the atoms from collapse) the general expectation is that this encompassing theory is a *quantum* theory. It must be emphasised that *if* gravity is quantised, the kinematical non-separability of quantum theory demands that the whole universe must be described in quantum terms. This leads to the concepts of quantum cosmology and the wave function of the universe, see below.
- **Problem of time.** Quantum theory and general relativity (in fact, every general covariant theory) contain drastically different concepts of time (and spacetime). Strictly speaking, they are incompatible. In quantum theory, time is an external (absolute) element, *not* described by an operator (in special relativistic quantum field theory, the role of time is played by the external Minkowski spacetime). In contrast, spacetime is a dynamical object in general relativity. It is clear that a unification with quantum theory must lead to modifications of the concept of time. Related problems concern the role of background structures in quantum gravity, the role of the diffeomorphism group (Poincaré invariance, as used in ordinary quantum field theory, is no longer a symmetry group), and the notion of 'observables'.

What are the relevant scales on which effects of quantum gravity should be unavoidable? As has already been shown by Max Planck in 1899, the fundamental constants speed of light (c), gravitational constant (G), and quantum of action ($\hbar$) can be combined in a unique way (up to a dimensionless factor) to yield units of length, time, and mass. In Planck's honour they are called Planck length, $l_{\rm P}$, Planck time, $t_{\rm P}$, and Planck mass, $m_{\rm P}$, respectively. One can argue on general grounds that a fundamental theory should contain three dimensionful constants, cf. Okun's contribution to this volume. The Planck units read

$$l_{\rm P} = \sqrt{\frac{\hbar G}{c^3}} \approx 1.62 \times 10^{-33}\ {\rm cm}\ , \tag{3}$$

$$t_{\rm P} = \frac{l_{\rm P}}{c} = \sqrt{\frac{\hbar G}{c^5}} \approx 5.40 \times 10^{-44}\ {\rm s}\ , \tag{4}$$

$$m_{\mathrm{P}} = \frac{\hbar}{l_{\mathrm{P}} c} = \sqrt{\frac{\hbar c}{G}} \approx 2.17 \times 10^{-5}\ \mathrm{g} \approx 1.22 \times 10^{19}\ \mathrm{GeV}\ . \tag{5}$$

The Planck mass seems to be a rather large quantity by microscopic standards. One has to keep in mind, however, that this mass (energy) must be concentrated in a region of linear dimension l_{P} in order to see direct quantum-gravity effects. In fact, the Planck scales are attained for an elementary particle whose Compton wavelength is (apart from a factor of 2) equal to its Schwarzschild radius,

$$\frac{\hbar}{m_{\mathrm{P}} c} \approx R_{\mathrm{S}} \equiv \frac{2 G m_{\mathrm{P}}}{c^2}\ ,$$

which means that the spacetime curvature of an elementary particle is non-negligible. A truly unified theory may contain different parameters. An example is string theory (see below) where the fundamental 'string length' l_{s} appears.

A quantity expressing the ratio of atomic scales to the Planck scale is the 'fine-structure constant of gravity' defined by

$$\alpha_{\mathrm{g}} = \frac{G m_{\mathrm{pr}}^2}{\hbar c} \equiv \left(\frac{m_{\mathrm{pr}}}{m_{\mathrm{P}}} \right)^2 \approx 5.91 \times 10^{-39}\ , \tag{6}$$

where m_{pr} denotes the proton mass. Its smallness is responsible for the unimportance of quantum-gravitational effects on laboratory and astrophysical scales, and for the separation between micro- and macrophysics. It is interesting that structures in the universe occur for masses which can be expressed as simple powers of α_{g} in units of m_{pr}, cf. [3]. For example, stellar masses are of the order $\alpha_{\mathrm{g}}^{-3/2} m_{\mathrm{pr}}$, while stellar lifetimes are of the order $\alpha_{\mathrm{g}}^{-3/2} t_{\mathrm{P}}$. It is an open question whether a fundamental theory of quantum gravity can provide an explanation for such values, e.g. for the ratio $m_{\mathrm{pr}}/m_{\mathrm{P}}$, or not. If not, only an anthropic principle could yield an 'explanation', but this is usually not considered as satisfactory.

What are the major expectations from quantum gravity [1]?

- As just mentioned, it should give an explanation for the origin of masses and coupling constants. It should, in particular, give an explanation for the observed small positive value of the cosmological constant Λ (or of the underlying field, if applicable). The theory could also give an explanation for a possible spatiotemporal variation of these variables.
- Quantum gravity should provide a full quantum description of black holes. This includes both the final evaporation process due to Hawking radiation and a quantum-statistical foundation of the Bekenstein-Hawking entropy

$$S_{\mathrm{BH}} = \frac{k_{\mathrm{B}} c^3 A}{4 G \hbar}\ , \tag{7}$$

  where A is the surface of the event horizon.
- Quantum gravity should yield an explanation for the origin of the universe, providing a singularity-free description. It should also shed some light on the 'probability' of inflation.

- It should exhibit the structure of spacetime at the smallest scales. Preliminary approaches in fact indicate that this structure is discrete. This could be experimentally testable. Quantum gravity should also provide information about the number of spacetime dimensions.

There are at present two main approaches in constructing a theory of quantum gravity. The first originates from the idea that one can construct separately a quantum theory of the gravitational field, without necessarily invoking a unification of interactions. Methodologically, one attempts to quantise general relativity. The second approach assumes that the problem of quantum gravity can only be solved within a unified theory of all interactions—this is string theory. I will briefly review these approaches in the next two sections, respectively.

2 Quantum General Relativity

Approaches to quantise general relativity are usually divided into 'covariant' and 'canonical' formulations. Whereas in the former a four-dimensional viewpoint is preserved in the formalism, the latter employ a split into space and time already at the classical level. A particular covariant approach is the path-integral formulation, in which all information about the quantum theory can be obtained from a sum over four-dimensional metrics,

$$Z[g] = \int \mathcal{D}g_{\mu\nu}(x)\ \mathrm{e}^{\mathrm{i}S[g_{\mu\nu}(x)]/\hbar}\ . \tag{8}$$

Here, the sum runs over all metrics on a four-dimensional manifold $\mathcal{M}$ divided by the diffeomorphism group Diff$\mathcal{M}$. A concrete evaluation can be tried, for example, by the method of dynamical triangulation, cf. Loll's contribution to [2]. Based on path-integral methods one can also attempt to study renormalisation-group equations for quantum gravity. This can render G and Λ 'running coupling constants', which would make them dependent on spacetime distance. Preliminary results even indicate how the observed small non-vanishing value of Λ could emerge from such considerations, cf. [4].

In this section I shall concentrate myself on the canonical approach. Since this requires a Hamiltonian formulation of the theory, one has to split the classical spacetime into space and time in order to define canonical momenta. This spoils the explicit four-dimensional covariance of general relativity—the theory is reformulated in such a way that the dynamics of *three*-dimensional hypersurfaces plays a central role. It is then not surprising that the configuration variable is the *three*-dimensional metric, $h_{ab}(\vec{x})$, on such hypersurfaces. This three-metric has six independent degrees of freedom. The remaining four components of the spacetime metric play the role of non-dynamical Lagrange multipliers called lapse function, $N^{\perp}(\vec{x})$, and shift vector, $N^a(\vec{x})$—they parametrise, respectively, the way in which consecutive hypersurfaces are chosen and how the coordinates are selected *on* a hypersurface. The momenta canonically conjugated to the three-metric, $p^{ab}(\vec{x})$, form a tensor which is linearly related to the second fundamental form associated with a hypersurface, which specifies the way in which the

hypersurface is embedded into the fourth dimension. In the quantum theory, the canonical variables are formally turned into operators obeying the commutation relations

$$[\hat{h}_{ab}(\vec{x}), \hat{p}^{cd}(\vec{y})] = \mathrm{i}\hbar\delta^{c}_{(a}\delta^{d}_{b)}\delta(\vec{x},\vec{y}) \ . \tag{9}$$

In a (formal) functional Schrödinger representation, the canonical operators act on wave functionals Ψ depending on the three-metric,

$$\hat{h}_{ab}(\vec{x})\Psi[h_{ab}(\vec{x})] = h_{ab}(\vec{x})\Psi[h_{ab}(\vec{x})] \ , \tag{10}$$

$$\hat{p}^{cd}(\vec{x})\Psi[h_{ab}(\vec{x})] = \frac{\hbar}{\mathrm{i}}\frac{\delta}{\delta h_{cd}(\vec{x})}\Psi[h_{ab}(\vec{x})] \ . \tag{11}$$

A central feature of canonical gravity is the existence of constraints. Because of the four-dimensional diffeomorphism invariance of general relativity, these are four constraints per space point, one Hamiltonian constraint,

$$\hat{\mathcal{H}}_{\perp}\Psi = 0 \ , \tag{12}$$

and three diffeomorphism constraints,

$$\hat{\mathcal{H}}_{a}\Psi = 0 \ . \tag{13}$$

These diffeomorphism constraints guarantee that the wave functional is independent of the choice of coordinates on space. The total Hamiltonian is obtained by integration (here I restrict myself for simplicity to closed compact spaces; otherwise, the Hamiltonian has to be augmented by surface terms containing the total (ADM) energy),

$$\hat{H} = \int \mathrm{d}^3x \ (N^{\perp}\hat{\mathcal{H}}_{\perp} + N^{a}\hat{\mathcal{H}}_{a}) \ , \tag{14}$$

where $N^{\perp}$ and N^{a} denote again lapse function and shift vector, respectively. The constraints then enforce that the wave functional be annihilated by the total Hamiltonian,

$$\hat{H}\Psi = 0 \ . \tag{15}$$

The *Wheeler–DeWitt equation* (15) is the central equation of canonical quantum gravity. It has here been formulated in the traditional version using the three-metric as the configuration variable. This is not mandatory, since one can perform a canonical transformation at the classical level leading to a new configuration variable that is a mixture of the three-metric and its canonical momentum. A popular version uses an $SU(2)$ connection variable which is similar in structure to a Yang-Mills connection. A closely related version employs the integral of this connection along loops, similar to a Wilson loop. This 'loop quantum gravity' leads to intriguing results such as the existence of a discrete spectrum for geometric operators like area and volume, cf. Thiemann's contribution to [2]. It turns out that scales below the Planck length (3) have no operational meaning. In this way the theory sheds light on the fundamental structure of spacetime.

The Wheeler–DeWitt equation (15) possesses the remarkable property that it does not depend on any external time parameter—the t of the time-dependent Schrödinger equation has totally disappeared, and (15) looks like a stationary zero-energy Schrödinger equation. How can this be understood? In classical canonical gravity, a spacetime can be represented as a 'trajectory' in configuration space—the space of all three-metrics. Although time coordinates have no intrinsic meaning in classical general relativity either, they can nevertheless be used to parametrise this trajectory in an essentially arbitrary way. Since no trajectories exist anymore in quantum theory, no spacetime exists at the most fundamental level, and therefore also no time coordinates to parametrise any trajectory. A simple analogy is provided by the relativistic particle: in the classical theory there is a trajectory which can be parametrised by some essentially arbitrary parameter, e.g. the proper time. Reparametrisation invariance leads to one constraint, $p^2 + m^2 = 0$. In the quantum theory, no trajectory exists anymore, the wave function obeys the Klein–Gordon equation as an analogue of (15), and any trace of a classical time parameter is lost (although, of course, for the relativistic particle the background Minkowski spacetime is present, which is not the case for gravity).

Since the presence of an external time parameter is very important in quantum mechanics—giving rise to such important notions as the unitarity of states—it is a priori not clear how to interpret a 'timeless' equation of the form (15). This reflects the *problem of time* listed in the introduction, cf. [1] for details and references. A related issue is the *Hilbert-space problem*: what is the appropriate inner product that encodes the probability interpretation and that is conserved in time? These questions still lack an answer at the most fundamental level. It is not clear, for example, whether (15) can sensibly be interpreted only as an eigenvalue equation for eigenvalue zero or whether it can have a more general meaning. The present options are thus to study either a semiclassical approximation and to aim at a consistent treatment of conceptual issues only at that level, or to look for sensible boundary conditions for the Wheeler–DeWitt equation and to discuss directly solutions of this equation. This is usually employed in a cosmological context, one example being the 'no-boundary condition' of Hartle and Hawking.

How are fundamental constants treated in canonical quantum gravity? Since this approach does not include a unification of interactions it seems that these constants play the same role as in classical gravity coupled to the standard model of particle physics. Apart from the above-mentioned possibility of a discrete spacetime structure at the Planck scale, mathematical consistency may lead, however, to additional restrictions on the fundamental constants, for example in the form of 'quantisation conditions'. Consider, e.g., a simple model for the Wheeler–DeWitt equation (15). As in classical cosmology, a considerable simplification is achieved by imposing homogeneity and isotropy. If a denotes the scale factor of the universe and χ a homogeneous scalar field, one can obtain a Wheeler–DeWitt equation of the following form of an 'indefinite oscillator', cf. [5],

$$\hat{H}\psi(a,\chi) \equiv \left(\frac{\partial^2}{\partial a^2} - \frac{\partial^2}{\partial \chi^2} - \omega_a^2 a^2 + \omega_\chi^2 \chi^2 \right) \psi(a,\chi) = 0 \ , \tag{16}$$

where ω_a and ω_χ are parameters of this model (mimicking fundamental constants) and a convenient choice of units has been made. In such a cosmological context, ψ is often referred to as 'wave function of the universe'. As long as (16) is only treated as a differential equation without additional constraints, these parameters are arbitrary. If one demands, on the other hand, that the solution be normalisable with respect to the usual quantum-mechanical Hilbert space (this of course is connected with the unsolved Hilbert-space problem mentioned above), the following commensurability condition results for the parameters,

$$\frac{\omega_\chi}{\omega_a} = \frac{2n+1}{2k+1} \ , \tag{17}$$

where n and k are natural numbers. In a more realistic model, a quantisation of this kind may be obtained e.g. for the cosmological constant, explaining perhaps its small non-vanishing value. This conclusion would follow, for example, from a quantisation condition of the type

$$\sqrt{\Lambda} L = n + 1 \ , \ n \in \mathbf{N}_0 \ , \tag{18}$$

where L could be the length scale of a field in the Hamiltonian. (For example, it could correspond to the Compton wavelength of a massive neutrino, which could have the correct magnitude to explain the observed value of Λ). However, the details remain so far in the realm of speculation.

3 Superstring Theory ('M-theory')

Superstring theory is perhaps the most ambitious approach towards a unified theory of all interactions including gravity. In contrast to quantum general relativity, this theory is *not* obtained by quantising a classical theory (except in a purely methodological sense). Quantum gravity as such emerges only in a low-energy approximation. In the following I shall briefly review the quantum-gravitational aspects of string theory, see also [1] and Mohaupt's contribution [2]. A comprehensive discussion of string theory is [6].

String theory transcends the level of local field theory because its fundamental objects are one-dimensional entities ('strings') instead of fields defined at spacetime points. More recently, it has turned out that higher-dimensional objects ('branes') appear within string theory in a natural way and on an equal footing with strings. From the existence of branes one can infer that there must be non-perturbative dualities relating different string theories and describing their strong-coupling behaviour. This led to the speculation that a fundamental 'M-theory' might exist from which the different possible string theories can be obtained in appropriate limits.

The main features of string theory are the following.

- String theory necessarily contains gravity. The graviton appears as an excitation of closed strings. Open strings do not contain the graviton by themselves, but since they contain closed strings as virtual contributions, the appearance of the graviton is unavoidable there, too.

- String theory necessarily leads to gauge theories since the corresponding gauge bosons are found in the string spectrum.
- String theory seems to need supersymmetry for a consistent formulation. Fermions are therefore an essential ingredient.
- All 'particles' arise from string excitations. Therefore, they are no longer fundamental and their masses should in principle be fixed with respect to the string mass scale.
- Higher spacetime dimensions appear in a natural way, thus implementing the old idea by Kaluza and Klein.
- String theory leads to a unified quantum description of all interactions.
- Since one can get chiral gauge couplings from string theory, the hope is raised that one can derive the standard model of elementary particles from it (although one is still very far from having achieved this).

Heuristically, one starts from a classical string propagating in a general curved D-dimensional spacetime. The corresponding action is taken to read (for simplicity only the bosonic string (no supersymmetry) is considered)

$$S_{\rm s} = -\frac{1}{4\pi\alpha'}\int {\rm d}^2\sigma \ \Big(\sqrt{h}h^{\alpha\beta}\partial_\alpha X^\mu \partial_\beta X^\nu g_{\mu\nu}(X) \\ -\alpha'\sqrt{h}\,{}^{(2)}R\Phi + \epsilon^{\alpha\beta}\partial_\alpha X^\mu \partial_\beta X^\nu B_{\mu\nu}(X)\Big) \ , \tag{19}$$

where the integration runs over the string worldsheet. Here, α' is the fundamental parameter of string theory and is related to the string length $l_{\rm s}$ by

$$l_{\rm s} = \sqrt{\alpha'\hbar c} \ . \tag{20}$$

The fields $g_{\mu\nu}$ (D-dimensional metric of the embedding space), Φ ('dilaton'), and $B_{\mu\nu}$ (antisymmetric tensor field) are *background fields*, i.e. they will not be integrated over in the path integral. They are included in this action because they automatically arise as string excitations. The fields X^μ define the embedding of the worldsheet into the D-dimensional space which is also called 'target space'. The metric on the worldsheet is denoted by $h_{\alpha\beta}$, with h denoting its determinant and ${}^{(2)}R$ the corresponding Ricci scalar. I emphasise that (19) leads to a quantum field theory on the worldsheet, not the target space. For the latter one uses the effective action discussed below.

In string theory it has been proven fruitful to employ a path-integral approach, cf. (8). In the Euclidean formulation the starting point would be (setting $\hbar = c = 1$ from now on)

$$Z = \int \mathcal{D}X\mathcal{D}h \ {\rm e}^{-S_{\rm P}} \ , \tag{21}$$

where X and h are a shorthand for the embedding variables and the worldsheet metric, respectively. Only these variables are to be integrated over. The sum over all configurations in (21) contains a sum over all worldsheets, i.e. a sum over all Riemann surfaces, where all topologies have to be taken into account. In this way the string interactions arise—as amplitudes in the path integral.

An important criterion for the consistency of string theory is the absence of quantum anomalies, being here the preservation of Weyl symmetry at the quantum level (the classical action (19) is invariant under a Weyl transformation on the worldsheet, i.e. under the transformation $h_{\alpha\beta} \to \Omega^2(\sigma) h_{\alpha,\beta}$). For a bosonic string propagating in flat spacetime this leads to the restriction $D = 26$. What about the string in a curved spacetime as described by (19)? The requirement that no Weyl anomaly be present on the worldsheet leads to the following set of consistency equations,

$$0 = R_{\mu\nu} - \frac{1}{4} H_{\mu}{}^{\lambda\rho} H_{\nu\lambda\rho} + 2\nabla_\mu \nabla_\nu \Phi + \mathcal{O}(\alpha') \ , \tag{22}$$

$$0 = \nabla_\lambda H^{\lambda}{}_{\mu\nu} - 2\nabla^\lambda \Phi H_{\lambda\mu\nu} + \mathcal{O}(\alpha') \ , \tag{23}$$

$$0 = \frac{D-26}{6\alpha'} + \nabla_\mu \nabla^\mu \Phi - \frac{1}{2} \nabla_\mu \Phi \nabla^\mu \Phi - \frac{1}{24} H_{\mu\nu\rho} H^{\mu\nu\rho} + \mathcal{O}(\alpha') \ . \tag{24}$$

Here we have introduced the field strength $H_{\mu\nu\rho}$ connected with the antisymmetric tensor field,

$$H_{\mu\nu\rho} = 3!\ \partial_{[\mu} B_{\nu\rho]} \ .$$

The consistency equations (22)–(24) follow as field equations from an *effective action* in D spacetime dimensions,

$$S_{\text{eff}} = \frac{1}{2\kappa_0^2} \int \mathrm{d}^D x\ \sqrt{-g}\ \mathrm{e}^{-2\Phi} \Big(R - \frac{2(D-26)}{3\alpha'} - \frac{1}{12} H_{\mu\nu\rho} H^{\mu\nu\rho} + 4\nabla_\mu \Phi \nabla^\mu \Phi + \mathcal{O}(\alpha') \Big) \ . \tag{25}$$

Higher orders exhibit powers of α' and are thus genuine string corrections. They contain, for example, higher-curvature terms, which means that one has to replace in (22)

$$R_{\mu\nu} \to R_{\mu\nu} + \frac{\alpha'}{2} R_{\mu\kappa\lambda\tau} R_{\nu}{}^{\kappa\lambda\tau} + \dots \ .$$

The approximation of a classical spacetime metric is well-defined only if the curvature scale r_c associated with it obeys

$$r_c \gg l_{\text{s}} \ .$$

The expansion of the effective action into powers of α' is thus a low-energy expansion.

As emphasised above, spacetime metric, dilaton, and the antisymmetric tensor field play the role of background fields. The simplest solution for them is

$$g_{\mu\nu} = \eta_{\mu\nu} \ , \quad B_{\mu\nu} = 0 \ , \quad \Phi = \text{const.} = \lambda \ .$$

It is usually claimed that quite generally the stationary points of S_{eff} correspond to possible ground states ('vacua') of the theory. It is clear from (24) that $D = 26$ is a necessary condition for the solution with constant background fields. Thus we have recovered the old consistency condition for the string in flat spacetime. There are now, however, solutions of (24) with $D \neq 26$ and $\Phi \neq$ constant,

which would correspond to a solution with a large cosmological constant $\propto (D-26)/6\alpha'$, in conflict with observation.

The parameter κ_0 in (25) does not have a physical significance by itself since it can be changed by a shift in the dilaton (assumed to be constant). The physical gravitational constant (in D dimensions) reads

$$8\pi G_D = 2\kappa_0^2 \mathrm{e}^{2\lambda} \ . \tag{26}$$

Apart from α'-corrections one can also consider loop corrections to (25). It turns out that the loop expansion of (21) is an expansion in terms of the string coupling defined by $g \equiv \exp(\lambda)$. Since g is determined by the value of the dilaton, the tree-level action (25) is of order g^{-2}. The one-loop approximation is obtained at order g^0, the two-loop approximation at order g^2, and so on (this holds for closed strings. For open strings also odd orders of the coupling appear).

Since the string amplitude contains the parameter α' and the effective action contains the gravitational constant G_D, a comparison of scattering amplitudes yields a connection between both,

$$G_D \sim g^2 l_{\mathrm{s}}^{D-2} \ . \tag{27}$$

An analogous relation holds between gauge couplings for grand unified theories and the string length.

Since we do not live in 26 dimensions (or 10 dimensions in case of the superstring), a connection must be made to the four-dimensional world. This is usually done through compactification of the additional dimensions. In this way one obtains a relation between the four-dimensional gravitational constant and the string length,

$$G \sim g^2 l_{\mathrm{s}}^2 \ , \tag{28}$$

in which the details of the compactification enter (geometric factors). Ideally, one would like to recover in this way other parameters such as particle masses or the number of families. One is, however, still far away from this goal.

4 Kaluza–Klein Theories

An important ingredient in string theory is the existence of higher dimensions. This is in fact an old idea, going back to the pioneering papers of Kaluza and Klein in the 1920s. It is still used independently of string theory and is of special relevance for the possibility of varying fundamental constants. For this reason the basic ideas of the Kaluza–Klein scenario will be discussed in this section. A collection of reviews and translation of the original papers can be found in [7]. In the simplest version there is one additional space dimension, which is compactified to a circle with circumference $2\pi R \equiv L$. Spacetime would in this case be five-dimensional. We label the usual four dimensions by coordinates x^μ and the fifth dimension by y. One can easily get in such a scenario particle masses in four dimensions from a massless five-dimensional field. Assuming for

simplicity that the metric is flat, the dynamical equation for a massless scalar field Φ in five dimensions is given by the wave equation

$$\Box_5 \Phi(x^\mu, y) = 0 \ , \tag{29}$$

where $\Box_5$ is the five-dimensional d'Alembert operator. Making a Fourier expansion with respect to the fifth dimension,

$$\Phi(x^\mu, y) = \sum_n \varphi_n(x^\mu) \mathrm{e}^{\mathrm{i}ny/R} \ , \ n \in \mathbf{Z} \ , \tag{30}$$

one obtains for the $\varphi_n(x^\mu)$ an effective equation of the form

$$\left(\Box_4 - \frac{n^2}{R^2} \right) \varphi_n(x^\mu) = 0 \ , \tag{31}$$

where $\Box_4$ is the four-dimensional d'Alembert operator. Equation (31) is nothing but the four-dimensional Klein-Gordon equation for a massive scalar field $\varphi_n(x^\mu)$ with mass

$$m_n = \frac{|n|}{R} \ . \tag{32}$$

From the four-dimensional point of view one thus has a whole 'Kaluza–Klein tower' of particles with increasing masses. For low energies $E < 1/R$ the massive Kaluza–Klein modes remain unexcited and only the massless mode for $n = 0$ remains. The higher dimensions only show up for energies beyond $1/R$. Since no evidence has been seen yet at accelerators for the massive modes, the size of the fifth dimension must be very small, definitely smaller than about 10^{-17} cm.

The reduction from five to four dimensions for the action proceeds as follows. One starts from the Einstein–Hilbert action in five dimensions,

$$S_5 = \frac{1}{\kappa^2 L} \int \mathrm{d}^5 x \ \sqrt{-g_5} R_5 \ , \tag{33}$$

where $\kappa^2 = 16\pi G$, and the index '5' describes the corresponding quantity in five dimensions (the cosmological constant is neglected here). Writing the five-dimensional line element in the form

$$\phi^{1/3} \mathrm{d}s_5^2 = g_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu - \phi \left(\mathrm{d}y + \kappa A_\mu \mathrm{d}x^\mu \right)^2 \ , \tag{34}$$

where all functions depend on the x^μ only, one recognises that a coordinate transformation in the fifth dimension of the form

$$y \longrightarrow y + \lambda(x^\mu) \ ,$$

just corresponds to a 'gauge transformation' for the field A_μ,

$$A_\mu \longrightarrow A_\mu + \kappa^{-1} \partial_\mu \lambda \ .$$

It is in this sense that Kaluza had intended to unify gravity with electromagnetism, since one can identify A_μ with the electromagnetic 'vector potential'.

Inserting (34) into (33) and integrating over the fifth dimension, one obtains the following four-dimensional action,

$$S = \int \mathrm{d}^4x \, \sqrt{-g} \left(\frac{R}{\kappa^2} + \frac{1}{4}\phi F_{\mu\nu}F^{\mu\nu} - \frac{(\nabla\phi)^2}{6\kappa^2\phi^2} \right) . \tag{35}$$

For $\phi =$ constant this would describe the coupling of the electromagnetic field to gravity (one has to set $\phi = -e^{-2}$, where e is the electric charge). In general, however, the scalar field is not expected to be constant. This would then lead to an electric charge varying in space and time and could be the reason for the observed cosmic variation of the fine-structure constant, cf. [8] and the references therein, as well as other contributions to this volume.

Note that (35) would lead to a violation of the weak equivalence principle, since the coupling to the electromagnetic field is ϕ-dependent instead of being a universal constant such as e. Such a violation could be within experimental reach [9]. The action (35) is the prototype of an action describing 'scalar-tensor theories of gravity', cf. [10] and the references therein. Among the early motivations for a scalar-tensor theory were the possible presence of higher dimensions, but also the possibility to get a time-varying gravitational constant ('Brans-Dicke theory').

Recent developments in the study of higher dimensions have circled around the 'branes' mentioned above. Many models have been discussed, which may originate from string theory or not. A simple example is a four-dimensional brane, representing our observable universe, which is embedded into a five-dimensional spacetime with a negative cosmological constant (the 'Randall-Sundrum model', see e.g. [11] and the references therein). Whereas gravity is present in all five dimensions, it turns out that the non-gravitational interactions are confined to the brane. For this reason it is possible that the size of the extra dimensions is much higher than in the original Kaluza–Klein scenario. The masses of the corresponding higher-dimensional modes could then be much smaller than the Planck mass and therefore susceptible to experimental observation. One effect out of many would be the mixing between the usual graviton and a massive graviton leading to oscillations analogous to neutrino oscillations [11].

5 Conclusion

Quantum theory in an external background cannot be compatible with general relativity at the most fundamental level. The search for a quantum theory of gravity is among the biggest open problems in physics. In this contribution I have briefly summarised the main motivations and reviewed the currently most popular approaches.

The situation with respect to fundamental constants can be summarised as follows. The fundamental theories of quantum gravity seem to contain three dimensionful constants, which can be $\hbar, c, l_{\mathrm{P}}$ for quantum general relativity and $\hbar, c, l_{\mathrm{s}}$ for string theory. One can introduce a time dependence for these constants by hand, but this is rather ad hoc and not motivated by the structure of those theories. In quantised general relativity, no unification occurs, so the physical constants appear in the same form as without quantum gravity. However,

mathematical consistency of the quantum theory may lead to constraints on the constants, which could occur in the form of quantisation conditions. On the other hand, in string theory all physical 'constants' except the mentioned three are fundamentally described by quantum fields. In the classical limit they may be time-and space-dependent and could lead, for example, to the phenomenon of a varying fine-structure constant. It is at present an open issue whether the ensuing 'constants' can be calculated from the fundamental theory or whether the anthropic principle has to be invoked.

An interesting feature is the quantum-to-classical transition: how do the fundamental quantum fields describing physical 'constants' assume classical properties? The key to the answer is 'decoherence'—the irreversible emergence of classical properties through interaction with irrelevant degrees of freedom, see [12]. In this manner one could get, for example, the classical field representing the gravitational constant G out of a fundamental Brans-Dicke type of field [13]. Even after decoherence has occurred, a finite quantum variance could still remain for such a field. One might speculate that this lies behind the fact that G is only poorly known, with some measurements giving a value lying outside the range of previous measurements, see [9] for a review of the experimental situation. The future will see interesting developments, both experimentally and theoretically, in understanding the relation between quantum gravity and fundamental constants.

References

1. C. Kiefer. *Quantum Gravity*. Oxford University Press, Oxford, to appear (2004).
2. D. Giulini, C. Kiefer, and C. Lämmerzahl, (ed.). *Quantum Gravity: From Theory to Experimental Search*. Springer, Berlin (2003).
3. M. Rees. *Perspectives in astrophysical cosmology*. Cambridge University Press, Cambridge (1995).
4. M. Reuter and F. Saueressig, Phys. Rev. D **66**, 125001 (2002).
5. C. Kiefer, Nucl. Phys. B **341**, 273 (1990).
6. J. Polchinski. *String theory. Vols I and II*. Cambridge University Press, Cambridge (1998).
7. H. C. Lee (ed.). *An introduction to Kaluza–Klein theories*. World Scientific, Singapore (1984).
8. J. D. Barrow, Astrophys. Space Sci. **283**, 645 (2003).
9. J.-P. Uzan, Rev. Mod. Phys. **75**, 403 (2003).
10. Y. Fujii and K. Maeda. *The Scalar-Tensor Theory of Gravitation*. Cambridge University Press, Cambridge (2003).
11. A. O. Barvinsky, A. Yu. Kamenshchik, A. Rathke, and C. Kiefer, Ann. Phys. (Leipzig) **12**, 343 (2003).
12. E. Joos, H. D. Zeh, C. Kiefer, D. Giulini, J. Kupsch, and I.-O. Stamatescu. *Decoherence and the appearance of a classical world in quantum theory*, (2nd edn). Springer, Berlin (2003). See also `http://www.decoherence.de`.
13. C. Kiefer and E. A. Martínez, Class. Quantum Grav. **10**, 2511 (1993).

Constraining Variations in the Fine-Structure Constant, Quark Masses and the Strong Interaction

Michael T. Murphy[1], Victor V. Flambaum[2], John K. Webb[2], Vladimir V. Dzuba[2], Jason X. Prochaska[3], and Arthur M. Wolfe[4]

[1] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
[2] School of Physics, University of New South Wales, Sydney N.S.W. 2052, Australia
[3] UCO-Lick Observatory, University of California, Santa Cruz, Santa Cruz, CA 95464, USA
[4] Department of Physics and Centre for Astrophysics and Space Sciences, University of California, San Diego, C-0424, La Jolla, CA 920923, USA

Abstract. We present evidence for variations in the fine-structure constant from Keck/HIRES spectra of 143 quasar absorption systems over the redshift range $0.2 < z_{\rm abs} < 4.2$. This includes 15 new systems, mostly at high-z ($z_{\rm abs} > 1.8$). Our most robust estimate is a weighted mean $\Delta\alpha/\alpha = (-0.57 \pm 0.11) \times 10^{-5}$. We respond to recent criticisms of the many-multiplet method used to extract these constraints. The most important potential systematic error at low-z is the possibility of very different Mg heavy isotope abundances in the absorption clouds and laboratory: *higher* abundances of ^{25,26}Mg in the absorbers may explain the low-z results. Approximately equal mixes of ^{24}Mg and ^{25,26}Mg are required. Observations of Galactic stars generally show *lower* ^{25,26}Mg isotope fractions at the low metallicities typifying the absorbers. Higher values can be achieved with an enhanced population of intermediate mass stars at high redshift, a possibility at odds with observed absorption system element abundances. At present, all observational evidence is consistent with the varying-α results.

Another promising method to search for variation of fundamental constants involves comparing different atomic clocks. Here we calculate the dependence of nuclear magnetic moments on quark masses and obtain limits on the variation of α and $m_{\rm q}/\Lambda_{\rm QCD}$ from recent atomic clock experiments with hyperfine transitions in H, Rb, Cs, Hg^+ and an optical transition in Hg^+.

1 Introduction

The last decade has seen the idea of varying fundamental constants receive unprecedented attention. The historical and modern theoretical motivations for varying constants, as well as the current experimental constraints, are reviewed in [1], the many articles in [2] and this proceedings. Here we set experimental constraints on variations in two fundamental quantities, the fine-structure constant ($\alpha \equiv e^2/\hbar c$; Sect. 2) and the ratio of quark masses to the quantum chromodynamic (QCD) scale ($m_{\rm q}/\Lambda_{\rm QCD}$; Sect. 3), from optical quasar absorption spectra and laboratory atomic clocks respectively.

M.T. Murphy et al., Constraining Variations in the Fine-Structure Constant, Quark Masses and the Strong Interaction, Lect. Notes Phys. **648**, 131–150 (2004)
http://www.springerlink.com/

2 Varying α from Quasar Absorption Lines

To serve the broad readership of this proceedings, we first outline the salient features of optical quasar absorption spectroscopy. Section 2.2 discusses in some detail the recently introduced many-multiplet (MM) method of constraining varying α from optical quasar absorption spectra. All optical data studied here were obtained at the Keck I 10-m telescope on Mauna Kea, Hawaii, with the High Resolution Echelle Spectrograph [3]. When applied to these data, this is the only method to date to yield internally robust evidence for a varying-α. We summarize our spectral analysis techniques and present this evidence in Sect. 2.3. After responding to some recent criticisms of the MM method in Sect. 2.4, we discuss the most important potential systematic error for the MM results – cosmological isotopic abundance variations – in Sect. 2.5.

2.1 Quasar Absorption Lines

The optical spectra of quasars are rich with absorption lines arising from gas clouds along the line of sight to Earth. For those unfamiliar with the anatomy of a quasar spectrum, Fig. 1 provides a tutorial. Of particular interest are those absorption clouds with sufficiently high hydrogen column density [number of absorbing atoms per area along the line of sight, $N(\mathrm{H\,I})$] to make metal absorption lines detectable. These are classified as Lyman-limit and damped Lyman-α systems (LLSs and DLAs): LLSs have $N(\mathrm{H\,I}) > 2 \times 10^{17}\mathrm{cm}^{-1}$ and DLAs have $N(\mathrm{H\,I}) > 2 \times 10^{20}\mathrm{cm}^{-1}$. It is the pattern of metal absorption lines – the relative separation between the different transitions – which carries information about the value of α in the clouds. The upper and lower panels of Fig. 1 detail some of the metal lines, those in the upper panel being of particular interest for the many-multiplet method in Sect. 2.2. Note the 'velocity structure' of the absorption cloud: each transition comprises many 'velocity components', each of which probably corresponds to a separate absorption cloud, all components probably being associated with a single high redshift galaxy or dark matter halo.

2.2 The Many-Multiplet (MM) Method

Initial attempts at constraining α-variability with quasar absorption spectra [5–10] used the alkali doublet (AD) method: for small variations in α, the relative wavelength separation between the transitions of an AD is proportional to α. While the AD method is simple, it is relatively insensitive to α-variations. The s ground state is most sensitive to changes in α (i.e. it has the largest relativistic corrections) but is common to both transitions of the AD. A more sensitive method is to compare transitions from different multiplets and/or atoms, allowing the ground states to constrain α, i.e. the many-multiplet (MM) method introduced in [11,12]. We summarize the advantages of the MM method in [13].

We first illustrate the MM method with a semi-empirical equation for the relativistic correction, Δ, for a transition from the ground state with total angular

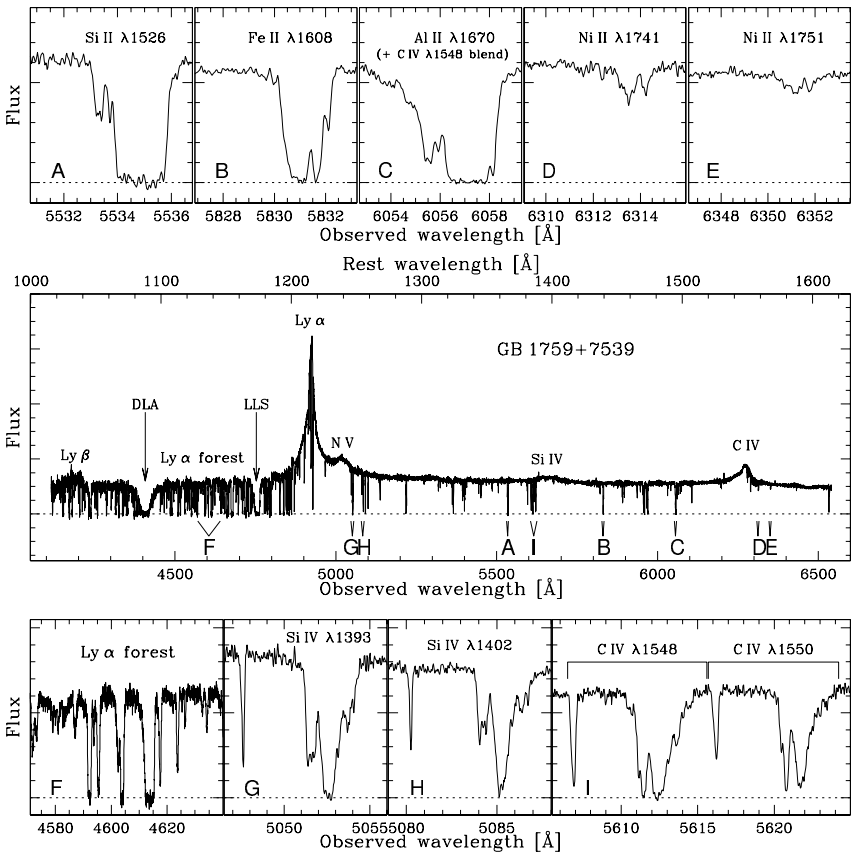


Fig. 1. Keck/HIRES spectrum of quasar GB 1759+7539 [4]. The full spectrum (middle panel) shows several broad emission lines (Ly-α, Ly-β, N V, C IV, Si IV) intrinsic to the quasar in the observer's rest-frame (i.e. vacuum heliocentric; bottom) and the quasar rest-frame (top). The dense 'Lyman-α forest' blue-wards of the Ly-α emission line is caused by cosmologically distributed low column-density hydrogen clouds along the line of sight to the quasar. The damped Lyman-α system (DLA) at $z_{\rm abs} = 2.625$ and the Lyman-limit system (LSS) at $z_{\rm abs} = 2.910$ give rise to heavy-element absorption lines red-wards of the Ly-α emission line (away from the confusing Ly-α forest). Some $z_{\rm abs} = 2.625$ transitions are detailed in panels A–E & G–I. Even though the transitions in the top panels have very different line-strengths, the velocity structures clearly follow each other closely. Detection of many such transitions facilitates determination of the velocity structure and allows easy detection of random blends. For example, the blue portion of the Al II λ1670 profile is blended with C IV λ1548 in the $z_{\rm abs} = 2.910$ LLS

momentum, j:

$$\Delta \propto (Z\alpha)^2 \left[\frac{1}{j+1/2} - C\right], \qquad (1)$$

where Z is nuclear charge and many-body effects are described by $C \sim 0.6$. To obtain strong constraints on α-variability one can (a) compare transitions of light

($Z \sim 10$) atoms/ions with those of heavy ($Z \sim 30$) ones and/or (b) compare s–p and d–p transitions of heavy elements. For the latter, the relativistic corrections will be of opposite sign which further increases sensitivity to α-variation and strengthens the MM method against systematic errors in the quasar spectra.

In practice, we express the rest-frequency, ω_z, for any transition observed in the quasar spectra at a redshift z, as

$$\omega_z = \omega_0 + q\left[\left(\frac{\alpha_z}{\alpha}\right)^2 - 1\right] \approx \omega_0 + 2q\Delta\alpha/\alpha\,, \tag{2}$$

where α_z is α in the absorption cloud. For most metal transitions observed in quasar absorption spectra, the laboratory wavenumber, ω_0, is measured with low precision compared with that achievable from the quasar spectra (!) since, in the laboratory, the transitions fall in the UV. For example, despite a recent order of magnitude precision gain [14], the C IV λ1548 and 1550 wavenumbers carry formal errors $> 0.04\,\mathrm{cm}^{-1}$. Compare this with the precision of $\approx 0.02\,\mathrm{cm}^{-1}$ available from absorption lines in a high resolution quasar spectrum (see Sect. 2.4). Dedicated laboratory measurements [15,16,14] of ω_0 for many transitions now reach an accuracy of $< 0.004\,\mathrm{cm}^{-1}$ allowing a precision of $\Delta\alpha/\alpha \sim 10^{-7}$ to be achieved. Updated values of ω_0 are given in Table 2 of [17].

The q coefficient of each transition contains all the relativistic corrections and measures the sensitivity of the transition frequency to changes in α. These have been calculated in [11,18–20] using many-body techniques. The accuracy of these calculations is given by how well various observable quantities (e.g. spectrum, g-factors etc.) of the ion in question are reproduced. To account for all dominant relativistic effects, the Dirac-Hartree-Fock approximation is used as a starting point. The accuracy is improved using many-body perturbation theory and/or the configuration-interaction method. For most transition combinations used in the MM method, the accuracy of these calculations is better than 10%. Note that in the absence of systematic effects in the quasar spectra, the form of (2) ensures one cannot infer a non-zero $\Delta\alpha/\alpha$ due to errors in the q coefficients. The q coefficients used in this paper are compiled in Table 2 of [17].

Figure 2 shows the distribution of q coefficients in (rest) wavelength space. Our sample conveniently divides into low- and high-z subsamples with very different properties (throughout this work we define $z < 1.8$ as low-z and $z > 1.8$ as high-z). Note the simple arrangement for the low-z Mg/Fe II systems: the Mg transitions are used as anchors against which the large, positive shifts in the Fe II transitions can be measured. Compare this with the complex arrangement for the high-z systems: low-order distortions in the wavelength scale of the quasar spectra will have a varied and complex effect on $\Delta\alpha/\alpha$ depending on which transitions are fitted in a given absorption system. This complexity at high-z may yield more robust estimates of $\Delta\alpha/\alpha$. The right panel quantifies this using simulations of the original 128 absorption system sample (see next section) which have been artificially compressed (see [17] for details). Even though the systems at high-z each respond differently to the compression of the wavelength scale, the binned plot reveals the average response is opposite to that in the low-z Mg/Fe II systems. This is an important strength of the MM method: the low-

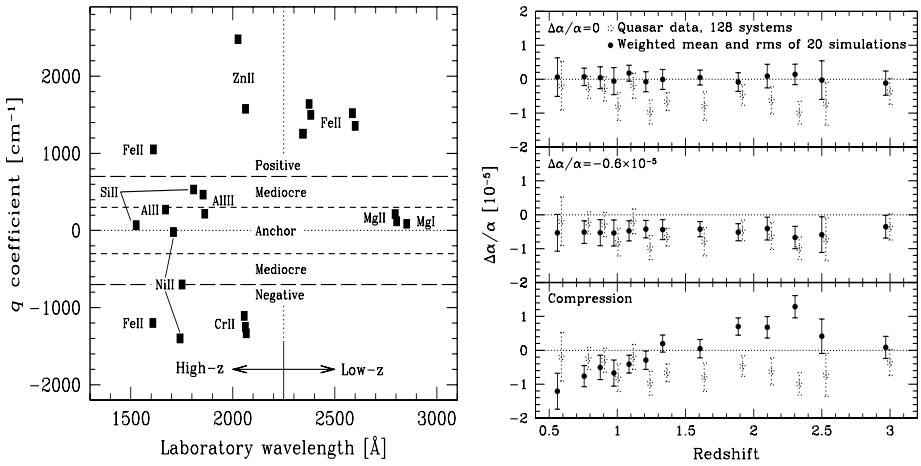


Fig. 2. (**Left**) Distribution of q coefficients for transitions used in the MM method. For low-z Mg/Fe II systems, a compression of the spectrum can mimic $\Delta\alpha/\alpha < 0$. However, the complex arrangement at high-z indicates resistance to such systematics. We define several 'q-types' by the horizontal bands and labels shown. (**Right**) Binned measurements of $\Delta\alpha/\alpha$ from 20 simulations of 128 absorption systems (*solid*) and the same, real quasar absorption systems (*dotted*). For the top and middle panels we input the indicated values of $\Delta\alpha/\alpha$. The values and errors are recovered reliably. A wavelength compression is introduced for the bottom panel to reproduce the low-z quasar results. At high-z, the variety of q coefficients causes the expected large scatter but the average effect on $\Delta\alpha/\alpha$ is opposite to that in the low-z systems. Simple distortions of the quasar spectra cannot explain the results

and high-z samples respond differently to simple systematic errors due to their different arrangement of q coefficients in wavelength space.

2.3 Spectral Analysis and Updated Results

Sample Definition

Our data set comprises three samples of Keck/HIRES spectra, each observed independently by different groups. The first sample [21] contains 27 low-z Mg/Fe II systems. The second sample [22] contains transitions from a wide variety of ionic species (though mostly singly ionized; Al II, Al III, Si II, Cr II, Fe II, Ni II and Zn II) in 19 high-z DLAs and 3 low-z Mg/Fe II systems. An additional high-z DLA is from [4]. The third sample was graciously provided by W. L. W. Sargent and collaborators and comprises 78 absorption systems over a wide redshift range. Together, these samples comprise 128 absorption systems and form the total sample presented in [17]. To illustrate many points in the following sections, we provide example spectra of a low-z Mg/Fe II system and a high-z DLA in Fig. 3.

In this work we update the second sample with 15 additional systems observed and reduced by two of the authors (JXP & AMW) and collaborators [23], containing a mix of low- and high-z systems.

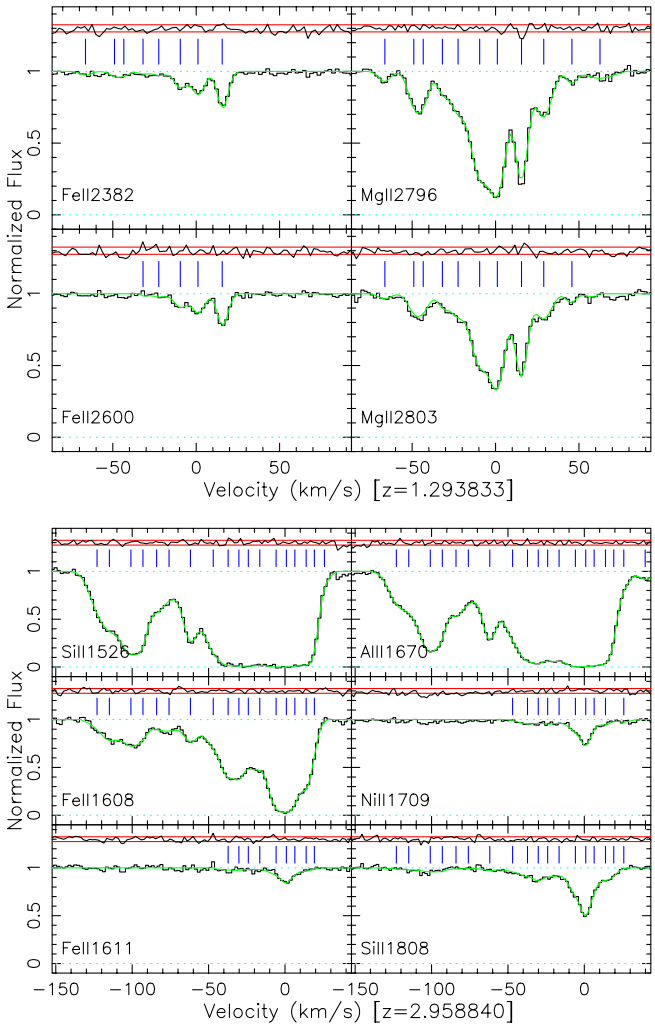


Fig. 3. Selected systems and transitions registered on a common (arbitrary) velocity scale. Data (*histogram*) are normalized by a fitted continuum. Our Voigt profile fit (*solid curve*) and residuals, normalized to the 1σ errors (*horizontal solid lines*), are also shown. Tick-marks indicate individual velocity components. (**Top**) $z_{\rm abs} = 1.2938$ Mg/Fe II system towards Q0636+6801. (**Bottom**) $z_{\rm abs} = 2.9587$ DLA towards Q1011+4315. Note that only optically thin components constrain $\Delta\alpha/\alpha$ strongly.

From the fully reduced spectra we select all systems which contain at least 2 transitions of different q-type (defined in Fig. 2), thereby potentially providing a tight constraint on $\Delta\alpha/\alpha$. Only in cases where all selected transitions have very low signal-to-noise ratio (SN) do we not attempt a fit. Only in very high SN cases have we selected systems where only transitions of the same q-type are detected. Apart from the obvious issues of line-strength and possible random

blends (Sect. 2.4), many well known instrumental limitations prevent us from detecting all MM transitions in every system. For example, the throughput of the telescope/spectrograph and detector sensitivity drops sharply below 4000 Å and above 6000 Å. Typically, the spectrum is not recorded below 3500 Å and above 7000 Å. Also, gaps in the wavelength coverage appear, particularly towards the red, since the spectrograph is an echelle–cross-disperser combination. Echelle orders cover ~ 60 Å and inter-order gaps can be up to ~ 20 Å.

Profile Fitting

For each system, the available transitions are fitted with multiple velocity component Voigt profiles. Each velocity component is described by three parameters: the absorption redshift $z_{\rm abs}$, the Doppler broadening or b parameter and the column density N. We reduce the number of free parameters by assuming either a completely turbulent or completely thermal broadening mechanism: corresponding components in all transitions have equal b or their bs are related by the inverse square-root of the ion masses. To apply the MM method one must assume that corresponding velocity components in all fitted ions have the same redshift. This further reduces the number of free parameters. We discuss this assumption in Sect. 2.4. To each *system* a single extra parameter is added, $\Delta\alpha/\alpha$. This allows all velocity components to shift in concert according to their q coefficients.

All free parameters are determined simultaneously using VPFIT, a non-linear least-squares χ^2 reduction algorithm written specifically for analysis of quasar absorption spectra. The $1\,\sigma$ parameter uncertainties are determined in the usual way from the diagonal terms of the final parameter covariance matrix. The assumption that off-diagonal terms are small (that parameters are not closely correlated) is a good one for $\Delta\alpha/\alpha$: redshift and $\Delta\alpha/\alpha$ are not correlated (Fig. 2). Monte Carlo simulations with 10000 realisations confirm the reliability of the parameter and error estimates [24]. It is important to realise that this numerical method ensures that constraints on $\Delta\alpha/\alpha$ are derived in a natural way from *optically thin* lines and not from saturated ones. The derivatives of χ^2 with respect to the saturated component redshifts are very small compared to the optically thin case and so only the optically thin lines strongly constrain $\Delta\alpha/\alpha$. If the two broadening mechanisms mentioned above result in significantly different $\Delta\alpha/\alpha$, the system is rejected. Otherwise, the broadening mechanism giving the lowest χ^2 fit is selected. We also require that χ^2 per degree of freedom, χ^2_ν, is ≈ 1.

Results

The distribution of $\Delta\alpha/\alpha$ with redshift and look-back time as a fraction of the age of the Universe is shown in Fig. 4. We also provide basic statistics for the different samples and the total, raw sample as a whole in Table 1. Note that all three samples give consistent, significantly smaller values of α in the absorption clouds compared to the laboratory. Breaking the sample down into low- and high-z subsamples also yields consistent results despite the very different q coefficient combinations used (Fig. 2, left) and overall reaction to simple systematic errors

Table 1. Statistics for different samples. χ^2_ν is $\chi^2/(N_{\rm sys}-1)$ about the weighted mean

Sample	$N_{\rm sys}$	$\langle z_{\rm abs}\rangle$	$\Delta\alpha/\alpha$ (10^{-5})	χ^2_ν
Churchill	27	1.00	-0.531 ± 0.223	1.109
Prochaska & Wolfe	38	2.27	-0.664 ± 0.219	2.024
Sargent	78	1.76	-0.620 ± 0.129	1.182
Low-z ($z < 1.8$)	77	1.07	-0.537 ± 0.124	1.064
High-z ($z > 1.8$)	66	2.55	-0.744 ± 0.167	1.739
Raw total	143	1.75	-0.611 ± 0.100	1.373
Fiducial[a]	**143**	**1.75**	$\mathbf{-0.573 \pm 0.113}$	**1.023**

[a] Low-z sample + low-contrast sample + high-contrast sample with increased errors.

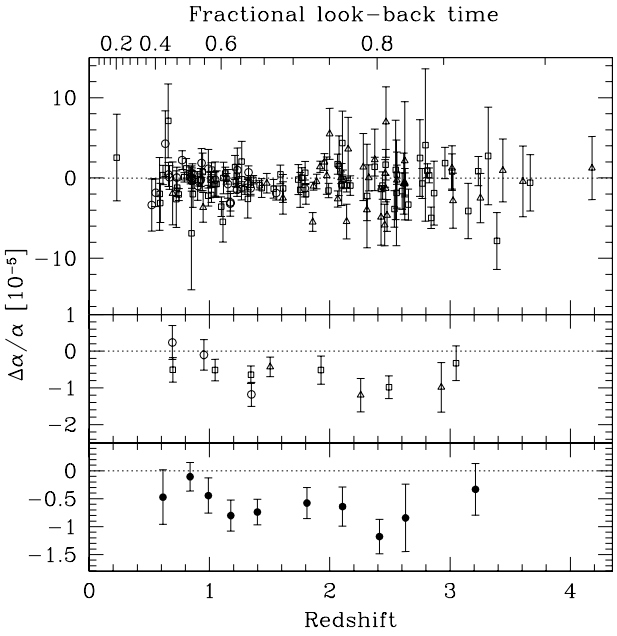


Fig. 4. $\Delta\alpha/\alpha$ and $1\,\sigma$ errors from the many-multiplet method for three Keck/HIRES samples: Churchill (*hollow circles*), Prochaska & Wolfe (*triangles*), Sargent (*squares*). Upper panel: unbinned individual values. Middle panel: binned results for each sample. Lower panel: binned over the whole sample. To calculate the fractional look-back time we use $H_0 = 70\,{\rm km\,s^{-1}Mpc^{-1}}$, $\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$, implying an age of 13.47 Gyr.

(Fig. 2, right). We have conducted numerous internal consistency checks on these results, including direct tests of the wavelength calibration of the quasar spectra and the effect of removing individual transitions or entire ionic species from our fits. These are described in detail in [13,17].

Extra Scatter at High z

Note that the scatter in the total low-z sample is consistent with that expected from the size of the error bars (i.e. $\chi^2_\nu \approx 1$). However, at high-z, Fig. 4 shows significant extra scatter. This is reflected in the high χ^2_ν values in Table 1. The weighted mean therefore exaggerates the true significance of $\Delta\alpha/\alpha$ at high-z.

We have identified the major source of this extra scatter at high z. Consider fitting two transitions with very different line-strengths (e.g. Al II λ1670 and Ni II λ1709 in Fig. 3). Weak components near the high optical depth edges of the strong transition's profile are not necessary to obtain a good fit to the data. Even though the VPFIT χ^2 minimization ensures that constraints on $\Delta\alpha/\alpha$ derive primarily from the optically thin velocity components, these weak components missing from the fit will cause small line shifts. The resulting shift in $\Delta\alpha/\alpha$ is random from component to component and from system to system: the effect of missing components will be to increase the random scatter in the individual $\Delta\alpha/\alpha$ values. This effect will be far larger in the high-z sample since only there do we fit transitions of such different line-strengths.

We form a 'high-contrast' sample from the high-z (i.e. $z_{\rm abs} > 1.8$) sample by selecting systems in which we fit both strong and weak lines, i.e. any of the Al II, Si II or Fe II transitions *and* any of the Cr II, Ni II or Zn II ones. To obtain a more robust estimate of the significance of $\Delta\alpha/\alpha$ in this high-contrast sample, we have increased the individual $1\,\sigma$ errors until $\chi^2_\nu = 1$ about the weighted mean. We achieve this by adding 1.75×10^{-5} in quadrature to the error bars of the 27 relevant systems. Other procedures for estimating the significance at high-z are discussed in [17]. Figure 5 identifies the high-contrast sample and presents the binned results with increased error bars. Table 1 presents the relevant statistics. The above procedure results in our most robust estimate from the 143 absorption systems over the redshift range $0.2 < z_{\rm abs} < 4.2$: $\Delta\alpha/\alpha = (-0.57 \pm 0.11) \times 10^{-5}$.

2.4 Recent Criticisms of the MM Method

Bekenstein [25] has pointed out that the form of the Dirac Hamiltonian for the hydrogen atom changes in a theory where α is considered a dynamical field. He warns that the resulting shifts in energy levels could be important for the MM method. Reference [26] have extended Bekenstein's model to many-electron atoms. They find that the energy shift within the Bekenstein model is proportional to $\Delta\alpha/\alpha$ defined in (2), up to corrections of order 1%. Thus, modifications to the energy shifts discussed by [25] for the hydrogen atom are not important for the MM method applied to many-electron metal ions.

Bahcall et al. [27] criticized the MM method on many points, summarized in their Table 3. Though none of these are real candidate explanations of our results, we address each criticism below to avoid future confusion:

- **Theory:** Since one must calculate the q coefficients using sophisticated many-body techniques, [27] argue that the MM method is less reliable than, say, the AD method. The likely sources of error are discussed in detail by [20] and we give a flavour of them in Sect. 2.2. The q coefficients are known

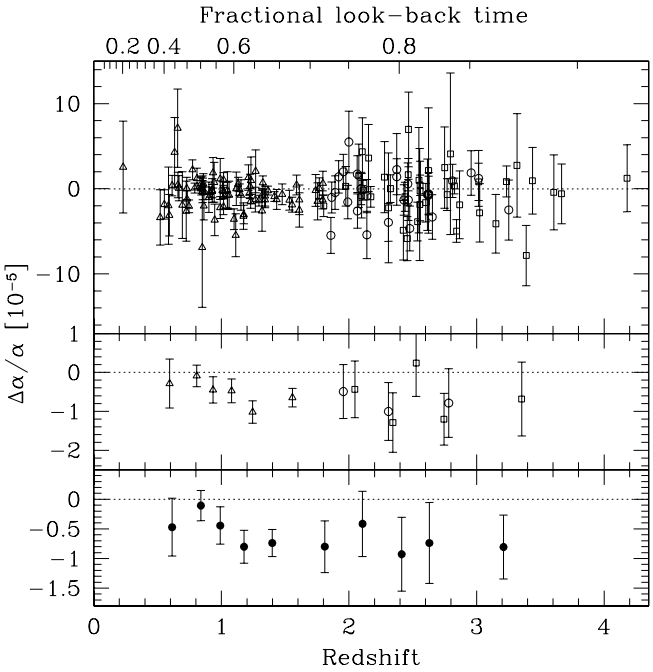


Fig. 5. The fiducial sample. Low-z sample (*triangles*), low-contrast sample (*squares*) and high-contrast sample with increased error bars (*circles*). The weighted mean, $\Delta\alpha/\alpha = (-0.57 \pm 0.11) \times 10^{-5}$, is our most robust estimate for the HIRES spectra

to high enough accuracy given our sample precision (discussed below). We again stress that if $\Delta\alpha/\alpha$ is really zero, one cannot manufacture a non-zero value through errors in the q coefficients (2) if systematic errors in the quasar spectra are not important.

- **Absolute or relative wavelengths?** Reference [27] argued that the MM method requires the measurement of absolute wavelengths in the quasar spectra[1]. This is incorrect. Since we simultaneously determine the redshifts of the absorption components and $\Delta\alpha/\alpha$ (these are not degenerate parameters; Fig. 2) for each system, any velocity shift may be applied to the spectra and $\Delta\alpha/\alpha$ will be unaffected. A velocity-space shift is a systematic error for the absorption redshifts, not for $\Delta\alpha/\alpha$.
- **Sample precision:** It is trivial to estimate the expected precision available from our observational sample. We explained this calculation in [13] and repeat it here. The Keck/HIRES pixels cover $\sim 3\,\mathrm{km\,s^{-1}}$ and so one reasonably expects to centroid barely resolved features (with SN $\sim$ 30–50) to $0.3\,\mathrm{km\,s^{-1}}$. One expects ~ 4 features per absorption system to be well-centroided in this way, providing a velocity precision of $\Delta v \approx 0.15\,\mathrm{km\,s^{-1}}$ or $\Delta\omega \approx 0.02\,\mathrm{cm^{-1}}$

[1] This suggestion is present only in preprint versions 1 & 2 of [27] (astro-ph/0301507). It is removed from later versions. Despite this, we address this point here to avoid further confusion in the literature.

for an $\omega_0 \approx 40000\,\mathrm{cm}^{-1}$ transition. The typical difference in q coefficient between the Mg I/II and Fe II lines is $\sim 1000\,\mathrm{cm}^{-1}$ and so, for a single Mg/Fe II absorption system, (2) implies a precision of $|\Delta\alpha/\alpha| \sim 1 \times 10^{-5}$. With ~ 50 such systems, one expects a precision $|\Delta\alpha/\alpha| \sim 0.14 \times 10^{-5}$ (cf. Table 1). As shown in Fig. 2 (and Figs. A.2 & A.3 in [24]), simulations also provide a simple "reality-check" on the sample precision.

- **Line misidentification and blending:** Reference [27] argue we may have misidentified many absorption features. In high resolution ($R \sim 50000$) spectra we largely resolve the velocity structure of absorption systems. Misidentifying transitions is highly improbable since, even by eye, the profiles of different species follow each other to within $|\Delta v| < 1\,\mathrm{km\,s}^{-1}$. Confirming this, we obtain good fits to the absorption profiles with the number of free parameters restricted by physical considerations (Sect. 2.3). Detecting blends from absorption at other redshifts is also greatly facilitated by high resolution. See [13,17] for thorough discussions of blending and Fig. 1 for an example. Even if we misidentified a small number of transitions in our sample and they miraculously mimicked the velocity structure of other detected transitions (thereby allowing a good fit), this would have a random, not systematic, effect on $\Delta\alpha/\alpha$. Indeed, compared with the AD method, the MM method is distinctly robust against misidentifications and blends: many transitions constrain the velocity structure so identifying blends and misidentifications is all the more trivial. Also, any blends that are not identified have a smaller effect on $\Delta\alpha/\alpha$ since many other transitions contribute to the constraints.
- **Velocity structure:** The MM method assumes that corresponding velocity components in different ions have the same redshift. Most transitions used are from ionic species with very similar ionization potentials and so absorption from these species should arise co-spatially. Consider a Mg II velocity component *blue*shifted with respect to the *corresponding* Fe II component by some kinematic effect in the absorption cloud. Clearly, this mimics $\Delta\alpha/\alpha < 0$ for that component. However, kinematic effects would equally well *red*shift the Mg II components. Thus, the effect on $\Delta\alpha/\alpha$ is random from component to component and absorption system to system. This argument is misunderstood in [27]: they feel it is unlikely that such random effects "average out to an accuracy of $0.2\,\mathrm{km\,s}^{-1}$ over a velocity range of more than $10^2\,\mathrm{km\,s}^{-1}$", the latter quantity referring to the total velocity extent of a typical absorption *system*. Inspecting Fig. 4, it is clear that any extra scatter in $\Delta\alpha/\alpha$ from kinematic effects derives only from the gas properties on velocity scales less than typical b parameters, i.e. $< 5\,\mathrm{km\,s}^{-1}$. That we obtain excellent agreement between the velocity component redshifts in different species to a precision $|\Delta v| \sim 0.3\,\mathrm{km\,s}^{-1}$ illustrates this. If kinematic effects were important, they would be most prominent in the low-z values of $\Delta\alpha/\alpha$ from the Mg/Fe II systems, appearing as an extra scatter beyond that expected from the $1\,\sigma$ errors. This is not observed. We discuss kinematic effects in more detail in [28] and [17].

2.5 Isotopic Abundance Variations

As we discussed in [29,30] and emphasised in [17,31], the main possible systematic error for the low-z results is that relative isotopic abundances may differ between the absorption clouds and terrestrial environment/laboratory, as summarized in Fig. 6. The isotopic structures used for Mg transitions are discussed in [17] whereas those for Si are from calculations similar to [32]. To our knowledge, isotopic structures for the transitions of Fe II are not known. However, these should be far more 'compact' than those of, say, Mg since the normal mass shift will be > 5 times smaller. The results in Figs. 4 and 5 were obtained by fitting the quasar spectra with terrestrial isotopic abundance ratios. For those systems where Mg lines and no Si lines are fitted, these results correspond to the large square on Fig. 6, *vice versa* for the large triangle. The Mg II and Si II systems approximately correspond to the low- and high-z samples respectively. If we change the heavy isotope abundances we note the marked change in $\Delta\alpha/\alpha$ for the Mg systems and the comparative insensitivity for the Si systems. This is expected given the distribution of q coefficients in Fig. 2 and the diversity of transitions fitted at high-z (see [17] for further discussion).

In previous works we argued that the heavy isotope fraction for Mg in our absorbers is likely to be significantly less than the terrestrial value. This is based on observations of low metallicity[2], Z, stellar environments in our Galaxy [35,36] and theoretical models of Galactic chemical evolution [37,38] where significantly sub-solar heavy isotope fractions are observed/expected at the low Zs of our absorption clouds. However, some stars in [35,36] and those in some globular clusters [39,40] are found to have super-solar values.

At low Z, the heavy Mg isotopes are thought to be primarily produced by intermediate mass (IM; $\sim$2–8 $M_\odot$) stars in their asymptotic giant branch (AGB) phase [41–43]. Reference [38] included a contribution from IM AGB stars in their chemical evolution model, finding sub-solar heavy Mg isotope abundances at low Z, consistent with the above observations. Recently, [44] noted that enhanced heavy Mg isotope abundances might be produced at low Z if one assumes an IM-enhanced stellar initial mass function (IMF) at high z. However, such an IMF seems incompatible with current constraints. See [45] for general discussion of the observational constraints on the IMF. More recently, [46] find Ar, S and O abundances consistent with a normal IMF in a high-z DLA. See also [47]. For example, an IM-enhanced IMF could produce vast amounts of Fe via type Ia supernovae [48], in disagreement with Galactic and DLA abundance studies. Moreover, AGB enrichment levels are constrained by the relative element abundances in the absorption clouds. For example, AGB stars produce large amounts of N relative to other enrichment processes (e.g. supernovae types Ia & II; e.g. [49]). However, the abundance of N relative to H is very low in DLAs, i.e. $10^{-3.8}$–$10^{-1.5}$ solar [50,51]. Overall, these points are a barrier to ad-hoc IMF changes such as those suggested by [44]. Though enhanced heavy Mg isotope

[2] Metallicity is the relative metal abundance with respect to that in the Solar System environment. For LSSs and DLAs, $\log_{10} Z$ typically ranges from -2.5 to -0.5.

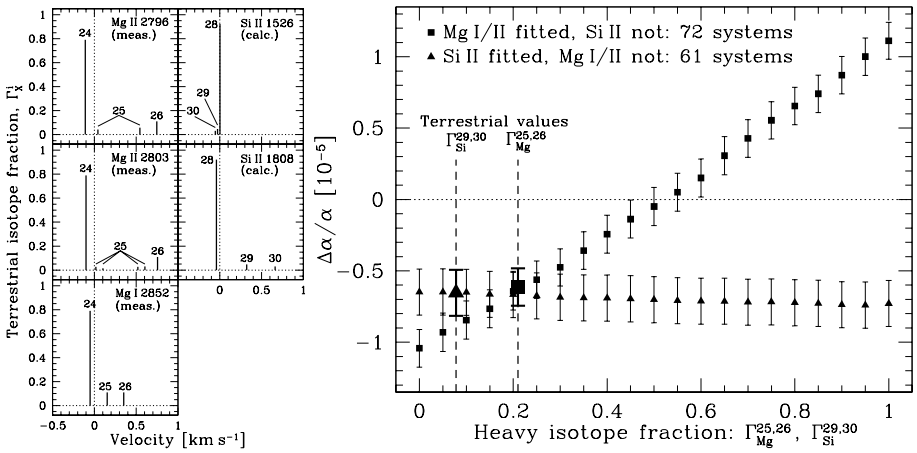


Fig. 6. (**Left**) Isotopic structures for relevant Mg/Si transitions from measurements [33,34,15] or calculations [32]. Zero velocity corresponds to the structure's centre of gravity. (**Right**) Sensitivity of $\Delta\alpha/\alpha$ to isotopic abundance variations. We alter the heavy isotope abundances proportionately: the heavy element fraction for Mg is $\Gamma_{\rm Mg}^{25,26}$ = const. $\times$ (0.10 + 0.11) where the numbers in parentheses are the terrestrial isotopic fractions of ^{25}Mg and ^{26}Mg. Much higher relative ^{25,26}Mg abundances in the absorbers can explain the low-z results but the high-z results (containing the Si II systems shown) are insensitive to ^{29,30}Si abundances

fractions are a possible explanation of the low-z varying-α results, we again conclude that they are an unlikely one.

3 Varying α and $m_{\rm q}/\Lambda_{\rm QCD}$ from Atomic Clocks

3.1 Introduction

The hypothetical unification of all interactions implies that variation of the electromagnetic interaction constant α should be accompanied by variation of masses and the strong interaction constant. Specific predictions require a model. For example, the grand unification model discussed in [52] (see also [53,54]) predicts that the quantum chromodynamic (QCD) scale, $\Lambda_{\rm QCD}$, is modified as

$$\frac{\delta\Lambda_{\rm QCD}}{\Lambda_{\rm QCD}} \approx 34\frac{\delta\alpha}{\alpha}\,. \tag{3}$$

The variation of quark and electron masses in this model is given by

$$\frac{\delta m}{m} \sim 70\frac{\delta\alpha}{\alpha}\,, \tag{4}$$

resulting in an estimate for the variation of the dimensionless ratio

$$\frac{\delta(m/\Lambda_{\rm QCD})}{(m/\Lambda_{\rm QCD})} \sim 35\frac{\delta\alpha}{\alpha}\,. \tag{5}$$

The large coefficients in these expressions are generic for grand unification models, in which modifications come from high energy scales: they appear because the running strong coupling constant and Higgs constants (related to mass) run faster than α. This means that if these models are correct the variation of masses and strong interaction may be easier to detect than the variation of α.

For the strong interaction there is generally no direct relation between the coupling constants and observable quantities, unlike the case for the electroweak forces. Since one can only measure variations in dimensionless quantities, we must extract from the measurements constraints on variation of $m_{\rm q}/\Lambda_{\rm QCD}$, a dimensionless ratio, where $m_{\rm q}$ is the quark mass (with dependence on the normalization point removed). A number of limits on variation of $m_{\rm q}/\Lambda_{\rm QCD}$ have been obtained recently from consideration of Big Bang Nucleosynthesis, quasar absorption spectra and the Oklo natural nuclear reactor which was active about 1.8 billion years ago [9,55–64].

Below we consider the limits which follow from laboratory atomic clock comparisons. Laboratory limits with a time base of several years are especially sensitive to oscillatory variation of fundamental constants. A number of relevant measurements have been performed already and many more have been started or planned. The increase in precision is very fast.

3.2 Nuclear Magnetic Moments, α and $m_{\rm q}/\Lambda_{\rm QCD}$

As pointed out by [65], measurements of the ratio of hyperfine structure intervals in different atoms are sensitive to variation of nuclear magnetic moments. The first rough estimates of the dependence of nuclear magnetic moments on $m_{\rm q}/\Lambda_{\rm QCD}$ and limits on time variation of this ratio were obtained in [55]. Using H, Cs and Hg^+ measurements [66], [55] limited the variation of $m_{\rm q}/\Lambda_{\rm QCD}$ to about $5\times10^{-13}\,{\rm yr}^{-1}$. Below we calculate the dependence of nuclear magnetic moments on $m_{\rm q}/\Lambda_{\rm QCD}$ and obtain the limits from recent atomic clock experiments with hyperfine transitions in H, Rb, Cs, Hg^+ and optical transitions in Hg^+. It is convenient to assume that the strong interaction scale $\Lambda_{\rm QCD}$ does not vary, so we will speak about variation of masses.

The hyperfine structure constant can be presented in the following form,

$$A = \text{const.} \times \left[\frac{m_{\rm e}e^4}{\hbar^2}\right] \left[\alpha^2 F_{\rm rel}(Z\alpha)\right] \left[\mu\frac{m_{\rm e}}{M_{\rm p}}\right] . \tag{6}$$

The factor in the first bracket is an atomic unit of energy. The second, 'electromagnetic', bracket determines the dependence on α. An approximate expression for the relativistic correction factor (Casimir factor) for the s-wave electron is

$$F_{\rm rel} = \frac{3}{\gamma(4\gamma^2-1)}\,, \tag{7}$$

where $\gamma=\sqrt{1-(Z\alpha)^2}$ and Z is the nuclear charge. Variation of α leads to the following variation of $F_{\rm rel}$ [66]:

$$\frac{\delta F_{\rm rel}}{F_{\rm rel}} = K\frac{\delta\alpha}{\alpha}\,, \tag{8}$$

$$K = \frac{(Z\alpha)^2(12\gamma^2 - 1)}{\gamma^2(4\gamma^2 - 1)}\,. \tag{9}$$

More accurate numerical many-body calculations [18] of the dependence of the hyperfine structure on α have shown that the coefficient K is slightly larger than that given by this formula. For Cs (Z=55) K=0.83 (instead of 0.74), for Rb K=0.34 (instead of 0.29) and for Hg^+ K=2.28 (instead of 2.18).

The last bracket in (6) contains the dimensionless nuclear magnetic moment, μ, in nuclear magnetons (the nuclear magnetic moment $M = \mu \times e\hbar/2M_{\rm p}c$) and the electron–proton mass ratio, $m_{\rm e}/M_{\rm p}$. We may also include a small correction due to the finite nuclear size. However, its contribution is insignificant.

Recent experiments measured time dependence of the ratios of hyperfine structure intervals of $^{199}Hg^+$ and H [66], ^{133}Cs and ^{87}Rb [67], and the ratio of the optical frequency in Hg^+ and the ^{133}Cs hyperfine frequency [68]. In the ratio of two hyperfine structure constants for different atoms, time dependence may appear from the ratio of the factors $F_{\rm rel}$ (depending on α) and the ratio of nuclear magnetic moments (depending on $m_{\rm q}/\Lambda_{\rm QCD}$). Magnetic moments in a single-particle approximation (one unpaired nucleon) are

$$\mu = \left[g_s + (2j-1)g_l\right]/2 \tag{10}$$

for $j = l + 1/2$ and

$$\mu = \frac{j}{2(j+1)}\left[-g_s + (2j+3)g_l\right] \tag{11}$$

for $j = l - 1/2$. Here, the orbital g-factors are $g_l = 1$ for valence protons and $g_l = 0$ for valence neutrons. Present values of the spin g-factors, g_s, are $g_{\rm p} = 5.586$ and $g_{\rm n} = -3.826$ for the proton and neutron. These depend on $m_{\rm q}/\Lambda_{\rm QCD}$. The light quark masses are only about 1% of the nucleon mass [$m_{\rm q} = (m_{\rm u} + m_{\rm d})/2 \approx 5\,{\rm MeV}$]. The nucleon magnetic moment remains finite in the chiral limit of $m_{\rm u} = m_{\rm d} = 0$. Therefore, one may think that the corrections to g_s due to the finite quark masses are very small. However, there is a mechanism which enhances the quark mass contribution: π-meson loop corrections to the nucleon magnetic moments which are proportional to the π-meson mass $m_\pi \sim \sqrt{m_{\rm q}\Lambda_{\rm QCD}}$. m_π=140 MeV is not so small.

According to [69], the dependence of the nucleon g-factors on the π-meson mass can be approximated as

$$g(m_\pi) = \frac{g(0)}{1 + am_\pi + bm_\pi^2}\,, \tag{12}$$

where $a = 1.37\,{\rm GeV}^{-1}$, $b = 0.452\,{\rm GeV}^{-2}$ for the proton and $a = 1.85\,{\rm GeV}^{-1}$, $b = 0.271\,{\rm GeV}^{-2}$ for the neutron. This leads to the following estimate:

$$\frac{\delta g_{\rm p}}{g_{\rm p}} = -0.174\frac{\delta m_\pi}{m_\pi} = -0.087\frac{\delta m_{\rm q}}{m_{\rm q}}\,, \tag{13}$$

$$\frac{\delta g_{\rm n}}{g_{\rm n}} = -0.213\frac{\delta m_\pi}{m_\pi} = -0.107\frac{\delta m_{\rm q}}{m_{\rm q}}\,. \tag{14}$$

Equations (10,11,13,14) give variation of nuclear magnetic moments. For the hydrogen nucleus (proton) ,

$$\frac{\delta\mu}{\mu} = \frac{\delta g_{\rm p}}{g_{\rm p}} = -0.087\frac{\delta m_{\rm q}}{m_{\rm q}}. \tag{15}$$

For ^{199}Hg we have the valence neutron (no orbital contribution), and so

$$\frac{\delta\mu}{\mu} = \frac{\delta g_{\rm n}}{g_{\rm n}} = -0.107\frac{\delta m_{\rm q}}{m_{\rm q}}\,. \tag{16}$$

For ^{133}Cs we have the valence proton with j=7/2 and l=4, giving

$$\frac{\delta\mu}{\mu} = 0.22\frac{\delta m_\pi}{m_\pi} = 0.11\frac{\delta m_{\rm q}}{m_{\rm q}}\,. \tag{17}$$

For ^{87}Rb we have the valence proton with j=3/2 and l=1, resulting in

$$\frac{\delta\mu}{\mu} = -0.128\frac{\delta m_\pi}{m_\pi} = -0.064\frac{\delta m_{\rm q}}{m_{\rm q}}\,. \tag{18}$$

Deviation of the single-particle nuclear magnetic moment values from the measured values is about 30% and so we have attempted to refine them. If we neglect the spin-orbit interaction, the total spin of nucleons is conserved. The magnetic moment of the nucleus changes due to the spin-spin interaction because the valence proton transfers part of its spin, $\langle s_z\rangle$, to core neutrons (transfer of spin from the valence proton to core protons does not change the magnetic moment). In this approximation, $g_s = (1-b)g_{\rm p} + bg_{\rm n}$ for the valence proton (or $g_s = (1-b)g_{\rm n} + bg_{\rm p}$ for the valence neutron). We can use the coefficient b as a fitting parameter to reproduce nuclear magnetic moments exactly. The signs of $g_{\rm p}$ and $g_{\rm n}$ are opposite, therefore a small mixing, $b \sim 0.1$, is enough to eliminate the deviation of the theoretical value from the experimental one. Note also that it follows from (13,14) that $\delta g_{\rm p}/g_{\rm p} \approx \delta g_{\rm n}/g_{\rm n}$. This produces an additional suppression of the mixing's effect, indicating that the actual accuracy of the single-particle approximation for the effect of the spin g-factor variation may be as good as 10%. Note, however, that we neglected variation of the mixing parameter b. This is difficult to estimate.

3.3 Results

We can now estimate the sensitivity of the ratio of the hyperfine transition frequencies to variations in $m_{\rm q}/\Lambda_{\rm QCD}$. For ^{199}Hg and hydrogen we have

$$\frac{\delta[A({\rm Hg})/A({\rm H})]}{[A({\rm Hg})/A({\rm H})]} = 2.3\frac{\delta\alpha}{\alpha} - 0.02\frac{\delta[m_{\rm q}/\Lambda_{\rm QCD}]}{[m_{\rm q}/\Lambda_{\rm QCD}]}\,. \tag{19}$$

Therefore, the measurement of the ratio of Hg and hydrogen hyperfine frequencies is practically insensitive to the variation of light quark masses and

the strong interaction. Measurements [66] constrain variations in the parameter $\tilde{\alpha} = \alpha[m_{\rm q}/\Lambda_{\rm QCD}]^{-0.01}$:

$$\left|\frac{1}{\tilde{\alpha}}\frac{{\rm d}\tilde{\alpha}}{{\rm d}t}\right| < 3.6\times10^{-14}\,{\rm yr}^{-1}\,. \tag{20}$$

Other ratios of hyperfine frequencies are more sensitive to $m_{\rm q}/\Lambda_{\rm QCD}$. For ^{133}Cs and ^{87}Rb we have

$$\frac{\delta[A({\rm Cs})/A({\rm Rb})]}{[A({\rm Cs})/A({\rm Rb})]} = 0.49\frac{\delta\alpha}{\alpha} + 0.17\frac{\delta[m_{\rm q}/\Lambda_{\rm QCD}]}{[m_{\rm q}/\Lambda_{\rm QCD}]}\,. \tag{21}$$

Therefore, measurements [67] constrain variations in the parameter $X = \alpha^{0.49}[m_{\rm q}/\Lambda_{\rm QCD}]^{0.17}$:

$$\frac{1}{X}\frac{{\rm d}X}{{\rm d}t} = (0.2\pm7)\times10^{-16}\,{\rm yr}^{-1}\,. \tag{22}$$

Note that if the relation (5) is correct, variation of X would be dominated by variation of $m_{\rm q}/\Lambda_{\rm QCD}$: (5) would give $X \propto \alpha^7$.

For ^{133}Cs and H we have

$$\frac{\delta[A({\rm Cs})/A({\rm H})]}{[A({\rm Cs})/A({\rm H})]} = 0.83\frac{\delta\alpha}{\alpha} + 0.2\frac{\delta[m_{\rm q}/\Lambda_{\rm QCD}]}{[m_{\rm q}/\Lambda_{\rm QCD}]}\,. \tag{23}$$

Therefore, measurements [70,71] constrain variations of the parameter $X_H = \alpha^{0.83}[m_{\rm q}/\Lambda_{\rm QCD}]^{0.2}$:

$$\left|\frac{1}{X_H}\frac{{\rm d}X_H}{{\rm d}t}\right| < 5.5\times10^{-14}\,{\rm yr}^{-1}\,. \tag{24}$$

If we assume the relation (5), we would have $X_H \propto \alpha^8$.

The optical clock transition energy, $E({\rm Hg})$, at $\lambda = 282\,{\rm nm}$ in the Hg^+ ion, can be presented as

$$E({\rm Hg}) = {\rm const.}\times\left[\frac{m_{\rm e}e^4}{\hbar^2}\right]F_{\rm rel}(Z\alpha)\,. \tag{25}$$

Note that the atomic unit of energy (first bracket) is cancelled out and so we do not consider its variation. Numerical calculation of the relative variation of $E({\rm Hg})$ has given [18]

$$\frac{\delta E({\rm Hg})}{E({\rm Hg})} = -3.2\frac{\delta\alpha}{\alpha}\,. \tag{26}$$

Variation of the ratio of the Cs hyperfine splitting $A({\rm Cs})$ to this optical transition energy is equal to

$$\frac{\delta[A({\rm Cs})/E({\rm Hg})]}{[A({\rm Cs})/E({\rm Hg})]} = 6.0\frac{\delta\alpha}{\alpha} + \frac{\delta[m_{\rm e}/\Lambda_{\rm QCD}]}{[m_{\rm e}/\Lambda_{\rm QCD}]} + 0.11\frac{\delta[m_{\rm q}/\Lambda_{\rm QCD}]}{[m_{\rm q}/\Lambda_{\rm QCD}]}\,. \tag{27}$$

Here we have taken into account that the proton mass $M_{\rm p} \propto \Lambda_{\rm QCD}$. The factor 6.0 before $\delta\alpha$ appeared from $\alpha^2 F_{\rm rel}$ in the Cs hyperfine constant (i.e. 2+0.83)

and α-dependence of $E(\mathrm{Hg})$ (i.e. 3.2). Therefore, the results of [68] give the limit on variation of the parameter $U = \alpha^6[m_\mathrm{e}/\Lambda_\mathrm{QCD}][m_\mathrm{q}/\Lambda_\mathrm{QCD}]^{0.1}$:

$$\left|\frac{1}{U}\frac{\mathrm{d}U}{\mathrm{d}t}\right| < 7 \times 10^{-15}\,\mathrm{yr}^{-1}\,. \tag{28}$$

If we assume the relation (5), we would have $U \propto \alpha^{45}$. Note that we present such limits on $|(\mathrm{d}\alpha/\mathrm{d}t)/\alpha|$ as illustrations only since they are strongly model-dependent.

4 Conclusions

We have presented evidence for a varying α based on many-multiplet measurements in 143 Keck/HIRES quasar absorption systems covering the redshift range $0.2 < z_\mathrm{abs} < 4.2$: $\Delta\alpha/\alpha = (-0.57 \pm 0.11) \times 10^{-5}$. Three independent observational samples give consistent results. Moreover, the low- and high-z samples are also consistent, which cannot be explained by simple systematic errors (Fig. 2). Our results therefore seem internally robust. The possibility that the isotopic abundances are very different in the absorption clouds and the laboratory is a potentially important systematic effect. A high heavy isotope fraction for Mg ($\mathit{\Gamma}^{25,26}_\mathrm{Mg} \approx 0.5$) compared with the terrestrial value ($\mathit{\Gamma}^{25,26}_\mathrm{Mg} \approx 0.21$) may explain the low-$z$ results (Fig. 6). However, observations of low-metallicity stars and Galactic chemical evolution (GCE) models suggest sub-solar values of $\mathit{\Gamma}^{25,26}_\mathrm{Mg}$ in the quasar absorption systems. GCE models with a stellar initial mass function greatly enhanced at intermediate masses may produce large quantities of heavy Mg isotopes via asymptotic giant branch stars. However, such models disagree with the observed element abundances in quasar absorption systems. The high-z results are insensitive to the isotopic fraction of ^{29,30}Si. However, we stress the need for theoretical calculations and laboratory measurements of isotopic structures for other elements/transitions observed in quasar absorption systems.

Aside from a varying α, no explanation of our results currently exists which is consistent with the available observational evidence. The results can best be refuted with detailed many-multiplet analyses of quasar absorption spectra from different telescopes/instruments now available (e.g. VLT/UVES, Subaru/HDS).

We have calculated the dependence of nuclear magnetic moments on quark masses. This leads to limits on possible variations in $m_\mathrm{q}/\Lambda_\mathrm{QCD}$ from recent laboratory atomic clock experiments involving hyperfine transitions in H, Rb, Cs, Hg^+ and an optical transition in Hg^+. These limits can be compared with limits on α-variation within the context of grand unification theories. Unfortunately, this comparison is strongly model-dependent. See, for example, [72].

References

1. J. Uzan: Rev. Mod. Phys. **75**, 403 (2003)
2. C.J.A.P. Martins, ed.: *Proc. JENAM2002: The cosmology of extra dimensions and varying fundamental constants* (Kluwer, Netherlands, 2003)

3. S.S. Vogt et al.: 'HIRES: The high-resolution echelle spectrometer on the Keck 10-m telescope'. In: *Instrumentation in Astronomy VIII*, ed. by D.L. Crawford, E.R. Craine (SPIE 2198, 1994) p. 362
4. P.J. Outram, F.H. Chaffee, R.F. Carswell: Mon. Not. Roy. Soc. **310**, 289 (1999)
5. M.P. Savedoff: Nature **178**, 689 (1956)
6. J.N. Bahcall, E.E. Salpeter: Astrophys. J. **142**, 1677 (1965)
7. J.N. Bahcall: Astrophys. J. **149**, L7 (1967)
8. A.M. Wolfe, R.L. Brown, M.S. Roberts: Phys. Rev. Lett. **37**, 179 (1976)
9. L.L. Cowie, A. Songaila: Astrophys. J. **453**, 596 (1995)
10. D.A. Varshalovich, V.E. Panchuk, A.V. Ivanchik: Astron. Lett. **22**, 6 (1996)
11. V.A. Dzuba, V.V. Flambaum, J.K. Webb: Phys. Rev. Lett. **82**, 888 (1999)
12. J.K. Webb, V.V. Flambaum, C.W. Churchill, M.J. Drinkwater, J.D. Barrow: Phys. Rev. Lett. **82**, 884 (1999)
13. M.T. Murphy et al.: Mon. Not. Roy. Soc. **327**, 1208 (2001)
14. U. Griesmann, R. Kling: Astrophys. J. **536**, L113 (2000)
15. J.C. Pickering, A.P. Thorne, J.K. Webb: Mon. Not. Roy. Soc. **300**, 131 (1998)
16. J.C. Pickering et al.: Mon. Not. Roy. Soc. **319**, 163 (2000)
17. M.T. Murphy, J.K. Webb, V.V. Flambaum: Mon. Not. Roy. Soc. **345**, 609 (2003)
18. V.A. Dzuba, V.V. Flambaum, J.K. Webb: Phys. Rev. A **59**, 230 (1999)
19. V.A. Dzuba, V.V. Flambaum, M.T. Murphy, J.K. Webb: Phys. Rev. A **63**, 42509 (2001)
20. V.A. Dzuba, V.V. Flambaum, M.G. Kozlov, M. Marchenko: Phys. Rev. A **66**, 022501 (2002)
21. C.W. Churchill et al.: Astrophys. J. Supp. **130**, 91 (2000)
22. J.X. Prochaska, A.M. Wolfe: Astrophys. J. Supp. **121**, 369 (1999)
23. J.X. Prochaska et al.: Astrophys. J. Supp. **137**, 21 (2001)
24. M.T. Murphy: Probing variations in the fundamental constants with quasar absorption lines. PhD thesis, University of New South Wales (2002). Available at http://www.ast.cam.ac.uk/∼mim
25. J.D. Bekenstein: Phys. Rev. Lett., submitted, preprint (astro-ph/0301566) (2003)
26. E.J. Angstmann, V.V. Flambaum, S.G. Karshenboim: preprint (physics/0403009) (2004)
27. J.N. Bahcall, C.L. Steinhardt, D. Schlegel: Astrophys. J. **600**, 520 (2004)
28. J.K. Webb, M.T. Murphy, V.V. Flambaum, S.J. Curran: Astrophys. Space Sci. **283**, 565 (2003)
29. M.T. Murphy, J.K. Webb, V.V. Flambaum, C.W. Churchill, J.X. Prochaska: Mon. Not. Roy. Soc. **327**, 1223 (2001)
30. J.K. Webb et al.: Phys. Rev. Lett. **87**, 091301 (2001)
31. M.T. Murphy, J.K. Webb, V.V. Flambaum, S.J. Curran: Astrophys. Space Sci. **283**, 577 (2003)
32. J.C. Berengut, V.A. Dzuba, V.V. Flambaum: Phys. Rev. A **68**, 022502 (2003)
33. L. Hallstadius: Z. Phys. A **291**, 203 (1979)
34. R.E. Drullinger, D.J. Wineland, J.C. Bergquist: Appl. Phys. **22**, 365 (1980)
35. P.L. Gay, D.L. Lambert: Astrophys. J. **533**, 260 (2000)
36. D. Yong, D.L. Lambert, I.I. Ivans: Astrophys. J. **599**, 1357 (2003)
37. F.X. Timmes, S.E. Woosley, T.A. Weaver: Astrophys. J. Supp. **98**, 617 (1995)
38. Y. Fenner et al.: Publ. Astron. Soc. Australia **20**, 340 (2003)
39. M.D. Shetrone: Astron. J. **112**, 2639 (1996)
40. D. Yong, F. Grundahl, D.L. Lambert, P.E. Nissen, M.D. Shetrone: Astron. Astrophys. **402**, 985 (2003)

41. M. Forestini, C. Charbonnel: Astron. Astrophys. Supp. **123**, 241 (1997)
42. L. Siess, M. Livio, J. Lattanzio: Astrophys. J. **570**, 329 (2002)
43. A.I. Karakas, J.C. Lattanzio: Publ. Astron. Soc. Australia **20**, 279 (2003)
44. T. Ashenfelter, G.J. Mathews, K.A. Olive: Phys. Rev. Lett. **92**, 041102 (2004)
45. R.C. Kennicutt Jr.: 'Overview: The Initial Mass Function in Galaxies'. In: *The Stellar Initial Mass Function (38th Herstmonceux Conference)*, ed. by G. Gilmore, D. Howell (Astron. Soc. Pac., San Francisco, CA, U.S.A, 1998) p. 1
46. P. Molaro, S.A. Levshakov, S. D'Odorico, P. Bonifacio, M. Centurión: Astrophys. J. **549**, 90 (2001)
47. M. Pettini, C.C. Steidel, K.L. Adelberger, M. Dickinson, M. Giavalisco: Astrophys. J. **528**, 96 (2000)
48. K. Nomoto et al.: Nucl. Phys. A **621**, 467 (1997)
49. L.M. Dray, C.A. Tout, A.I. Karakas, J.C. Lattanzio: Mon. Not. Roy. Soc. **338**, 973 (2003)
50. M. Pettini, S.L. Ellison, J. Bergeron, P. Petitjean: Astron. Astrophys. **391**, 21 (2002)
51. M. Centurión, P. Molaro, G. Vladilo, C. Péroux, S.A. Levshakov, V. D'Odorico: Astron. Astrophys. **403**, 55 (2003)
52. P. Langacker, G. Segrè, M.J. Strassler: Phys. Lett. B **528**, 121 (2002)
53. W.J. Marciano: Phys. Rev. Lett. **52**, 489 (1984)
54. X. Calmet, H. Fritzsch: Euro. Phys. J. C **24**, 639 (2002)
55. V.V. Flambaum, E.V. Shuryak: Phys. Rev. D **65**, 103503 (2002)
56. K.A. Olive, M. Pospelov: Phys. Rev. D **65**, 085044 (2002)
57. V.F. Dmitriev, V.V. Flambaum: Phys. Rev. D **67**, 063513 (2003)
58. V.V. Flambaum, E.V. Shuryak: Phys. Rev. D **67**, 083507 (2003)
59. M.T. Murphy, J.K. Webb, V.V. Flambaum, M.J. Drinkwater, F. Combes, T. Wiklind: Mon. Not. Roy. Soc. **327**, 1244 (2001)
60. A.I. Shlyakhter: Nature **264**, 340 (1976)
61. T. Damour, F. Dyson: Nucl. Phys. B **480**, 37 (1996)
62. Y. Fujii et al.: Nucl. Phys. B **573**, 377 (2000)
63. H. Oberhummer, A. Csótó, M. Fairbairn, H. Schlattl, M.M. Sharma: Nucl. Phys. A **719**, 283 (2003)
64. S.R. Beane, M.J. Savage: Nucl. Phys. A **713**, 148 (2003)
65. S.G. Karshenboim: Canadian J. Phys. **318**, 680 (1997)
66. J.D. Prestage, R.L. Tjoelker, L. Maleki: Phys. Rev. Lett. **74**, 3511 (1995)
67. H. Marion et al.: Phys. Rev. Lett. **90**, 150801 (2003)
68. S. Bize et al.: Phys. Rev. Lett. **90**, 150802 (2003)
69. D.B. Leinweber, D.H. Lu, A.W. Thomas: Phys. Rev. D **60**, 034014 (1999)
70. N.A. Demidov, E.M. Ezhov, B.A. Sakharov, B.A. Uljanov, A. Bauch, B. Fisher: 'Investigations of the frequency instability of CH1-75 hydrogen masers'. In: *Proc. of the 6th European Frequency and Time Forum, Noordwijk, Netherlands 1992.* (European Space Agency, Noordwijk, 1993), p. 409
71. L.A. Breakiron: 'A Comparative Study of Clock Rate and Drift Estimation'. In: *Proc. of the 25th Annual Precise Time Interval Applications and Planning Meeting.* (U.S. Naval Observatory Time Service Department, Washington DC, 1993) p. 401
72. T. Dent: Nucl. Phys. B, accepted, preprint (hep-ph/0305026) (2003)

Astrophysical Constraints on Hypothetical Variability of Fundamental Constants

Sergei A. Levshakov

Department of Theoretical Astrophysics, Ioffe Physico-Technical Institute,
194021 St. Petersburg, Russia

Abstract. Applied to three inhomogeneous samples of quasar absorption-line spectra, the many-multiplet method gives a shift in the value of the fine-structure constant of $\Delta\alpha/\alpha = (-5.4 \pm 1.2) \times 10^{-6}$ in the redshift range $0.2 < z < 3.7$ (Sydney group). The 1σ error claimed in these calculations is, however, much too small and cannot be maintained by current observations of quasars. We present a modified many-multiplet method to set an upper limit on $\Delta\alpha/\alpha$ from a homogeneous sample of Fe II lines identified in the up-to-date best quality spectrum of the quasar HE 0515–4414. Our result is $\Delta\alpha/\alpha = (1.1 \pm 1.1) \times 10^{-5}$ at $z = 1.149$. Theoretical models of the fundamental physical interactions predict that the proton-to-electron mass ratio ($\mu = m_{\rm p}/m_{\rm e}$) may relate to the shift in $\Delta\alpha/\alpha$ as $\Delta\mu/\mu = \mathcal{R}\,\Delta\alpha/\alpha$. We use high-resolution observations of molecular hydrogen H_2 ultraviolet absorption lines at $z = 3.025$ toward the quasar Q 0347–3818 to bound the value of $\mathcal{R}$. The obtained constraints on $\Delta \ln \mu = (2.1 \pm 3.6) \times 10^{-5}$ and on $\Delta \ln \alpha = (1.1 \pm 1.1) \times 10^{-5}$ rule out vary large values of $|\mathcal{R}| > 6$. Future observations with a new High Accuracy Radial velocity Planet Searcher spectrograph may provide a crucial test for the $\Delta\alpha/\alpha$ measurements at a level of 10^{-6}.

1 Introduction

The purpose of this paper is to outline some problems related to the measurements of the dimensionless fundamental constants from quasar (QSO) absorption-line spectra: the fine-structure constant $\alpha = e^2/\hbar c$ and the proton-to-electron mass ratio $\mu = m_{\rm p}/m_{\rm e}$, which characterize the strength of electromagnetic and strong interactions, respectively.

The current interest in this field stems from the prospect of using the high-resolution spectra of extragalactic objects to measure with high accuracy the transition energies of a certain ionic species and molecules at different cosmological redshifts

$$z = \frac{\lambda - \lambda_0}{\lambda_0} \,,$$

where λ is the observed wavelength of a given atomic or molecular transition and λ_0 is the corresponding rest frame (laboratory) value. Modern observational data are ranging from $z = 0$ (the Milky Way) up to $z \sim 6$ (distant QSOs).

If physical constants varied in the past, then the comparison of the transition energies measured from QSO spectra with their laboratory values provides direct

S.A. Levshakov, Astrophysical Constraints on Hypothetical Variability of Fundamental Constants, Lect. Notes Phys. **648**, 151–166 (2004)
http://www.springerlink.com/

astrophysical constraints on the ratios α_z/α and μ_z/μ over a cosmological timescale ($\Delta t \sim 10^{10}$ yr). We define $\Delta\alpha/\alpha = (\alpha_z - \alpha)/\alpha$ and $\Delta\mu/\mu = (\mu_z - \mu)/\mu$, where index z marks the values of the fine-structure constant and the proton-to-electron mass ratio at redshift z, whereas their laboratory values at $z = 0$ are denoted by α and μ, respectively[1]. At present, these constants are measured with the relative accuracies $\delta_\alpha/\alpha = 4 \times 10^{-9}$ and $\delta_\mu/\mu = 2 \times 10^{-9}$ and their values are equal to $\alpha = 1/137.035\,999\,76(50)$ and $\mu = 1836.152\,667\,5(39)$ [3].

The variability of the fundamental constants was firstly suggested by Milne [4] and Dirac [5] who assumed that the Newtonian gravitational constant G may change in cosmic time. Different functional forms for the time variations of G and other constants were later considered in a series of publications (see [6], for a review). The variability of α caused by a coupling between electromagnetism and gravity is considered within the framework of different modifications of the Kaluza-Klein theory (e.g., [7–12]). The number of theoretical publications on this topic is considerably increased in the last two years.

For astronomical observations, it is more convenient to deal with dimensionless constants since they are independent of the choice of physical units. That is why most efforts to reveal a hypothetical variability of constants are connected with measurements of α (see [2,13] and references cited therein).

The first observational study (1967) of the time dependence of α from QSO absorption spectra was based on the analysis of alkali doublets (AD method) and yielded $\Delta\alpha/\alpha = -0.02 \pm 0.05$ at $z = 1.95$ [14]. Since then, there have been many other AD studies of $\Delta\alpha/\alpha$ with increasing accuracy in accord with increasing quality of observations: $\Delta\alpha/\alpha = (0 \pm 1.5) \times 10^{-3}$ at $z = 3.2$ [15], $\Delta\alpha/\alpha = (-4.5 \pm 4.3 \pm 1.4) \times 10^{-5}$ at $z = 2 - 4$ [16], $\Delta\alpha/\alpha = (-0.5 \pm 1.3) \times 10^{-5}$ at $z = 2 - 3$ [17]. However, the use of absorption lines from different multiplets in different ions, the so-called many-multiplet (MM) method, may in principle increase the accuracy of these $\Delta\alpha/\alpha$ estimations [1,2].

Applying the MM method to three samples of Keck/HIRES QSO data, Murphy et al. [2] obtained a statistically significant negative value of $\Delta\alpha/\alpha = (-5.4 \pm 1.2) \times 10^{-6}$ in the redshift interval $0.2 < z < 3.7$. In spite of all advantages of the MM method, the result is nevertheless surprising since it implicitly assumes extremely stable conditions during the observing nights and unrealistically homogeneous physical parameters in the absorption clouds. We consider computational caveats of the AD and MM methods in Sect. 2.

If α is supposed to be time dependent, the other gauge coupling constants should also depend on time as predicted by theoretical models of the fundamental physical interactions. For instance, within the framework of a unified theory (e.g., supersymmetric SU(5)) the time variation of α at low energies is caused by a time variation of the unification scale Λ ($= 213^{+38}_{-35}$ MeV), which, in turn, leads to a time variation of the proton mass m_p :

$$\frac{\Delta m_\mathrm{p}}{m_\mathrm{p}} = \frac{\Delta\Lambda}{\Lambda} = \mathcal{R}\,\frac{\Delta\alpha}{\alpha}\,, \tag{1}$$

[1] In the literature μ is sometimes referred to its inverse value, $\mu^* = m_\mathrm{e}/m_\mathrm{p}$. In this case $\Delta\mu^*/\mu^* = -\Delta\mu/\mu$.

where the numerical coefficient $\mathcal{R}$ ($\simeq 40$) is known with an uncertainty of about 20% [18].

Assuming that the possible time changes of the electron mass or the quark masses are negligible, the relation (1) suggests that a variation of $\Delta\alpha/\alpha \simeq -5.4\times 10^{-6}$ should be accompanied by a shift in the the proton-to-electron mass ratio of $\Delta\mu/\mu \sim -2\times 10^{-4}$. This prediction can be tested by measuring the electronic-vibrational-rotational lines of the H_2 molecule observed in QSO spectra at high-z as described in [19]. We discuss this test in Sect. 3. Our conclusions and remarks for future prospects are summarized in Sect. 4.

2 Methods to Constrain $\Delta\alpha/\alpha$ from QSO Absorption Spectra

2.1 The Alkali-Doublet (AD) Method

The value of the fine-structure constant α_z at redshift z can be directly estimated from the relative wavelength separation between the fine-splitting lines of an alkali doublet (AD method) [20]:

$$\frac{\alpha_z}{\alpha} = \left(\frac{\Delta\lambda_z}{\langle\lambda\rangle_z} \Big/ \frac{\Delta\lambda}{\langle\lambda\rangle} \right)^{1/2} , \tag{2}$$

where $\Delta\lambda_z$ and $\langle\lambda\rangle_z$ are, respectively, the fine-structure separation and the weighted mean wavelength for a given doublet from a QSO absorption system at redshift z, while $\Delta\lambda$ and $\langle\lambda\rangle$ denote the same quantities from laboratory data. The uncertainty $\sigma_{\Delta\alpha/\alpha}$ in the individual measurement can be estimated from [21]:

$$\sigma_{\Delta\alpha/\alpha} \simeq \frac{1}{\sqrt{2}} \left(\delta_{\lambda_z}^2 + \delta_\lambda^2\right)^{1/2} , \tag{3}$$

where $\delta_{\lambda_z} = \sigma_{\lambda_z}/\Delta\lambda_z$, $\delta_\lambda = \sigma_\lambda/\Delta\lambda$, and σ_{λ_z}, σ_λ are, respectively, the errors of the wavelengths in the absorption system and laboratory.

Equation (2) shows that the AD method uses only the wavelength ratios and thus it does not require the measurements of absolute wavelengths. This fact excludes large systematic errors caused by the wavelength calibration in a wide spectral range. Therefore the accuracy of the AD measurements is mainly restricted by statistical fluctuations in the recorded counts at different wavelengths within the line profile.

The measurement of the line profile parameters and their errors are thoroughly described in [22]. The midpoint λ_c of an interval over M pixels covering the line profile can be calculated from the equations

$$\lambda_c = \frac{a}{W_\lambda} \sum_{i=1}^{M} \lambda_i (1 - I_i/C_i) , \tag{4}$$

and

$$W_\lambda = a \sum_{i=1}^{M} (1 - I_i/C_i) , \tag{5}$$

where W_λ is the equivalent width, a is the wavelength interval between pixels, I_i is the observed count rate at pixel i, C_i is the continuum count rate at pixel i, and λ_i is the wavelength at pixel i.

The errors in λ_c and W_λ caused by counting statistic and the uncertainty of the overall height of the continuum level are given by [22]:

$$\sigma_{\lambda_c} = \frac{a}{W_\lambda}\left[\left\langle\frac{\sigma_I}{C}\right\rangle^2\left(A - 2\lambda_c B + M\lambda_c^2\right) + \left\langle\frac{\sigma_C}{C}\right\rangle^2\left(B - M\lambda_c\right)^2\right]^{1/2}, \quad (6)$$

and

$$\sigma_{W_\lambda} = a\left[\langle\sigma_I/C\rangle^2\, M + \langle\sigma_C/C\rangle^2\left(M - W_\lambda/a\right)^2\right]^{1/2}, \quad (7)$$

where $A = \sum_i \lambda_i^2$, $B = \sum_i \lambda_i$, $\langle\sigma_I/C\rangle$ is the inverse mean signal-to-noise $(S/N)^{-1}$ ratio over the line measurement interval, and $\langle\sigma_C/C\rangle$ is the mean accuracy of the continuum level.

If the rms error of the wavelength scale calibration is $\sigma_{\rm sys}$, then the total error is given by the sum

$$\sigma_{\rm tot}^2 = \sigma_{\lambda_c}^2 + \sigma_{\rm sys}^2. \quad (8)$$

The error of the wavelength scale calibration is related to the radial velocity accuracy as

$$\sigma_v = c\,\sigma_{\rm sys}/\lambda\,, \quad (9)$$

where c is the speed of light.

For example, $\sigma_{\rm sys} \simeq 0.002$ Å for VLT/UVES data [23] corresponds at 5000 Å to $\sigma_{\rm sys} = 0.12$ km s^{-1}. If $\sigma_{\lambda_c} = 0.002$ Å, then for the Mg II $\lambda\lambda 2796, 2803$ Å doublet (3) gives $\sigma_{\Delta\alpha/\alpha} = 1.3 \times 10^{-4}$ at $z = 2$, which is a typical accuracy for the individual AD estimations from isolated doublets with $W_\lambda \sim 100$ mÅ. Equation (6) shows that σ_{λ_c} decreases linearly with W_λ, meaning that, not surprisingly, the error for weak absorption lines ($W_\lambda \ll 1$ Å) may be very large. For example, the rest-frame equivalent widths for Fe II $\lambda 1608$ Å and $\lambda 1611$ Å at $z = 3.39$ toward Q0000–2620 measured from the Keck/HIRES data are $W_{1608} \simeq 0.178$ Å and $W_{1611} \simeq 0.003$ Å [24], i.e. the measured errors σ_{λ_c} for these lines should differ by ~ 50 times. It is clear that to achieve the highest accuracy in the $\Delta\alpha/\alpha$ estimations by means of the the AD method only well separated and unsaturated doublets with $W_\lambda \sim 100$ mÅ should be selected for the analysis.

2.2 The Many-Multiplet (MM) Method

The many-multiplet (MM) method is a generalization of the AD method (see [2] and references therein). The method uses many transitions from different multiplets and different ions observed in a QSO absorption system. The line profiles selected for the MM analysis are represented by a sum of the same number of components (Voigt profiles) each of which is defined by three individual parameters (the column density, the Doppler width or b-parameter, and the redshift

z_i) and a common value of $\Delta\alpha/\alpha$. The rest wavenumbers in the MM method can be altered according to [2]

$$\omega_z = \omega_0 + q\, x_z \ , \tag{10}$$

where

$$x_z = (\alpha_z/\alpha)^2 - 1 \ , \tag{11}$$

and ω_z, ω_0 are the rest-frame wavenumbers of a transition in a QSO system with redshift z and measured in the laboratory, respectively. The sensitivity coefficients q (determining the sensitivity of ω_0 to the variation of α) are listed in Table 2 in [2].

All parameters are estimated through the minimization of χ^2 simultaneously for the whole set of lines. A significant improvement in the accuracy of the $\Delta\alpha/\alpha$ estimations by the MM method compared to the AD method is due to a wide range of the q values (see Sect. 2.3 below). Nevertheless, the MM method has some immanent shortcomings which could affect the claimed accuracy $\sigma_{\langle\Delta\alpha/\alpha\rangle} = 1.2 \times 10^{-6}$.

First of all, the weakest point of the MM method is the multiple velocity component Voigt profile deconvolution. From the mathematical point of view, any deconvolution itself is a typical *ill-posed* problem [25]. When several smoothing operators (like convolution with the spectrograph point-spread function, summing, and integrating) are involved in the minimization procedure, the ill-posing may cause quite ambiguous results. In case of complex line profiles when many components are required to describe the whole profile, the fitting parameters cannot be estimated with high accuracy because of strong *inter-parameter correlations*. For example, the value of $\Delta\alpha/\alpha$ can be artificially constrained by the increase of the component number or by variations of b-values. Thus, to enhance the reliability of the results only simple line profiles (requiring a minimum number of components) should be taken into the MM analysis.

From the physical point of view, the Voigt profile deconvolution assumes that the complex profile is caused by separate clouds lined along the line of sight with different radial velocities, each having its own constant gas density, kinetic temperature and, hence, its own ionization state. However, observations of QSOs have shown that with increasing spectral resolution more and more subcomponents appear in the line profile. This complexity indicates that, in general, metal lines arise in continuous turbulent media with fluctuating velocity and density fields.

We note in passing that the highest spectral resolution achieved in observations of QSOs at modern giant telescopes is ~ 5 km s^{-1} (FWHM), whereas the expected values for minimum (thermal) widths of metal lines are less or about 1 km s^{-1} for typical kinetic temperatures 100 K $< T_{\rm kin} < 10^4$ K. This means that we are still not able to measure directly the true intensity in the line profile but observe only the apparent intensity which is a convolution of the true spectrum and the spectrograph point-spread function.

Thus, to measure the transition energies, the distribution function of the radial velocities within the intervening cloud must be known for the AD method,

whereas the MM method requires in addition the knowledge of the density distribution, for it deals with ions of different elements which may not arise cospatially [26]. Since the velocity and density distributions are determined from the same line profiles which are used in the measurements of $\Delta\alpha/\alpha$, it is not possible to discriminate among kinematic effects and possible changes in α on the basis of only one absorption system. But if we have many absorption systems, a statistical approach may probably solve this problem if shifts in the measured $\Delta\alpha/\alpha$ values are of random nature and, thus, can be averaged out in a large ensemble of absorption systems where any deviations in velocity distributions of different ions are equally probable.

Shifts in $\Delta\alpha/\alpha$ may also arise if the main assumption of the MM method that different ionic species trace each other exactly in their spatial distribution is not adequate. Dissimilarities in the line profiles of different ions may be caused by density fluctuations leading to inhomogeneous photoionization in the absorption cloud.

The QSO absorption clouds form a very inhomogeneous population. Their linear sizes, L, can be estimated from the observed neutral hydrogen column densities, $N(\mathrm{H\,I}) \sim 10^{20} - 10^{21}$ cm^{-2}, and the estimated gas number densities, $n_{\mathrm{H}} \sim 10^{-2} - 1$ cm^{-3}, giving $L = N(\mathrm{H\,I})/n_{\mathrm{H}} \sim$ 10s pc – 10s kpc. The intervening clouds may also contain small dense filaments with much higher densities ($n_{\mathrm{H}} \sim 10^2$ cm^{-3}) and the line-of-sight sizes less than 1 pc [27].

It is usually assumed that the metal absorbers in QSO spectra are in photoionization equilibrium with the ambient metagalactic UV background. To describe the thermal and the ionization state of the gas in the diffuse clouds it is convenient to introduce the dimensionless 'ionization parameter' $U = n_{\mathrm{ph}}/n_{\mathrm{H}}$, equal to the ratio of the number of photons in 1 cm^3 with energies above one Rydberg to the number density of the gas.

For an illustrative purpose, we will assume that the UV background is not time or space dependent and that its shape and intensity is given by the metagalactic spectrum at $z = 3$ presented in [28] ($n_{\mathrm{ph}} \simeq 2 \times 10^{-5}$ cm^{-3}). Then, fractional ionizations $\Upsilon_{a,i} = n_{a,i}/n_a$ (with n_a being the local number density of element a and $n_{a,i}$ the density of ions in the ith ionization state) of different elements can be calculated as functions of U using the photoionization code CLOUDY [29].

Figure 1 shows these results for a metallicity $Z = 0.1Z_{\odot}$ (typical for the QSO metal absorbers) and for the elements involved in the MM analysis in [2]. Similar profiles for different ions can only be observed when the ratios $\Upsilon_{a,i}/\Upsilon_{b,j}$ are constant, i.e., the curves in Fig. 1 are parallel. This condition is realized for Mg I, Mg II, Al II, Si II, Fe II, Ni II, and Zn II in the range $\log U < -4.2$, i.e., for the most dense volumes of the cloud. Such dense clumps (filaments) have, however, a negligible filling factor since their sizes are small, and, thus, most of the diffuse cloud is filled up by a rarefied, low density gas characterized by higher values of $\log U$. In the range $\log U > -4$, the diversity of the Υ_X values is considerable and one may expect to observe nonsimilar profiles of the above mentioned ions. Under these circumstances, in order to be more secure the MM method should

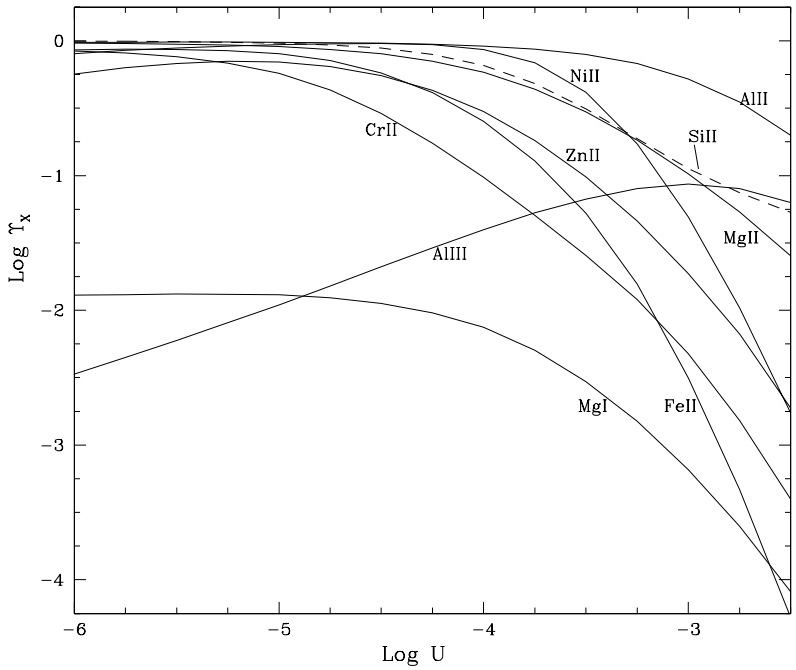


Fig. 1. Fractional ionizations $\log \Upsilon_X$ vs. the ionization parameter $\log U$ in case of the Haardt-Madau [28] metagalactic UV background at $z = 3$ and metallicity $Z = 0.1 Z_\odot$. Similarity of the line profiles of a pair of ions a and b requires the fractional ionization ratio Υ_a/Υ_b to be constant for different U

work with homogeneous samples of ions which have undoubtly the same volume distribution like, e.g., Fe II lines.

Note that Al III is obviously not a good candidate for MM calculations since the shape of the $\Upsilon_{\mathrm{Al\,III}}$ curve clearly differs from the others. The fact that no additional scatter in $\Delta\alpha/\alpha$ for systems containing Al III was found in [2] may imply that the available spectral resolution is not high enough to obtain the true profiles of the metal lines.

To conclude this section we comment on the accuracy $\sigma_{\Delta\alpha/\alpha}$ which can in principle be achieved from the up-to-date QSO observations. Using the method of error propagation, one finds from (10) the relative error of ω_z:

$$\delta_{\omega_z} \equiv \sigma_{\omega_z}/\omega_z = 2\,|Q|\,\sigma_{\Delta\alpha/\alpha}\ . \tag{12}$$

Here $Q = q/\omega_0$ is the dimensionless sensitivity coefficient.

For the lines listed in Table 2 in [2], the Q values are ranging from -0.027 (Cr II λ2066 Å) to 0.050 (Zn II λ2026 Å), however, for the most frequently used transitions in the MM calculations (Table 1 in [2]), this range is narrower: Q(Mg II λ2803 Å) = 0.0034, and Q(Fe II λ2600 Å) = 0.035. These numbers show that the error of the sample mean $\sigma_{\langle\Delta\alpha/\alpha\rangle} = 1.2 \times 10^{-6}$ found in [2] requires the relative error $\langle\delta_{\omega_z}\rangle \simeq 8.6 \times 10^{-9} - 8.4 \times 10^{-8}$, meaning that the

sample mean wavelength accuracy should be $\langle\sigma_\lambda\rangle \simeq 2\times10^{-5} - 2\times10^{-4}$ Å at 2000 Å ($\sigma_\lambda = \lambda\,\delta_{\omega_z}$), which is equal to the radial velocity accuracy of $\Delta v \simeq 2.5 - 25$ m s^{-1} ($\Delta v = c\,\delta_{\omega_z}$). Even if the individual measurement is an order of magnitude less accurate as compared to the sample mean, the uncertainty of $\Delta v \simeq 25 - 250$ m s^{-1} seems to be hardly realized for all systems collected in [2]. Namely, thermal shifts of order 300 m s^{-1} are expected if the temperature changes by 0.5 K between the science and calibration spectra exposures [23].

2.3 The Regression MM Method

The standard MM method based on the multiple velocity component Voigt profile fit to the complex spectra when $\Delta\alpha/\alpha$ is treated as an additional free parameter [2] can be easily modified to be free from the uncertainties mentioned in the previous section. By its very nature, the MM method is similar to that developed in [19] for the $\Delta\mu/\mu$ calculations. So, by analogy, we can re-write (10) in the linear approximation ($|\Delta\alpha/\alpha| \ll 1$) in the form

$$z_i = z_0 + \kappa_\alpha Q_i \,, \tag{13}$$

where the slope parameter κ_α is given by

$$\kappa_\alpha = 2\,(1+z_0)\,\frac{\Delta\alpha^*}{\alpha} \,. \tag{14}$$

Here $\Delta\alpha^*/\alpha = -\Delta\alpha/\alpha$.

If $\Delta\alpha/\alpha$ is not zero, there should be a correlation between z_i ($= \lambda_{c,i}/\lambda_{0,i} - 1$) and Q_i, and the regression parameters z_0 and κ_α can be estimated from the measured centers of the isolated metal lines.

For instance, Fe II lines listed in Table 2 in [2], for which the Q values are ranging in the broad interval between -0.019 and 0.035, may be suitable for such calculations. Equation (13) shows that the broader the interval of Q-values, the higher the accuracy of the slope parameter κ_α (see, e.g., [30] where the errors of the linear regression are calculated).

It is also an advantage that Fe II lines arise mainly in low ionized regions, where their thermal width is the smallest among other elements (Mg, Al, Si) proposed for the MM method (we do not consider Cr II, Ni II, and Zn II as appropriate for the MM analysis because they are rare in QSO spectra and, when observed, are usually very week) since the narrow lines allow to measure the transition energies with highest accuracy. Below we consider an example of such a Fe II sample.

Figure 2 shows isolated and unsaturated Fe II lines from the metal system at $z = 1.149$ toward the bright quasar HE 0515–4414. The spectra of this quasar were obtained with the UV-Visual Echelle Spectrograph (UVES) installed at the VLT/Kueyen telescope. The VLT/UVES data have a very high signal-to-noise ratio ($S/N \simeq 90 - 150$ per resolution element) and a high resolution power of $\lambda/\Delta\lambda \simeq 55\,000$ (FWHM $\simeq 5.5$ km s^{-1}). The absorption systems from this data were analyzed in [27] and [31,32].

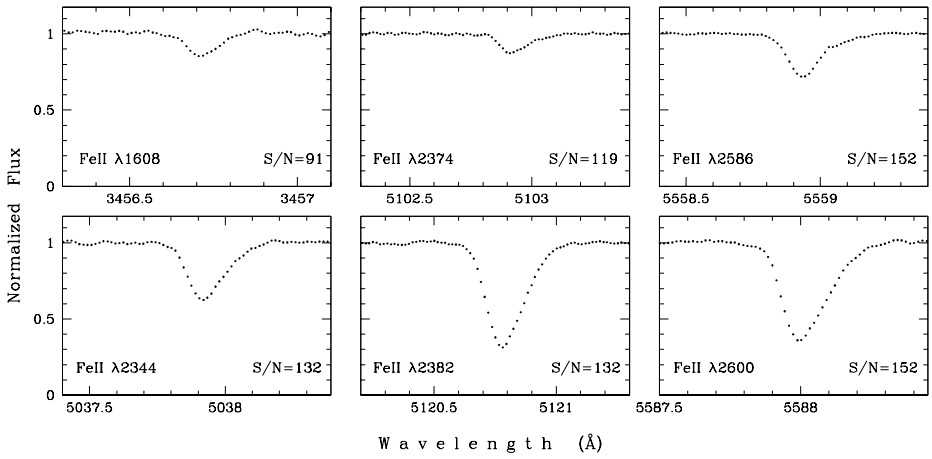


Fig. 2. Fe II lines associated with the $z = 1.149$ absorption system toward the quasar HE 0515–4414 (normalized intensities are shown by dots). The mean signal-to-noise (S/N) ratios per resolution element are indicated

Table 1. Fe II lines at $z = 1.149$ toward the quasar HE 0515–4414 and sensitivity coefficients Q

λ_0^a, Å	f	λ_c, Å	W_λ, mÅ	z	Q
1608.45085(8)	0.0580[b]	3456.7176(46)	7.8(1.8)	1.1490975(28)	−0.0193(48)
2344.2130(1)	0.1140[b]	5037.9398(23)	27.5(1.6)	1.1490964(10)	0.0294(35)
2374.4603(1)	0.0313[b]	5102.9319(79)	8.4(1.5)	1.1490913(33)	0.0389(35)
2382.7642(1)	0.3006[c]	5120.8015(22)	61.0(1.9)	1.1491012(9)	0.0357(36)
2586.6496(1)	0.0684[d]	5558.9439(29)	20.1(1.4)	1.1490904(11)	0.0393(39)
2600.1725(1)	0.2239[c]	5588.0269(21)	59.1(2.0)	1.1490985(8)	0.0353(39)

[a]Vacuum rest wavelengths are taken from [2].

REFERENCES: [b]Welty et al. [33]; [c]Morton [34]; [d]Cardelli & Savage [35].

We used (4)-(8) to estimate the rest-frame equivalent widths W_λ, the line centers λ_c, and their errors which are listed in Table 1. In this table, column (2) presents the oscillator strengths for absorption. The sensitivity coefficients Q_i and their errors σ_{Q_i} are calculated from the values q_i, σ_{q_i}, and $\omega_{0,i}$ listed in Table 2 in [2].

Our result of the linear regression analysis is shown in Fig. 3 by the solid line, while two dotted lines correspond to the 1σ deviations of the slope parameter κ_α. We find $z_0 = 1.1490988 \pm 0.0000016$ and $\Delta\alpha/\alpha = (1.1 \pm 1.1) \times 10^{-5}$ ($1\ \sigma$). The uncertainties of the z_0 and $\Delta\alpha/\alpha$ values are calculated through statistical Monte Carlo simulations assuming that the deviations in the z_i and Q_i values

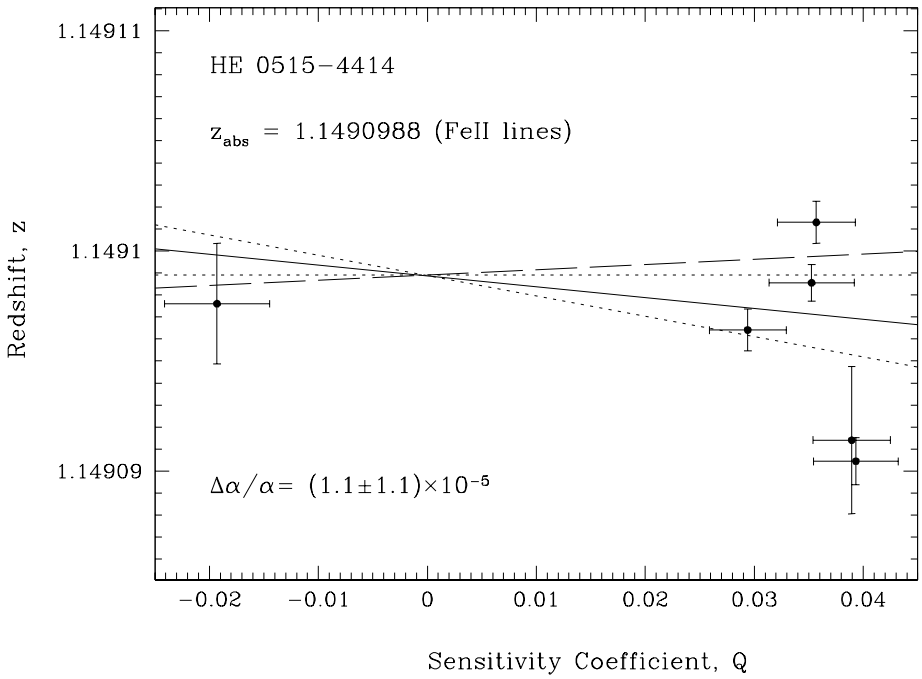


Fig. 3. Relation between the redshifts z_i calculated for individual Fe II lines listed in Table 1 and their sensitivity coefficients Q_i. The solid line is the linear regression $z_i = z_0 + \kappa_\alpha Q_i$. The dotted lines representing the 1σ deviations from the optimal slope demonstrate that $\Delta\alpha/\alpha = 0$ at the level $\sim 2 \times 10^{-5}$. The dashed line is drawn for $\Delta\alpha/\alpha = -5.4 \times 10^{-6}$ found in the MM analysis in [2]

are equally likely in the intervals $[z_i - \sigma_{z_i}, z_i + \sigma_{z_i}]$ and $[Q_i - \sigma_{Q_i}, Q_i + \sigma_{Q_i}]$. The result obtained shows that $\Delta\alpha/\alpha = 0$ within the 1σ interval. For comparison, the linear regression for $\Delta\alpha/\alpha = -5.4 \times 10^{-6}$ based on the inhomogeneous sample containing 127 absorption systems [2] is shown by the dashed line in Fig. 3. Although the sign of $\Delta\alpha/\alpha$ is opposite, the value of $\Delta\alpha/\alpha = -5.4 \times 10^{-6}$ is consistent with our data within the 2σ interval. The accuracy of the regression MM analysis may be increased if a few homogeneous samples are combined. The regression analysis is easily generalized for this case (see, e.g., [36]).

3 Constraints on the Proton-to-Electron Mass Ratio

The proton-to-electron mass ratio μ can be estimated from high redshift molecular hydrogen systems. With some modifications, such measurements were performed for the $z = 2.811$ H_2-bearing cloud toward the quasar PKS 0528–250 in [19] and [37] setting the limit of $|\Delta\mu/\mu| < 1.8 \times 10^{-4}$ (1σ). Later on higher resolution spectra of this QSO obtained with the VLT/UVES [38] revealed that the H_2 profiles at $z = 2.811$ are complex and consist of at least two subcomponents with $\Delta v \simeq 10$ km s^{-1}. Lines arising from the low rotational levels with

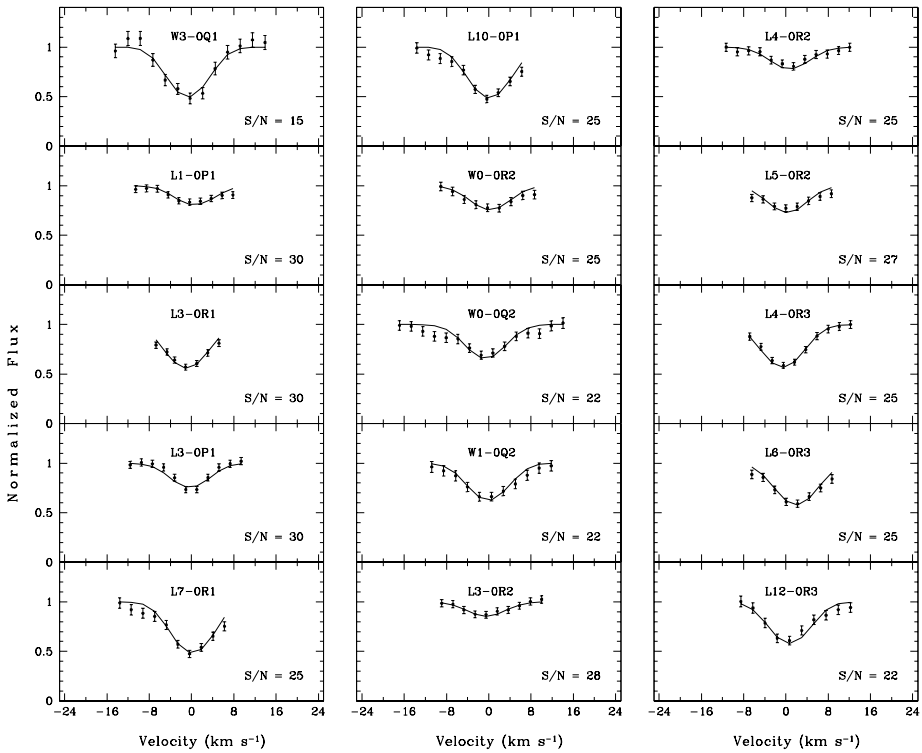


Fig. 4. H_2 absorption features identified at $z = 3.025$ toward the quasar Q 0347–3819 (normalized intensities are shown by dots with 1σ error bars). The zero radial velocity is fixed at $z_{H_2} = 3.024895$. Smooth curves are the synthetic H_2 profiles found by the least-squares procedure. The mean S/N values per resolution element are indicated

$J = 0$ and $J = 1$ show a red-side asymmetry – contrary to those arising from $J = 2$ and $J = 3$ where a blue-side asymmetry is clearly seen. This complexity makes any higher accuracy measurements of $\Delta\mu/\mu$ in this system impossible. Again, as in the case with the $\Delta\alpha/\alpha$ estimations, one needs a single component narrow H_2 absorption-line system to restrict the observed wavelengths within the uncertainty interval of a few mÅ.

Such a system has been found at $z = 3.025$ toward the quasar Q0347–3819 [39]. The absorption spectrum of this QSO was obtained with FWHM$\simeq$ 7 km s^{-1} and $S/N \sim 15 - 30$ per resolution element at the VLT/UVES. 15 unblended H_2 lines (selected from ~ 80 identified H_2 transitions) which provide the most accurate line center measurements are shown in Fig. 4.

For a given H_2 line, the number of pixels involved in the analysis corresponds to the number of points in the line profile shown in Fig. 4. To measure the line centers we used a method that matches the observed profiles with the synthesized ones to estimate a set of model parameters (for details, see [23]). The results obtained are presented in Table 2. The measured individual z values and

Table 2. H_2 lines at z = 3.025 toward Q 0347–3819 and sensitivity coefficients $\mathcal{K}$

J	Line	λ_0^a, Å	λ_c, Å	z	$\mathcal{K}$
1 . . .	W3-0Q	947.4218(5)	3813.2653(41)	3.024887(5)	0.0217427(8)
	L1-0P	1094.0522(51)	4403.4575(77)	3.02491(2)	−0.0023282(1)
	L3-0R	1063.4603(1)	4280.3051(30)	3.024885(3)	0.0112526(4)
	L3-0P	1064.6056(5)	4284.9249(33)	3.024894(4)	0.0102682(4)
	L7-0R	1013.4412(20)	4078.9785(25)	3.024898(9)	0.03050(1)
	L10-0P	982.8340(6)	3955.8049(51)	3.024896(6)	0.04054(6)
2 . . .	W0-0R	1009.0233(7)	4061.2194(61)	3.024901(7)	−0.0050567(7)
	W0-0Q	1010.9380(1)	4068.9088(53)	3.024885(6)	−0.0068461(2)
	W1-0Q	987.9744(20)	3976.4943(49)	3.02490(1)	0.0039207(2)
	L3-0R	1064.9935(9)	4286.4755(98)	3.02488(1)	0.0097740(5)
	L4-0R	1051.4981(4)	4232.1793(85)	3.024904(8)	0.015220(1)
	L5-0R	1038.6855(32)	4180.6048(70)	3.02490(1)	0.020209(3)
3 . . .	L4-0R	1053.9770(11)	4242.1419(28)	3.024890(5)	0.012837(2)
	L6-0R	1028.9832(16)	4141.5758(31)	3.024921(7)	0.022332(7)
	L12-0R	967.6752(21)	3894.7974(39)	3.02490(1)	0.0440(2)

[a]Listed wavelengths are from [40,41].

their standard deviations are listed in column (5). Using these redshifts we can constrain possible changes of μ.

The rest-frame wavelengths, $\lambda_{0,i}$, of different electronic-vibrational-rotational transitions of H_2 depend in a different way on the reduced mass of the molecule. The sensitivity of $\lambda_{0,i}$ to variation of μ is given by the sensitivity coefficient $\mathcal{K}_i$ defined as [19]

$$\mathcal{K}_i = \frac{\mu}{\lambda_{0,i}} \frac{d\lambda_{0,i}}{d\mu} \; . \tag{15}$$

Following the procedure described in [37] we calculate the coefficients $\mathcal{K}_i$ listed in column (6) of Table 2. Their errors are obtained from coefficients Y_{mn} (presented in Table 1 in [37]) using the method of error propagation (Y_{mn} values are considered to be accurate to k decimal places and their rounding errors are set to 0.5×10^{-k}).

In linear approximation, (15) can be re-written in the form

$$\frac{\lambda_{i,z}/\lambda_{j,z}}{\lambda_{i,0}/\lambda_{j,0}} = 1 + (\mathcal{K}_i - \mathcal{K}_j)\, \Delta\mu/\mu \tag{16}$$

or

$$z_i = z_0 + \kappa_\mu\, (\mathcal{K}_i - \bar{\mathcal{K}}) \; , \tag{17}$$

where $\lambda_{i,z}$ and $\lambda_{j,z}$ are the line centers measured in a quasar spectrum, and

$$\kappa_\mu = (1 + z_0)\,\frac{\Delta\mu}{\mu}\,, \tag{18}$$

with z_0 and $\bar{\mathcal{K}}$ being the mean redshift and the mean sensitivity coefficient, respectively.

The linear regression (17) calculated for the complete sample of the H_2 lines from Table 2 provides $(\Delta\mu/\mu)_{J=1+2+3} = (5.0 \pm 3.2) \times 10^{-5}$. However, if we consider H_2 transitions from individual J levels, then the weighted mean redshifts reveal *a gradual shift in the radial velocity* for features arising from progressively higher rotational levels : $z(J{=}1) = 3.024890(2)$, $z(J{=}2) = 3.024895(3)$, and $z(J{=}3) = 3.024904(4)$. The H_2 lines with changing profiles and small velocity shifts as J increases were also observed in our Galaxy in the direction of ζ Ori A. At present these H_2 lines are interpreted as being formed in *different zones* of a postshock gas [42].

Although the $z(J{=}1)$ and $z(J{=}2)$ values are consistent within 1σ intervals, the difference between $z(J{=}1)$ and $z(J{=}3)$ is essential and equals (1.0 ± 0.3) km s^{-1}. This small change in the radial velocity with increasing J may mimic a shift in $\Delta\mu/\mu$. If we exclude from the regression analysis the rotational levels with $J = 3$, then $(\Delta\mu/\mu)_{J=1+2} = (2.1 \pm 3.6) \times 10^{-5}$. The use of lines arising from the same rotational levels would be more reasonable to estimate $\Delta\mu/\mu$, but in our case the sample size is rather small and we have to combine the $J = 1$ and $J = 2$ levels to increase statistics. The linear regression is shown by the solid line in Fig. 5 while two dashed lines correspond to the 1σ deviations of the slope parameter κ_μ.

Thus, the 1σ confidence interval to $\Delta\mu/\mu$ is $-1.5{\times}10^{-5} < \Delta\mu/\mu < 5.7{\times}10^{-5}$. For a cosmological model with $\Omega_{\rm M} = 0.3$, $\Omega_\Lambda = 0.7$, and $H_0 = 72$ km s^{-1} Mpc^{-1}, the look-back time for $z = 3.025$ is 11.2 Gyr [43]. This leads to the restriction $|\dot{\mu}/\mu| < 5 \times 10^{-15}$ yr^{-1} on the variation rate of μ.

4 Conclusions and Future Prospects

We have considered modern experimental investigations of the hypothetical variations of the fundamental physical constants from absorption-line spectra of distant quasars. The subject was limited to the properties of the two dimensionless quantities – the fine-structure constant α and the proton-to-electron mass ratio μ. Our main conclusions are as follows:

- The accuracy $\sigma_{\langle\Delta\alpha/\alpha\rangle} = 1.2 \times 10^{-6}$ claimed in [2] is an order of magnitude higher compared to what can be provided by the wavelength measurements of isolated and unsaturated QSO absorption lines. Taking into account all shortcomings of the MM method as it was used in [2], the quoted value of $\sigma_{\langle\Delta\alpha/\alpha\rangle}$ is likely to be overestimated.
- A modification of the MM method (regression MM method) which makes it more reliable is presented. This modified MM method applied to the homogeneous sample of the isolated and unsaturated Fe II lines yields $\Delta\alpha/\alpha =$

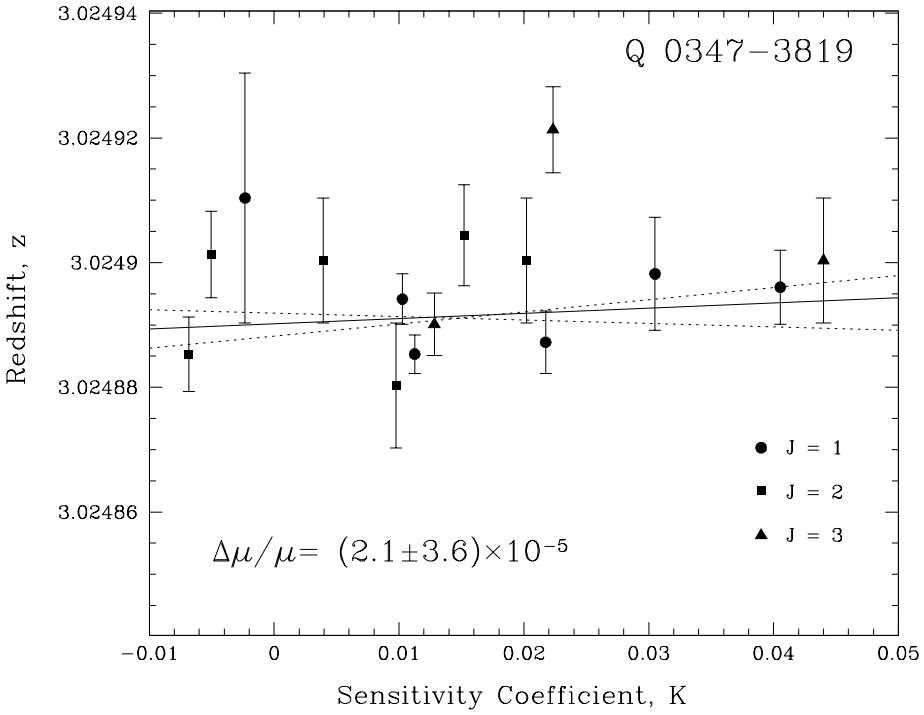


Fig. 5. Relation between the redshift values z_i calculated for individual H_2 lines shown in Fig. 4 and their sensitivity coefficients $\mathcal{K}_i$. The solid line is the linear regression $z_i = z_0 + \kappa_\mu(\mathcal{K}_i - \bar{\mathcal{K}})$. The dotted lines representing the 1σ deviations from the optimal slope κ_μ demonstrate that $\Delta\mu/\mu = 0$ at the level $\sim 5 \times 10^{-5}$.

$(1.1 \pm 1.1) \times 10^{-5}$. The uncertainty of this value is in agreement with the available accuracy of the wavelength measurements.

- If the fine-structure constant varies at the level of $\Delta\alpha/\alpha = (-5.4 \pm 1.2) \times 10^{-6}$ [2], then the link between α_z/α and the proton-to-electron mass ratio μ_z/μ expected within the grand unification theories (e.g., [18]) leads to the shift $\Delta\mu/\mu \sim -2 \times 10^{-4}$. This prediction is *not confirmed*. The variability of μ is constrained by the values $-1.5 \times 10^{-5} < \Delta\mu/\mu < 5.7 \times 10^{-5}$ at redshift $z = 3.025$ [23].

As stated above, the accuracy of $\sigma_\lambda \sim 0.2$ mÅ is required to achieve a level higher than 10^{-5} in the $\Delta\alpha/\alpha$ measurements. This accuracy is ~ 10 times higher than available at modern giant telescopes. In 2003, a new High Accuracy Radial velocity Planet Searcher (HARPS) spectrograph [44] started to operate at the 3.6 m telescope (ESO, La Silla, Chile). Observations with HARPS may in principle provide this desirable accuracy. With a resolving power of $\lambda/\Delta\lambda = 120\,000$ and a projection of the fiber over 4 pixels, one HARPS pixel corresponds to $\simeq 600$ m s^{-1}. For an isolated and unsaturated absorption line, the line center may be measured with an accuracy of about 1/20 of the pixel size (~ 30 m s^{-1}).

This corresponds to $\sigma_\lambda = 0.4$ mÅ at 4000 Å. If isotopic and hyperfine structure systematics as well as the others discussed in [2] are known, then we may hope to reach a level of $\sim 10^{-6}$ in future astronomical measurements of $\Delta\alpha/\alpha$.

Acknowledgments

This work is supported in part by the RFBR grant No. 03-02-17522. I thank the organizing committee for providing me with the opportunity to attend. I especially thank Sandro D'Odorico, Luca Pasquini, Dieter Reimers, and Robert Baade for helpful discussions.

References

1. V. A. Dzuba, V. V. Flambaum, J. K. Webb: Phys. Rev. Lett. **82**, 888 (1999)
2. M. T. Murphy, J. K. Webb, V. V. Flambaum: Mon. Not. R. Astron. Soc. **345**, 609 (2003)
3. P. J. Mohr, B. N. Tailor: Rev. Mod. Phys. **72**, 351 (2000)
4. E. A. Milne: Proc. R. Soc. A **158**, 324 (1937)
5. P. A. M. Dirac: Nature **139**, 323 (1937)
6. J. Uzan: Rev. Mod. Phys. **75**, 403 (2003)
7. A. Chodos, S. Detweiler: Phys. Rev. D **21**, 2167 (1980)
8. W. J. Marciano: Phys. Rev. Lett. **52**, 489 (1984)
9. Y. S. Wu, Z. Wang: Phys. Rev. Lett. **57**, 1978 (1986)
10. P. D. B. Collins, A. Martin, E. J. Squires: *Particle Physics and Cosmology* (Wiley, New York 1989)
11. J. D. Barrow, J. Maguejo, H. B. Sandvik: Phys. Rev. D **66**, 3515 (2002)
12. J. P. Mbelek, M. Lachièze-Rey: Astron. Astrophys. **397**, 803 (2003)
13. J. N. Bahcall, C. L. Steinhardt, D. Schlegel: Astrophys. J. **600**, 520 (2004)
14. J. N. Bahcall, W. L. W. Sargent, M. Schmidt: Astrophys. J. **149**, L11 (1967)
15. D. A. Varshalovich, A. Y. Potekhin: Astron. Lett. **20**, 771 (1994)
16. D. A. Varshalovich, A. Y. Potekhin, A. V. Ivanchik: 'Testing cosmological variability of fundamental constants'. In: *X-ray and Inner-Shell Processes*. ed. by R. W. Danford, D. S. Gemmel, E. P. Kauter, B. Kraessig, S. H. Southworth, L. Yong (AIP, Melville 2000) pp. 503–511
17. M. T. Murphy, J. K. Webb, V. V. Flambaum, J. X. Prochaska, A. M. Wolfe: Mon. Not. R. Astron. Soc. **327**, 1237 (2001)
18. P. Langacker, G. Segrè, M. J. Strassler: Phys. Lett. B **528**, 121 (2002)
19. D. A. Varshalovich, S. A. Levshakov: JETP Lett. **58**, 237 (1993)
20. J. N. Bahcall, M. Schmidt: Phys. Rev. Lett. **19**, 1294 (1967)
21. S. A. Levshakov: Mon. Not. R. Astron. Soc. **269**, 339 (1994)
22. R. C. Bohlin, J. K. Hill, E. B. Jenkins, B. D. Savage, T. P. Snow, L. Spitzer, D. G. York: Astrophys. J. Suppl. Ser. **51**, 277 (1983)
23. S. A. Levshakov, M. Dessauges-Zavadsky, S. D'Odorico, P. Molaro: Mon. Not. R. Astron. Soc. **333**, 373 (2002)
24. L. Lu, W. L. W. Sargent, T. A. Barlow, C. W. Churchill, S. S. Vogt: Astrophys. J. Suppl. Ser. **107**, 475 (1996)
25. S. A. Levshakov, F. Takahara, I. I. Agafonova: Astrophys. J. **517**, 609 (1999)
26. S. A. Levshakov, I. I. Agafonova, W. H. Kegel: Astron. Astrophys. **360**, 833 (2000)

27. D. Reimers, R. Baade, R. Quast, S. A. Levshakov: Astron. Astrophys. **410**, 785 (2003)
28. F. Haardt, P. Madau: Astrophys. J., **461**, 20 (1996)
29. G. Ferland: *A Brief Introduction to Cloudy* (Internal Rep., Lexington, Univ. Kentucky, 1997)
30. W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery: *Numerical Recipes in C* (Cambridge Univ. Press, Cambridge 1992)
31. R. Quast, R. Baade, D. Reimers: Astron. Astrophys. **386**, 796 (2002)
32. S. A. Levshakov, I. I. Agafonova, D. Reimers, R. Baade: Astron. Astrophys. **404**, 449 (2003)
33. D. E. Welty, L. M. Hobbs, J. T. Lauroesch, et al.: Astrophys. J. Suppl. Ser. **124**, 465 (1999)
34. D. C. Morton: Astrophys. J. Suppl. Ser. **77**, 119 (1991)
35. J. Cardelli, B. D. Savage: Astrophys. J. **452**, 275 (1995)
36. A. V. Ivanchik, E. Rodriguez, P. Petitjean, D. A. Varshalovich: Astron. Lett. **28**, 423 (2002)
37. A. Y. Potekhin, A. V. Ivanchik, D. A. Varshalovich, K. M. Lanzetta, J. A. Baldwin, G. M. Williger, R. F. Carswell: Astrophys. J. **505**, 523 (1998)
38. S. A. Levshakov, M. Centurión, W. H. Kegel, P. Molaro: in preparation (2003)
39. S. A. Levshakov, M. Dessauges-Zavadsky, S. D'Odorico, P. Molaro: Astrophys. J. **565**, 696 (2002)
40. H. Abgrall, E. Roueff, F. Launay, J.-Y. Roncin, J.-L. Subtil: Astron. Astrophys. Suppl. Ser. **101**, 273 (1993)
41. H. Abgrall, E. Roueff, F. Launay, J.-Y. Roncin, J.-L. Subtil: Astron. Astrophys. Suppl. Ser. **101**, 323 (1993)
42. E. B. Jenkins, A. Peimbert: Astrophys. J. **477**, 265 (1997)
43. S. M. Carroll, W. H. Press, E. L. Turner: Ann. Rev. Astron. Astrophys. **30**, 499 (1992)
44. F. Pepe, M. Mayor, G. Rupprecht et al.: ESO Messenger **110**, 9 (2002)

Oklo Constraint on the Time-Variability of the Fine-Structure Constant

Yasunori Fujii

Advanced Research Institute for Science and Engineering, Waseda University, Tokyo 169-8555, Japan

Abstract. The Oklo phenomenon, natural fission reactors which had taken place in Gabon about 2 billion years ago, provides one of the most stringent constraints on the possible time-variability of the fine-structure constant α. We first review briefly what it is and how reliable it is in constraining α. We then compare the result with a more recent result on the nonzero change of α obtained from the observation of the QSO absorption lines. We suggest a possible way to make these results consistent with each other in terms of the behavior of a scalar field which is expected to be responsible for the acceleration of the universe.

1 What Is the Oklo Phenomenon?

Oklo is the name of the place of a uranium mine in Gabon, West Africa, near the equator (Fig. 1). The mining company would supply the uranium ore to the French government. But in June of 1972, something unusual was noticed on the ore from Oklo; the abundance of ^{235}U was somewhat below the world standard, 0.7202%, well beyond the limit of permissible range. This might have undermined the company's reputation about the quality of their uranium. But finally after a few months of serious effort, French scientists came up with an unexpected, startling conclusion:

The deficit of ^{235}U was a real effect of that a self-sustained fission reaction took place *naturally* in Oklo about 1.8 Gyr ago, during the period of Proterozoic, part of Precambrian. In other words, natural reactors did exist well before 1942 when Enrico Fermi invented the artificial reactor for the first time in Chicago. This has been called the "Oklo phenomenon", since then [1]. The result of their work was published in many ways, including [2].

Fig. 1. Oklo in Gabon, West Africa, near the equator

Y. Fujii, Oklo Constraint on the Time-Variability of the Fine-Structure Constant, Lect. Notes Phys. **648**, 167–185 (2004)
http://www.springerlink.com/

There was a big press campaign, including the Le Monde article, for example, saying that Fermi was not an "innovator," but was only an "imitator" of Nature.

Even more surprising was that the occurrence of this "natural reactors" had been predicted much earlier, 17 years earlier by a nuclear geochemist, Paul K. Kuroda in 1955 [3]. He discussed several conditions. But the most important was that the ratio of ^{235}U, currently 0.70%, used to be much higher because of the different lifetimes of ^{235}U and ^{238}U; 7.038×10^8 yr and 4.468×10^9 yr, respectively. One can easily calculate the ratio 1.8 Gyr ago to be as high as 3.2%. We note that 3% is a goal of most of today's enrichment facilities. Another condition was the presence of water which served as a moderator.

So far 16 "reactor zones" (RZ) have been discovered in the Oklo area. In each of them, extensive and detailed measurements have been made on the leftover fission products.

2 How Did Shlyakhter Probe $\Delta\alpha$?

Under this circumstance, in 1976, Alex Shlyakhter [4,5] then in Leningrad proposed to look at ^{149}Sm, which is present naturally at the ratio 13.8%, but should be depleted in the reactor zones because it had absorbed neutrons strongly in the reactors 2 Gyr ago, according to the reaction

$$n + {}^{149}\mathrm{Sm} \to {}^{150}\mathrm{Sm} + \gamma\ . \tag{1}$$

One measures the abundance in Oklo reactor zones to estimate the cross section of this process, and compares the result with today's laboratory value. In this way one can tell how much nuclear physics 2 Gyr ago could have been different from what it is.

What is unique with this particular process (1) is that it is dominated by a resonance that lies as low as $E_r = 97.3$ meV, while we know that a typical energy scale of nuclear physics is $\sim$ MeV. Compared with this, the above value is very small, nearly 7 orders of magnitude too small. This must be due to a nearly perfect cancellation between two effects; repulsive Coulomb force which is proportional to the fine-structure constant α, and attractive nuclear force which depends on the strong-interaction coupling constant α_s. We are left with a very small leftover for the resonance energy, as illustrated Fig. 2. Suppose we change one of the coupling constants, α, say, only slightly. Then the strength of the Coulomb energy will change also slightly, and so will E_r. However, the *relative* change may not be so small, because the starting value E_r was already small. If this really happens, then the cross section may change rather significantly. This is a kind of amplification mechanism, which Shlyakhter exploited.

For the sake of illustration, we plot in Fig. 3 the cross section (based on the Breit-Wigner formula) as a function of ΔE_r, the fictitious change of the resonance energy from today's value. We also assume thermal equilibrium of the neutron flux; assuming a temperature corresponding to one of the curves shown. We find a sharp peak obviously coming from the resonance. Suppose E_r at 2 Gyr ago were smaller by 10 meV, a tiny amount. Suppose also $T = 300^\circ$ C, for

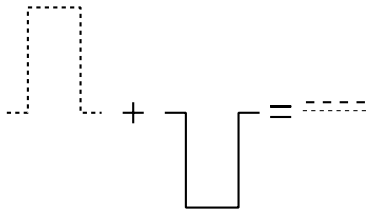


Fig. 2. Nearly complete cancellation between the repulsive Coulomb energy and the attractive nuclear energy, leaving a small leftover for the resonance energy

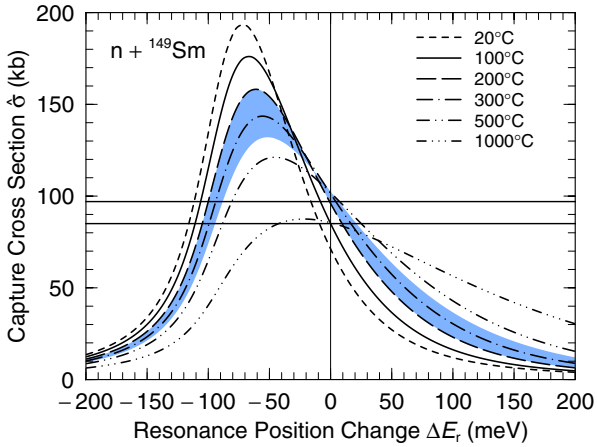


Fig. 3. The cross section $\hat{\sigma}_{149}$ for the process (1) as a function of $\Delta E_r = E_r - 97.3\text{meV}$, assuming thermal equilibrium of the neutron flux. The ranges of the observed cross section and the temperature are also shown, as given later by (6) and (7), respectively

example. Then we find the cross section bigger than today's value by a few %, a significant change.

Incidentally, we are going to use the hat symbol attached to the cross section rather frequently. But this is only a technical convention of normalization, which is particularly popular among the researchers of the Oklo phenomenon. We do not worry too much at this moment.

Taking advantage of this strong dependence of the cross section σ_{149} on ΔE_r, Shlyakhter gave the upper bound

$$|\Delta E_r| \lesssim 50\,\text{meV}. \tag{2}$$

It is not very clear how he derived this result, particularly how much the uncertainties in the data affected the conclusion. This is one of the points to be re-examined later.

He still went on to discuss how this change of E_r corresponds first to the change of the strong-interaction coupling constant, α_s. He considered the resonance as a single-particle excitation in the potential, with its depth $V_0 \sim 50$ MeV, which he assumed to be proportional to α_s. If α_s changes, V_0 changes, and so

does E_r. Substituting from (2), he obtained the result

$$\left|\frac{\Delta\alpha_s}{\alpha_s}\right| = \left|\frac{\Delta E_r}{V_0}\right| \lesssim \frac{50\ \text{meV}}{50\ \text{MeV}} = 10^{-9}, \tag{3}$$

leading to the value 10^{-9}. Further dividing by 2×10^9 yr, he arrived at

$$\left|\frac{\dot{\alpha}_s}{\alpha_s}\right| \lesssim 5 \times 10^{-19}\ \text{yr}^{-1}. \tag{4}$$

As for the electromagnetic interaction, he apparently replaced α_s by α, resulting in dividing these by $\alpha/\alpha_s \sim 1/20$, giving

$$\left|\frac{\Delta\alpha}{\alpha}\right| \lesssim 2 \times 10^{-8}, \quad \left|\frac{\dot{\alpha}}{\alpha}\right| \lesssim 1 \times 10^{-17}\ \text{yr}^{-1}, \tag{5}$$

This last value has been several orders of magnitude more stringent than any other estimates, a kind of "champion result" for many years. However, one may raise a question against the argument from α_s to α, and the suspicion may go further back to the derivation (3).

For better understanding, we re-examined the whole analysis [6] by forming a team, which includes theorists of nuclear physics, reactor scientists and geologists. Among them, Hiroshi Hidaka is an expert nuclear chemist who has been specialized to the Oklo phenomenon.

3 How Good Is It?

We soon realized that a major error source of the data comes from the "post-reactor contamination", implying that certain amount of ^{149}Sm present in the outside environment flowed into the reactor core after the end of the reactor activity that is believed to have lasted several 10^5 years. This amount has nothing to do with what happened inside the reactors, so it is a contamination for our purpose. This inflow was in fact the gradual mixture between inside and outside prompted by repeated successions of dissolution and precipitation of Sm, caused essentially by weathering, namely being exposed to the air. To minimize this embarrassing effect, we looked for samples in the reactor zones 10–16 discovered later than 1984, deep underground, as shown in Fig. 4. Finally we decided to collect five samples taken from RZs 10 and 13 below the surface, with enough care of geologists expertise.

The measured isotopic ratios related to ^{149}Sm for five samples are shown in Table 1. We see how small the abundances of ^{149}Sm are compared with the natural abundance 13.8%. We did not show the errors, but they are simply small. t_1 is the time of the end of the reactor activity that started at $t = 0$. Also in the first line, we entered what is called "fluence", denoted by $\hat{\phi}t_1$, but actually a time-integrated neutron flux $\hat{\phi}$ during the whole duration of reactor activity.

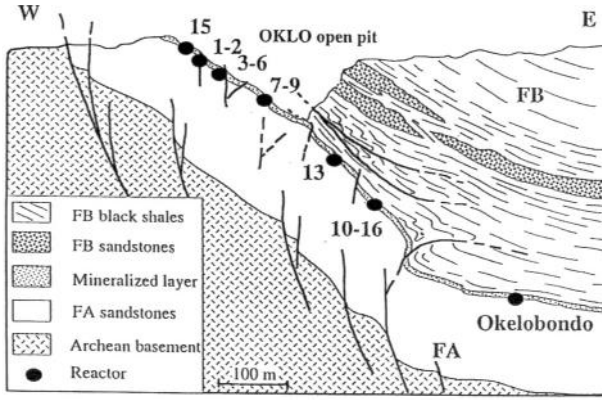


Fig. 4. Geological vertical cross section of the Oklo area.

Table 1. Measured isotopic ratios related to ^{149}Sm obtained from five samples. The fluence and cross section are also listed

	Sample				
	SF84-1469	SF84-1480	SF84-1485	SF84-1492	SD.37
$\hat{\phi}t_1$ (1/kb)	0.525	0.798	0.622	0.564	0.780
$N_{144}(t_1)$ (%)	0.1052	0.2401	0.2073	0.1619	0.06909
$N_{147}(t_1)$ (%)	55.34	53.23	54.03	54.81	52.74
$N_{148}(t_1)$ (%)	2.796	3.468	3.079	2.890	4.694
$N_{149}(t_1)$ (%)	0.5544	0.2821	0.4466	0.4296	0.3088
$N_{235}(t_1)/N_{238}(t_1)$	0.03181	0.02665	0.02971	0.03047	0.02435
$\hat{\sigma}_{149}$ (kb)	85.6	96.5	83.8	99.0	89.5

We then solved the evolution equations

$$dN_{147}(t)/dt = -\hat{\sigma}_{147}\hat{\phi}N_{147}(t) + N^0_{235}\exp(-\hat{\sigma}_{\mathrm{a}}\hat{\phi}t)\hat{\sigma}_{\mathrm{f}235}\hat{\phi}Y_{147},$$
$$dN_{148}(t)/dt = \hat{\sigma}_{147}\hat{\phi}N_{147}(t),$$
$$dN_{149}(t)/dt = -\hat{\sigma}_{149}\hat{\phi}N_{149}(t) + N^0_{235}\exp(-\hat{\sigma}_{\mathrm{a}}\hat{\phi}t)\hat{\sigma}_{\mathrm{f}235}\hat{\phi}Y_{149}.$$

for the related isotopes to calculate the cross section $\hat{\sigma}_{149}$ for the process (1). The result is summarized,

$$\hat{\sigma}_{149} = (91 \pm 6)\,\mathrm{kb}, \tag{6}$$

corresponding to the narrow horizontal band shown in Fig. 3.

We also made an estimate of the temperature, by the traditional way supplemented by a latest technique, giving

$$T = (200 - 400)^{\circ}\mathrm{C}, \tag{7}$$

corresponding to the shaded area in Fig. 3. We find two intersections, and the corresponding two separated ranges of ΔE_r.

$$\Delta E_r = \begin{cases} (9 \pm 11)\,\text{meV}, & \text{right-branch, null} \\ (-97 \pm 8)\,\text{meV}, & \text{left-branch, non-null} \end{cases} \tag{8}$$

The right-branch range covers zero, so that it is a null result in the usual sense, while the other implies that E_r was different from today's value by more than 10 standard deviations. Does this really imply an evidence for a difference in coupling constants 2 billion years ago? We tried to see if the non-null result can be eliminated by looking at other isotopes like ^{155,157}Gd, but have not obtained a final conclusion so far.

At this point we compare our result with those by Damour and Dyson [7] (DD), who used the samples obtained mainly from near the surface, giving the cross section

$$\hat{\sigma}_{149} = (75 \pm 9)\,\text{kb}, \tag{9}$$

somewhat smaller than our result (6). This seems consistent with our suspicion that their data suffered from contamination. Also, they did not come to separating the two ranges. They could have done it, though the "right-branch" range would have failed to cover zero even at the level of 2 standard deviations. Instead, they gave only a combined range, $-120\,\text{meV} \lesssim \Delta E_r \lesssim 90\,\text{meV}$, which covers more than our two ranges.

We admit that we are still short of determining which range is correct. Then one might say that we should also be satisfied with the combined range. But we still insist that it is a progress to have established a disallowed range in between.

Now we move on to discuss how ΔE_r is translated into $\Delta\alpha$ following DD. First they ignored the contribution from the strong interaction entirely, focusing on the first "term" coming from the Coulomb contribution in Fig. 2.

Consider the energy $\mathcal{M}_c$, given by the difference of the Coulomb energies between the states with mass numbers 150 and 149. They paid special attention to the fact that the resonance in the ^{150}Sm is excited. But we simplify the analysis at this moment by appealing to the semi-empirical mass formula due to Weizsäcker, finding $\mathcal{M}_c \approx -1.1$ MeV. In this calculation, one has to allow an error perhaps within the factor 2 or 1/2. Notice also that the above result is negative, apparently in contradiction to the illustration in Fig. 2. Obviously we dealt basically with a repulsive force, but we calculated the difference, which turns out to be negative. Nothing is wrong, but we would better put the two terms upside-down on the left-hand side, but keeping right-hand side still positive. We may also assume that $\mathcal{M}_c \propto \alpha$. We then obtain $\Delta E_r = \Delta\mathcal{M}_c = (\Delta\alpha/\alpha)\mathcal{M}_c$, thus giving

$$\frac{\Delta\alpha}{\alpha} = \frac{\Delta E_r}{\mathcal{M}_c} = \begin{cases} (-0.8 \pm 1.0) \times 10^{-8}, & \text{null, upper bound} \\ (0.88 \pm 0.07) \times 10^{-7}, & \text{non-null} \end{cases}$$

Divide by -2×10^9 yr to get

$$\frac{\dot{\alpha}}{\alpha} = \begin{cases} (0.4 \pm 0.5) \times 10^{-17}\,\text{yr}^{-1}, & \text{null, upper bound} \\ (-0.44 \pm 0.04) \times 10^{-16}\,\text{yr}^{-1}, & \text{non-null} \end{cases}$$

This upper bound happens to agree quite well with Shlyakhter's result 1×10^{-17} yr^{-1}. The agreement to this extent seems, however, rather accidental, because, among other things, it is unlikely that the data as good as ours was available in 1976.

We emphasize here that the simple estimate by DD, as described here, might be called "Coulomb-only estimate", which serves as a basis for more general analyses.

In fact what really happens might be a combined result of both interactions, and one wishes if one could include the strong interaction as well. But then everything is going to be complicated, for example, like the QCD analysis in [8]. But there is something independent of such complications as long as we appreciate the condition that ΔE_r is much smaller than either of the mass scales, $\mathcal{M}_c$ and its strong-interaction counterpart $\mathcal{M}_s$. First, we find from Fig. 2 that the mass scale of $\mathcal{M}_s$ is nearly equal in its size to that of $\mathcal{M}_c$, obviously much smaller than Shlyakhter's "50 MeV." It then also follows that $\Delta\alpha_s/\alpha_s$ should be nearly of the same size as $\Delta\alpha/\alpha$. Of course there are some differences from the Coulomb case; $\mathcal{M}_s$ may not be simply proportional to α_s. This may result in a revision of a factor, but certainly not of an order of magnitude. Then we go through a bit of analysis to conclude finally that it is unlikely that, by the strong interaction, $\Delta\alpha/\alpha$ deviates from the Coulomb-only estimate by more than an order of magnitude, no matter how complicated the exact analysis might be. It can be smaller. See Appendix A for more details.

Then, as always, there is a possibility of an exception, no matter how remote. This allows, in principle, that both of $\Delta\alpha/\alpha$ and $\Delta\alpha_s/\alpha_s$ are quite large, in fact without limit, but cancel each other leaving a small value of ΔE_r. At this moment, however, we assume that no such fine-tuning nor coincidence occurs in the real world.

Let us summarize what the situation is with the question of "uncertainties". In the theoretical aspect as a whole, we say again that the Coulomb-only estimate of the relative change of α is likely correct within an order of magnitude, for whatever the complication of the effect of strong interaction. We only add a few related remarks:

- Only ΔE_r and the related cross section σ_{149} are sensitive to $\Delta\alpha$. No other quantities are.
- Our formulation is such that, to a good approximation in practice, the neutron flux can be any function of time. For example, the reactor activity can be even "sporadic". The only thing that counts is the fluence, no matter how long it took.
- Estimating fluence is complicated but is a standard estimate and is reliable.
- In principle we may not rule out one of the higher resonances to come down near the threshold, giving much larger value of ΔE_r. We have a reason, however, to believe this to be highly unlikely. See Appendix B.

In the observational aspect, we repeat our previous statement; post-reactor contamination is the largest error source, with a few more comments:

- Nothing is serious for ^{149}Sm in our samples. (A few percent contamination seem even better because then the range of ΔE_r for the null result covers zero more in the middle.)
- This is not the case for ^{155,157}Gd which enjoy even lower resonance energies. The absorption cross sections are also larger. But the effects are too strong to the extent that the residual abundances are too little, so they are too sensitive to contamination, even with our samples with minimized effect of weathering. Shlyakhter was clever when he chose ^{149}Sm. This is also precisely why we reached short of complete elimination of the non-null result, as mentioned before.

4 How Can It Be Consistent with the QSO Result?

According to V. Flambaum and M. Murphy at ACFC meeting, the latest version of their result on the time-variation of α from spectroscopy of QSO absorption systems is [9] (see also their contribution to this volume [10])

$$\frac{\Delta\alpha}{\alpha} = (-0.54 \pm 0.12) \times 10^{-5}.$$

We show in Fig. 5, taken from Fig. 8 of [9], our own plot as a function of the fractional look-back time u defined by $u = (t_0 - t)/t_0$, with t_0 the present age of the universe. Their weighted mean can be viewed as a fit by a horizontal straight line at -0.54, as also shown in Fig. 5. We will call this a "1-parameter fit" for later convenience. Notice that $\chi^2_{\rm red} = 1.06$.

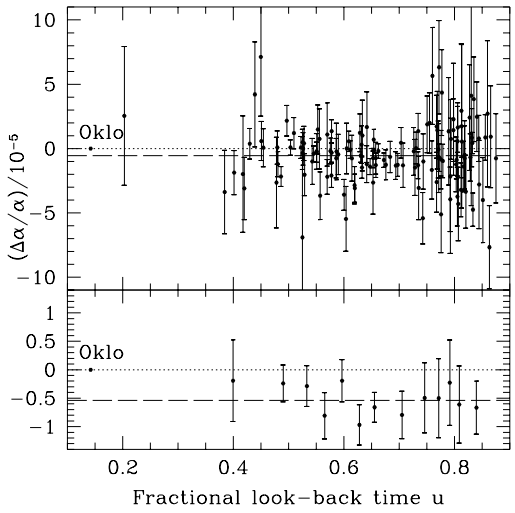


Fig. 5. QSO result from the 128 data points is shown in the upper panel, while the 13 binned data might provide an easier view in the lower panel [9]. The long-dashed line is for the weighted mean -0.54. The Oklo time $u_{\rm oklo} = 0.142$ is also shown

We now include the data from the Oklo phenomenon, at $u_{\rm oklo} = 0.142$, barely outside the QSO range, as also shown in Fig. 5. In this sense they are different things. However, we should put error-bars, which are invisibly small in this plot; less than 10^{-4} and $10^{-2} - 10^{-3}$ in the horizontal and vertical directions, respectively. Remember that $\Delta\alpha/\alpha$ from the Oklo constraint is $10^{-7} - 10^{-8}$. If we extend the straight line naively down to the Oklo time, it will miss the point off $(10^2 - 10^3)$ standard deviations, resulting in an enormous value of χ^2, too large to be acceptable. One wants to bend the line to pass the point almost exactly, but one needs a physical reason. What is that? This is the issue. Already there have been several attempts [11–16]. But we stick to our own idea that this issue has something to do with the accelerating universe, another big issue in today's cosmology.

Now probably everyone knows that our universe is accelerating [17]. This behavior is best described in terms of a positive cosmological constant, whose size is given usually by the parameter $\Omega_\Lambda = \Lambda/\rho_{\rm cr} \sim 0.7$, where the critical density $\rho_{\rm cr}$ is given by $\sim t_0^{-2}$. The coefficient here is of the order one if we use the reduced Planckian unit system in which $c = \hbar = M_{\rm P}(= \sqrt{8\pi G/(c\hbar)}) = 1$. In this unit, the present age $t_0 \sim 1.4 \times 10^{10}$ yr is about 10^{60}. So we find $\rho_{\rm cr} \sim \Lambda \sim 10^{-120}$.

Today's cosmological constant problem has two faces or questions: Why is it so small? Why is it still nonzero? The first question can be replied by the "scenario of a decaying cosmological constant": Λ is not a true constant but decays like $\sim t^{-2}$ [18,19]. This simple idea can be implemented by the "scalar-tensor theory". We expect that a scalar field plays an important role. This scalar field may have its origin in string theory in which a graviton has a spinless companion called the "dilaton".

The second question seems to require a deviation from the simplest version of the scalar-tensor theory. As one of the possible ways we call for another scalar field, called χ, in addition to the dilaton σ. These two fields comprise what is called the "dark energy", and their energy density $\rho_s = \rho(\sigma, \chi)$ is interpreted as an effective cosmological constant $\Lambda_{\rm eff}$. There are many details involved, though we are not going into any details. Readers are advised to refer to our recent book [20].

Skipping all the details, we show in Fig. 6 an example of our solutions in the Friedmann universe with flat 3-space. The horizontal axis is $\log t$. In the Planckian unit system, the present time is somewhere around 60. In the lower panel, ρ is the usual matter energy density, which falls off roughly as t^{-2}. The energy density of the scalar fields ρ_s is the effective cosmological constant, $\Lambda_{\rm eff}$, also falling off like t^{-2} as an overall behavior, thus respecting the scenario of a decaying cosmological constant.[1] But the plot also shows occasional deviations, notably the plateau behaviors. Obviously each plateau mimics a cosmological "constant". Furthermore, it comes to a crossing with the ordinary matter energy density. One of them is expected to occur around the present epoch. Nearly in

[1] As emphasized in [20], the gravitational "constant" G in the *physical* conformal frame identified (nearly) with the Einstein frame in this model is (nearly) time-independent, instead of decaying with time as in [19] presented in the Jordan frame.

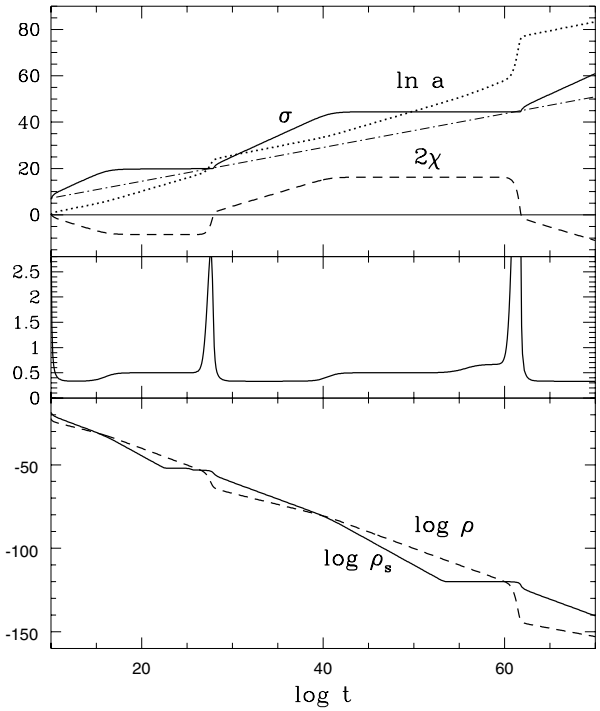


Fig. 6. An example of our cosmological solutions in the Friedmann universe with $k = 0$. The scale factor a, the scalar fields σ and χ, the energy densities ρ and ρ_s of the ordinary matter and the scalar fields, respectively, are plotted against $\log t$, in the reduced Planckian unit system. The present time is around $\log t \approx 60$. The middle panel shows an effective exponent of the scale factor, $\ln a / \ln t$

coincidence with this crossing, we find a "mini-inflation" of the scale factor; a bit of sharp increase in $\ln a$ shown in the upper panel. This nicely fits the observed acceleration of the universe.

Also shown in the upper panel are the sudden changes of the scalar fields σ and χ, again in coincidence with the crossing between ρ and ρ_s. However, the most interesting is to take a close-up view of what appears to be a simple and small jump of σ. With the magnification rate as large as 330 in the vertical direction, we find a surprising behavior shown in Fig. 7, something like a damped oscillation.

We know what the underlying mechanism is at the deeper level. But we still say that this damped-oscillation-like behavior is in fact "the heart and soul" of the entire dynamics that eventually brings about the acceleration as we see it. On the other hand, the acceleration itself does not care how an invisibly tiny oscillation is taking place behind the scene. There are many variations in the way of oscillation. In this sense, we have a degeneracy, which the cosmological acceleration does not resolve. However, this invisibly small oscillation may show up through the time-variation of α.

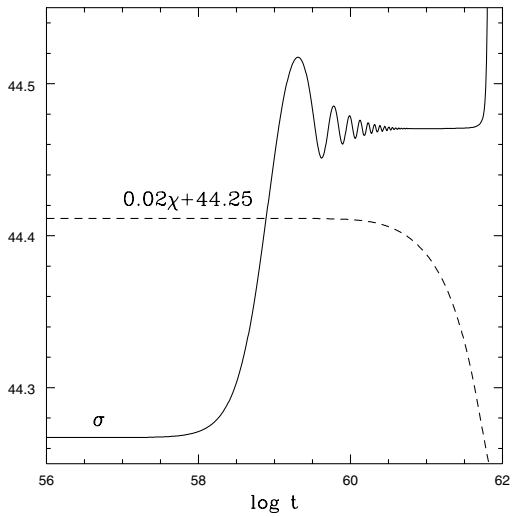


Fig. 7. A close-up view of σ and χ in the upper panel of Fig. 6 around the present epoch. The magnification rate in the vertical direction is 330.

This is an expectation based on a general view that changing α, if any, is due to the changing scalar field, expressed symbolically as

$$\frac{\Delta\alpha}{\alpha} \propto \Delta\sigma. \tag{10}$$

String theory suggests this dependence for the gauge coupling constant. We ourselves derived a relation of this type, based on QED, featuring a quantum-anomaly type of calculation. But we do not want to be too specific on these theoretical details, nor to depend heavily on the choice of the solutions, like the one in Fig. 7. This is particularly crucial because we have many different solutions for a given cosmological behavior we want to fit. Rather, we are going to follow a phenomenological approach which we describe briefly, leaving more details to [21].

Let y denote $\Delta\alpha/\alpha$ in units of 10^{-5}. Then we assume a dependence on the fractional look-back time u in the way of a damped oscillation

$$y = ae^{bu} \sin\left(2\pi \frac{u - u_{\rm oklo}}{T}\right), \tag{11}$$

where the parameters are going to be determined to fit the QSO data as well as the Oklo constraint.

We first choose $u_{\rm oklo} = 0.142$ corresponding to the Oklo time of 1.95 Gys ago. The Oklo constraint, to be $10^{-2} - 10^{-3}$ in terms of y, is approximately zero in this scale. The remaining parameters a, b, T are determined by minimizing χ^2 for the QSO data. In this sense we call this a 3-parameter fit. We do not include the Oklo data in computing χ^2, because we consider the Oklo has been already fitted approximately by choosing a zero of the function as above.

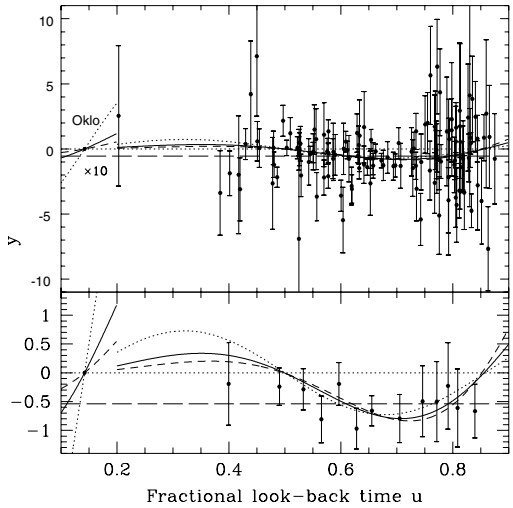


Fig. 8. The solid curves are for the 3-parameter fit with the least χ^2 ($a = 0.151, b = 2.4, T = 0.714, \chi^2_{\rm red} = 1.09$), to be compared with the 1-parameter fit, represented by a horizontal straight (long-dashed) line at -0.54 in Fig. 5. Note the 10 times magnification for $u < 0.2$. The dotted and dashed curves are for $b = 0.0$ and $b = 4.0$, respectively

We limit ourselves to a region of a, b, T in a manner roughly consistent with the theoretical model of the accelerating universe. In this range we searched for local minima of χ^2. Among several of them we find the least minimum which is given by $b = 2.4, a = 0.151, T = 0.714$ resulting in $\chi^2_{\rm red} = 1.09$. This $\chi^2_{\rm red}$ is similar to $\chi^2_{\rm red} = 1.06$ obtained for their 1-parameter fit. In this sense our 3-parameter fit is nearly as good as the fit in [9]. The solid curve in Fig. 8 shows the actual plot, probably better shown in the binned plot in the lower panel. We magnified the curves below $u = 0.2$ by 10 times.

One might ask us why we are satisfied with $\chi^2_{\rm red} = 1.09$ which is not smaller than 1.06 for the 1-parameter fit, in spite of the fact that we have more degrees of freedom. We answer the question by pointing out the following:

- Our $\chi^2_{\rm red} = 1.09$ is for the whole data including Oklo, because, as we noted, the Oklo constraint has been already "included" in a sense. In comparison, however, the 1-parameter fit gives an unacceptably large χ^2 when we include the Oklo data, which was the starting point of the whole discussion.
- Our 3-parameter fit was motivated originally by a theoretical "prejudice". There was no guarantee that it fits the reality. We are relieved to find that our prejudice somehow survived a realistic test.

We now discuss what the confidence region for 68% is like for this 3-parameter fit. We imagine a confidence volume in three-dimensional space of b, a, T. Figure 9 shows, however, two-dimensional cross sections for three different values of b. The contour labeled by 2.4 in this figure shows the one for $b = 2.4$. We also show

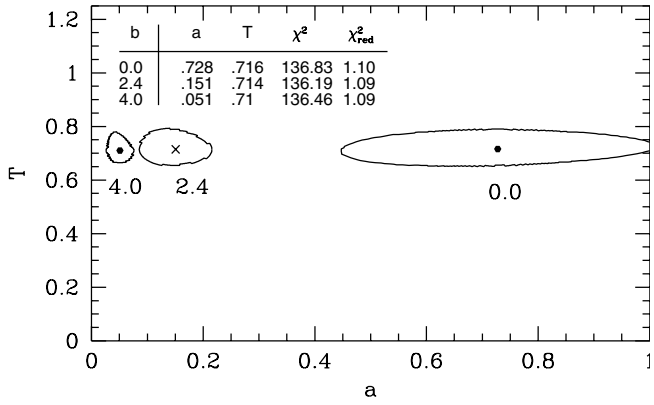


Fig. 9. The 3-dimensional 68% confidence region is illustrated in terms of three cross sections for $b = 0.0, 2.4, 4.0$, as marked beside each contour, shown in 2-dimensional $a - T$ space

Table 2. Comparison between the damped-oscillator-like fit with an example of the cosmological solution

	a	b	T
Cosmological solution in Fig. 7	≈ 2.4	≈ 2.5	≈ 0.22
Fit in Fig. 8	0.15 ± 0.05	2.4 ± 2	0.71 ± 0.06

other two cross sections corresponding to $b = 0.0$ and $b = 4.0$, respectively. They give only slightly larger χ^2 than that for $b = 2.4$. In this way one imagines what the three-dimensional volume looks like. The curves for these b are also plotted by the dotted and dashed curves, respectively. in Fig. 8. They are different from each other only in the lower-u region, $u \lesssim 0.5$.

We further add that we obtained several other solutions with other values of b, a, T which give local minima of χ^2, as we indicated before. As it turned out, however, they tend to give $\chi^2_{\rm red} \gtrsim 1.2$. This is a number which is nearly comparable with $\chi^2_{\rm red} = 1.24$, which we would obtain by fitting the QSO data by a horizontal straight line $y = 0$, namely the u-axis itself. We may have a good reason to exclude these fits.

We may compare the result shown in Fig. 8 with the cosmological solution which we started from, as we showed in Fig. 7. For the latter we may estimate the parameters approximately, which will be shown in Table 2, together with the corresponding ones for the former.

Agreement in the values of b is obvious. We have to have more theoretical details in order for the comparison of a to make sense, though we may reasonably find a consistency. On the other hand, there is a discrepancy between the values of T. The available QSO data shows a rather flat distribution of $\Delta\alpha/\alpha$, which

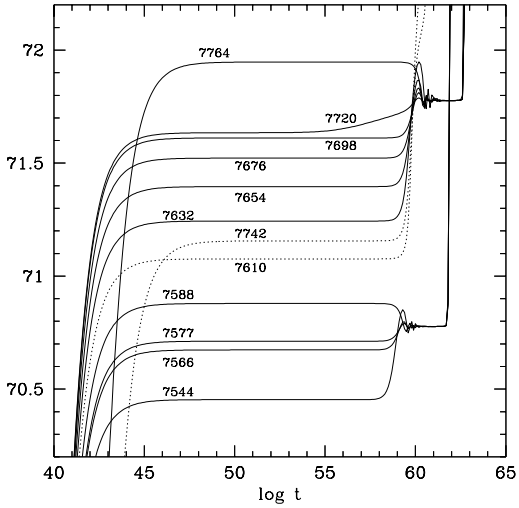


Fig. 10. Another magnified view of the behavior of σ. The previous Fig. 7 is only a small part of the present figure, corresponding to a curve with the attached number 7544. These numbers show the last 4 digits in the initial value at $t = 10^{10}$, as explained in detail in [20]. The nucleosynthesis and CMB times are ≈ 45 and ≈ 55, respectively. The vertical scale may be accepted as arbitrary at this moment.

favors a "larger" T. In this connection we point out, however, that we have chosen the solution in Fig. 7 rather arbitrarily. In fact Fig. 10 indicates that it happened that we have come across a relatively small T. We only conclude that we have to look for other solutions of the cosmological equations which still fit the way of the cosmological acceleration.

Finally, we add the following comments for further improvements of the fit.

- We show Fig. 10, taken from Chap. 5 of [20], which includes Fig. 7 as a small portion for a special choice of the initial value. This figure also demonstrates that the detailed behavior of the small oscillation depends heavily on the early history, depending sensitively on the behavior of the scalar fields at the initial times, particularly on those around the time of nucleosynthesis. In this sense determining $\Delta\alpha/\alpha$ in this epoch as well as in the CMB era is crucially important in this approach.
- In our approach in terms of (11) we assumed that the present time corresponds to the limiting, still transient, behavior toward the common flat value immediately prior to the big and sudden jump of σ. It seems better to consider that we are in the middle of the oscillation behavior, in general. Taking this possibility into account will make it more likely to satisfy the natural condition $y(0) = 0$, which should be true by definition. This type of the fit will be discussed in Appendix C.
- Our analysis is based on the simplest assumption on the $\sigma - \chi$ interaction, as given by (5.58) of [20]. This might be modified to improve the fit.

As the very last comment, we hope that natural reactors will be discovered somewhere else, thus providing us with additional constraints, hopefully at different times.

Appendix

A Bound on $\Delta\alpha/\alpha$ from the Coulomb-Only Estimate

The situation described in Fig. 2 may be given the expression

$$E_r = \mathcal{M}_c + \mathcal{M}_s, \tag{A.1}$$

with the condition

$$|E_r| \ll |\mathcal{M}_c(\alpha)| \sim |\mathcal{M}_s(\alpha_s)|. \tag{A.2}$$

We then obtain

$$\begin{aligned} \Delta E_r &= \frac{\partial E_r}{\partial \alpha}\Delta\alpha + \frac{\partial E_r}{\partial \alpha_s}\Delta\alpha_s \\ &= \frac{\mathcal{M}_c}{\alpha}\Delta\alpha + \frac{\mathcal{M}_s}{\alpha_s}\Delta\alpha_s. \end{aligned} \tag{A.3}$$

In deriving the second equation we assumed

$$\frac{\partial \mathcal{M}_c}{\partial \alpha} = \frac{\mathcal{M}_c}{\alpha} \quad \text{and} \quad \frac{\partial \mathcal{M}_s}{\partial \alpha_s} = \frac{\mathcal{M}_s}{\alpha_s}, \tag{A.4}$$

to simplify the equations, for the moment.

According to (A.1) and (A.2) we put (A.3) into

$$\Delta E_r \approx \mathcal{M}_c \left(\frac{\Delta\alpha}{\alpha} - \frac{\Delta\alpha_s}{\alpha_s} \right). \tag{A.5}$$

Ignoring the second term yields the Coulomb-only estimate

$$\frac{\Delta\alpha}{\alpha} \approx \frac{\Delta E_r}{\mathcal{M}_c} \equiv D_{c0}. \tag{A.6}$$

On the other hand, we notice that the right-hand side of (A.5) happens to vanish if

$$\frac{\Delta\alpha_s}{\alpha_s} = \frac{\Delta\alpha}{\alpha}, \tag{A.7}$$

leaving $\Delta\alpha/\alpha$ undetermined in terms of ΔE_r. (We then have to bring E_r back again on the right-hand side of (A.5), as discussed at the end of Sect. 5 of [6], thus corresponding to the "exception" mentioned toward the end of Sect. 3. We ignore this case at this moment.) We may assume, however,

$$\frac{\Delta\alpha_s}{\alpha_s} = \xi\frac{\Delta\alpha}{\alpha}, \tag{A.8}$$

where ξ is to be determined based on the more fundamental laws of physics, as attempted in [8],[22],[23]. By using this in (A.5), we obtain

$$\frac{\Delta\alpha}{\alpha} \approx (1-\xi)^{-1} D_{c0}. \tag{A.9}$$

Suppose a special relation (A.7) holds true within the accuracy of 10%, for example. This implies that (A.8) holds true for

$$\xi = 1 + \delta\xi \quad \text{with} \quad |\delta\xi| \lesssim 0.1. \tag{A.10}$$

Then (A.9) implies

$$\left| \frac{1}{D_{c0}} \frac{\Delta\alpha}{\alpha} \right| = |\delta\xi|^{-1} \gtrsim 10. \tag{A.11}$$

In other words, $\Delta\alpha/\alpha$ should remain close to the Coulomb-only estimate within an order of magnitude, unless the equality $\Delta\alpha/\alpha = \Delta\alpha_s/\alpha_s$ holds true to the accuracy better than 10%.

This result may be extended to more general situations, in which $\mathcal{M}_c$ and $\mathcal{M}_s$ depend on α_s and α, respectively, though then separating into $\mathcal{M}_c$ and $\mathcal{M}_s$ in (A.1) may not be unique. A certain relation like (A.7) is expected to result in the vanishing right-hand side of an equation corresponding to (A.5). It is unlikely that the relation of this kind holds true exactly in practice.[2] Unless it does within the accuracy of 10%, we should always expect $\Delta\alpha/\alpha$ to remain less than an order of magnitude of the Coulomb-only estimate.

B Distant Migration of the Higher Resonances

We have so far assumed that ΔE_r is very small, much smaller than $E_r = 93.7$ meV for Sm. In fact we obtained ΔE_r as small as 10 meV, thus giving $|\Delta\alpha/\alpha| \sim 10^{-8}$. This is the right attitude as long as we try to find as small an upper bound as we can. Now, however, the QSO result indicates a much larger value, up to 2 or 3 orders of magnitude larger. This might raise a question if the Oklo phenomenon does in fact yield a correspondingly larger ΔE_r. One may suggest that we are looking in the remnants of Oklo RZ at a distant "migration" of a higher resonance down nearly to the threshold of $n + {}^{149}\mathrm{Sm}$.

This possibility was already discussed by Shlyakhter [5], based on a statistical argument. Inspired by his approach, Akira Iwamoto and the present author attempt here a similar analysis on the issue, reaching a rather negative conclusion by including the observation of Gd. We start, however, with discussing Sm first, for which the energies and the widths of the first four resonances are shown in Table 3.

[2] From the minimal supersymmetric standard model follows $\xi \approx 6$ at $\mu = M_Z$ [23], though this result is not readily extrapolated to a much smaller μ in QCD, making it even unlikely to derive a value anywhere near unity.

Table 3. First four resonances in $n + {}^{149}\mathrm{Sm}$. Γ_n is the elastic width. The last line represents the approximate expected time-variability of α obtained by the Coulomb-only estimate in units of 10^{-5}

Resonances	1st	2nd	3rd	4th
E_r (meV)	97	872	4950	6430
Γ_n (meV)	0.53	0.74	2.13	0.72
Γ_{tot} (meV)	30	30	31	30
$\Delta\alpha/\alpha$ (10^{-5})	$\sim 10^{-3}$	0.08	0.45	0.58

Suppose ΔE_r was negative and so large 2 Gys ago that one of the higher levels came down nearly to the same position as the first resonance. We may expect a significant effect if its (shifted) energy minus 97 meV happens to be close to the threshold. We may ignore the thermal energy $(60-90)$ meV corresponding to the estimated temperature of $(200-400)^{\circ}$C in the natural reactors. According to Table 3, the elastic width Γ_n is found to be proportional to the center-of-mass momentum within an order of magnitude. We then expect the Coulomb-only estimate $\Delta\alpha/\alpha = \Delta E_r/\mathcal{M}_c \approx -E_r/\mathcal{M}_c$ to be roughly correct with a common value $\mathcal{M}_c \approx -1.1$ MeV for $E_r \ll$ MeV. This is the way we have obtained the last line of Table 3. Interesting enough, the values for the last two resonances turn out to be comparable with those reported by the QSO observations [9], but with the wrong sign.

We notice, however, that Γ_{tot} is considerably smaller than the energy difference required for the shift, implying that the "probability" of finding a shifted level that falls in the range of $\sim \Gamma_{\mathrm{tot}} \sim 30$ meV around the threshold is rather small. This conclusion seems to be corroborated by taking ${}^{155,157}\mathrm{Gd}$ into account as will be discussed.

There are many excited levels also in $n + {}^{155,157}\mathrm{Gd}$, some of which are illustrated in Fig. 11. The first resonance levels appear at exceptionally low energies, 26.8meV and 31.4meV, respectively. They are nearly degenerate. This unique feature is shared by none of the higher resonances, though all of them show remarkably similar widths, around 100 meV.

The analysis of the Oklo natural reactors shows, however, that considerable enhancement near the threshold appears to occur for both isotopes. This conclusion seems to remain true even if a possible significant effect of "contamination" is included [6]. What we observe is reasonably interpreted as coming from both of the two resonances.

The inherent ambiguity coming from higher resonances as encountered for $n + {}^{149}\mathrm{Sm}$ is also relevant here. This time, however, we must expect the levels to land near the thresholds "simultaneously" in both reactions. This requires "squared" smallness of the probability in view of Fig. 11. Note that the ratios of the widths to the level spacings are even smaller than those in Sm. Small probability will be "cubic" if we combine all of the results of Sm and Gd.

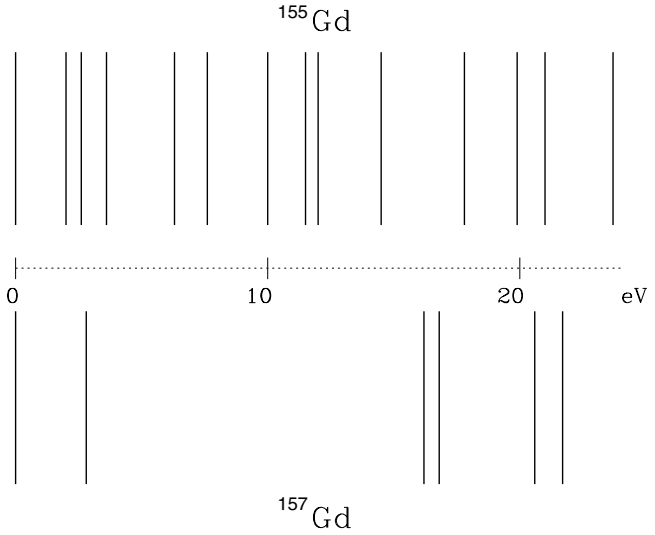


Fig. 11. Levels of resonances of $n + {}^{155}\mathrm{Gd}$ and $n + {}^{157}\mathrm{Gd}$, shown above and below the horizontal axis, respectively. The scale of the energy levels is eV. They appear to be distributed rather randomly, except for the first resonances which are nearly degenerate.

All in all, it is highly unlikely that the "discrepancy" between the Oklo constraint and the QSO result can be removed by assuming "distant migration" of higher resonance levels of the relevant isotopes.

C Another 3-Parameter Fit with an Offset

We have recently found a fit with an offset parameter included, parametrized by

$$y(u) = a\left(e^{bu}\cos\left(v - v_1\right) - \cos\left(v_1\right)\right), \tag{C.1}$$

where $v/u = v_{\mathrm{oklo}}/u_{\mathrm{oklo}} = 2\pi T^{-1}$ with v_1 determined by

$$v_1 = \tan^{-1}\left(\left(e^{-bu_{\mathrm{oklo}}} - \cos(v_{\mathrm{oklo}})\right)/\sin(v_{\mathrm{oklo}})\right). \tag{C.2}$$

We easily find that y defined this way vanishes both at $u = 0$ and $u = u_{\mathrm{oklo}}$. The three parameters a, b, T are then determined to minimize χ^2 for the QSO data. The result is for $a = 0.046, b = 4.0, T = 1.307$ with the fit shown in Fig. 12. The resulting $\chi^2_{\mathrm{red}} = 1.071$ is even smaller than 1.09 for our previous 3-parameter fit in Fig. 8.

Through these analyses we also find it unlikely that the current QSO result respects another constraint at the "meteorite time" around 4.6 Gys ago, or $u \approx 0.33$, requiring $|y| \lesssim 0.025$ [8]. The basic formulation used in analyzing the decay of ${}^{187}\mathrm{Re} \rightarrow {}^{187}\mathrm{Os}$ has been re-examined [24],[25].

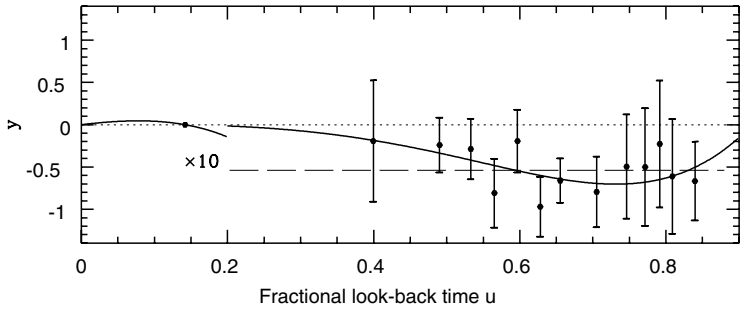


Fig. 12. The best fit of (C.1) for $a = 0.046, b = 4.0, T = 1.307$ giving $\chi^2_{\rm red} = 1.071$ which is comparable with 1.06 for the weighted mean of [9]

References

1. See, for example, R. Naudet, *Oklo: des réacteurs nucléaires fossiles*, Collection CEA, Eyrolles, Paris, 1991
2. The Oklo Phenomenon, Proc. of a Symposium, Libreville, June, 1975 (IAEA, Vienna, 1975)
3. P.K. Kuroda, J. Chem. Phys. **25**, 781, 1295 (1956)
4. A.I. Shlyakhter, Nature **264**, 340 (1976)
5. A.I. Shlyakhter, ATOMKI Report A/1, unpublished (1983), physics/0307023
6. Y. Fujii, A. Iwamoto, T. Fukahori, T. Ohnuki, M. Nakagawa, H. Hidaka, Y. Oura and P. Möller, Nucl. Phys. **B573**, 377 (2000). See also, hep-ph/0205206
7. T. Damour and F.J. Dyson, Nucl. Phys. **B480**, 37 (1996)
8. K.A. Olive, M. Pospelov, Y.-Z. Qian, A. Coc, M. Cassé and E. Vangioni-Flam, Phys. Rev. **D66**, 045022 (2002)
9. M.T. Murphy, J.K. Webb, V.V. Flambaum, MNRAS **345**, 609 (2003); astro-ph/0306483
10. M.T. Murphy et al., Constraining Variations in the Fine-Structure Constant, Quark Masses and the Strong Interaction, Lect. Notes Phys. **648**, 131–150 (2004)
11. J.D. Barrow and J. Magueijo, Astrophys. J., **532**, L87 (2000)
12. J.D. Barrow and C. O'Toole, astro-ph/9904116
13. C. Wetterich, Phys. Lett. **B561**, 10 (2003)
14. J.D. Bekenstein, Phys. Rev. **D25**, 1527 (1982): gr-qc/0208081
15. L. Anchordoqui and H. Goldberg, Phys. Rev. D **68**, 083513 (2003)
16. C.L. Gardner, Phys. Rev. D **68**, 043513 (2003)
17. A.G. Riess *et al.*, Astgron. J. **116**, 1009 (1998); S. Perlmutter *et al.*, Nature **391**, 51 (1998)
18. Y. Fujii, Phys. Rev. **D26**, 2580 (1982)
19. D. Dolgov, Proc. Nuffield Workshop, ed. G.W. Gibbons and S.T. Siklos, Cambridge University Press, 1982
20. Y. Fujii and K. Maeda, *The scalar-tensor theory of gravitation*, Cambridge University Press, 2003
21. Y. Fujii, Phys. Lett. **B573**, 39 (2003)
22. X. Calmet and H. Fritzsch, Eur. Phys. J. **C24**, 639 (2002)
23. P. Langacker, G. Segrè and M. Strassler, Phys. Lett. **B528**, 121 (2002)
24. K.A. Olive, M. Pospelov, Y.-Z. Qian, G. Manhès, E. Vangioni-Flam, A. Coc and M. Cassé, astro-ph/0309252
25. Y. Fujii and A. Iwamoto, Phys. Rev. Lett. **91**, 261101 (2003)

Cold Atom Clocks, Precision Oscillators and Fundamental Tests

S. Bize[1], P. Wolf[1,2], M. Abgrall[1], L. Cacciapuoti[1], A. Clairon[1], J. Grünert[1], Ph. Laurent[1], P. Lemonde[1], A.N. Luiten[4], I. Maksimovic[1], C. Mandache[3], H. Marion[1], F. Pereira Dos Santos[1], P. Rosenbusch[1], C. Salomon[5], G. Santarelli[1], Y. Sortais[1], M.E. Tobar[4], C. Vian[1], and S. Zhang[1]

[1] BNM-SYRTE, Observatoire de Paris, 61 avenue de l'Observatoire, 75014 Paris, France
[2] Bureau International des Poids et Mesures, Pavillon de Breteuil, 92312 Sèvres Cedex, France
[3] Institutul National de Fizica Laserilor, Plasmei si Radiatiei, P.O. Box MG36, Bucaresti, Magurele, Romania
[4] University of Western Australia, School of Physics, Nedlands 6907 WA, Australia
[5] Laboratoire Kastler Brossel, 24 rue Lhomond, 75231 Paris cedex 05, France

Abstract. We describe two experimental tests of the Equivalence Principle that are based on frequency measurements between precision oscillators and/or highly accurate atomic frequency standards. Based on comparisons between the hyperfine frequencies of ^{87}Rb and ^{133}Cs in atomic fountains, the first experiment constrains the variability of fundamental constants. The second experiment is based on a comparison between a cryogenic sapphire oscillator and a hydrogen maser. It tests Local Lorentz Invariance. In both cases, we report recent results which improve significantly over previous experiments.

1 Introduction

Einstein's Equivalence Principle (EP) is at the heart of special and general relativity and a cornerstone of modern physics. The central importance of this postulate in modern theory has motivated tremendous work to experimentally test EP [1]. Additionally, nearly all unification theories (in particular string theories) violate EP at some level [2–4] which further motivates experimental searches for such violations. A third motivation comes from a recent analysis of absorption systems in the spectra of distant quasars [8] which seems to indicate a variation of the fine-structure constant α over cosmological timescale, in violation of EP.

EP is made of three constituent elements. The Weak Equivalence Principle (WEP) postulates that *trajectories of neutral freely falling bodies are independent of their structure and composition.* Local Lorentz Invariance (LLI) postulates that *in any local freely falling reference frame, the result of a non-gravitational measurement is independent of the velocity of the frame.* Finally, Local Position Invariance (LPI) postulates that *in any local freely falling reference frame, the result of a non-gravitational measurement is independent of where and when it is performed.* The experiments described here use precision oscillators and atomic clocks to test LLI and LPI.

S. Bize et al., Cold Atom Clocks, Precision Oscillators and Fundamental Tests, Lect. Notes Phys. **648**, 189–207 (2004)
http://www.springerlink.com/

In a first section, experiments testing LPI are described. In these experiments, frequencies of atomic transitions are compared to each other in a local atomic clock comparison. The measurements are repeated over a few years. LPI implies that these measurements give consistently the same answer, a prediction which is directly tested. These experiments can be further interpreted as testing the stability of fundamental constants, if one recognizes that any atomic transition frequency can (at least in principle) be expressed as a function of properties of elementary particles and parameters of fundamental interactions. Such interpretation requires additional input from theoretical calculations of atomic frequencies. In this first section, after reviewing some of these calculations, we describe a comparison between ^{87}Rb and ^{133}Cs hyperfine frequencies in atomic fountains. We give new and significant constraints to the variability of fundamental constants based on the results of these experiments.

In a second part, a test of LLI is described. The frequency of a macroscopic cryogenic sapphire resonator is measured against a hydrogen maser as a function of time. We look for sidereal and and semi-sidereal modulations of the measured frequency which would indicate a violation of LLI depending on the speed and orientation of the laboratory frame with respect to a preferred frame. First the Robertson, Mansouri and Sexl theoretical framework is described as a basis to interpret the experiments. Then new results improving on previous experiments are reported.

2 Test of Local Position Invariance. Stability of Fundamental Constants

2.1 Theory

Local Position Invariance can be verified experimentally by repeating the same local experiment at different locations and/or different times. The most fundamental such experiments are measurements of the variation in time and space of fundamental constants, like the ones presented here. In general, putative violations of LPI are expected to be linked to couplings between the non-gravitational fields of the local experiment and the external gravitational field [5,6] or to slow time drifts of fundamental constants of cosmological origin [4]. The former can be tested by searching for a variation of fundamental constants with changes of the external gravitational field (e.g. due to the periodic movement of the laboratory in the sun's gravitational field [7]), the latter by searching for a linear time variation of the fundamental constants. Here we present experiments of the second kind, but in principle the same experiments can also be used to search for the periodic variations (at diurnal and annual frequencies) that would be induced by the varying local gravitational potential, provided that sufficient data was available.

Tests described here are based on highly accurate comparisons of atomic energies. In principle at least, it is possible to express any atomic energy as a function of the elementary particle properties and the coupling constants of

fundamental interactions using Quantum Electro-Dynamics (QED) and Quantum Chromo-Dynamics (QCD). As a consequence, it is possible to deduced a constraint to the variation of fundamental constants from a measurement of the stability of the ratio between various atomic frequencies.

Different types of atomic transitions are linked to different fundamental constants. The hyperfine frequency in a given electronic state of alkali-like atoms (involved for instance in ^{133}Cs, ^{87}Rb [9], ^{199}Hg$^+$ [10,11], ^{171}Yb$^+$ microwave clocks) can be approximated by:

$$\nu_{\rm hfs}^{(i)} \simeq R_\infty c \times \mathcal{A}_{\rm hfs}^{(i)} \times g^{(i)} \left(\frac{m_e}{m_p}\right) \alpha^2 \, F_{\rm hfs}^{(i)}(\alpha), \tag{1}$$

where R_∞ is the Rydberg constant, c the speed of light, $g^{(i)}$ the nuclear g-factor, m_e/m_p the electron to proton mass ratio and α the fine structure constant. In this equation, the dimension is given by $R_\infty c$, the atomic unit of frequency. $\mathcal{A}_{\rm hfs}^{(i)}$ is a numerical factor which depends on each particular atom. $F_{\rm hfs}^{(i)}(\alpha)$ is a relativistic correction factor to the motion of the valence electron in the vicinity of the nucleus. This factor strongly depends on the atomic number Z and has a major contribution for heavy nuclei. The superscript (i) indicates that the quantity depends on each particular atom. Similarly, the frequency of an electronic transition (involved in H [12], ^{40}Ca [13], ^{199}Hg$^+$ [14], ^{171}Yb$^+$ [15,16] optical clocks) can be approximated by

$$\nu_{\rm elec}^{(i)} \simeq R_\infty c \times \mathcal{A}_{\rm elec}^{(i)} \times F_{\rm elec}^{(i)}(\alpha). \tag{2}$$

Again, the dimension is given by $R_\infty c$. $\mathcal{A}_{\rm elec}^{(i)}$ is a numerical factor. $F_{\rm elec}^{(i)}(\alpha)$ is a function of α which accounts for relativistic effects, spin-orbit couplings and many-body effects [1]. From (1) and (2), it is possible to calculate the ratio between the frequencies in atomic species (i) and (ii), depending on the type of transition involved:

$$\frac{\nu_{\rm elec}^{(ii)}}{\nu_{\rm elec}^{(i)}} \propto \frac{F_{\rm elec}^{(ii)}(\alpha)}{F_{\rm elec}^{(i)}(\alpha)} \, , \tag{3}$$

$$\frac{\nu_{\rm hfs}^{(ii)}}{\nu_{\rm elec}^{(i)}} \propto g^{(ii)} \frac{m_e}{m_p} \alpha^2 \frac{F_{\rm hfs}^{(ii)}(\alpha)}{F_{\rm elec}^{(i)}(\alpha)} \, , \tag{4}$$

$$\frac{\nu_{\rm hfs}^{(ii)}}{\nu_{\rm hfs}^{(i)}} \propto \frac{g^{(ii)}}{g^{(i)}} \frac{F_{\rm hfs}^{(ii)}(\alpha)}{F_{\rm hfs}^{(i)}(\alpha)}. \tag{5}$$

The product of constants $R_\infty c$ cancels out in the ratio of two atomic frequencies and only dimensionless factors are left. Also, numerical factors that are irrelevant

[1] It should be noted that in general the energy of an electronic transition has in fact a contribution from the hyperfine interaction. However, this contribution is a small fraction of the total transition energy and thus carries no significant sensitivity to a variation of fundamental constants. The same applies to higher order terms in the expression of the hyperfine energy (1). A precision of 1 to 10 % on the sensitivity is sufficient to interpret current experiments.

to the present discussion have been omitted. Already, the different sensitivity of the various type of comparison can be seen in these equations. Comparisons between electronic transitions (3) are only sensitive to α. Comparisons between hyperfine transitions (5) are sensitive both to α and the strong interaction through the nuclear g-factors. Comparisons between an electronic transition and a hyperfine transition (4) are sensitive to α, to the strong interaction and to the electron mass.

The sensitivity of a given atomic transition to the variation of fundamental constants can be derived from (1) and (2):

$$\delta \ln \left(\frac{\nu_{\mathrm{hfs}}^{(i)}}{R_\infty c} \right) \simeq \frac{\delta g^{(i)}}{g^{(i)}} + \frac{\delta(m_e/m_p)}{(m_e/m_p)} + \left(2 + \alpha \frac{\partial}{\partial \alpha} \ln F_{\mathrm{hfs}}^{(i)}(\alpha) \right) \times \frac{\delta \alpha}{\alpha} \,, \tag{6}$$

$$\delta \ln \left(\frac{\nu_{\mathrm{elec}}^{(i)}}{R_\infty c} \right) \simeq \left(\alpha \frac{\partial}{\partial \alpha} \ln F_{\mathrm{elec}}^{(i)}(\alpha) \right) \times \frac{\delta \alpha}{\alpha} \,. \tag{7}$$

In recent work [17], it has been suggested that this analysis can be pushed one step further by linking the g-factors $g^{(i)}$ and the proton mass m_p to fundamental parameters, namely the mass scale of QCD Λ_{QCD} and the quark mass $m_q = (m_u + m_d)/2$. Within this framework, any atomic frequency comparison can be interpreted as testing the stability of three dimensionless fundamental constants: α, $m_q/\Lambda_{\mathrm{QCD}}$ and $m_e/\Lambda_{\mathrm{QCD}}$. For any transition we can write

$$\delta \ln \left(\frac{\nu^{(i)}}{R_\infty c} \right) \simeq K_\alpha^{(i)} \times \frac{\delta \alpha}{\alpha} + K_q^{(i)} \times \frac{\delta(m_q/\Lambda_{\mathrm{QCD}})}{(m_q/\Lambda_{\mathrm{QCD}})} + K_e^{(i)} \times \frac{\delta(m_e/\Lambda_{\mathrm{QCD}})}{(m_e/\Lambda_{\mathrm{QCD}})} \tag{8}$$

where the superscript (i) again indicates that the coefficient depends on each particular transition. Hyperfine transitions are sensitive to all three fundamental constants ($K_\alpha^{(i)}, K_q^{(i)} \neq 0$; $K_e^{(i)} \simeq 1$). For electronic transitions, we have $K_e^{(i)}, K_q^{(i)} \simeq 0$ and therefore they are essentially sensitive to a variation of α. Four well chosen atomic transitions constraining the stability of three independent frequency ratios are enough to constrain independently the stability of the three fundamental constants α, $m_q/\Lambda_{\mathrm{QCD}}$ and $m_e/\Lambda_{\mathrm{QCD}}$. From these equations, it can be seen that at least two different hyperfine transitions must be involved to independently constrain $m_q/\Lambda_{\mathrm{QCD}}$ and $m_e/\Lambda_{\mathrm{QCD}}$ which emphasizes the need to maintain and improve highly accurate microwave atomic clocks[2]. With more than four well chosen atomic clocks, redundancy is achieved which means that a non-vanishing variation of fundamental constants can be identified by a clear signature.

Calculations of the coefficients have now been done for a large number of atomic species [10,17–22]. For hyperfine transitions in ^{87}Rb, ^{133}Cs and ^{199}Hg$^+$ the most recent calculation gives [17]

$$\delta \ln \left(\frac{\nu_{\mathrm{hfs}}(^{87}\mathrm{Rb})}{R_\infty c} \right) \simeq 2.34 \frac{\delta \alpha}{\alpha} - 0.064 \frac{\delta(m_q/\Lambda_{\mathrm{QCD}})}{(m_q/\Lambda_{\mathrm{QCD}})} + \frac{\delta(m_e/\Lambda_{\mathrm{QCD}})}{(m_e/\Lambda_{\mathrm{QCD}})} \,, \tag{9}$$

[2] In principle, it is also possible to use a vibrational molecular transition with $K_\alpha^{(i)} \simeq 0$ and $K_e^{(i)} \simeq 1/2$.

$$\delta \ln \left(\frac{\nu_{\rm hfs}(^{133}{\rm Cs})}{R_\infty c} \right) \simeq 2.83 \frac{\delta\alpha}{\alpha} + 0.11 \frac{\delta(m_q/\Lambda_{\rm QCD})}{(m_q/\Lambda_{\rm QCD})} + \frac{\delta(m_e/\Lambda_{\rm QCD})}{(m_e/\Lambda_{\rm QCD})} \ , \quad (10)$$

$$\delta \ln \left(\frac{\nu_{\rm hfs}(^{199}{\rm Hg}^+)}{R_\infty c} \right) \simeq 4.3 \frac{\delta\alpha}{\alpha} - 0.02 \frac{\delta(m_q/\Lambda_{\rm QCD})}{(m_q/\Lambda_{\rm QCD})} + \frac{\delta(m_e/\Lambda_{\rm QCD})}{(m_e/\Lambda_{\rm QCD})} \ . \quad (11)$$

In each case, the most recent and precise value for the $K_\alpha^{(i)}$ coefficient given here is in good agreement with earlier calculations [10]. For electronic transitions in H, ^{40}Ca and $^{199}{\rm Hg}^+$, we have

$$\delta \ln \left(\frac{\nu_{\rm elec}({\rm H})}{R_\infty c} \right) \simeq 0 \ , \quad (12)$$

$$\delta \ln \left(\frac{\nu_{\rm elec}(^{40}{\rm Ca})}{R_\infty c} \right) \simeq 0.03 \frac{\delta\alpha}{\alpha} \ , \quad (13)$$

$$\delta \ln \left(\frac{\nu_{\rm elec}(^{199}{\rm Hg}^+)}{R_\infty c} \right) \simeq -3.2 \frac{\delta\alpha}{\alpha} \ . \quad (14)$$

A new generation of laser cooled optical clocks is now under development in several groups (^{87}Sr [23–25], ^{171}Yb, $^{27}{\rm Al}^+$ [26],...). This work will significantly improve the stringency and redundancy of this test of LPI.

2.2 Experiments with ^{87}Rb and ^{133}Cs Fountain Clocks

In these experiments, three atomic fountains are compared to each other, using a hydrogen maser (H-maser) as a flywheel oscillator (Fig. 1). Two fountains, a transportable fountain FOM, and FO1 [27] are using cesium atoms. The third fountain is a dual fountain (FO2) [28], operating alternately with rubidium ($\rm FO2_{Rb}$) and cesium ($\rm FO2_{Cs}$). These fountains have been continuously upgraded in order to improve their accuracy from 2×10^{-15} in 1998 to 8×10^{-16} for cesium and from 1.3×10^{-14} [29] to 6×10^{-16} for rubidium.

Fountain clocks operate as follows. First, atoms are collected and laser cooled in an optical molasses or in a magneto-optical trap in 0.3 to 0.6 s. Atoms are

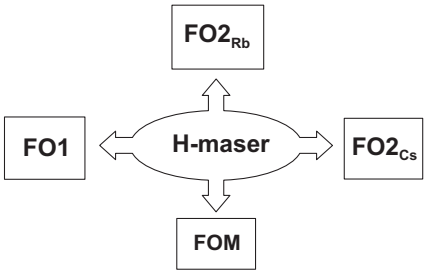


Fig. 1. BNM-SYRTE clock ensemble. A single 100 MHz signal from a H-maser is used for frequency comparisons and is distributed to each of the microwave synthesizers of the ^{133}Cs (FO1, FOM, $\rm FO2_{Cs}$) and ^{87}Rb fountain clocks. In 2001, the Rb fountain has been upgraded and is now a dual fountain using alternately rubidium ($\rm FO2_{Rb}$) or cesium atoms ($\rm FO2_{Cs}$).

launched upwards, and selected in the clock level ($m_F = 0$) by a combination of microwave and laser pulses. Then, atoms interact twice with a microwave field tuned near the hyperfine frequency, in a Ramsey interrogation scheme. The microwave field is synthesized from a low phase noise 100 MHz signal from a quartz oscillator, which is phase locked to the reference signal of the H-maser (Fig. 1). After the microwave interactions, the population of each hyperfine state is measured using light induced fluorescence. This provides a measurement of the transition probability as a function of microwave detuning. Successive measurements are used to steer the average microwave field to the frequency of the atomic resonance using a digital servo system. The output of the servo provides a direct measurement of the frequency difference between the H-maser and the fountain clock.

The three fountains have different geometries and operating conditions: the number of detected atoms ranges from 3×10^5 to 2×10^6 at a temperature of $\sim 1\,\mu$K, the fountain cycle duration from 1.1 to 1.6 s. The Ramsey resonance width is between 0.9 and 1.2 Hz. In measurements reported here the fractional frequency instability is $(1-2) \times 10^{-13}\tau^{-1/2}$, where τ is the averaging time in seconds. Fountain comparisons have a typical resolution of $\sim 10^{-15}$ for a 12 hour integration, and each of the four data campaigns lasts from 1 to 2 months during which an accuracy evaluation of each fountain is performed.

The 2002 measurements are presented in Fig. 2, which displays the maser fractional frequency offset, measured by the Cs fountains FOM and $\mathrm{FO2_{Cs}}$. Also shown is the H-maser frequency offset measured by the Rb fountain $\mathrm{FO2_{Rb}}$ where the Rb hyperfine frequency is conventionally chosen to be $\nu_{\mathrm{Rb}}(1999) = 6\,834\,682\,610.904\,333$ Hz, our 1999 value. The data are corrected for the systematic frequency shifts listed in Table 1. The H-maser frequency exhibits fractional fluctuations on the order of 10^{-14} over a few days, ten times larger than the typical statistical uncertainty resulting from the instability of the fountain clocks. In order to reject the H-maser frequency fluctuations, the fountain data are recorded simultaneously (within a few minutes). The fractional frequency differences plotted in Fig. 2 b illustrate the efficiency of this rejection. FO2 is operated alternately with Rb and Cs, allowing both Rb-Cs comparisons and Cs-Cs comparisons (central part of Fig. 2) to be performed.

Systematic effects shifting the frequency of the fountain standards are listed in Table 1. The quantization magnetic field in the interrogation region is determined with a 0.1 nT uncertainty by measuring the frequency of a linear field-dependent "Zeeman" transition. The temperature in the interrogation region is monitored with 5 platinum resistors and the uncertainty on the black-body radiation frequency shift corresponds to temperature fluctuations of about 1 K [30]. Clock frequencies are corrected for the cold collision and cavity pulling frequency shifts using several methods [31,32]. For Rb, unlike [32], an optical molasses with a small number of atoms ($\sim 5.4 \times 10^6$) is used. We thus estimate that these two shifts are smaller than 5×10^{-17}. All other effects do not contribute significantly and their uncertainties are added quadratically. We searched for the influence of synchronous perturbations by changing the timing sequence and the atom launch height. To search for possible microwave leakage, we changed the power

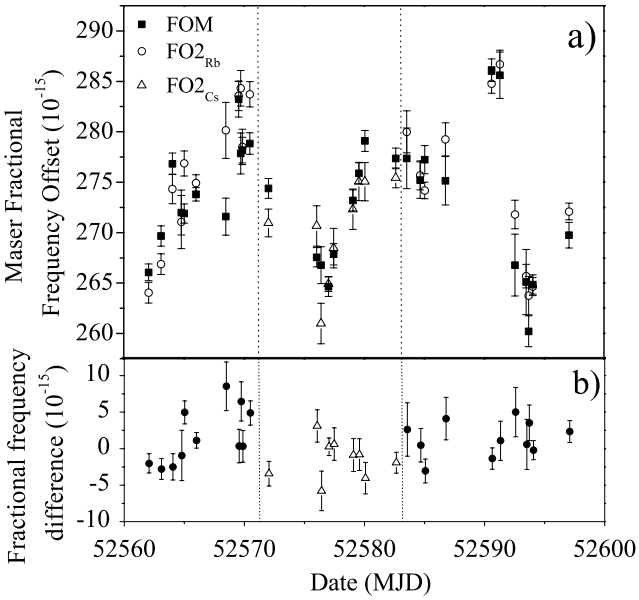


Fig. 2. The 2002 frequency comparison data. a) H-maser fractional frequency offset versus FOM (black squares), and alternately versus $FO2_{Rb}$ (open circles) and $FO2_{Cs}$ (open triangles between dotted lines). b) Fractional frequency differences. Between dotted lines, Cs-Cs comparisons, outside Rb-Cs comparisons. Error bars are purely statistical. They correspond to the Allan standard deviation of the comparisons and do not include contributions from fluctuations of systematic shifts of Table 1.

Table 1. Accuracy budget of the fountains involved in the 2002 measurements (FO2 et FOM).

Fountain	**$FO2_{Cs}$**	**$FO2_{Rb}$**	**FOM**
Effect	Value & Uncertainty (10^{-16})		
2^{nd} order Zeeman	1773.0 ± 5.2	3207.0 ± 4.7	385.0 ± 2.9
Blackbody Radiation	-173.0 ± 2.3	-127.0 ± 2.1	-186.0 ± 2.5
Cold collisions + cavity pulling	-95.0 ± 4.6	0.0 ± 1.0	-24.0 ± 4.8
others	0.0 ± 3.0	0.0 ± 3.0	0.0 ± 3.7
Total uncertainty	**8**	**6**	**8**

($\times 9$) in the interrogation microwave cavity. No shift was found at a resolution of 10^{-15}. The shift due to residual coherences and populations in neighboring Zeeman states is estimated to be less than 10^{-16}. As shown in [33], the shift due to the microwave photon recoil is very similar for Cs and Rb and smaller than $+1.4 \times 10^{-16}$. Relativistic corrections (gravitational redshift and second order Doppler effect) contribute to less than 10^{-16} in the clock comparisons.

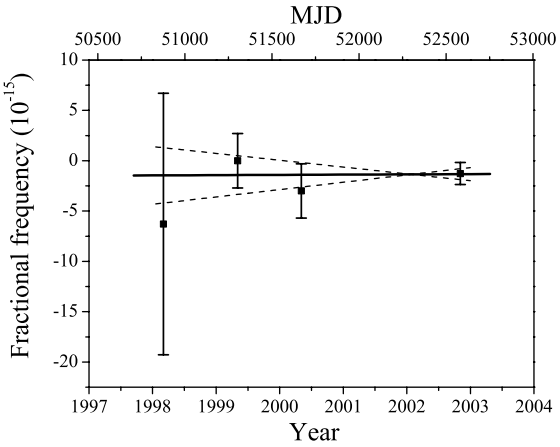


Fig. 3. Measured ^{87}Rb frequencies referenced to the ^{133}Cs fountains over 57 months. The 1999 measurement value ($\nu_{\mathrm{Rb}}(1999) = 6\,834\,682\,610.904\,333\,\mathrm{Hz}$) is conventionally used as reference. A weighted linear fit to the data gives $\frac{d}{dt}\ln\left(\frac{\nu_{\mathrm{Rb}}}{\nu_{\mathrm{Cs}}}\right) = (0.2 \pm 7.0) \times 10^{-16}\,\mathrm{yr}^{-1}$. Dotted lines correspond to the 1σ slope uncertainty.

For the Cs-Cs 2002 comparison, we find

$$\frac{\nu_{\mathrm{Cs}}^{\mathrm{FO2}}(2002) - \nu_{\mathrm{Cs}}^{\mathrm{FOM}}(2002)}{\nu_{\mathrm{Cs}}} = +12(6)(12) \times 10^{-16} \tag{15}$$

where the first parenthesis reflects the 1σ statistical uncertainty, and the second the systematic uncertainty, obtained by adding quadratically the inaccuracies of the two Cs clocks (see Table 1). The two Cs fountains are in good agreement despite their significantly different operating conditions (see Table 1.), showing that systematic effects are well understood at the 10^{-15} level.

In 2002, the ^{87}Rb frequency measured with respect to the average ^{133}Cs frequency is found to be

$$\nu_{\mathrm{Rb}}(2002) = 6\,834\,682\,610.904\,324(4)(7)\,\mathrm{Hz} \tag{16}$$

where the error bars now include $\mathrm{FO2}_{\mathrm{Rb}}$, $\mathrm{FO2}_{\mathrm{Cs}}$ and FOM uncertainties. This is the most accurate frequency measurement to date.

In Fig. 3 are plotted all our Rb-Cs frequency comparisons. Except for the less precise 1998 data [29], two Cs fountains were used together to perform the Rb measurements. The uncertainties for the 1999 and 2000 measurements were 2.7×10^{-15}, because of lower clock accuracy and lack of rigorous simultaneity in the earlier frequency comparisons [28]. A weighted linear fit to the data in Fig. 3 determines how our measurements constrain a possible time variation of $\nu_{\mathrm{Rb}}/\nu_{\mathrm{Cs}}$. We find

$$\frac{d}{dt}\ln\left(\frac{\nu_{\mathrm{Rb}}}{\nu_{\mathrm{Cs}}}\right) = (0.2 \pm 7.0) \times 10^{-16}\,\mathrm{yr}^{-1} \tag{17}$$

which represents a 5-fold improvement over our previous results [28] and a 100-fold improvement over the Hg^+-H hyperfine energy comparison [10]. Using (6) and the sensitivity to α in (9) and (10), we find that this results implies the following constraint:

$$\frac{d}{dt}\ln\left(\frac{g_{\rm Cs}}{g_{\rm Rb}}\alpha^{0.49}\right) = (0.2 \pm 7.0) \times 10^{-16}\,{\rm yr}^{-1}. \tag{18}$$

Using the link between g-factors, m_q and $\Lambda_{\rm QCD}$ ([17], 9 and 10), we get

$$\frac{d}{dt}\ln\left(\alpha^{0.49}\,[m_q/\Lambda_{\rm QCD}]^{0.17}\right) = (0.2 \pm 7.0) \times 10^{-16}\,{\rm yr}^{-1}. \tag{19}$$

A comparison between the single mercury $^{199}Hg^+$ ion optical clock and the ^{133}Cs hyperfine splitting has been recently reported by the NIST group [14] which (according to 10 and 14) constrains the stability of $\alpha^{6.0}[m_e/\Lambda_{\rm QCD}][m_q/\Lambda_{\rm QCD}]^{0.1}$ at the level of 7×10^{-15} yr^{-1}.

3 Tests of Local Lorentz Invariance

Local Lorentz Invariance is one of the constituent elements of the Einstein Equivalence Principle (see Sect. 1) and therefore one of the cornerstones of modern physics. It is the fundamental hypothesis of special relativity and is related to the "constancy of the speed of light". The central importance of this postulate in modern physics has motivated tremendous work to experimentally test LLI [1,34]. Additionally, nearly all unification theories (in particular string theory) violate the EP at some level [35] which further motivates experimental searches for such violations of the universality of free fall [3] and of Lorentz invariance [36,37].

We report here on experimental tests of LLI using a cryogenic sapphire oscillator and a hydrogen maser. The relative frequency of the two clocks is monitored looking for a Lorentz violating signal which would modulate that frequency at, typically, sidereal and semi-sidereal periods due to the movement of the lab with the rotation of the Earth. We set limits on parameters that describe such Lorentz violating effects, improving our previous results [42] by a factor 2 and the best other results [41,44] by up to a factor 70.

Many modern experiments that test LLI rely essentially on the stability of atomic clocks and macroscopic resonators [38–44], therefore improvements in oscillator technology have gone hand in hand with improved tests of LLI. Our experiment is no exception, the large improvements being a direct result of the excellent stability of our cryogenic sapphire oscillator. Additionally, its operation at a microwave frequency allows a direct comparison to a hydrogen maser which provides a highly stable and reliable reference frequency.

Numerous test theories that allow the modeling and interpretation of experiments that test LLI have been developed. Kinematical frameworks [45,46] postulate a simple parametrisation of the Lorentz transformations with experiments

setting limits on the deviation of those parameters from their special relativistic values. A more fundamental approach is offered by theories that parametrise the coupling between gravitational and non-gravitational fields (TH$\epsilon\mu$ [5,1,6] or χg [47] formalisms) which allow the comparison of experiments that test different aspects of the EP. Finally, formalisms based on string theory [34,35,3,36] have the advantage of being well motivated by theories of physics that are at present the only candidates for a unification of gravity and the other fundamental forces of nature.

3.1 Theory

Owing to their simplicity the kinematical frameworks of [45,46] have been widely used to model and interpret many previous experiments testing LLI [38,40–42,48,49] and we will follow that route. An analysis based on the more fundamental "Standard Model Extension" (SME) [36,50] is under way and will be published shortly.

By construction kinematical frameworks do not allow for any dynamical effects on the measurement apparatus. This implies that in all inertial frames two clocks of different nature (e.g. based on different atomic species) run at the same relative rate and two length standards made of different materials keep their relative lengths. Coordinates are defined by the clocks and length standards, and only the transformations between those coordinate systems are modified. In general this leads to observable effects on light propagation in moving frames but, by definition, to no observable effects on clocks and length standards. In particular, no attempt is made at explaining the underlying physics (e.g. modified Maxwell and/or Dirac equations) that could lead to Lorentz violating light propagation but leave e.g. atomic energy levels unchanged. On the other hand dynamical frameworks (e.g. the TH$\epsilon\mu$ formalism or the SME) in general use a modified general Lagrangian that leads to modified Maxwell and Dirac equations and hence to Lorentz violating light propagation and atomic properties, which is why they are considered more fundamental and more complete than the kinematical frameworks. Furthermore, as shown in [50], the SME is kept sufficiently general to, in fact, encompass the kinematical frameworks and some other dynamical frameworks (in particular the TH$\epsilon\mu$ formalism) as special cases, although there are no simple and direct relationships between the respective parameters.

Concerning our experiment the SME allows the calculation of Lorentz violating effects on the fields inside the sapphire resonator, on the properties of the sapphire crystal itself and on the hydrogen maser transition. As shown in [59] the effect on the sapphire crystal amounts to only a few percent of the direct effect on the fields, and [60] show that the hydrogen $m_F = 0 \rightarrow m'_F = 0$ clock transition is not affected to first order. Hence the total effect is dominated by the Lorentz violating properties of the electromagnetic fields inside the resonator which can be calculated following the principles laid down in [50]. An analysis of our experiment in that framework is currently being carried out, and the results will be published elsewhere in the near future. In this paper we concentrate on the analysis using the kinematical framework of Mansouri and Sexl [46].

The Robertson, Mansouri and Sexl Framework

Kinematical frameworks for the description of Lorentz violation have been pioneered by Robertson [45] and further refined by Mansouri and Sexl [46] and others. Fundamentally the different versions of these frameworks are equivalent, and relations between their parameters are readily obtained. As mentioned above these frameworks postulate generalized transformations between a preferred frame candidate $\Sigma(T, \mathbf{X})$ and a moving frame $\mathrm{S}(t, \mathbf{x})$ where it is assumed that in both frames coordinates are realized by identical standards. We start from the transformations of [46] (in differential form) for the case where the velocity of S as measured in Σ is along the positive X-axis, and assuming Einstein synchronization in S (we will be concerned with signal travel times around closed loops so the choice of synchronization convention can play no role),

$$dT = \frac{1}{a}\left(dt + \frac{vdx}{c^2}\right); dX = \frac{dx}{b} + \frac{v}{a}\left(dt + \frac{vdx}{c^2}\right);\ dY = \frac{dy}{d};\ dZ = \frac{dz}{d} \tag{20}$$

with c the velocity of light in vacuum in Σ. Using the usual expansion of the three parameters ($a \approx 1+\alpha_{\mathrm{MS}}v^2/c^2+\mathcal{O}\left(v^4/c^4\right); b \approx 1+\beta_{\mathrm{MS}}v^2/c^2+\mathcal{O}\left(v^4/c^4\right); d \approx 1+\delta_{\mathrm{MS}}v^2/c^2+\mathcal{O}\left(v^4/c^4\right)$), setting $c^2dT^2 = dX^2+dY^2+dZ^2$ in Σ, and transforming according to (20) we find the coordinate travel time of a light signal in S,

$$dt = \frac{dl}{c}\left(1 - (\beta_{\mathrm{MS}} - \alpha_{\mathrm{MS}} - 1)\frac{v^2}{c^2} - \left(\frac{1}{2} - \beta_{\mathrm{MS}} + \delta_{\mathrm{MS}}\right)\sin^2\theta\frac{v^2}{c^2}\mathcal{O}\left(\frac{v^4}{c^4}\right)\right), \tag{21}$$

where $dl = \sqrt{dx^2 + dy^2 + dz^2}$ and θ is the angle between the direction of light propagation and the velocity $\mathbf{v}$ of S in Σ. In special relativity $\alpha_{\mathrm{MS}} = -1/2; \beta_{\mathrm{MS}} = 1/2; \delta_{\mathrm{MS}} = 0$ and (20) reduces to the usual Lorentz transformations. Generally, the best candidate for Σ is taken to be the frame of the cosmic microwave background (CMB) [51,52] with the velocity of the solar system in that frame taken as $v_\odot \approx 377$ km/s, decl. $\approx -6.4°$, $RA \approx 11.2$h.

Michelson-Morley type experiments [53,38] determine the coefficient $P_{MM} = (1/2 - \beta_{\mathrm{MS}} + \delta_{\mathrm{MS}})$ of the direction dependent term. For many years the most stringent limit on that parameter was $|P_{MM}| \leq 5 \times 10^{-9}$ determined over 23 years ago in an outstanding experiment [38]. Our experiment confirms that result with roughly equivalent uncertainty (2.2×10^{-9}). Recently an improvement to $|P_{MM}| \leq 1.5 \times 10^{-9}$ has been reported [44]. Kennedy-Thorndike experiments [39–41] measure the coefficient $P_{KT} = (\beta_{\mathrm{MS}} - \alpha_{\mathrm{MS}} - 1)$ of the velocity dependent term. The most stringent limit [41] on $|P_{KT}|$ has been recently improved from [40] by a factor 3 to $|P_{KT}| \leq 2.1\times10^{-5}$. We improve this result by a factor of 70 to $|P_{KT}| \leq 3.0\times10^{-7}$. Finally clock comparison and Doppler experiments [48,54,49] measure α_{MS}, currently limiting it to $|\alpha_{\mathrm{MS}} + 1/2| \leq 8 \times 10^{-7}$. The three types of experiments taken together then completely characterize any deviation from Lorentz invariance in this particular test theory, with present limits summarized in Table 2.

Our cryogenic oscillator consists of a sapphire crystal of cylindrical shape operating in a whispering gallery mode (see Fig. 4 for a schematic drawing and

Table 2. Present limits on Lorentz violating parameters in the framework of [46]

Reference	$\alpha_{\rm MS} + 1/2$	$1/2 - \beta_{\rm MS} + \delta_{\rm MS}$	$\beta_{\rm MS} - \alpha_{\rm MS} - 1$
[48,54,49]	$\leq 8 \times 10^{-7}$	-	-
[38]	-	$\leq 5 \times 10^{-9}$	-
[44]	-	$(2.2 \pm 1.5) \times 10^{-9}$	-
[41]	-	-	$(1.9 \pm 2.1) \times 10^{-5}$
our previous results [42]	-	$(1.5 \pm 4.2) \times 10^{-9}$	$(-3.1 \pm 6.9) \times 10^{-7}$
this work	-	$(1.2 \pm 2.2) \times 10^{-9}$	$(1.6 \pm 3.0) \times 10^{-7}$

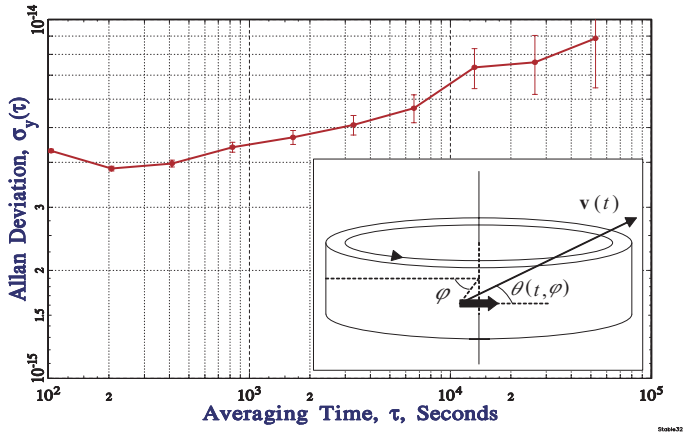


Fig. 4. Typical relative frequency stability of the CSO - H-maser difference after removal of a linear frequency drift. The inset is a schematic drawing of the cylindrical sapphire oscillator with the Poynting vector **P** in the whispering gallery (WG) mode, the velocity $\mathbf{v}(t)$ of the cylinder with respect to the CMB, and the relevant angles for a photon in the WG mode.

[55,56] for a detailed description). Its coordinate frequency can be expressed by $\nu = m/t_c$ where t_c is the coordinate travel time of a light signal around the circumference of the cylinder (of radius r) and m is a constant. From (21) the relative frequency difference between the sapphire oscillator and the hydrogen maser (which, by definition, realizes coordinate time in S [57]) is

$$\frac{\Delta\nu(t)}{\nu_0} = P_{KT}\frac{v(t)^2}{c^2} + P_{MM}\frac{v(t)^2}{c^2}\frac{1}{2\pi}\int_0^{2\pi}\sin^2\theta(t,\varphi)d\varphi + \mathcal{O}(3) \qquad (22)$$

where $\nu_0 = m/(2\pi r/c)$, $v(t)$ is the (time dependent) speed of the lab in Σ, and φ is the azimuthal angle of the light signal in the plane of the cylinder. The periodic time dependence of v and θ due to the rotation and orbital motion of the Earth with respect to the CMB frame allow us to set limits on the two

parameters in (22) by adjusting the periodic terms of appropriate frequency and phase (see [58] for calculations of similar effects for several types of oscillator modes). Given the limited durations of our data sets (≤ 15 days) the dominant periodic terms arise from the Earth's rotation, so retaining only those we have $\mathbf{v}(t) = \mathbf{u} + \mathbf{\Omega} \times \mathbf{R}$ with $\mathbf{u}$ the velocity of the solar system with respect to the CMB, $\mathbf{\Omega}$ the angular velocity of the Earth, and $\mathbf{R}$ the geocentric position of the lab. We then find after some calculation.

$$\begin{aligned} \Delta\nu/\nu_0 = & P_{KT}(H\sin\lambda) \\ & + P_{MM}(A\cos\lambda + B\cos(2\lambda) + C\sin\lambda + D\sin\lambda\cos\lambda + E\sin\lambda\cos(2\lambda)) \end{aligned} \tag{23}$$

where $\lambda = \Omega t + \phi$, and $A - E$ and ϕ are constants depending on the latitude and longitude of the lab ($\approx 48.7°$N and $2.33°$E for Paris). Numerically $H \approx -2.6 \times 10^{-9}$, $A \approx -8.8 \times 10^{-8}$, $B \approx 1.8 \times 10^{-7}$, $C - E$ of order 10^{-9}. We note that in (23) the dominant time variations of the two combinations of parameters are in quadrature and at twice the frequency which indicates that they should decorrelate well in the data analysis allowing a simultaneous determination of the two (as confirmed by the correlation coefficients given in Sect. 3.2). Adjusting this simplified model to our data we obtain results that differ by less than 10% from the results presented in Sect. 3.2 that were obtained using the complete model ((22) including the orbital motion of the Earth).

3.2 Experimental Results

The cryogenic sapphire oscillator (CSO) is compared to a commercial (Datum Inc.) active hydrogen maser whose frequency is also regularly compared to cesium and rubidium atomic fountain clocks in the laboratory [28]. The CSO resonant frequency at 11.932 GHz is compared to the 100 MHz output of the hydrogen maser. The maser signal is multiplied up to 12 GHz of which the CSO signal is subtracted. The remaining ≈ 67 MHz signal is mixed to a synthesizer signal at the same frequency and the low frequency beat at ≈ 64 Hz is counted, giving access to the frequency difference between the maser and the CSO. The instability of the comparison chain has been measured and does not exceed a few parts in 10^{16}. The typical stability of the measured CSO - maser frequency after removal of a linear frequency drift is shown in Fig. 4. Since September 2002 we are taking continuous temperature measurements on top of the CSO dewar and behind the electronics rack. Starting January 2003 we have implemented an active temperature control of the CSO room and changed some of the electronics. As a result the diurnal and semi-diurnal temperature variations during measurement runs (≈ 2 weeks) were greatly reduced to less than $0.025°$ C in amplitude (best case), and longer and more reliable data sets became available.

Our previously published results [42] are based on data sets taken between Nov. 2001 and Sep. 2002 which did not all include regular temperature monitoring and control. Here we use only data that was permanently temperature controlled, 13 data sets in total spanning Sept. 2002 to Aug. 2003, of differing lengths (5 to 16 days, 140 days in total). The sampling time for all data sets was 100 s except two data sets with $\tau_0 = 12$ s. To make the data more manageable we

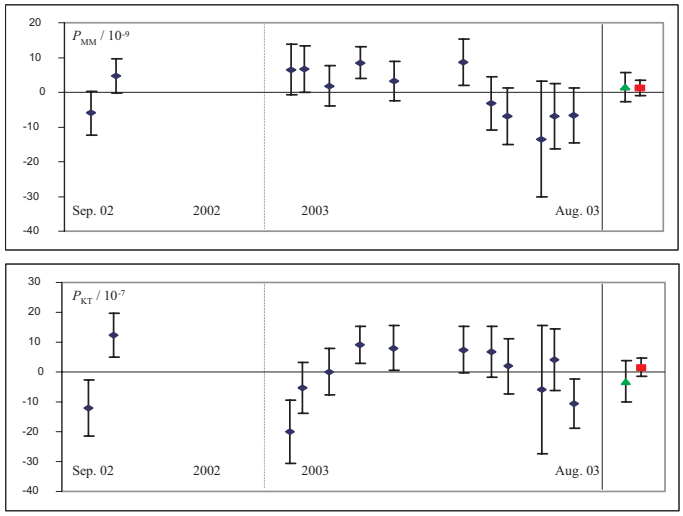


Fig. 5. Values of the two parameters (P_{KT} and P_{MM}) from a fit to each individual data set (diamonds) and a global fit to all the data (squares). For comparison our previously published results [42] are also shown (triangles). The error bars indicate the combined uncertainties from statistics and systematic effects.

first average all points to $\tau_0 = 2500$ s. For the data analysis we simultaneously adjust an offset and a rate (natural frequency drift, typically $\approx 1.7 \times 10^{-18}$ s^{-1}) per data set and the two parameters of the model (22). In the model (22) we take into account the rotation of the Earth and the Earth's orbital motion, the latter contributing little as any constant or linear terms over the durations of the individual data sets are absorbed by the adjusted offsets and rates.

When carrying out an ordinary least squares (OLS) adjustment we note that the residuals have a significantly non-white behavior as one would expect from the slope of the Allan deviation of Fig. 4. The power spectral density (PSD) of the residuals when fitted with a power law model of the form $S_y(f) = kf^{\mu}$ yields typically $\mu \approx -1.5$ to -2. In the presence of non-white noise OLS is not the optimal regression method [61,62] as it can lead to significant underestimation of the parameter uncertainties [61].

An alternative method is weighted least squares (WLS) [62] which allows one to account for non-white noise processes in the original data by pre-multiplying both sides of the design equation (our (22) plus the offsets and rates) by a weighting matrix containing off diagonal elements. To determine these off diagonal terms we first carry out OLS and adjust the $S_y(f) = kf^{\mu}$ power law model to the PSD of the post-fit residuals determining a value of μ for each data set. We then use these μ values to construct a weighting matrix following the method of fractional differencing described, for example, in [61]. Figure 5 shows the resulting values of the two parameters (P_{KT} and P_{MM}) for each individual data set. A global WLS fit of the two parameters and the 13 offsets and drifts yields $P_{MM} = (1.2 \pm 1.9) \times 10^{-9}$ and $P_{KT} = (1.6 \pm 2.3) \times 10^{-7}$ (1σ uncertainties),

with the correlation coefficient between the two parameters less than 0.01 and all other correlation coefficients < 0.06. The distribution of the 13 individual values around the ones obtained from the global fit is well compatible with a normal distribution ($\chi^2 = 10.7$ and 14.6 for P_{MM} and P_{KT} respectively).

Systematic effects with diurnal or semi-diurnal period and appropriate phase could mask a putative sidereal signal. The statistical uncertainties of P_{MM} and P_{KT} obtained from the WLS fit above correspond to sidereal and semi-sidereal terms (from (23)) of $\approx 7 \times 10^{-16}$ and $\approx 4 \times 10^{-16}$ respectively so any systematic effects exceeding these limits need to be taken into account in the final uncertainty. We expect the main contributions to such effects to arise from temperature, pressure and magnetic field variations that would affect the hydrogen maser, the CSO and the associated electronics, and from tilt variations of the CSO which are known to affect its frequency. Measurements of the tilt variations of the CSO with diurnal and semi-diurnal periods show amplitudes of 4.6 μrad and 1.6 μrad.

To estimate the tilt sensitivity we have intentionally tilted the oscillator by ≈ 5 mrad off its average position which led to relative frequency variations of $\approx 3 \times 10^{-13}$ from which we deduce a tilt sensitivity of $\approx 6 \times 10^{-17} \mu\text{rad}^{-1}$. This value corresponds to a worst case scenario as we expect a quadratic rather than linear frequency variation for small tilts around the vertical. Even with this pessimistic estimate diurnal and semi-diurnal frequency variations due to tilt do not exceed 3×10^{-16} and 1×10^{-16} respectively and are therefore negligible with respect to the statistical uncertainties.

In December 2002 we implemented an active temperature stabilization inside an isolated volume (≈ 15 m^3) that included the CSO and all the associated electronics. The temperature was measured continously in two fixed locations (behind the electronics rack and on top of the dewar). For the best data sets the measured temperature variations did not exceed 0.02/0.01 °C in amplitude for the diurnal and semi-diurnal components. In the worst cases (the two 2002 data sets and some data sets taken during a partial air conditioning failure) the measured temperature variations could reach 0.26/0.08 °C. When intentionally heating and cooling the CSO lab by ≈ 3°C we see frequency variations of $\approx 5 \times 10^{-15}$ per °C. This is also confirmed when we induce a large sinusoidal temperature variation (≈ 1.5°C amplitude). Using this we can calculate a value for temperature induced frequency variations with diurnal and semi-diurnal periods for each data set, obtaining values that range from $\approx 1.3 \times 10^{-15}$ to $\approx 5 \times 10^{-17}$.

The hydrogen maser is kept in a dedicated, environmentally controlled clock room. Measurements of magnetic field, temperature and atmospheric pressure in that room and the maser sensitivities as specified by the manufacturer allow us to exclude any systematic effects on the maser frequency that would exceed the statistical uncertainties above and the systematic uncertainties from temperature variations in the CSO lab.

Our final uncertainties (the error bars in Fig. 5) are the quadratic sums of the statistical uncertainties from the WLS adjustment for each data set and the systematic uncertainties calculated for each data set from (23) and the measured temperature variations. For the global adjustment we average the

systematic uncertainties from the individual data sets obtaining $\pm 1.2 \times 10^{-9}$ on P_{MM} and $\pm 1.9 \times 10^{-7}$ on P_{KT}. Adding these quadratically to the WLS statistical uncertainties of the global adjustment we obtain as our final result $P_{MM} = (1.2 \pm 2.2) \times 10^{-9}$ and $P_{KT} = (1.6 \pm 3.0) \times 10^{-7}$ (1σ uncertainties).

4 Conclusion and Outlook

We have reported on two different tests of the Einstein Equivalence Principle (EP) using the comparison of atomic clocks with different atomic species on one hand, and the comparison of an atomic clock and a cryogenic sapphire cavity oscillator on the other. The two experiments are interpreted as testing Local Position Invariance (LPI) and Local Lorentz Invariance (LLI) respectively which are both constituent elements of the EP.

The test of LPI is based on the comparison of the hyperfine transitions in ^{87}Rb and ^{133}Cs using atomic fountains that presently reach uncertainties of $(6-8)\times 10^{-16}$. Such measurements were repeated over the last 5 years to search for a time variation that would indicate a violation of LPI. Our present results limit a linear variation to $\frac{d}{dt}\ln\left(\frac{\nu_{\mathrm{Rb}}}{\nu_{\mathrm{Cs}}}\right) = (0.2 \pm 7.0)\times 10^{-16}\,\mathrm{yr}^{-1}$ which represents a 5-fold improvement over our previous results [28] and a 100-fold improvement over the Hg^+-H hyperfine energy comparison [10]. When interpreting the results as a limit on the time variation of fundamental constants (c.f. Sect. 2.1) we obtain

$$\left|0.49\frac{\dot{\alpha}}{\alpha} + 0.17\frac{d}{dt}\ln\left(\frac{m_q}{\Lambda_{\mathrm{QCD}}}\right)\right| \leq 7 \times 10^{-16}\,\mathrm{yr}^{-1} \tag{24}$$

By itself this experiment limits the time variation of a combination of two of the three fundamental constants of Sect. 2.1. The $^{199}Hg^+$ to ^{133}Cs comparisons by the NIST group [14] provide

$$\left|6.0\frac{\dot{\alpha}}{\alpha} + 0.1\frac{d}{dt}\ln\left(\frac{m_q}{\Lambda_{\mathrm{QCD}}}\right) + \frac{d}{dt}\ln\left(\frac{m_e}{\Lambda_{\mathrm{QCD}}}\right)\right| \leq 7 \times 10^{-15}\,\mathrm{yr}^{-1} \tag{25}$$

Combining these two results we have two constraints on the variation of the three fundamental constants, with the missing third constraint requiring the comparison over time with a fourth atomic transition (c.f. Sect. 2.1).

The test of LLI is based on the comparison of a hydrogen maser clock to a cryogenic sapphire microwave cavity. This experiment simultaneously constrains two combinations of the three parameters of the Mansouri and Sexl test theory (previously measured individually by Michelson-Morley and Kennedy-Thorndike experiments). We obtain $\delta_{\mathrm{MS}} - \beta_{\mathrm{MS}} + 1/2 = 1.2(1.9)(1.2) \times 10^{-9}$ which is of the same order as the best previous results [38,44], and $\beta_{\mathrm{MS}} - \alpha_{\mathrm{MS}} - 1 = 1.6(2.3)(1.9)\times 10^{-7}$ which improves the best previous limit [41] by a factor of 70 (the first bracket indicates the 1σ uncertainty from statistics the second from systematic effects). We improve our own previous results [42] by about a factor 2 due to more and longer data sets and to improved temperature control of the experiment. We note that our value on $\delta_{\mathrm{MS}} - \beta_{\mathrm{MS}} + 1/2$ is compatible with the slightly significant recent result of [44] who obtained $\delta_{\mathrm{MS}} - \beta_{\mathrm{MS}} + 1/2 = (2.2 \pm 1.5)\times 10^{-9}$.

As a result of our experiment the Lorentz transformations are confirmed in this particular test theory with an overall uncertainty of $\leq 8 \times 10^{-7}$ limited now by the determination of $\alpha_{\rm MS}$ from Doppler and clock comparison experiments [48,49]. This is likely to be improved in the coming years by experiments such as ACES (Atomic Clock Ensemble in Space [63]) that will compare ground clocks to clocks on the international space station aiming at a 10-fold improvement on the determination of $\alpha_{\rm MS}$.

In the future, the two tests of LPI and LLI presented here are expected to further improve due to improvements in the accuracies of the atomic clocks involved and due to new experimental strategies, ultimately leading to space-borne versions of the experiments.

Ongoing efforts are expected to improve the accuracy of both ^{87}Rb and ^{133}Cs to the 10^{-16} level. The corresponding limit to variations of fundamental constants will then be decreased to $\sim 10^{-16}$ yr^{-1} or less. Recent advances in the field of optical frequency metrology will probably lead optical frequency standards to surpass microwave clocks. Comparing such standards to each other will provide very stringent limits to the variation of the fine structure constant α. To keep the constraints to the variation of $m_q/\Lambda_{\rm QCD}$ and $m_e/\Lambda_{\rm QCD}$ at the same level, further efforts and new methods will have to be invented to improve microwave clocks. These tests will also greatly benefit from a new generation of time/frequency transfer at the 10^{-16} level which is currently under development for the ESA space mission ACES which will fly ultra-stable clocks on board the international space station in 2006 [63] and a similar project conducted by NASA. These missions will allow highly precise comparisons between clocks developed in distant laboratories and based on different atomic species and/or different technologies.

Concerning the test of LLI, new proposals have been made to use two orthogonal resonators or two orthogonal modes in the same sapphire resonator placed on a rotating platform [58]. Such a set-up is likely to improve the tests of LLI by several orders of magnitude as the relevant time variations will now be at the rotation frequency ($\approx 0.01 - 0.1$ Hz) which is the range in which such resonators are the most stable (≈ 100 fold better stability). Additionally many systematic effects are likely to cancel between the two orthogonal oscillators or modes and the remaining ones are likely to be less coupled to the rotation frequency than to the sidereal frequencies used in our experiment. Ultimately, it has been proposed [64] to conduct these tests on board of an Earth orbiting satellite, again with a potential gain of several orders of magnitudes over current limits.

References

1. Will C.M., *Theory and Experiment in Gravitational Physics, revised edition*, Cambridge U. Press, (1993).
2. W.J. Marciano, Phys. Rev. Lett. **52**, 489 (1984).
3. T. Damour and A. Polyakov, Nucl. Phys. B **423**, 532 (1994).
4. T. Damour, F. Piazza, and G. Veneziano, Phys. Rev. Lett. **89**, 081601 (2002).
5. A.P. Lightman and D.L. Lee, Phys. Rev. D **8**, 364 (1973).

6. L. Blanchet, Phys. Rev. Lett. **69**, 559 (1992).
7. A. Bauch, S. Weyers, Phys. Rev. D **65**, 081101 (2002).
8. J.K. Webb *et al.*, Phys. Rev. Lett. **87**, 091301 (2001).
9. H. Marion *et al.*, Phys. Rev. Lett **90**, 150801 (2003).
10. J.D. Prestage, R.L. Tjoelker and L. Maleki, Phys. Rev. Lett. **74**, 3511 (1995).
11. D.J. Berkeland *et al.*, Phys. Rev. Lett. **80**, 2089 (1998).
12. M. Niering *et al.*, Phys. Rev. Lett. **84**, 5496 (2000).
13. J. Helmcke *et al*, IEEE Trans. Instrum. Meas. **52**, 250 (2003).
14. S. Bize *et al.*, Phys. Rev. Lett **90**, 150802 (2003).
15. J. Stenger *et al.*, Opt. Lett. **26**, 1589 (2001).
16. E. Peik *et al.*, in *Proc. of the Joint Mtg. IEEE Intl. Freq. Cont. Symp. and EFTF Conf.*, (2003).
17. V.V. Flambaum, ArXiv:physics/0302015 (2003).
18. V.A. Dzuba, V.V. Flambaum and J.K. Webb, Phys. Rev. A **59**, 230 (1999).
19. V.A. Dzuba, V.V. Flambaum and J.K. Webb, Phys. Rev. Lett. **82**, 888 (1999).
20. V.A. Dzuba, V.V. Flambaum and J.K. Webb, Phys. Rev. A **61**, 034502 (2000).
21. S.G. Karshenboim, Can. J. Phys. **47**, 639 (2000).
22. V.A. Dzuba, V.V. Flambaum and M.V. Marchenko, ArXiv:physics/0305066 (2003).
23. I. Courtillot *et al*, Phys. Rev. A **68**, 030501 (2003).
24. H. Katori *et al*, ArXiv:physics/0309043 (2003).
25. M. Takamoto and H. Katori, ArXiv:physics/0309004 (2003).
26. D.J. Wineland *et al.*, in *Proc. of the Joint Mtg. IEEE Intl. Freq. Cont. Symp. and EFTF Conf.*, p 68. (2003).
27. A. Clairon *et al.*, in *Proc. of the 5^{th} Symposium on Frequency Standards and Metrology*, ed. J. Bergquist (World Scientific, Singapore, 1995), p. 49.
28. S. Bize *et al.*, in *Proc. of the 6^{th} Symposium on Frequency Standards and Metrology* (World Scientific, Singapore, 2001), p 53.
29. S. Bize *et al.*, Europhys. Lett. **45**, 558 (1999).
30. E. Simon *et al.*, Phys. Rev. A **57**, 436 (1998).
31. F. Pereira Dos Santos *et al.*, Phys. Rev. Lett. **89**, 233004 (2002).
32. Y. Sortais *et al.*, Phys. Rev. Lett. **85**, 3117 (2000).
33. P. Wolf *et al.*, in *Proc. of the 6^{th} Symposium on Frequency Standards and Metrology* (World Scientific, Singapore, 2001), p 593.
34. V.A. Kostelecky ed., *CPT and Lorentz Symmetry II*, World Scientific, Singapore, (2002).
35. T. Damour, gr-qc/9711060 (1997).
36. D. Colloday and V.A. Kostelecky, Phys. Rev. D **55**, 6760 (1997).
37. R. Bluhm et al., Phys. Rev. Lett. **88**, 090801 (2002).
38. A. Brillet and J.L. Hall, Phys. Rev. Lett. **42**, 549 (1979).
39. R.J. Kennedy and E.M. Thorndike, Phys. Rev. B **42**, 400 (1932).
40. D. Hils and J.L. Hall, Phys. Rev. Lett., **64**, 1697 (1990).
41. C. Braxmaier et al., Phys. Rev. Lett. **88**, 010401 (2002).
42. P. Wolf et al., Phys. Rev. Lett. **90**, 060402 (2003).
43. J.A. Lipa et al. Phys. Rev. Lett. **90**, 060403 (2003).
44. H. Müller et al., Phys. Rev. Lett. **91**, 020401 (2003).
45. H.P. Robertson, Rev. Mod. Phys. **21**, 378 (1949).
46. R. Mansouri and R.U. Sexl, Gen. Rel. Grav. **8**, 497, 515, 809 (1977).
47. W.-T. Ni, Phys. Rev. Lett. **38**, 301 (1977).
48. E. Riis et al., Phys. Rev. Lett. **60**, 81 (1988).

49. P. Wolf and G. Petit, Phys. Rev. **A56**, 4405 (1997).
50. A.V. Kostelecky and M. Mewes, Phys. Rev. D **66**, 056005 (2002).
51. D.J. Fixsen et al., Phys. Rev. Lett. **50**, 620 (1983).
52. P.M. Lubin et al., Phys. Rev. Lett. **50**, 616 (1983).
53. A.A. Michelson and E.W. Morley, Am. J. Sci. **34**, 333 (1887).
54. R. Grieser et al., Appl. Phys. B **59**, 127 (1994).
55. S. Chang, A.G. Mann and A.N. Luiten, Electron. Lett. **36**, 480 (2000).
56. A.G. Mann, S. Chang and A.N. Luiten, IEEE Trans. Instrum. Meas. **50**, 519 (2001).
57. We assume here that local position invariance (cf. Sect. 2) is sufficiently verified so that the variation of the maser frequency due to the diurnal variation of the local gravitational potential is negligible. Indeed the results of [7] imply that such variations should not exceed 2 parts in 10^{-17} which is significantly below our noise level.
58. M.E. Tobar et al., Phys. Lett. A **300**, 33 (2002).
59. H. Müller et al., Phys. Rev. D **67**, 056006 (2003).
60. A.V. Kostelecky and C.D. Lane, Phys. Rev. D **60**, 116010 (1999).
61. L.S. Schmidt, Metrologia **40**, in press, (2003).
62. N.R. Draper and H. Smith, *Applied Regression Analysis*, Wiley, (1966).
63. C. Salomon, et al., C.R. Acad. Sci. Paris **2**, 1313 (2001).
64. C. Lämmerzahl et al., Class. Quant. Grav. **18**, 2499 (2001).

Precision Spectroscopy of Atomic Hydrogen and Variations of Fundamental Constants

M. Fischer[1], N. Kolachevsky[1,2], M. Zimmermann[1], R. Holzwarth[1], Th. Udem[1], T.W. Hänsch[1,3], M. Abgrall[4], J. Grünert[4], I. Maksimovic[4], S. Bize[4], H. Marion[4], F. Pereira Dos Santos[4], P. Lemonde[4], G. Santarelli[4], P. Laurent[4], A. Clairon[4], and C. Salomon[5]

[1] Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Straße 1, 85748 Garching, Germany
[2] P.N. Lebedev Physics Institute, Moscow, Russia
[3] Ludwig-Maximilians-University, Munich, Germany
[4] BNM-SYRTE, Observatoire de Paris, 61 Avenue de l'Observatoire, 75014 Paris, France
[5] Laboratoire Kastler Brossel, ENS, 24 rue Lhomond, 75005 Paris, France

Abstract. In 2003 we have measured the absolute frequency of the $(1S, F = 1, m_F = \pm 1) \rightarrow (2S, F' = 1, m'_F = \pm 1)$ two-photon transition in atomic hydrogen. By comparison with the earlier measurement in 1999 we can set an upper limit on its variation of (-29 ± 57) Hz within 44 months. We have combined this result with recently published results of optical transition frequency measurements in the $^{199}Hg^+$ ion and comparison between clocks based on ^{87}Rb and ^{133}Cs. From this combination we deduce the limits for fractional time variations of the fine structure constant $\dot{\alpha}/\alpha = \partial/\partial t(\ln \alpha) = (-0.9 \pm 2.9) \times 10^{-15}$ yr^{-1} and for the ratio of ^{87}Rb and ^{133}Cs nuclear magnetic moments $\partial/\partial t(\ln[\mu_{\mathrm{Rb}}/\mu_{\mathrm{Cs}}]) = (-0.5 \pm 1.7) \times 10^{-15}$ yr^{-1}. This is the first precise restriction for the fractional time variation of α made without assumptions about the relative drifts of the constants of electromagnetic, strong and weak interactions.

1 Introduction

The question of constancy of fundamental constants was first raised in Dirac's "Large Number hypothesis" (1937) which aimed for a harmonization of basic laws of physics [1]. Since then, this hypothesis has been reviewed and extended by many other scientists opening a broad field of theoretical and experimental investigations. As there is no accepted theory predicting the values of fundamental constants, the question of their possible time variation belongs mostly to the field of experimental physics. The last decades saw a number of different astrophysical, geological, and laboratory tests searching for their possible variation in different time epochs with an ever increasing accuracy. From the point of view of its importance for physics in general, this problem stays at the same level as the test of CPT-symmetry and the search for an electric dipole moment of elementary particles.

In all metric theories of gravity including general relativity any drift of non-gravitational constants is forbidden. This statement bases on Einstein's Equivalence Principle (EEP) postulating that (i) the weight of a body is proportional

M. Fischer et al., Precision Spectroscopy of Atomic Hydrogen and Variations of Fundamental Constants, Lect. Notes Phys. **648**, 209–227 (2004)
http://www.springerlink.com/

to its mass, (ii) the result of any non-gravitational measurement is independent of the velocity of the laboratory rest-frame (local Lorentz invariance), and (iii) the result of a non-gravitational measurement is independent of its time and position in this frame (local time and position invariance). On the other hand, theories towards a unified description of quantum mechanics and gravity allow for, or even predict some violations of EEP [2]. In this sense, any experimental search for a drift of fundamental constants tests the validity of EEP as well as it provides important constraints on new theoretical models.

The basic principle of all tests of the stability of fundamental constants is the investigation of time variations of some stable physical value Θ. Usually, Θ is a dimensionless value which can be the ratio of reaction cross-sections, the distances, masses, magnetic moments, frequencies and so on. In an experiment one measures the value Θ at two different times t_1 and t_2 and compares $\Theta(t_1)$ with $\Theta(t_2)$. The value of Θ may depend on a number of fundamental constants α_i $(i = 1, \ldots, n)$ and the conclusion about drifts of α_i originate from the analysis of $\Theta(t_1) - \Theta(t_2)$. The functional connection between Θ and α_i can include rather complicated theoretical models and assumptions which make the results somehow unclear and strongly model-dependent. Even if the dependence $\Theta(\alpha_i)$ is straightforward, it is difficult to separate the contributions from individual α_i drifts if $n > 1$. As mentioned in [3], all the relative drifts of fundamental constants, if existing, should be on the same order of magnitude which can result in a cancelation of the drift of Θ as well as in its amplification. For example, according to an elaborate scenario in the framework of a Grand Unification Theory, the fractional time variation of hadron masses and their magnetic moments should change about 38 times faster than the fractional time variation of the fine structure constant α [4].

Astrophysical and geological methods test the stability of fundamental constants over very long time intervals of 1–10 Gyr. Due to the large difference of $|t_1 - t_2|$, the sensitivity of these methods to a monotonic long-time drift is very high but they are insensitive to more rapid fluctuations. A recent analysis of quasar absorption spectra by Murphy *et al.* with redshifted UV transition lines indicates a variation of α on the level of $\Delta\alpha/\alpha = (-0.54 \pm 0.12) \times 10^{-5}$ in the first half of the evolution of the universe (5–11 Gyr ago) [5]. There are also indications that in this period the electron to proton mass ratio was different from its contemporary value on the same level of 10^{-5} [6]. The analysis of astrophysical data requires a number of model assumptions which include not only the well-established scenarios of the evolution of the universe, but also assumptions about the isotopic abundance in interstellar gas clouds, the presence of magnetic fields and others (see e.g. the review [7]) which are difficult to prove. More recent observation of quasar absorption spectra, performed by different groups, seem to rule out a variation of α on the level observed by Murphy *et al.* [8,9].

A very stringent limit for the time variation of α on geological timescales follows from the analysis of isotope abundance ratios in the natural fission reactor of Oklo, Gabon, which operated about 2 Gyr ago. A recent re-analysis of the data of the $^{149}\mathrm{Sm}/^{147}\mathrm{Sm}$ isotope abundance ratio sets a limit of $\Delta\alpha/\alpha = (-0.36 \pm 1.44) \times 10^{-8}$ [10]. The interpretation of the data is not unambiguous, as the

result strongly depends on reactor operating conditions which are not exactly known. Selecting another possible reaction branch yields a value of $\Delta\alpha/\alpha = (9.8 \pm 0.8) \times 10^{-8}$ [10]. In contrast to the first one, this result indicates a non-zero drift.

Laboratory experiments are sensitive to variations of fundamental constants during the last few years and typically base on precise frequency measurements in atomic or molecular systems. In comparison to astrophysical and geological ones, laboratory measurements considerably win in relative accuracy which, in spite of much shorter $|t_1 - t_2|$ time intervals, leads to a competitive sensitivity on drifts. Moreover, in this case systematic effects can be well controlled and the dependence of the transition frequencies on fundamental constants is straightforward.

Any absolute frequency measurement of some transition in an atomic system is a comparison of this frequency with the frequency of the ground state hyperfine transition of ^{133}Cs. Such a measurement of *one* transition frequency in *one* atomic system imposes a limit on the variation of some simple combinations of α, nuclear and electron magnetic moments and/or their masses [3]. To separate the drift of the fine-structure constant one needs either to impose some restricting assumptions on the time dependence of the coupling constants of the strong (α_S) and electroweak (α_W) interaction [4,11] or make absolute measurements of *two or more* transition frequencies possessing *different functional dependencies* on the fundamental constants. The second method does not include any model parameters or additional assumptions which favorably distinguishes it. It is also possible to make such a model-independent evaluation by directly comparing e.g. gross- and fine structure or two gross-structure frequencies without comparison to a primary frequency standard and thus excluding the corresponding dependence on the nuclear magnetic moments. To our knowledge, such laboratory experiments still have not been done with a level of accuracy competing modern absolute frequency measurements.

In this work we deduce separate stringent limits for the relative drifts of the fine structure constant α and the ratio μ_{Cs}/μ_B by combining the results of two optical frequency measurements in the hydrogen atom and in the mercury ion relative to the ground state hyperfine splitting of ^{133}Cs. The measurements of the hydrogen transition frequency have been carried out at MPQ, Garching, Germany and are described below. The experiments on the drift of the $5d^{10}6s\ ^2S_{1/2}(F=0) \rightarrow 5d^9 6s^2\ ^2D_{5/2}(F'=2, m'_F=0)$ electric quadrupole transition frequency ν_{Hg} in $^{199}\mathrm{Hg}^+$ have been performed by the group of J. Bergquist at NIST, Boulder CO, USA between July 2000 and December 2002. They are described in detail elsewhere [12].

From 1999 to 2003, the ratio of the ground state hyperfine splittings of ^{87}Rb and ^{133}Cs has been determined from a comparison between several simultaneously running atomic fountain clocks in BNM-SYRTE and ENS, Paris, France [13]. Using this result, we can also set a limit for the fractional time variation of the Rb and Cs nuclear magnetic moment ratio $\mu_{\mathrm{Rb}}/\mu_{\mathrm{Cs}}$.

As the measurements were performed at different places and at different times we have to use the hypothesis, that the results are independent of the place on

the Earth's orbit, at least within the last 4 years. In other words, we have to assume a validity of local Lorentz invariance (LLI) and local position invariance (LPI) as well as to make the additional hypothesis, that the constants change on a cosmological time scale and do not oscillate within a few years (linear drifts). With this exceptions, our results are independent of any further model assumptions like any form of correlation between the constants or constancy of a particular set of constants.

2 Hydrogen Spectrometer

In 1999 [14] and 2003, the frequency of the $(1S, F = 1, m_F = \pm 1) \rightarrow (2S, F' = 1, m'_F = \pm 1)$ two-photon transition in atomic hydrogen has been phase coherently compared to the frequency of the ground state hyperfine splitting in ^{133}Cs using a high-resolution hydrogen spectrometer and a frequency comb technique [15]. In 1999, the accuracy of the evaluation of the transition frequency was 1.8×10^{-14}. The setup of the hydrogen spectrometer used during this measurement has been described previously in [16]. We have introduced a number of improvements in the spectroscopic setup which will be described in the following. A sketch of the actual setup is shown in Fig. 1.

A cw dye laser emitting near 486 nm is locked to an external reference cavity. The cavity used during the 1999 measurement was made from Zerodur and had a typical drift of 25 Hz s^{-1} at the fundamental frequency. The new cavity made from Ultra Low Expansion (ULE) glass for the 2003 measurement is better

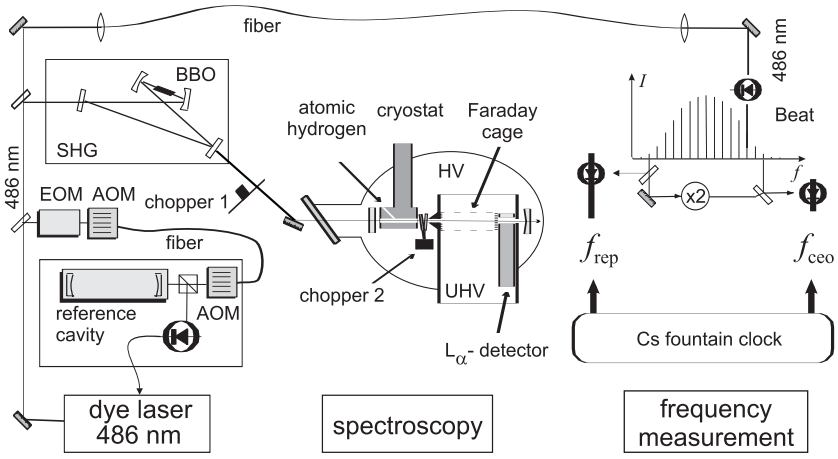


Fig. 1. Experimental setup for the comparison of the hydrogen $1S$–$2S$ transition frequency with a primary frequency standard. The 486 nm light is doubled in a Barium β-Borate crystal (BBO) in the second harmonic generation (SHG) stage. Resulting radiation is coupled to a linear enhancement cavity in a vacuum chamber with the pressure of about 10^{-5} mbar (HV), while the excitation and detection take place in an ultra-high vacuum (UHV) zone at the pressure of 10^{-8}–10^{-7} mbar. EOM and AOM denote electro- and acousto-optical modulators correspondingly.

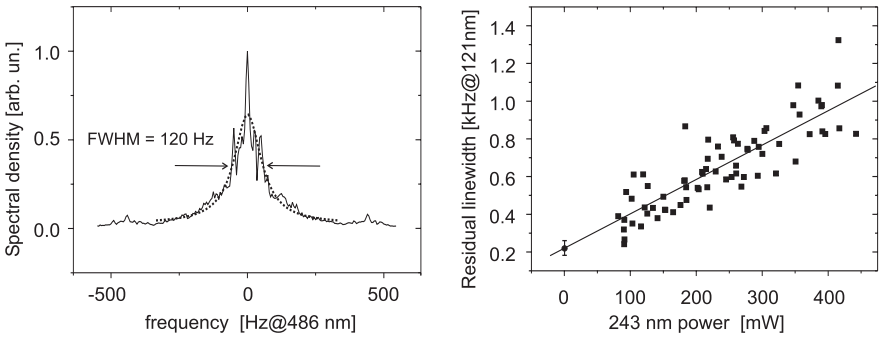


Fig. 2. (Left) spectrum of the beat signal between laser fields locked to two independent cavities. (Right) extrapolation of the ionization broadening of the $1S$–$2S$ transition spectra to zero excitation power circulating in the enhancement cavity.

shielded against the environment. Its drift has been less than 0.5 Hz s^{-1} for the entire time of the measurement. Due to the better thermal and acoustic isolation and improvements in the laser locking electronics, the laser linewidth is narrower than it has been in 1999. An upper limit for the laser linewidth has been deduced from an investigation of the beat signal between two laser fields locked separately to independent Zerodur and ULE cavities. The spectrum of 12 averaged scans, each taken in 0.2 s is represented in Fig. 2 (left). The width of this beat signal spectrum is about 120 Hz at a laser wavelength of 486 nm. Yet it is impossible to distinguish between the individual noise contributions from the two independent, but not equivalent cavities. Another restriction can be deduced from the analysis of the $1S$–$2S$ transition spectra. The linewidth of the transition is mainly defined by time-of-flight broadening and is between 1 kHz and 5 kHz at 121 nm. The residual linewidth obtained after subtracting the estimated contribution of time-of flight broadening is plotted on Fig. 2 (right) versus the excitation light power. The observed broadening is due to the ionization processes and corresponds to a reduced lifetime of the metastable excited atoms. Extrapolating the residual linewidth to zero intensity, we get 240(30) Hz at 121 nm. This can be considered as a contribution from laser frequency fluctuations. Thus, we evaluate the 486 nm laser linewidth as 60 Hz for averaging times of 0.5 s.

A small part of the laser light is transferred to the neighboring laboratory via a single mode fiber where its absolute frequency can be measured. The main part is frequency doubled in a BBO crystal. For higher conversion efficiency, the crystal is placed in a folded enhancement cavity. The resulting 20 mW of radiation near 243 nm (corresponding to half of the $1S$–$2S$ transition frequency) is coupled into a linear enhancement cavity inside the vacuum chamber of the hydrogen spectrometer.

Molecular hydrogen is dissociated in a 15 W, 2.5 GHz radio-frequency gas discharge. The resulting flow of atomic hydrogen is cooled by inelastic collisions with the walls of a copper nozzle having the temperature of 5–7 K. The nozzle forms a beam of cold atomic hydrogen which leaves the nozzle collinearly

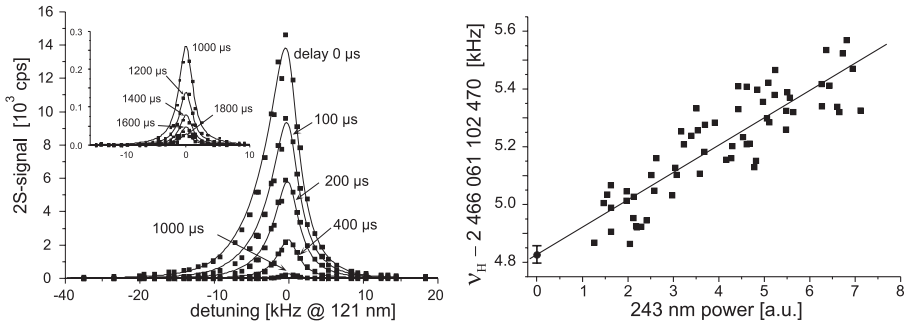


Fig. 3. (Left) simultaneous fit of a $1S$–$2S$ transition spectrum recorded at different delays Δt. The nozzle temperature was equal to 7 K. (Right) AC Stark shift extrapolation.

with the cavity axis and enters the interaction region between the nozzle and the L_α-detector. This region is shielded from stray electric fields by a Faraday cage. Some of the atoms are excited from the ground state to the metastable $2S$ state by Doppler-free absorption of two counter-propagating photons from the laser field in the enhancement cavity. After the 1999 measurement which had been performed at a background gas pressure of around 10^{-6} mbar in the interaction region, we have upgraded the vacuum system to a differential pumping configuration. This allows us to vary the background gas pressure between 10^{-8} and 10^{-7} mbar in 2003 and to reduce the background gas pressure shift and the corresponding uncertainty down to 2 Hz.

Due to small apertures, only atoms flying close to the cavity axis can enter the detection region where the $2S$ atoms are quenched in a small electric field and emit L_α-photons. The excitation light and the hydrogen beam are periodically blocked by two phase locked choppers operating at 160 Hz frequency and the L_α-photons are counted time-resolved only in the dark period of the cycle. This eliminates background counts from the excitation light. The delay Δt between blocking the 243 nm radiation and the start of counting sets an upper limit on the velocity of the atoms which contribute to the signal. For some definite Δt only atoms with velocities $v < d/\Delta t$ are selected, where d is the distance between nozzle and detector. Therefore, velocity dependent systematic effects such as the second-order Doppler shift and the time-of-flight broadening are smaller for spectra recorded at larger Δt. The hydrogen beam is blocked by a fork chopper in less then 200 μs after the blocking of the excitation light to prevent slow atoms from being blown away by fast atoms that emerge subsequently from the nozzle. With the help of a multi-channel scaler, we count all photons and sort them into 12 equidistant time bins. From each scan of the laser frequency over the hydrogen $1S$–$2S$ resonance we therefore get 12 spectra at different delays. To correct for the second order Doppler shift, we use an elaborated theoretical model [16] to fit all the delayed spectra of one scan simultaneously with one set of 7 fit parameters (see Fig. 3). The result of the fitting procedure is the $1S$–$2S$ transition frequency for the hydrogen atom at rest.

Besides the second order Doppler effect, the other dominating systematic effect is the dynamic AC Stark shift which shifts the transition frequency linearly with the excitation light intensity. We have varied the intensity and extrapolate the transition frequency to zero intensity to correct for it [14]. A typical set of data taken within one day of measurement in 2003 and the corresponding extrapolation is presented in Fig. 3 (right).

3 Frequency Measurement

For an absolute measurement of the $1S$–$2S$ transition frequency in units of Hz, the frequency of the dye laser near 616.5 THz (486 nm) was phase coherently compared with a cesium fountain clock [13]. To bridge the large gap between the optical- and radio-frequency (RF) domain we took advantage of the recently developed femtosecond laser frequency comb technique incorporating a highly nonlinear glass fiber, which allows for a further simplification of the experimental setup as compared to the measurement performed in 1999. In this section we give an introduction of the frequency comb technique and a description of the experimental setup, which was used for $1S$–$2S$ frequency measurement in 2003.

The pulse train emitted by a sufficiently stable mode locked femtosecond (fs) laser equals a comb of cw laser modes in the frequency domain. The frequency of each mode of this comb can be written as $f_n = nf_{\mathrm{rep}} + f_{\mathrm{ceo}}$, where f_{rep} is the pulse repetition rate of the fs laser, n is an integer number and f_{ceo} is the so-called carrier envelope offset frequency [17].

The fs laser emits a train of pulses with a repetition rate $f_{\mathrm{rep}} = 1/T$, where T is the time between consecutive pulses. The envelope function of the pulses has the periodicity of f_{rep}, but it does not necessarily mean that the electrical field of the pulses has the same periodicity. The pulses have identical field transients only when the laser cavity roundtrip phase delay of the fs laser pulse equals the group delay (Fig. 4 top). In this case not only the envelope function but also the electrical field has the periodicity of f_{rep}. This leads to a Fourier spectrum $f_n = nf_{\mathrm{rep}}$, where all the modes are exact multiples of f_{rep}. Generally, the group delay does not equal the phase delay inside the cavity and the frequencies f_n cannot be integer multiples of f_{rep} (Fig. 4 bottom). Denoting the phase shift between the envelope function and the carrier frequency of consecutive pulses as $\Delta\varphi$ one can show, that the frequencies can be written as

$$\begin{aligned} f_n &= nf_{\mathrm{rep}} + f_{\mathrm{ceo}} \qquad \text{with} \\ f_{\mathrm{ceo}} &= \frac{\Delta\varphi}{2\pi} f_{\mathrm{rep}}\,, \qquad f_{\mathrm{ceo}} < f_{\mathrm{rep}}\,. \end{aligned} \tag{1}$$

If f_{rep} and f_{ceo} are fixed, all the modes of the frequency comb are determined in their frequency and can be used for measuring the frequency of cw laser light via beat notes between the cw laser light and a nearby comb mode. The large gap between the RF and the optical domain is bridged due to the fact that n is a large integer number of the order of 10^6. To use the frequency comb for high precision optical frequency measurements one has to link f_{rep} and f_{ceo} phase coherently to

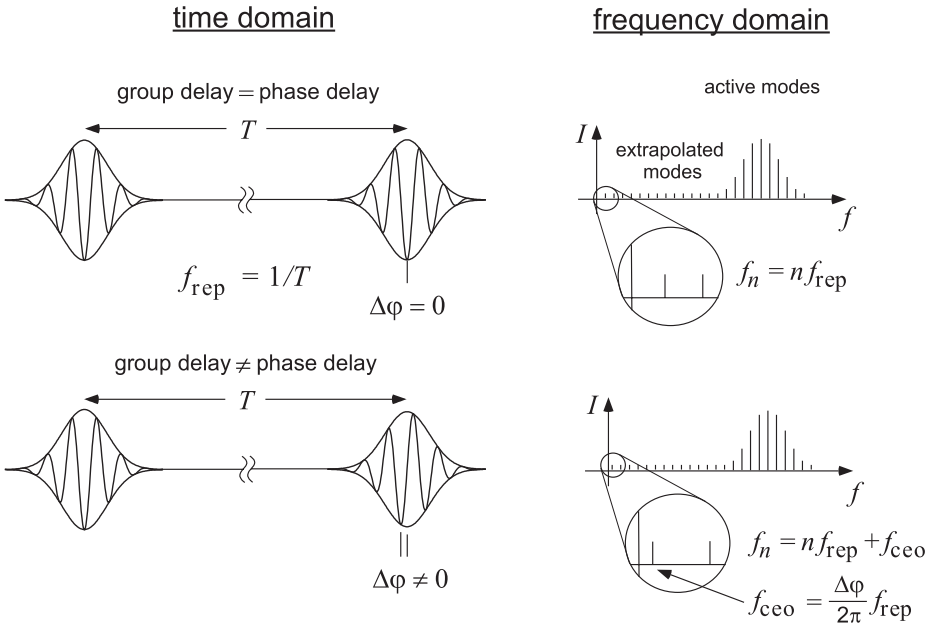


Fig. 4. Time- and frequency domain representation of a pulse train emitted by a mode-locked laser. If the phase delay is different from the group delay inside the laser cavity, this leads to the so-called carrier envelope offset frequency f_{ceo}, which shifts the frequency comb as a whole.

a Cs primary frequency standard. The Cs clock provides us with an extremely precise reference frequency to control f_{rep} and f_{ceo}. The pulse repetition rate f_{rep} is easily measured with a photodiode and controlled via the length of the fs laser cavity, which can be changed by means of a piezo-mounted cavity mirror. In general, f_{ceo} can be controlled by adjusting the pump power of the fs laser [18,19]. In the case of a linear laser cavity with a prism pair to compensate for the group velocity dispersion, f_{ceo} can also be *controlled* by tilting the end mirror of the dispersive arm of the laser cavity [17]. The challenging problem for some time was to *measure* f_{ceo}. If the spectrum of the optical frequency comb covers an entire octave, f_{ceo} is most conveniently determined by frequency doubling the mode f_n on the low frequency side of the comb spectrum and comparing the result with the mode f_{2n} on the high frequency side via a beat note measurement [19,20]:

$$2f_n - f_{2n} = 2(nf_{\mathrm{rep}} + f_{\mathrm{ceo}}) - (2nf_{\mathrm{rep}} + f_{\mathrm{ceo}}) = f_{\mathrm{ceo}}\,. \tag{2}$$

If the spectrum does not cover an entire octave, one can alternatively compare $3.5f_{8n}$ and $4f_{7n}$ to get $\frac{1}{2}f_{\mathrm{ceo}}$ [15,17] or $3f_{2n}$ with $2f_{3n}$ to obtain f_{ceo} [21–23]. The broad spectra needed for this technique are either directly emitted by the fs laser [24,25] or can be obtained by external broadening in a highly nonlinear medium such as a photonic crystal fiber (PCF) [26,27]. A PCF as pictured in Fig. 5 can be designed to have zero group velocity dispersion (GVD) at 800 nm, which

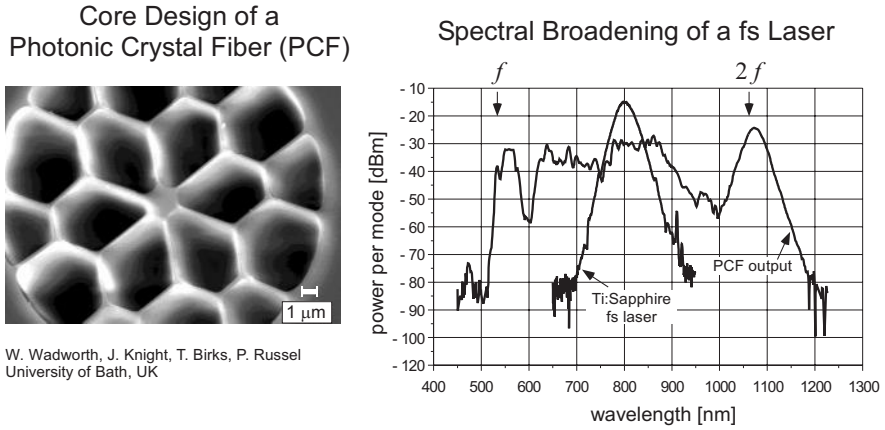


Fig. 5. Core design of a photonic crystal fiber (PCF) and spectral broadening of a fs laser. The PCF was seeded with 20 fs pulses and the average output power of the PCF was 180 mW.

is the central wavelength of commonly used Ti:sapphire fs lasers. Due to the vanishing GVD the pulse spreading within the PCF is lower than in usual single mode fibers. The resulting high peak intensity leads to self phase modulation and therefore efficient broadening of the initial frequency comb.

If f_{rep} and f_{ceo} are stabilized by phase coherently linking them to a RF reference, the accuracy of the RF reference is in one step transferred to all cw modes of the octave spanning optical frequency comb. Using state-of-the-art Cs fountain clocks, which already reach accuracies of 10^{-15} [13], the frequency of an unknown light field can in principle be measured with the same level of accuracy. The fs frequency comb technique was tested to be accurate at the $< 10^{-16}$ level by comparing two independent systems [19,28]. To determine an optical frequency f_{opt} of the unknown light field one needs to measure the frequency f_{beat} of the beat note between the unknown light field and the neighboring mode f_n of the frequency comb. The unknown frequency f_{opt} can then be written as

$$f_{\mathrm{opt}} = f_n + f_{\mathrm{beat}} = nf_{\mathrm{rep}} + f_{\mathrm{ceo}} + f_{\mathrm{beat}} \ . \tag{3}$$

The mode number n may be determined by a coarse measurement of f_{opt} with a commercial wavemeter. Using the fs frequency comb technique optical frequency measurements have been carried out on atoms and ions, demonstrating accuracies of up to 10^{-14} [14,29–31]. An experimental setup for detecting f_{ceo}, f_{rep} and f_{beat} using an octave broad frequency comb is shown in Fig. 6.

Another application of fs frequency combs is the determination of optical frequency ratios. As a frequency is dimensionless, no RF reference based on Cs is needed and one can take advantage of the high stability and accuracy of optical frequency standards, which should lead to an increased sensitivity to the drift of fundamental constants [31].

Due to the invention of photonic crystal fibers the complexity of the frequency measurement in 2003 has been considerably reduced as compared to the 1999

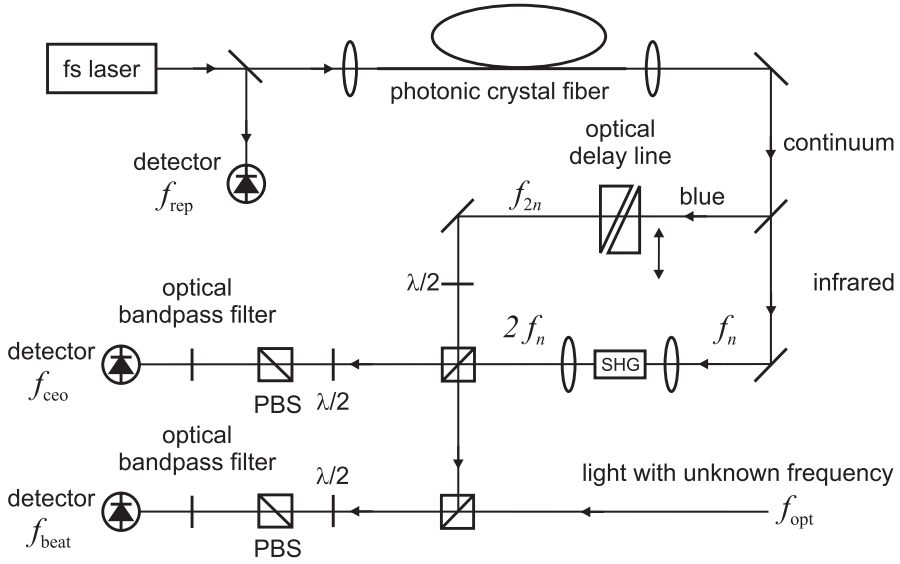


Fig. 6. Experimental setup for detecting f_{rep}, f_{ceo} and f_{beat}. An optical delay line is inserted into the "blue" arm of the nonlinear interferometer to match the optical path lengths. PBS denotes a polarizing beam splitter.

experiment, where a fs laser was already in use. The experimental setup used in 2003 to measure the frequency of the hydrogen spectroscopy dye laser was equivalent to that shown in Fig. 6 and employed a fs Ti:sapphire ring laser (GigaOptics, model GigaJet) with 800 MHz repetition rate. The spectrum of the fs laser was externally broadened with the help of a PCF to more than one octave including light from 946 nm to 473 nm. The detection of the repetition rate f_{rep} was placed in front of the microstructured fiber to not be affected by amplitude noise caused by imperfect fiber coupling. f_{rep} was phase locked to a 800 MHz signal which was directly derived from the transportable Cs fountain clock FOM. For both the 1999 and 2003 measurements, the transportable Cs fountain clock FOM has been installed at MPQ. Its instability is $1.8\times10^{-13}\tau^{-1/2}$ and its accuracy has been evaluated to be 8×10^{-16} [32] at BNM-SYRTE. During the experiments in Garching, only a verification at the level of 10^{-15} has been performed. Consequently we attribute a conservative FOM accuracy of 2×10^{-15} for these measurements.

To check for possible cycle slipping, the phase locked frequencies f_{ceo} and f_{rep} were additionally counted to verify consistency. The 486 nm dye laser and the blue part of the frequency comb were spatially overlapped, optically filtered around 486 nm, and directed onto an avalanche photodiode to measure the beat frequency with the neighboring mode of the frequency comb. The detected beat note was filtered, amplified and directed to three radio frequency counters (Hewlett Packard, models 53131A and 53132A) utilizing different detection bandwidth and power level. All counters were referenced to the Cs clock. To

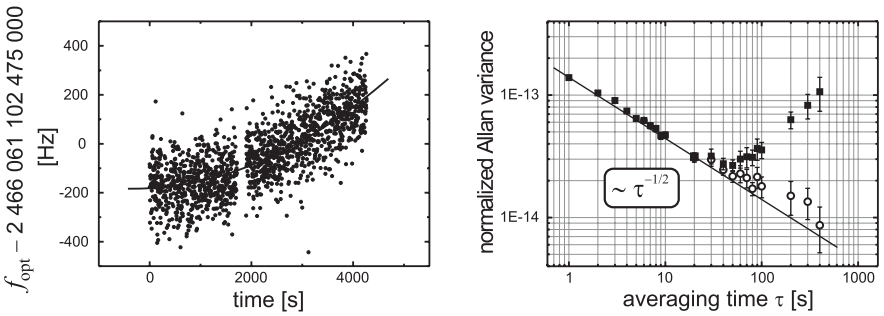


Fig. 7. (Left) beat frequency of the 486 nm dye laser relative to the stabilized fs frequency comb. The solid line is a parabolic fit to the data. (Right) normalized Allan variance vs. averaging time computed from a time series of 1 second counter readings with a considerable dead time. The straight line indicates the $\tau^{-1/2}$ dependence, which is the signature of the Cs fountain clock. The raw data analysis (squares) shows that the stability for averaging times longer than 20 s is limited by the drift of the ULE reference cavity. Open circles represent data corrected for the parabolic cavity drift.

check for errors in the counting process only data points were accepted where all three counter readings were consistent with each other. Additionally it was verified that the dye laser was successfully locked to the ULE reference cavity during the measurement time.

Figure 7 shows a typical beat note measurement (left) and the corresponding normalized Allan variance (right) of the dye laser locked to the reference cavity relative to the fs frequency comb which was locked to FOM. For longer averaging times, the plot of the Allan variance is generated by juxtaposing 1-s counter readouts. Whereas it is known that such a procedure can alter the functional dependence of the Allan variance [33], white frequency noise, as produced by the Cs fountain, is immune to this form of bias. The observed $\tau^{-1/2}$ dependence coincides with the independently measured fountain clock instability for averaging times shorter than $\approx$ 10 s. The short term stability of the laser system is better than the stability of the fountain clock. However, the long term stability is limited by the drift of the ULE reference cavity.

To compensate for the slow ULE cavity drift we fit a second-order polynomial to the measured beat note before averaging which significantly reduces the Allan variance for longer averaging times. To accurately determine the frequency of the dye laser, we first average the frequency of the ULE cavity with a polynomial such as the one shown on the left side of Fig. 6 with the consistent counter readout. Then we use this polynomial and the recorded AOM readings for each data point, that determine the cavity-laser detuning, to derive a highly stable value for the laser frequency. For the given stability of the Cs fountain clock and the cavity, the optimum record length is around 500 s. For longer averaging times the Cs fountain is more stable than the drift-corrected ULE cavity.

We have measured the $1S$–$2S$ transition in atomic hydrogen during 10 days in 1999 and during 12 days in 2003. Both data sets have been analyzed using

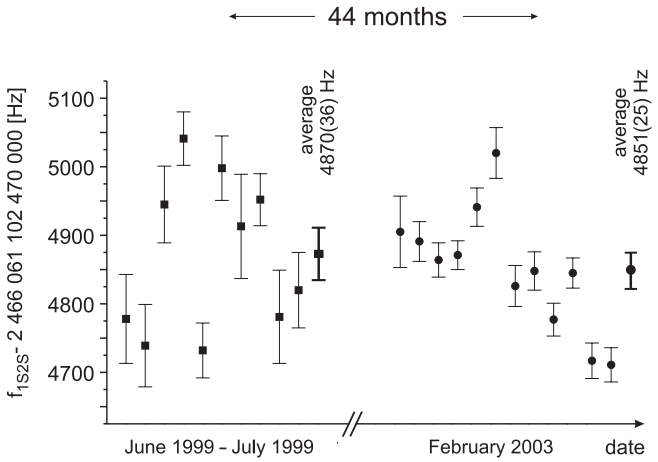


Fig. 8. Experimental results and averages for the 1999 and 2003 measurements of the ($1S$, $F = 1$, $m_F = \pm 1 \rightarrow 2S$, $F' = 1$, $m'_F = \pm 1$) transition frequency in atomic hydrogen.

the same theoretical line shape model and are therefore comparable. In Fig. 8, the results of the extrapolation to zero excitation light intensity and the respective statistical error bars for each day are presented. Since 1999, the statistical uncertainty for each day of measurement was significantly reduced due to the narrower laser linewidth and better signal-to-noise ratio, but the scatter of the day averages did not reduce accordingly. We have tested several possible reasons for this additional scatter including an intra-beam pressure shift, a background gas pressure shift, Stark effects due to the RF gas discharge, and DC Stark shift and have been able to exclude all these effects at least on a conservative level of 10–20 Hz. A possible origin of the observed scatter can be due to a residual first order Doppler effect arising from a violation of the axial symmetry of the enhancement cavity mode and the hydrogen atomic beam. The scattering of the excitation light on intra-cavity diaphragms can also cause slight changes of the field distribution and the corresponding first order Doppler effect. However, it should average to zero over multiple adjustments of the hydrogen spectrometer because the shifts can have both signs. As the scatter is the same for both the measurement sets, we believe them to be equivalent. The main statistical and systematic uncertainties of these measurements are collected in Table 1. The averaging of the 1999 and 2003 daily data points was performed without weighting them.[1] For both measurements the dominating resulting uncertainty arises from the day-to-day scatter, while the pure statistical uncertainty for each day is significantly smaller. In fact, weighting of the day data only slightly influences the results (on the level of $\sigma/2$).

[1] The result of 2 466 061 102 474 870 Hz was inadvertently described in [14] as "the weighted mean value" but was calculated without consideration of the daily statistical uncertainties.

Table 1. Results of the ($1S$, $F = 1$, $m_F = \pm 1 \rightarrow 2S$, $F' = 1$, $m'_F = \pm 1$) transition frequency measurement ($\nu_{H,1999}$, $\nu_{H,2003}$) and uncertainty budgets ($\sigma_{H,1999}$, $\sigma_{H,2003}$) for the 1999 and 2003 measurements correspondingly.

Contribution	$\nu_{H,1999}$ [Hz]	$\sigma_{H,1999}$ [Hz]	$\nu_{H,2003}$ [Hz]	$\sigma_{H,2003}$ [Hz]
Extrapolated value − 2 466 061 102 474 kHz	870	36	851	25
Background gas pressure shift	10	10	0	2
Intra-beam pressure shift	0	10	0	10
Lineshape model	0	20	0	20
DC Stark shift	0	5	0	5
Blackbody radiation	0	1	0	1
Standing wave effects	0	10	0	1
Intensity zero uncertainty	0	1	0	0
Fountain clock uncertainty	0	5	0	5
Total − 2 466 061 102 474 kHz	880	45	851	34

Comparing both measurements we deduce a difference of $\nu_{H,2003} - \nu_{H,1999}$ equal to (-29 ± 57) Hz within 44 months. This corresponds to a relative drift of ν_{H} against the ^{133}Cs ground state hyperfine splitting of $\partial_t(\ln(\nu_{\mathrm{Cs}}/\nu_{\mathrm{H}})) = (3.2 \pm 6.3) \times 10^{-15}$ per year.

4 Determination of Drift Rates

Despite the high sensitivities (less than 10^{-14} yr^{-1}), the accuracy of transition frequency drift measurements are rather low (uncertainty is typically over 100%), so that only the first order expansion in terms of the constants involved in the evaluation is sufficient. The frequency of any optical transition can be written as

$$\nu = const\, Ry\, F_{\mathrm{rel}}(\alpha), \tag{4}$$

where Ry is the Rydberg frequency expressed in hertz and $F_{\mathrm{rel}}(\alpha)$ takes into account relativistic and many-body effects. The Rydberg energy cancels in atomic frequency comparisons. Therefore the dependence of Ry on α ($Ry \sim \alpha^2$) and other fundamental constants contained in Ry is irrelevant[2]. The relativistic correction F_{rel} depends on the transition in the system considered and embodies additional dependence on α, while *const* is a numerical factor and is independent of any fundamental constants.

The frequency

$$\nu_{\mathrm{Hg}} = 1\ 064\ 721\ 609\ 899\ 143.7(10)\ \mathrm{Hz} \tag{5}$$

[2] The expansion of ν in terms of small changes of α as given in [34] are said to be derived assuming the constancy of the Rydberg frequency. However, no such restriction on the unit of frequency is necessary here, as any chosen unit will cancel out in the final result since only frequency ratios are used.

of the $5d^{10}6s\ ^2S_{1/2}(F=0) \rightarrow 5d^9 6s^2\ ^2D_{5/2}(F'=2, m'_F=0)$ electric quadrupole transition in $^{199}Hg^+$ was precisely measured at NIST between the years 2000 and 2002 [12]. Numerical calculations including relativistic and many-body effects for the dependence of $F_{\mathrm{rel,Hg}}(\alpha)$ for ν_{Hg} on the fine structure constant α yield [34]

$$\alpha \frac{\partial}{\partial \alpha} \ln F_{\mathrm{rel,Hg}}(\alpha) \approx -3.2\,. \tag{6}$$

In the light hydrogen atom, the relativistic correction for ν_{H} nearly vanishes ($F_{\mathrm{rel,H}}(\alpha) \approx$ const.):

$$\alpha \frac{\partial}{\partial \alpha} \ln F_{\mathrm{rel,H}}(\alpha) \approx 0 \tag{7}$$

or

$$\nu_{\mathrm{H}} \sim Ry\,. \tag{8}$$

The frequency of hyperfine transitions have a different functional dependence on α. For the ground state hyperfine transition in ^{133}Cs we have

$$\nu_{\mathrm{Cs}} = const'\ Ry\ \alpha^2\ \frac{\mu_{\mathrm{Cs}}}{\mu_B}\ F_{\mathrm{rel,Cs}}(\alpha) \tag{9}$$

with a relativistic correction $F_{\mathrm{rel,Cs}}(\alpha)$ of [34]

$$\alpha \frac{\partial}{\partial \alpha} \ln F_{\mathrm{rel,Cs}}(\alpha) \approx +0.8\,. \tag{10}$$

Combining these equations, we find that the comparison of the clock transition in Hg against a primary frequency standard tests the following fractional time variation [12]:

$$\begin{aligned}\frac{\partial}{\partial t} \ln \frac{\nu_{\mathrm{Cs}}}{\nu_{\mathrm{Hg}}} &= \frac{\partial}{\partial t} \ln \left(\frac{\alpha^2 \frac{\mu_{\mathrm{Cs}}}{\mu_B} F_{\mathrm{rel,Cs}}(\alpha)}{F_{\mathrm{rel,Hg}}(\alpha)} \right) = 2\frac{\partial \ln \alpha}{\partial t} + \frac{\partial}{\partial t} \ln \frac{\mu_{\mathrm{Cs}}}{\mu_B} + (0.8 + 3.2)\frac{\partial \ln \alpha}{\partial t} \\ &= 6\frac{\partial}{\partial t} \ln \alpha + \frac{\partial}{\partial t} \ln \frac{\mu_{\mathrm{Cs}}}{\mu_B} = (0.2 \pm 7) \times 10^{-15}\ \mathrm{yr}^{-1}. \end{aligned} \tag{11}$$

Likewise we derive for the fractional variation of $\nu_{\mathrm{Cs}}/\nu_{\mathrm{H}}$ from the hydrogen $1S$–$2S$ experiment [this work]:

$$\frac{\partial}{\partial t} \ln \frac{\nu_{\mathrm{Cs}}}{\nu_{\mathrm{H}}} = 2.8\frac{\partial}{\partial t} \ln \alpha + \frac{\partial}{\partial t} \ln \frac{\mu_{\mathrm{Cs}}}{\mu_B} = (3.2 \pm 6.3) \times 10^{-15}\ \mathrm{yr}^{-1}. \tag{12}$$

With $x = \partial_t \ln \alpha$ and $y = \partial_t \ln(\mu_{\mathrm{Cs}}/\mu_B)$ we can write the experimental results as

$$6x + y \quad = (0.2 \pm 7) \times 10^{-15}\ \mathrm{yr}^{-1} \qquad (\mathrm{Hg}^+), \tag{13}$$

$$2.8x + y = (3.2 \pm 6.3) \times 10^{-15}\ \mathrm{yr}^{-1} \qquad (\mathrm{H}). \tag{14}$$

These equations are easily solved, yielding the mean expectation values $\langle y \rangle$ and $\langle x \rangle$ without any assumptions of possible correlations between the drifts. In

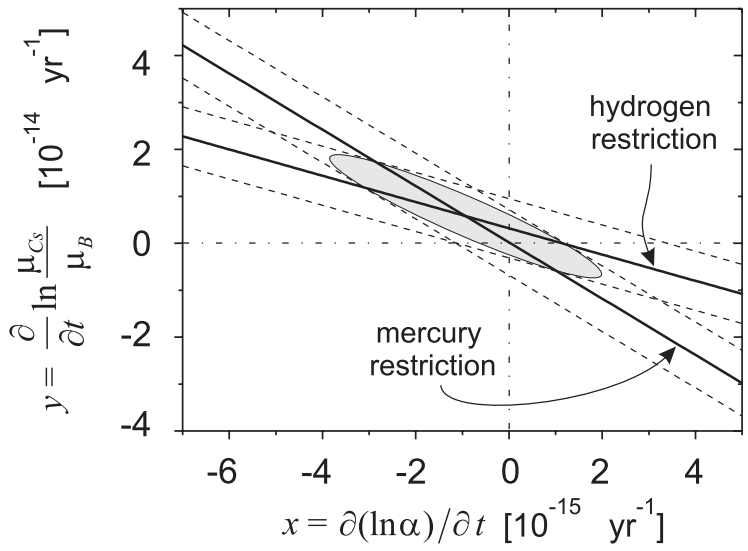


Fig. 9. Drifts of the ${}^2S_{1/2}(F=0) \rightarrow {}^2D_{5/2}(F'=2, m'_F=0)$ transition in $^{199}Hg^+$ and of the $1S(F=1, m_F=\pm 1) \rightarrow 2S(F'=1, m'_F=\pm 1)$ transition in H against the frequency of the ground-state hyperfine transition in ^{133}Cs. Dashed lines represent 1 σ experimental restrictions from the mean measured values. The elliptical region defined by $R(\Delta x, \Delta y) = 1$ gives the standard deviation for x and y when projected on corresponding axis by integration over the other.

Fig. 9, both equations and the graphical solution are shown. Obviously, testing the stability of α by monitoring only one transition frequency during a time period would require additional assumptions of the drift of other fundamental constants.

The uncertainties can be calculated by making two assumptions: (i) the experimental data are Gaussian distributed and (ii) the mercury (13) and the hydrogen (14) measurements are statistically independent. In this case normal Gaussian error propagation allows the calculation of the variances $\langle y^2\rangle - \langle y\rangle^2$ and $\langle x^2\rangle - \langle x\rangle^2$ even when the drift rates x and y are correlated [4]. This is because the covariance term $\langle xy\rangle$ does not appear when (13,14) are resolved for x and y.

For a graphical representation it is possible to calculate the two-dimensional probability density of x and y to be the true values:

$$P(x,y) = \frac{1}{2\pi\sqrt{\sigma_{\rm H}\,\sigma_{\rm Hg}}} \exp[-R(\Delta x, \Delta y)/2], \tag{15}$$

where Δx and Δy are the distances along the corresponding axes from the crossing point of the solid lines (Fig. 9) i.e. the solution of (13,14). The experimental uncertainties are $\sigma_{\rm H} = 6.3 \times 10^{-15}$ yr^{-1} and $\sigma_{\rm Hg} = 7 \times 10^{-15}$ yr^{-1} taken from (13) and (14), and the exponent function is given by:

$$R(\Delta x, \Delta y) = (\Delta y + 6\Delta x)^2/\sigma^2_{\rm Hg} + (\Delta y + 2.8\Delta x)^2/\sigma^2_{\rm H}\,. \tag{16}$$

We deduce the uncertainties for x and y as projections of the ellipse defined by $R(\Delta x, \Delta y) = 1$ on the corresponding axes (Fig. 9) by integration over the other dimension. For only two independent measurements this method is equivalent to performing simple Gaussian error propagation of uncertainties when resolving (13,14). However, the projection method can be generalized to more than two measurements, i.e. more than two equations for the two unknowns x and y (see contribution by E. Peik in this volume). The integration in both directions can be performed analytically to derive the uncertainties of x and y. Our evaluation is model-independent in the sense that we neither assume x and y to be uncorrelated nor that they are correlated in any way.

The relative drift of the fine structure constant α between July 2000 and the end of 2003 is

$$x = \frac{\partial}{\partial t} \ln \alpha = (-0.9 \pm 2.9) \times 10^{-15} \text{ yr}^{-1}. \tag{17}$$

For the limit on the relative drift of μ_{Cs}/μ_B, we find

$$y = \frac{\partial}{\partial t} \ln \frac{\mu_{\text{Cs}}}{\mu_B} = (0.6 \pm 1.3) \times 10^{-14} \text{ yr}^{-1}. \tag{18}$$

The given 1 σ uncertainties for x and y incorporate both the statistical and systematic uncertainties of the hydrogen and the mercury measurements. Both limits (17) and (18) are consistent with zero.

These results allow us to deduce a restriction for the relative drift of the ratio of the nuclear magnetic moments in ^{87}Rb and ^{133}Cs. From 1998 to 2003, the drift of the ratio of the ground state hyperfine frequencies in ^{87}Rb and ^{133}Cs has been measured to be [13]

$$\frac{\partial}{\partial t} \ln \frac{\nu_{\text{Rb}}}{\nu_{\text{Cs}}} = (0.2 \pm 7.0) \times 10^{-16} \text{ yr}^{-1}. \tag{19}$$

Substituting the corresponding dependencies $F_{\text{rel}}(\alpha)$ for these transitions [13,34], we can write

$$\frac{\partial}{\partial t} \ln \frac{\nu_{\text{Rb}}}{\nu_{\text{Cs}}} = \frac{\partial}{\partial t} \left(\ln \frac{\mu_{\text{Rb}}}{\mu_{\text{Cs}}} - 0.53 \ln \alpha \right). \tag{20}$$

Combining (17), (19), and (20) we deduce a restriction for the relative drift of the nuclear magnetic moments in ^{87}Rb and ^{133}Cs:

$$\frac{\partial}{\partial t} \ln \frac{\mu_{\text{Rb}}}{\mu_{\text{Cs}}} = (-0.5 \pm 1.7) \times 10^{-15} \text{ yr}^{-1}. \tag{21}$$

where the same procedure as in Fig. 9 was used with a diagram covering x and $z = \partial_t \ln(\mu_{\text{Rb}}/\mu_{\text{Cs}})$.

The values of the nuclear moments are determined by the strong and the electromagnetic interaction. If the latter is constant, the time changing nuclear moments point toward a variation of the strong coupling constant. Unfortunately, there is no simple scaling law such as (4) or (9) known for the nuclear moments. However, they can be approximated with the Schmidt model [35]. For ^{87}Rb and

^{133}Cs atoms the Schmidt nuclear magnetic moments μ^s depend only on the proton gyromagnetic ratio g_p. Using this model, one can get an approximate relation

$$\frac{\partial}{\partial \ln g_p} \ln \frac{\mu^s_{\mathrm{Rb}}}{\mu^s_{\mathrm{Cs}}} \simeq 2, \tag{22}$$

which, in combination with (21), yields a stringent upper bound for the drift of the proton gyromagnetic factor g_p:

$$\frac{\partial}{\partial t} \ln g_p = (-0.2 \pm 0.8) \times 10^{-15}\ \mathrm{yr}^{-1}. \tag{23}$$

5 Conclusion

In conclusion, we have determined limits for the drift of α, μ_{Cs}/μ_B and $\mu_{\mathrm{Rb}}/\mu_{\mathrm{Cs}}$ from laboratory experiments without any assumptions of their conceivable correlations. All these limits are consistent with zero drift. Table 2 represents some of

Table 2. Some of the precise recent measurements testing the relative changes of the fine-structure constant α over a time interval $(t_2 - t_1)$ where t_2 is the present time and t_1 corresponds to the past. The drift can be calculated as $\partial/\partial t(\ln \alpha) \simeq [\alpha(t_2) - \alpha(t_1)]\ \alpha^{-1}\ (t_2 - t_1)^{-1}$. Combining the results of absolute frequency measurements of the optical transitions in Hg^+ and H yields a restriction for the drift of α without assumptions of conceivable correlations between the constants.

Method, reference	$t_2 - t_1$	$[\alpha(t_1) - \alpha(t_2)]/\alpha$	Model assumptions
Geological (Oklo reactor) [10]	2 Gyr	$(-0.36 \pm 1.44) \times 10^{-8}$	fission conditions, $\dot{\alpha}_S = \dot{\alpha}_W = 0$
Astrophysical (absorption spectra) [5]	5–11 Gyr	$(-0.54 \pm 0.12) \times 10^{-5}$	astrophysical models
Astrophysical (absorption spectra) [8]	9.7 Gyr	$(-0.06 \pm 0.06) \times 10^{-5}$	astrophysical models
Astrophysical (absorption spectra) [9]	8 Gyr	$(0.1 \pm 1.7) \times 10^{-5}$	astrophysical models
Laboratory (Rb–Cs clocks comparison) [13]	4 yr	$(0.2 \pm 5.2) \times 10^{-15}$	$\dot{\alpha}_S = \dot{\alpha}_W = 0$
Laboratory (Hg^+ transition frequency measurement) [12]	3 yr	$(-0.1 \pm 3.5) \times 10^{-15}$	$\dot{\alpha}_S = \dot{\alpha}_W = 0$
Laboratory (H transition frequency measurement) [this work]	3.6 yr	$(-4.1 \pm 8.2) \times 10^{-15}$	$\dot{\alpha}_S = \dot{\alpha}_W = 0$
Combination of [12] and this work	3.5 yr	$(3.2 \pm 10.2) \times 10^{-15}$	LLI, LPI, linear drifts

the most accurate recent measurements of drifts of the fine structure constant α in different epochs. From all these data only the investigations of quasar absorption spectra measured with the Keck/HIRES spectrograph show a significant deviation between the values of α today and 10 Gyrs ago [5]. Considering the Oklo data as well as results of modern astrophysical and laboratory measurements one can suppose that the drift, if existent at all, is not linear and that α has reached an asymptotic value or is in the extremum of an oscillation or is simply too small to be detected yet. To make a definite conclusion additional independent astrophysical data as well as a further increase of the accuracy of laboratory methods are required.

Acknowledgements

We thank S.G. Karshenboim for the fruitful discussion of this work. N. Kolachevsky acknowledges support from Alexander von Humboldt Stiftung. The work was partly supported by the Deutsche Forschungsgemeinschaft (grant No. 436RUS113/769/0-1) and RFBR. The development of the FOM fountain was supported by Centre National d'études spatiales and Bureau National de Métrologie.

References

1. P.A.M. Dirac, Nature (London) **139**, 323 (1937).
2. T. Damour and A.M. Polyakov, Nucl. Phys. B **423**, 532 (1994).
3. S.G. Karshenboim, physics/0306180.
4. X. Calmet, H. Fritzsch, Eur. Phys. J. C **24**, 639 (2002).
5. M. T. Murphy, J. K. Webb, V. V. Flambaum, astro-ph/0306483, see also J. K. Webb *et al.*, Phys. Rev. Lett **87**, 091301 (2001).
6. A. Ivanchik, P. Petitjean, E. Rodriguez, and D. Varshalovich, astro-ph/0210299. A. Ivanchik, A. Potekhin, and D. Varshalovich, Astron. and Astroph. **343**, 439 (1999).
7. J.-P. Uzan, Rev. Mod. Phys. **75**, 403 (2003).
8. H. Chand, R. Srianand, P. Petitjean, and B. Aracil, Astron. and Astroph. **417**, 853 (2004).
9. R. Quast, D. Reimers, and S.A. Levshakov, Astron. and Astroph. **415**, 27 (2004).
10. Y. Fujii, *et al.*, Nucl. Phys. B, **573**, 377 (2000).
11. J.D. Prestage, R.L. Tjoelker and L. Maleki, Phys. Rev. Lett. **7**4, 3511 (1995).
12. S. Bize *et al.*, Phys. Rev. Lett. **90**, 150802 (2003).
13. H. Marion *et al.*, Phys. Rev. Lett. **90**, 150801 (2003).
14. M. Niering *et al.*, Phys. Rev. Lett. **84**, 5496 (2000).
15. J. Reichert *et al.*, Phys. Rev. Lett. **84**, 3232 (2000).
16. A. Huber *et al.*, Phys. Rev. A **59**, 1844 (1999).
17. J. Reichert, R. Holzwarth, Th. Udem and T.W. Hansch, Opt. Commun. **172**, 59 (1999).
18. H.A. Haus and E.P. Ippen, Opt. Lett. **26**, 1654 (2001).
19. R. Holzwarth *et al.*, Phys. Rev. Lett. **85**, 2264 (2000).
20. S.A. Diddams *et al.*, Phys. Rev. Lett. **84**, 5102 (2000).

21. T.M. Rammond, S.A. Diddams, L. Hollberg and A. Bartels, Opt. Lett. **27**, 1842 (2002).
22. U. Morgner *et al.*, Phys. Rev. Lett. **86**, 5462 (2001).
23. T.M. Fortier, D.J. Jones and S.T. Cundiff, Opt. Lett. **28**, 2198 (2003).
24. A. Bartels and H. Kurz, Opt. Lett. **27**, 1839 (2002).
25. R. Ell *et al.*, Opt. Lett. **26**, 373 (2001).
26. J.C. Knight, T.A. Birks, P.St.J. Russell and D.M. Atkin, Opt. Lett. **21**, 1547 (1996).
27. J.K. Ranka, R.S. Windeler and A.J. Stentz, Opt. Lett. **25**, 25 (2000).
28. S.A. Diddams, L. Hollberg, L.-Sh. Ma and L. Robertsson, Opt. Lett. **27**, 58 (2002).
29. Th. Udem *et al.*, Phys. Rev. Lett. **86**, 4996 (2001).
30. M. Eichenseer, A.Yu. Nevsky, Ch. Schwedes, J. von Zanthier and H. Walther, J. Phys. B: At. Mol. Opt. Phys. **36**, 553 (2003).
31. J. Stenger, H. Schnatz, C. Tamm and H. Telle, Phys. Rev. Lett. **88**, 073601 (2002).
32. M. Abgrall, Thèse de doctorat de l'université Paris VI (2003).
33. P. Lesage, IEEE Trans. Instrum. Meas. **32**, 204 (1983).
34. V.A. Dzuba, V.V. Flambaum, and J.K. Webb, Phys. Rev. A **59**, 230 (1999); V.V. Flambaum, physics/0302015.
35. S. G. Karshenboim, Can. J. Phys. **78**, 639 (2000).

An Optical Frequency Standard with Cold and Ultra-cold Calcium Atoms

Fritz Riehle[1], Carsten Degenhardt[1], Christian Lisdat[2], Guido Wilpers[1], Harald Schnatz[1], Tomas Binnewies[1], Hardo Stoehr[1], and Uwe Sterr[1]

[1] Physikalisch-Technische Bundesanstalt, Bundesallee 100, 38116 Braunschweig, Germany
[2] Institut für Quantenoptik, Universität Hannover, Welfengarten 1, 30167 Hannover, Germany

Dedicated to Jürgen Helmcke, who led the research on the Ca standard for two decades, on the occasion of his 65^{th} birthday.

Abstract. An optical frequency standard ($\lambda = 657$ nm) based on cold and ultra-cold neutral Ca atoms is operated at PTB and its frequency has been measured with respect to the Cs hyperfine transition over eight years. Since measurements of this kind can be used to detect possible variations of fundamental constants, we present the status of the Ca optical frequency standard with emphasis to the systematic contributions to the frequency that might mimick variations of the ratio of the Ca and Cs frequency. We show that the uncertainty resulting from the residual velocity of the cold atoms at 3 mK limits the fractional uncertainty to 2×10^{-14}. Applying a novel method for producing ultra-cold atoms ($T \approx 10$ μK) a first frequency measurement has been performed where the contributions from the residual velocity now become insignificant at that level of uncertainty. A residual fractional uncertainty below 10^{-15} is expected for the near future.

1 Introduction

Optical frequency standards find increasing applications to realize basic units such as the length unit [1], to establish optical clocks [2–4] that eventually might become superior to the currently best microwave clocks [5–7], to determine fundamental constants [8,9], or to test basic theories and possible variations of fundamental constants [7,10]. Compared to the present atomic clocks operating in the microwave region, the use of optical frequencies is advantageous since for a given linewidth, a higher quality factor $Q = \nu/\delta\nu$ is obtained with the potential to increase the stability and the accuracy.

Currently three different routes are followed to set up an optical clock of ultimate performance. The first and second are based on a *single* ion confined and cooled in a radio-frequency trap and on a large cloud of cold neutral atoms prepared in a magneto-optical trap, respectively. With a single ion, the achievable signal-to-noise ratio and hence the stability of the clock is limited. On the other

F. Riehle et al., An Optical Frequency Standard with Cold and Ultra-cold Calcium Atoms, Lect. Notes Phys. **648**, 229–244 (2004)
http://www.springerlink.com/

hand, in neutral atom standards the large number N of atoms allows one to reach excellent short-term stabilities, since the limit for the stability is ultimately given by the "quantum projection noise" [11] scaling with $N^{-1/2}$. However, compared to single ion standards, neutral atoms suffer from the fact that the preparation of the cold atoms, e.g., in a magneto-optical trap and the probing of the clock transition have to be performed sequentially. Consequently, the residual free motion of the atoms and the acceleration in the gravitational field limit the interaction time with the clock transition and introduce frequency shifts due to a residual first-order Doppler effect and cold collisions. On the third route, one tries to combine the advantages of the two former approaches where ultra-cold atoms are stored in an optical lattice of light with a well-defined "magic wavelength" which is chosen such that the huge associated light shifts for the two states connected by the clock transition exactly cancel [12].

Here, we report on the status of the optical frequency standard based on a ballistically expanding cloud of cold and ultra-cold calcium atoms. In the second section we describe the techniques used in our optical frequency standard where the Ca atoms are laser-cooled to about 3 mK and in a second stage to a few microkelvin. In the third section we present the results of the investigations that allowed us to set up the uncertainty budget. In the fourth section we discuss the measurements of the frequency of the Ca transition with respect to the frequency of the caesium microwave standard.

2 Methods and Experimental Realization

2.1 Properties of the Calcium Standard

The calcium optical frequency standard is based on the intercombination transition from the $^1S_0$ ground state to the lowest $^3P_1$ state (Fig. 1) of ^{40}Ca at $\lambda \approx 657$ nm which exhibits a narrow natural linewidth of 370 Hz. The frequency of the transition ($m_j = 0 \rightarrow m_j = 0$) is to first order insensitive to electric and

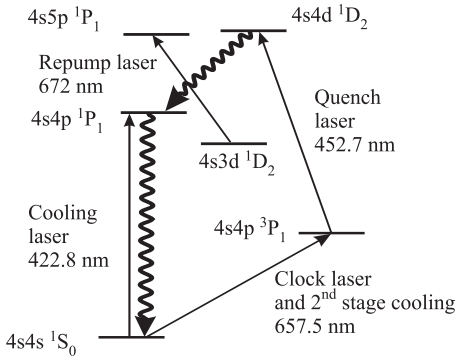


Fig. 1. Energy level diagram of ^{40}Ca with the laser transitions relevant for the optical frequency standard

magnetic fields. The quadratic dependencies are 1.6×10^{-6} Hz $(\mathrm{V/m})^{-2}$ and 6.4×10^{7} Hz T^{-2}, respectively [13].

2.2 Production of Cold Ca Atoms ($T \approx 3$ mK)

A standard magneto-optical trap (MOT) is loaded from an effusive beam (Fig. 2) by using the strong 423 nm transition from the $^1\mathrm{S}_0$ ground state to the $^1\mathrm{P}_1$ level. The 423 nm radiation is produced by the radiation of a Ti:sapphire laser frequency doubled in a resonant cavity with a LBO crystal. The available output power is about 500 mW with 1.6 W of infrared power. As there is no Zeeman slower used, only a small fraction of the atoms from the velocity distribution of the thermal beam can be captured by the trap. These are atoms from the Boltzman velocity distribution ($\bar{v} \approx 700$ m/s, $T_{\mathrm{oven}} \approx 900$ K) which are slower than the trap's maximum capture velocity of about 30 m/s. To obtain an acceptable loading rate, first, the oven was placed as close as possible to the trap center ($d \approx 15$ cm). Second, each of the horizontal trapping beams had a another laser beam superposed. The first beam was red detuned by 90 MHz and the second one by 120 MHz from resonance as to slow down faster atoms below the maximum capture velocity of the trap. We generated the laser beams with two frequencies using the first diffraction orders of two switching AOMs with different radio frequencies and overlaying the beams on a 50% beam splitter. The number of stored atoms was increased by this method by about an order of magnitude. To further increase the number of atoms a particular loss channel has to be closed. These losses occur since atoms in the $^1\mathrm{P}_1$ decay with a probability of 10^{-5} to the $4\mathrm{s}3\mathrm{d}\,^1\mathrm{D}_2$ state. From here they predominantly fall into the metastable $4\mathrm{s}4\mathrm{p}^3\mathrm{P}$ states and get lost from the trap before they can decay into the ground state. This bypass was closed by exciting the atoms on the $4\mathrm{s}3\mathrm{d}\,^1\mathrm{D}_2$–$4\mathrm{s}5\mathrm{p}\,^1\mathrm{P}_1$ transition, applying 2 mW of laser light at 672 nm (repump transition in Fig. 1). From the $4\mathrm{s}5\mathrm{p}\,^1\mathrm{P}_1$ state the atoms can decay into the ground state

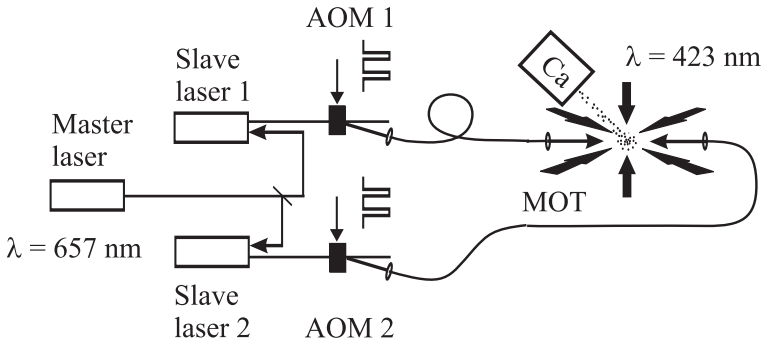


Fig. 2. Experimental setup. Ca atoms are trapped in a magneto-optical trap (MOT). The clock transition is interrogated by pulses cut from two cw clock-laser beams from two slave diode lasers injection locked by a common master of extended cavity diode laser (ECDL) configuration

and again take part in the cooling process. With these improvements the time constant of the loss of atoms from the trap was increased from 20 ms to more than 300 ms and the number of trapped atoms was increased by another factor of 5 leading to about 10^8 cold Ca atoms at a temperature of about 3 mK. After this loading the cold atoms can be either interrogated directly in a frequency standard or further cooled by a second-stage cooling technique.

2.3 Production of Ultra-cold Atoms

It has been shown that with cold atoms in the millikelvin range, given by the Doppler limit, the influence of the residual first-order Doppler shift ultimately limits the accuracy of a Ca frequency standard [16]. Even though a method has been devised to measure the associated spurious phase shifts and to reduce their influence [14,15], lower uncertainties can be obtained by a second-stage cooling on the intercombination transition. Since, however, the scattering rate on this transition is that low that the resulting force is of the order of the gravitational force we have devised a quench-cooling scheme [17,18]. There, the scattering rate on the intercombination transition can be increased by pumping the atoms with the help of a "quenching" laser (453 nm) back into the ground state via intermediate levels (see Fig. 1). An alternative quench transition was used in the group at the National Institute of Standards and Technology in Boulder (USA) [19,20]. With the quench cooling scheme we are able to achieve a temperature of the atomic cloud as low as 6 μK with up to 30 % of the atoms being transferred from the cloud at 3 mK to the ultracold cloud.

2.4 Interrogation of the Clock Transition

Cooling and trapping of the atoms and probing of the clock transition are performed sequentially. After the atoms are loaded for about twenty milliseconds, the trapping fields (laser beams and magnetic quadrupole field) are switched off and a small homogeneous magnetic quantization field is turned on in just 200 μs which separates the Zeeman components of the $^3\mathrm{P}_1$ state by 3.8 MHz each. The atomic cloud expands according to the temperature of 3 mK at a root-mean-square velocity of $v_{\mathrm{rms}} \approx 80$ cm/s, while the clock transition is probed. In the third step, the excitation to the $^3\mathrm{P}_1$ state is detected, e. g. by observing the fluorescence of its spontaneous decay to the $^1\mathrm{S}_0$ ground state.

The excitation of the clock transition is performed by using a pulsed excitation scheme that can be thought of as a time-domain Ramsey-Bordé atom interferometer [21,22] (Fig. 3). In the interaction with the first laser pulse of duration T_p an atomic wave packet that is initially in the ground state $\mid g >$ is coherently split into wave packets in the ground and excited state $\mid e >$. Due to the photon recoil the trajectories of the excited state wave packet (dashed lines) are deflected from the ground state ones (full lines). If the interaction represents a $\pi/2$ pulse it can be regarded as a 50% beam splitter for the atomic wave packet. After a time T another $\pi/2$ pulse splits the wave packets in the ground state and excited state. A series of two laser pulses applied after a dead time T' from

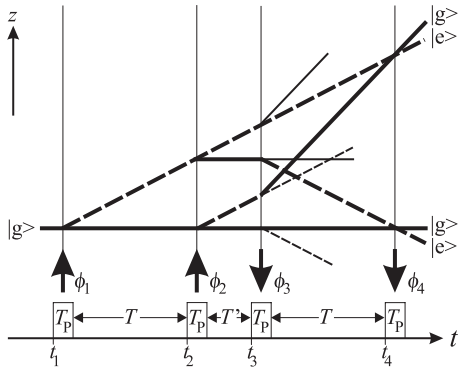


Fig. 3. Ramsey-Bordé atom interferometer with four pulses in the time domain

the opposite direction leads to a re-direction and mixing of partial waves where the two closed trapezoids each represents an atom interferometer. The number of atoms in the exit ports varies sinusoidally with the phase difference $\Delta\Phi$ of the partial waves aquired along the trajectories

$$\Delta\Phi = 2T(\omega - \omega_0 \pm \delta_{\mathrm{rec}} + \frac{\omega_0 v^2}{2c^2}) + \phi_2 - \phi_1 + \phi_4 - \phi_3. \tag{1}$$

In (1), $\omega - \omega_0$ represents the detuning of the laser frequency from the atomic transition frequency, $\omega_0 v^2/(2c^2)$ is the second-order Doppler shift, $\delta_{rec} = \hbar k^2/(2m_{\mathrm{Ca}})$ results from the photon recoil where k is the wave vector of the laser field and m_{Ca} the mass of a Ca atom and ϕ_i, with $i = 1, 2, 3, 4$, are the phases of the laser in the respective interaction zones.

The required two counter-propagating pairs of coherent short laser pulses are generated by cutting them from cw laser beams generated by two diode lasers injection-locked by a common extended cavity diode laser system (Fig. 2). This master laser is comprised of a laser in Littmann configuration that was locked by the Pound-Drever-Hall method [23] to a reference resonator for better short-term stability. The interrogation laser was tuned and later stabilized to the central fringe of the Ramsey-Bordé pattern (Fig. 4)

by adjusting the frequency, delivered to an acousto-optical modulator inserted between laser and reference cavity. The obtained spectra differ markedly depending on whether cold or ultra-cold atoms are used. With pulse widths of about 1 μs, up to 25 % of the atoms of the cold ensemble with a halfwidth of the inhomogeneously Doppler-broadened line of about 2 MHz can be excited (Fig. 4). The signals shown in Fig. 4 b) represent the superposition of the two interference signals caused by the two recoil components (the two atom interferometers of Fig. 3) which are separated by 23.1 kHz. The highest resolution obtained in our standard with cold atoms was 230 Hz.

With ultra-cold atoms the Doppler line broadening becomes much narrower ($\approx$ 200 kHz) and the interferences now spread all across the homogeneously broadened line (Fig. 5). Furthermore, the contrast of the interferences is much

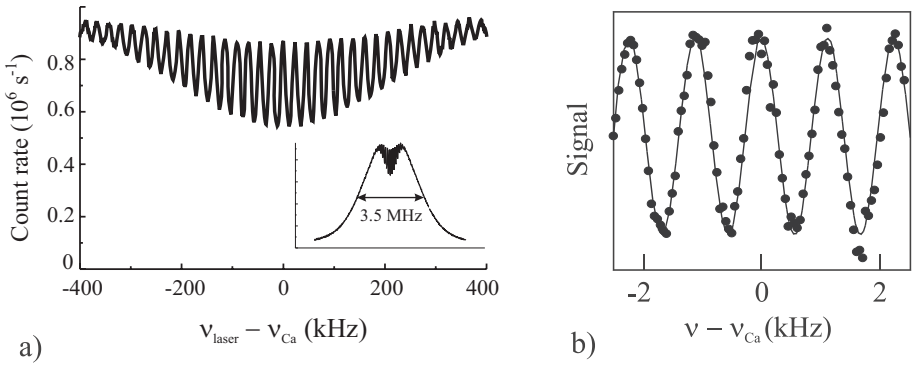


Fig. 4. Atom interferences vs. laser frequency excited in a cloud of Ca atoms released from a magneto-optical trap. **(a)** The atom interferences appear in the center of the Doppler-broadened fluorescence signal. **(b)** In the measurement the period of the fringes is chosen such that 21 fringes correspond to the recoil splitting

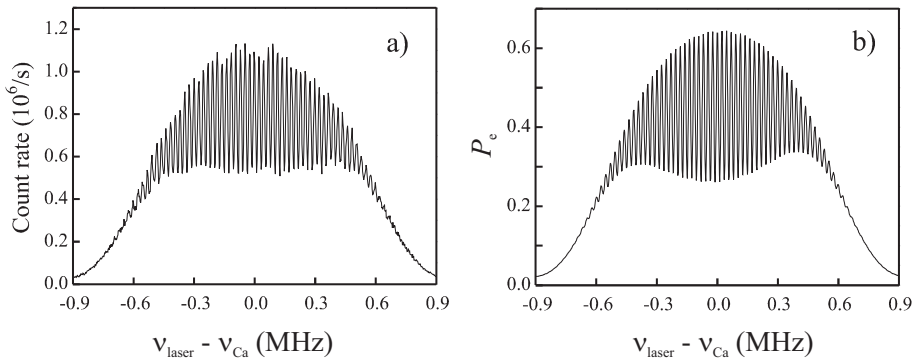


Fig. 5. Atom interferences vs. laser frequency excited in an expanding cloud of ultracold Ca atoms. **(a)** Experiment with atoms released from a magneto-optical trap. **(b)** Calculated interferences according to Bordé et al. [24]

more pronounced in Fig. 5 as compared to Fig. 4. The reason for these differences comes from the fact that with the cold atoms the Doppler broadening with the available power in each excitation pulse is larger than the Fourier broadening corresponding to the width of the laser pulses. Using the ultra-cold atoms the situation has changed and the envelope of the line is now determined by the interaction time broadening. The spectrum of Fig. 5 was recorded during 40 minutes without any normalization by counting the photons at 657 nm from the decay of the excited $^3P_1$ state. In this case, each excited atom contributes only a single photon which is detected in our set-up with a probability of about 10^{-3}. The ultracold atoms, however, allow one to apply a detection scheme similar to the shelved electron scheme which is capable to detect each excited atom [4] with almost unity probability. The increased signal-to-noise ratio possible with

the novel detection scheme now allows one to set up a frequency standard with improved short-term stability basically limited by the quantum projection noise [11]. It has been pointed out that the quantum limit for the Ca standard corresponds to an Allan standard deviation $\sigma_y(2,\tau) < 10^{-16}$ for $\tau = 1$ s [25]. Besides improving the stability the low velocities of ultracold atoms in the optical Ca frequency standard lead to a reduced influence of the Doppler effect on the shape of the spectrum. As a result, the shape of the spectrum can be calculated more precisely which allows one to locate the position of the true line center with higher accuracy. In the case of the Cs atomic clock the ability to accurately know the shape of the spectral line was prerequisite for the determination of the unperturbed line center with a precision equivalent to 10^{-5} of the measured line width. It is expected that these improvements will allow one to reach a relative uncertainty of the optical Ca standard of 1×10^{-15}, thereby competing with the best optical standards based on single ions and with the best microwave clocks.

3 Uncertainty of the Optical Ca Frequency Standard

The accuracy of an optical frequency standard as the Ca standard, where the frequency of a high-resolution laser spectrometer is stabilized to the central maximum or minimum of the interference pattern, is ultimately limited to the extent of how well all relevant offsets that shift this frequency from that of the unperturbed transition can be determined and corrected for. The accuracy is usually expressed [26] as an estimated standard uncertainty. In the following we give an uncertainty budget containing the relevant contributions for a standard with cold and ultra-cold atoms.

3.1 Residual First-order Doppler Shifts

It was shown [16] that a significant contribution to the uncertainty of the line center can be attributed to residual errors in the phase fronts of the laser beams probing the clock transition or from tilts with respect to the equipotential surface of the gravitational potential. These effects appear since the moving atoms are probed in different regions of the wave fronts for the different pulses of the atom interferometer (Fig. 6). According to (1) in atom interferometers the instantaneous laser phases ϕ_i in the interaction zones i are added to the phases of the corresponding deflected matter waves. These phase shifts show up as a shift of the measured transition frequency that to first order changes linearly with the pulse separation time T. Even a careful pre-alignment of the laser beams with optical interferometry is usually not sufficient to prevent these frequency shifts.

Up to an offset phase, constant in time, the spatial phase of the laser beam at the position $\vec{r}(t_l)$ at the time t_l can be approximated by

$$\phi_{l,j} = -\vec{k}_j \cdot \vec{r}(t_l) - \frac{|\vec{k}_j|}{2R_j} \cdot \vec{r}_\perp^{\,2}(t_l) \tag{2}$$

if the wave front is described as a spherical wave with large radius R_j and wave vektor $\vec{k}_j$ along the optical axis of the laser beam. The sign of the radius

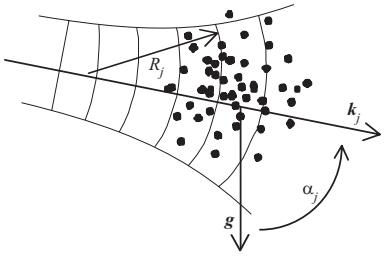


Fig. 6. The phases experienced by atoms in a ballistically expanding and falling cloud depend on the tilt angle α_j of the laser beam, its curvature R_j at the position of the atomic cloud and on the velocity of the atoms

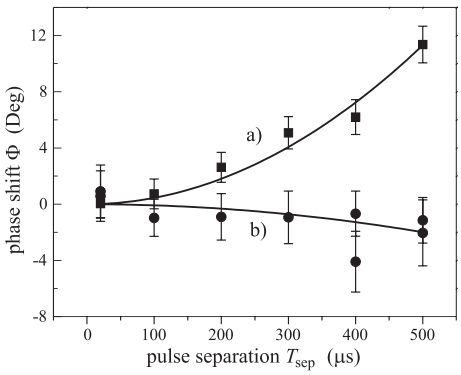


Fig. 7. Measured phase shift in an interferometer due to a misaligned laser beam ($\alpha = 1.6$ mrad, $R \approx 12$ m) [4]. **(a)** with cold atoms ($T \approx 2.8$ mK) and **(b)** with ultra-cold atoms ($T \approx 14$ μK)

R_j is defined to be positive for each laser beam divergent in its direction of propagation. Here $\vec{r}_\perp(t_l)$ denotes the distance vector of the atom from the optical axis of the laser beam. During the time the atomic wave packets in the free expanding ensemble travel through the atom interferometer they move along a trajectory $\vec{r}(t)$ through the wave fronts of the respective laser beams. Hence, in every interaction point ($l = 1, 2, 3, 4$) they are exposed to a different spatial phase of the respective laser beam. Inserting the accelerated motion of the atoms

$$\vec{r}(t_l) = \vec{r}_0 + \vec{v}_0 \cdot t_l + \frac{1}{2}\vec{g} \cdot t_l^2 \tag{3}$$

into (2) the phase difference relevant for (1) can be calculated and corrected for (Fig. 7). A method based on three pulse atom interferometers has been devised that allows one to determine and correct these influences [14,15].

3.2 Other Phase Shifts

Besides the influence of the Doppler effect on the resulting phase shifts a direct shift of the global phase of the laser beams has to be taken into account. As Oates et al. [2] have investigated such a phase shift can also occur from a phase chirp produced in the switching AOMs. In our experiment this influence contributes to a relative uncertainty of about 1×10^{-14} for the cold atoms (see Table 1).

3.3 Frequency Shifts Due to External Fields

External fields interacting with the atoms in general shift the frequency of the atoms during interrogation with respect to the unperturbed atoms. Besides con-

Table 1. Uncertainty budget for the measurement of the Ca frequency standard with cold atoms (col. 1), the currently achieved contributions with ultra-cold atoms (col. 2) and the contributions achievable according to [4] (col. 3).

Effect	uncertainty		
	$T = 3$ mK (8. Oct. 2001)	$T = 12$ μK (8. Oct. 2003)	$T = 6$ μK (achievable)
Residual first - order Doppler effect	2.6 Hz	1.0 Hz	0.15 Hz
Second-order Doppler effect	0.005 Hz	0.02 mHz	0.01 mHz
Other phase contributions	4 Hz	1.6 Hz	200 mHz
Asymmetry of line shape	0.05 Hz	0.05 Hz	50 mHz
Magnetic field (64 Hz/mT^2)	0.1 Hz	0.1 Hz	100 mHz
ac Stark effect	0.1 Hz	0.1 Hz	100 mHz
Quadratic Stark effect (E $<$ 2 V/cm)	0.02 Hz	0.02 Hz	20 mHz
Black - body radiation			
Oven	4.3 Hz	3.9 Hz	
Chamber (300 K)	0.07 Hz	0.07 Hz	70 mHz
Collisions of cold atoms	1.8 Hz	0.06 Hz	60 mHz
Influence of laser drift	3.2 Hz	0.1 Hz	100 mHz
Sum in quadrature	7 Hz	4.4 Hz	400 mHz
Statistical uncertainty of frequency meas.	3 Hz	3 Hz	100 mHz
Cs clock (1×10^{-15})	0.5 Hz	0.5 Hz	500 mHz
Total uncertainty	8 Hz	5.4 Hz	700 mHz
total relative uncertainty $\delta\nu/\nu$	$\mathbf{2 \times 10^{-14}}$	$\mathbf{1.2 \times 10^{-14}}$	$\mathbf{1.6 \times 10^{-15}}$

stant or slowly fluctuating magnetic and electric (dc) fields also the rapidly oscillating electric fields of electromagnetic radiation give rise to frequency shifts due to the ac-Stark effect induced by optical radiation or due to the so-called black-body shift induced by thermal radiation.

Magnetic Fields. Beverini et al. [29] have shown that due to the missing hyperfine splitting of the ground state of ^{40}Ca only the quadratic dependence of the excited clock state, 4s4p ^{3}P$_1$, on external magnetic fields has to be taken into account. Neglecting higher order contributions accounts for a relative error of 10^{-6} with the low magnetic fields applied. The influence of the quadratic Zeeman-effect was measured in a calcium atomic beam [29] to be $(6.4 \pm 0.1) \times 10^7\,\mathrm{HzT}^{-2}$. Oates et al. [30] measured the coefficient at an atomic beam and a cold ensemble to be $(6.1 \pm 0.4) \times 10^7\,\mathrm{HzT}^{-2}$. We use the weighted average of these results, $(6.4 \pm 0.1) \times 10^7\,\mathrm{HzT}^{-2}$, to correct the quadratic Zeeman shift $\Delta\nu_{\mathrm{qZ}}$ resulting from the homogeneous magnetic field, typically $B = (0.2 \pm 0.004)$ mT, which we apply to define the quantization axis. The magnitude of the remaining magnetic field after compensation with Helmholtz coils in three dimensions was measured using the cross-over effect of the Zeeman lines in saturation spectroscopy. For times $T_1 \geq 200$ μs after switching off the magnetic quadrupole field of the trap the value was (0 ± 4) μT. The average magnetic bias field was determined with a relative uncertainty of 0.6 % utilizing the two σ components of the ^{3}P$_1$ state.

Electric dc Fields. External electric fields can shift the transition frequency due to the Stark effect. The main contributions to an electric dc field E in our set-up originate from static charges on the surface of the stainless steel vacuum chamber or the Capton-coated magnetic field coils. We assume a maximum field of 200 V/m at the position of the trap. The atomic states that are involved in the clock transition are all of defined parity. Therefore, there are no odd order contributions to the Stark effect. Zeiske [31] showed that with these electric fields higher order contributions only lead to shifts in the microhertz range. Hence, only the quadratic Stark effect needs to be taken into account and the frequency shift of the clock transition is given by

$$\nu(E \neq 0) - \nu(E = 0) = -\frac{1}{2}\Delta\alpha E^2. \tag{4}$$

The difference of the polarizabilities of the two clock states $\Delta\alpha \equiv \alpha(^3\mathrm{P}_1) - \alpha(^1\mathrm{S}_0)$ was determined [31] to be $\Delta\alpha_\sigma = (3.37 \pm 0.04) \times 10^{-6}$ Hz/(V/m)^2 for an electric field perpendicular to the quantization axis of the atoms and $\Delta\alpha_\pi = (2.47 \pm 0.04) \times 10^{-6}$ Hz/(V/m)^2 for a field parallel to the quantization axis. As the direction of a static electric field is expected to be fairly constant but is unknown, a rectangular distribution for the shift due to an electric field $\vec{E}$ between 0 and 200 V/m is assumed resulting in a frequency shift between 0 and a minimum of -70 mHz. Hence, we would have to apply a correction of 35 mHz, which up to now has been neglected.

ac Stark Effect. There are numerous dipole transitions in the ^{40}Ca originating from the ground or the excited state of the clock transition (J) to higher states (J') that give rise to an ac Stark shift of the clock transition via their contributions $\alpha_{JJ'}$ to the polarizabilities of the two clock states [31]. Since the most important transition wavelengths range from 200 nm to 2 μm, direct light or stray light from the lasers used in the experiment can cause ac Stark shifts. For non-resonant laser light of frequency ν_{ac} the contributions $\Delta\nu(JJ')$ of the single lines to the shifts are given by

$$\Delta\nu(JJ') = -\frac{\alpha_{JJ'}}{2}\frac{I}{\epsilon_0 c}\frac{1}{1 - \nu_{ac}^2/\nu_{JJ'}^2} \tag{5}$$

where ϵ_0 is the permittivity of free space and I is the irradiance at the position of the atoms. An evaluation of the maximum effect of the laser light reaching the trapped ensemble can be performed using the polarizabilities given in [31]. First, one has to take into account scattered light reflected from different surfaces which was measured to be less than a total irradiance of 8 mW/m^2 [13] at the location of the MOT. Second, light scattered in the AOM crystals, even when the AOMs are switched off, is coupled directly into the trap. The calculated total shift due to the interaction of all laser frequencies with the 24 most important transitions was less than 2 mHz. This includes the light with 800 μW/m^2 from the spectroscopy laser, offset by 70 MHz from the clock resonance and scattered into the fibers. A larger shift may be due to the light of the cooling laser scattered

in the AOMs and reflected into the trap that is red shifted by -220 MHz. An irradiance of up to 80 mW/m^2 has been measured that could result in frequency shifts of up to 40 Hz. As the spatial intensity distribution of the scattered light was irregular an exact calculation of the shift was impossible. Therefore, in general we chop the light mechanically during the interrogation. Provided that the relevant experimental precautions are met, the remaining ac Stark effect of the laser light of a few millihertz can be neglected.

Influence of Blackbody Radiation. The influence of the radiation due to the ambient temperature of the water cooled copper coils with a temperature between 288 K and 298 K and the calcium oven at 883 K can be calculated using a quasi-static approach [31] and assuming that the spectral distribution follows Planck's law of a black-body radiator. At room temperature the peak spectral irradiance of the black-body radiation is near 1.7×10^{13} Hz while at the frequencies of the lowest lying dipole transitions contributing to the ac Stark effect of the clock transitions, i. e. two strong lines close to 1.54×10^{14} Hz and one weak line at 2×10^{14} Hz, the spectral irradiance is already a factor of 10^7 below the peak. This leads to a correction of 5% [32] if we assume that the mean square of the electric field $\langle E^2 \rangle$ of the radiation has the same effect as a static electric field. With

$$\langle E^2 \rangle = \frac{8\pi^5 k_B^4}{15 c_0^3 \epsilon_0 h^3} \vartheta^4 = 8.545 \times 10^{-5} \frac{\mathrm{V}^2}{\mathrm{m}^2\mathrm{K}^4} \vartheta^4 \tag{6}$$

the frequency shift as function of the temperature ϑ is

$$\Delta\nu_{bbr}(\vartheta) = -\frac{1}{2}\overline{\Delta\alpha}\langle E^2 \rangle = (-1.37 \pm 0.03) \times 10^{-10} \ \frac{\mathrm{Hz}}{\mathrm{K}^4} \vartheta^4 \tag{7}$$

with the average difference of the polarizabilities $\overline{\Delta\alpha} = 1/3(\Delta\alpha_\pi + 2\Delta\alpha_\sigma)$. The black-body shift for a temperature $\vartheta = (293 \pm 5)$ K is calculated to be (-1.02 ± 0.07) Hz.

In our current set-up the oven was only 15 cm away from the trapped ensemble and it could in principle lead to a large uncertainty due to the ac Stark shift, as a uniform temperature of $\vartheta = 883$ K would result in a shift of -80 Hz according to (7). Rather than modelling the influence in our set-up, we varied the oven temperature from 883 K to 783 K and measured the frequency difference as $(+6.3 \pm 11.3)$ Hz which is statistically not significant. The measurements of Zeiske [31] for the dc polarizability of the clock transition clearly showed that the shift should be negative. For our measurement we assume a Gaussian uncertainty distribution where the true value of the shift can only be below 0 Hz. Therefore, to estimate the true shift and uncertainty, we took the negative frequency shift tail and determined the first and second moment of the distribution. In combination with a more recent measurement this resulted in a shift of (-2.1 ± 3.9) Hz due to the black body radiation of the calcium oven.

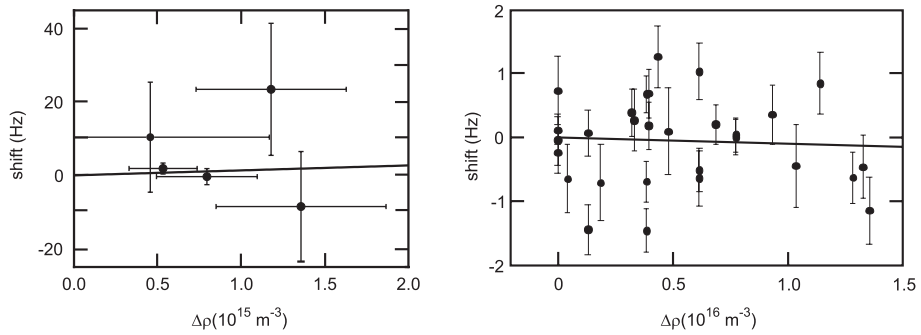


Fig. 8. Measurement of the density-dependent frequency shift **(a)** using cold atoms ($T \approx 3$ mK), **(b)** using ultra-cold atoms ($T \approx 20$ μK). The lines represent a linear regression to the data

3.4 Influence of Cold and Ultra-cold Atomic Collisions

Interactions between the atoms in the laser cooled ensemble, i. e. cold collisions, lead to a frequency shift of the clock transition. The magnitude of the shift depends on the states of the two colliding atoms and their distance and results in a mean frequency shift that depends on the temperature and the density of the atomic ensemble. For the low densities of about $< 10^{16}$ m^{-3} in our trap we assume the frequency shift to be a linear function of the collision rate and, hence, of the density ρ. In order to determine the density dependence of the frequency shift we have measured the frequency differences as a function of average density differences [15]. The density was varied by switching the repump laser at 672 nm (Fig. 1) which changed the number of trapped atoms while all other parameters like the magnetic field, loading time, frequency of the trapping light and laser beam profile were kept constant. We have checked that the radius and temperature of the trapped cloud did not change significantly when the average density $\Delta\rho$ was varied by a factor of five. The measurement of the collisional shift $\Delta\nu/\nu$ depending on the density ρ (Fig. 8 a) gave $\Delta\nu/\nu = (3 \pm 4.4) \times 10^{-30}$ m$^3 \times \rho$ [15]. The corresponding frequency shift for a resolution of 1.2 kHz ($T = 215$ μs) would be (1.2 ± 1.8) Hz. Within the uncertainty of this measurement no significant shift has been observed. The same was true when ultra-cold atoms were used (Fig. 8 b) where we obtained an even lower limit of $\Delta\nu/\nu = (-0.4 \pm 1.2) \times 10^{-31}$ m$^3 \times \rho$. The low density dependence allows one to use considerably higher densities and atom numbers compared to that in ultracold cesium ensembles used in Cs atomic clocks with temperatures of about 10 μK. For Cs the collisional shift is at least four orders of magnitude higher [33].

3.5 Uncertainty Budget

These contributions for the ultra-cold atoms, together with the results of previous work for cold atoms [4] and the achievable uncertainty are given in Table 1.

One of the largest contributions results from the black-body shift (4.3 Hz) which is a particular property of our setup with the oven close to the atomic ensemble. Other contributions to the uncertainty of the Ca standard in Table 1 include influences due to asymmetries in the line shape of the atom interferences, the relativistic Doppler-effect, as well as contributions of the stabilization electronics and the statistical uncertainty of the frequency comparison to the Cs fountain clock of PTB.

4 Frequency Measurements

In order to establish an optical frequency standard its frequency has to be determined with respect to the primary standard of time and frequency, i.e. the Cs atomic clock. For a long time, this used to be a complicated task as the the optical and microwave frequency differ by almost five orders of magnitude. Earlier concepts of optical frequency measurements relied on harmonic frequency chains employing a large number of oscillators [34,35]. The measurements of the Ca frequency with the harmonic chain (see 1995 to 1997 in Fig. 9) where the optical frequency could be phase-coherently determined gave a fractional uncertainty of 2.5×10^{-13}.

First measurements of the Ca standard with a femtosecond laser comb agreed within the combined uncertainties (2000) [27]. In the mean time further measurements have been performed. The uncertainties of these measurements were determined mainly by the uncertainties resulting from the corrections due to magnetic quadrupole fields of the MOT and ac Stark shifts (June 2001). Subsequent frequency measurements with reduced uncertainties employing cold atoms were performed [13] where the spurious phase shifts resulting from misalignment and curvature of wave fronts have been reduced as described before, the quadrupole MOT field was switched off and the ac Stark shifts due to residual cooling laser light was eliminated by using mechanical choppers. Recently a first frequency measurement using ultra-cold atoms has been performed where the

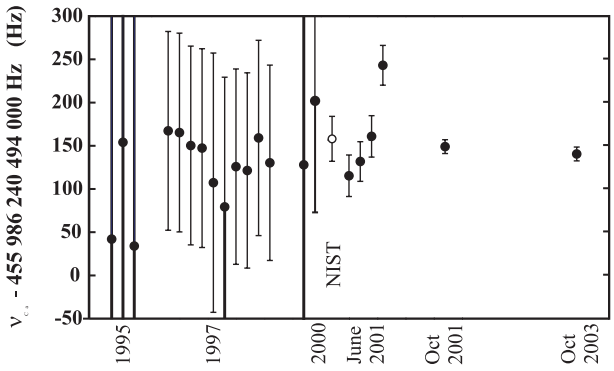


Fig. 9. Measured frequency values of the Ca optical frequency standard. Abszissa not to scale. The open circle represents the value obtained at NIST [4]

uncertainty budget of column 2 of Table 1 applies [36]. In this measurement a chirp in the optical pulses has been measured that lead to a resolution-dependent frequency shift of a few hertz which was corrected. Frequency ratios of ν_{Ca}/ν_{Cs} derived from the measurements presented in Fig. 9 in principle could be used to monitor possible variations of fundamental constants with time. As almost no sensitivity of the Ca intercombination transition is expected from theoretical reasons [37] any possible observed variation would have to be attributed to the Cs transition. From the data presented here, no significant effect can be deduced in the order of $\Delta(\nu_{Ca}/\nu_{Cs})/\Delta t \approx 10^{-14}/$ year. A more accurate evaluation is not possible as in the previous frequency measurements a chirp in the laser pulses might be present also, whose magnitude is not known and hence can not be corrected for. With the current uncertainty, however, the limit for a possible variation of the frequency ratio and hence the fine structure constant can be reduced by about two orders of magnitude taking measurements separated by a few years.

5 Prospects of the Ca Optical Frequency Standard

To further reduce the uncertainty of the Ca optical frequency standard, first of all the black-body shift has to be addressed. We currently set up an apparatus where the atoms are decelerated in a Zeeman slower before being deflected by a tilted two-dimensional molassis. In this case the oven temperature will no longer contribute and one ends up with the much lower uncertainty of 0.07 Hz given in the third column of Table 1. The influence of cold collisions is particularly low and it might even be reduced by using lower densities. The influence of the residual first-order Doppler effect seems to limit the fractional uncertainty to slightly below 10^{-15}. It remains to be explored if the use of atoms in an optical lattice operated at the magic wavelength where the first-order Doppler effect can be eliminated [12] will lead to lower uncertainties.

As has been pointed out before, the use of ultra-cold Ca atoms allows one to apply a detection scheme [4] similar to the shelved-electron scheme [28] to detect each excited atom with a probability close to unity. The increased signal-to-noise ratio possible with the novel detection scheme now allows us to set up a frequency standard with improved short-term stability and the prospects to reach the quantum projection-noise limit [11]. It has been pointed out [25] that the quantum limit for the Ca standard corresponds to an Allan standard deviation $\sigma_y(2,\tau) < 10^{-16}$ for $\tau = 1$ s. To approach this limit, a frequency instability of the spectroscopy laser of better than 3×10^{-16} for the duration of the atom interferometry $2T \approx 1.3$ ms is necessary. While laser instabilities better than 1 Hz in 1 s have already been demonstrated [38,39], to reach this value is still a challenging task. Higher stability indirectly contributes to higher accuracy. It allows one to detect systematic frequency shifts and fluctuations within short measuring times thereby identifying hitherto undetected error sources. It is expected that these improvements will allow us to reach a relative uncertainty of the optical Ca standard below 1×10^{-15} thereby competing with the best optical standards based on single ions and with the best microwave clocks.

Acknowledgement

Helpful contributions of Burghard Lipphardt, Nils Haverkamp, Harald Telle, Jörn Stenger, Stefan Weyers with the frequency measurements are gratefully acknowledged. The work was supported in part by the Deutsche Forschungsgemeinschaft under SFB 407 and by the European Commission through the Human Mobility Programme (Cold Atoms and Ultraprecise Atomic Clocks; CAUAC).

References

1. T. J. Quinn: Metrologia **40**, 103 (2003).
2. T. Udem, S. A. Diddams, K. R. Vogel, C. W. Oates, E. A. Curtis, W. D. Lee, W. M. Itano, R. E. Drullinger, J. C. Bergquist, L. Hollberg: Phys. Rev. Lett. **86**, 4996 (2001).
3. J. Stenger, C. Tamm, N. Haverkamp, S. Weyers, H. R. Telle: Opt. Lett. **26**, 1589 (2001).
4. G. Wilpers, T. Binnewies, C. Degenhardt, U. Sterr, J. Helmcke, F. Riehle: Phys. Rev. Lett. **89**, 230801 (2002).
5. S. Weyers, A. Bauch, R. Schröder, Chr. Tamm: 'The Atomic Caesium Fountain CSF1 of PTB'. In: *Frequency Standards and Metrology, Proceedings of the Sixth Symposium.* ed. by P. Gill (World Scientific, Singapore 2002) pp. 64–71.
6. S. R. Jefferts, J. Shirley, T. E. Parker, T. P. Heavner, D. M. Meekhof, C. Nelson, F. Levi, G. Costanzo, A. De Marchi, R. Drullinger, L. Hollberg, W. D. Lee, F. L. Walls: Metrologia **39**, 321 (2002).
7. H. Marion, F. Pereira Dos Santos, M. Abgrall, S. Zhang, Y. Sortais, S. Bize, I. Maksimovic, D. Calonico, J. Grünert, C. Mandache, P. Lemonde, G. Santarelli, Ph. Laurent, A. Clairon, C. Salomon: Phys. Rev. Lett. **90**, 150801 (2003).
8. C. Schwob, L. Jozefowski, B. de Beauvoir, L. Hilico, F. Nez, L. Julien, F. Biraben, O. Acef, A. Clairon: Phys. Rev. Lett. **82**, 4960 (1999).
9. A. Huber, T. Udem, B. Gross, J. Reichert, M. Kourogi, K. Pachucki, M. Weitz, T. W. Hänsch: Phys. Rev. Lett. **80**, 468 (1998).
10. S. Bize, S. A. Diddams, U. Tanaka, C. E. Tanner, W. H. Oskay, R. E. Drullinger, T. E. Parker, T. P. Heavner, S. R. Jefferts, L. Hollberg, W. M. Itano, J. C. Bergquist: Phys. Rev. Lett. **90**, 150802 (2003).
11. W. M. Itano, J. C. Bergquist, J. J. Bollinger, J. M. Gilligan, D. J. Heinzen, F. L. Moore, M. G. Raizen, D. J. Wineland: Phys. Rev. A **47**, 3554 (1993).
12. T. Ido, H. Katori: Phys. Rev. Lett. **91**, 053001 (2003).
13. G. Wilpers: Ein optisches Frequenznormal mit kalten und ultrakalten Atomen. PhD Thesis, University of Hannover, Hannover (2002) and PTB-Bericht PTB-Opt-66 (ISBN 3-89701-892-6), Physikalisch–Technische Bundesanstalt, Braunschweig (2002).
14. T. Trebst, T. Binnewies, J. Helmcke, F. Riehle: IEEE Trans. Instrum. Meas., **50**, 535 (2001).
15. G. Wilpers, C. Degenhardt, T. Binnewies, A. Chernyshov, F. Riehle, J. Helmcke, U. Sterr: Appl. Phys. B **76**, 149 (2003).
16. J. Helmcke, G. Wilpers, T. Binnewies, C. Degenhardt, U. Sterr, H. Schnatz, F. Riehle: IEEE Trans. Instrum. Meas. **52**, 250 (2003).
17. T. Binnewies, G. Wilpers, U. Sterr, F. Riehle, J. Helmcke, T. E. Mehlstäubler, E. M. Rasel, W. Ertmer: Phys. Rev. Lett. **87**, 123002 (2001).

18. U. Sterr, T. Binnewies, C. Degenhardt, G. Wilpers, J. Helmcke, Fritz Riehle: J. Opt. Soc. Am. B **20**, 985 (2003).
19. E. A. Curtis, C. W. Oates, L. Hollberg: Phys. Rev. A **64**, 031403(R) (2001).
20. E. A. Curtis, C. W. Oates, L. Hollberg: J. Opt. Soc. Am. B **20**, 977 (2003).
21. C. J. Bordé: Phys. Lett. A **140**, 10 (1989).
22. P. R. Berman, ed.: *Atom interferometry* (Academic Press, San Diego, 1997).
23. R. W. P. Drever, J. L. Hall, F. V. Kowalski, J. Hough, G. M. Ford, A. J. Munley, H. Ward: Appl. Phys. B **31**, 97 (1983).
24. C. J. Bordé, C. Salomon, S. Avrillier, A. Van Lerberghe, C. Bréant, D. Bassi, G. Scoles: Phys. Rev. A **30**, 1836 (1984).
25. L. Hollberg, C. W. Oates, E. A. Curtis, E. N. Ivanov, S. A. Diddams, T. Udem, H. G. Robinson, J. C. Bergquist, R. J. Rafac, W. M. Itano, R. E. Drullinger, D. J. Wineland: IEEE J. Quantum Electron. **37**, 1502 (2001).
26. *Guide to the Expression of Uncertainty in Measurement*, ISO/TAG 4. Published by ISO, 1993 (corrected and reprinted, 1995) in the name of the BIPM, IEC, IFCC, ISO, IUPAC, IUPAP and OIML, ISBN number: 92-67-10188-9, (1995).
27. J. Stenger, T. Binnewies, G. Wilpers, F. Riehle, H. R. Telle, J. K. Ranka, R. S. Windeler, A. J. Stentz: Phys. Rev. A **63**, 021802(R) (2001).
28. W. Nagourney, J. Sandberg, H. Dehmelt, Phys. Rev. Lett. **56**, 2797 (1986).
29. N. Beverini, F. Strumia: 'High precision measurements of the Zeeman effect in the Calcium metastable states'. In: *Interaction of Radiation with Matter, A Volume in Honour of A. Gozzini*, Quaderni della Scuola Normale Superiore de Pisa, Pisa 1987 pp. 361 - 373.
30. C. W. Oates, F. Bondu, R. W. Fox, L. Hollberg: Eur. Phys. J. D **7**, 449 (1999).
31. K. Zeiske: Atominterferometrie in statischen elektrischen Feldern. PhD Thesis, University of Hannover, Hannover (1995) and PTB-Ber. **PTB-Opt-48** Physikalisch–Technische Bundesanstalt Braunschweig, Braunschweig (1995).
32. J. W. Farley, W. H. Wing: Phys. Rev. A. **23**, 2397 (1981).
33. S. Ghezali, Ph. Laurent, S. Lea, A. Clairon: Europhys. Lett. **36**, 25 (1996).
34. D. A. Jennings, C. R. Pollock, F. R. Petersen, R. E. Drullinger, K. M. Evenson, J. S. Wells, J. L. Hall, H. P. Layer: Opt. Lett. **8**, 136 (1983).
35. H. Schnatz, B. Lipphardt, J. Helmcke, F. Riehle, G. Zinner: Phys. Rev. Lett. **76**, 18 (1996).
36. C. Degenhardt et al., to be published.
37. S. G. Karshenboim: Canad. J. Phys. **78**, 639 (2000).
38. C. Salomon, D. Hils, J. L. Hall: J. Opt. Soc. Am. B **5**, 1576 (1988).
39. B. C. Young, F. C. Cruz, W. M. Itano, J. C. Bergquist: Phys. Rev. Lett. **82**, 3799 (1999).

Trapped Ion Optical Frequency Standards for Laboratory Tests of Alpha-Variability

Christian Tamm, Tobias Schneider, and Ekkehard Peik

Physikalisch-Technische Bundesanstalt, Time and Frequency Department, Bundesallee 100, 38116 Braunschweig, Germany

Abstract. Optical transition frequencies of trapped laser-cooled ions have been measured so far with an uncertainty of 10^{-14} with reference to primary cesium clocks. Systematic frequency shifts in these systems should be controllable to significantly higher accuracy. We have performed comparisons between frequency standards based on two independently stored $^{171}Yb^{+}$ ions. The present experimental results indicate a relative instability (Allan standard deviation) of the optical frequency difference between the two systems of $1.0 \cdot 10^{-15}$ at an averaging time of 1000 s and a mean frequency difference of 0.2 Hz for the reference frequency of 688 THz. Monitoring the absolute transition frequency over a period of a few years at this level of accuracy will lead to a stringent limit on the present value of the temporal derivative of the fine-structure constant. First results have been obtained with a frequency standard based on $^{199}Hg^{+}$ at NIST and, more recently, also with Yb^{+}. The combination of results obtained with different transitions can be used to separate the electronic and nuclear contributions to a drift of a transition frequency. We propose optical spectroscopy of a nuclear transition in $^{229}Th^{3+}$ as a means for a precision monitoring of possible temporal variations in the strength of the nuclear forces.

1 Introduction

Among all physical quantities, time intervals and frequencies are the ones that can be measured to the highest precision, thanks to the progress in the development of atomic clocks and frequency standards. Cesium fountain clocks with laser cooled atoms realize the SI second to an uncertainty of $1 \cdot 10^{-15}$ only [1,2]. Apart from counting, frequency metrology can also provide the most precise technique to determine a dimensionless quantity, if it is available in the form of a frequency ratio: a predetermined ratio of two optical frequencies was recently measured with an error of only $7 \cdot 10^{-19}$ [3]. For these reasons, frequency standards and clocks are very well suited for highly sensitive tests of the foundations of physics like searches for violations of Local Lorentz Invariance or for a temporal variability of fundamental constants. Presently, a number of research groups is investigating a variety of atomic systems as candidates for future atomic clocks, focusing especially on forbidden electronic transitions in the optical frequency range. Precise frequency comparisons between different atoms over a time scale of a few years can be used for sensitive tests for a nonvanishing temporal derivative of the fine-structure constant α at the present epoch, complementing the geophysical and astrophysical searches, which are sensitive at geological or

C. Tamm, T. Schneider, E. Peik, Trapped Ion Optical Frequency Standards for Laboratory Tests of Alpha-Variability, Lect. Notes Phys. **648**, 247–261 (2004)
http://www.springerlink.com/

cosmological time scales [4]. The list of "recommended radiations for the practical realization of the definition of the metre" from 2001 includes seven optical transition frequencies with uncertainties in the range 10^{-13} and below in the positive ions of indium, mercury, ytterbium, and strontium, in hydrogen and calcium atoms, and in the OsO_4 molecule [5]. Several more measurements have been performed since then. The lowest relative uncertainty of $1 \cdot 10^{-14}$ in optical frequency measurements has been reached for the transitions $^2S_{1/2} \rightarrow {}^2D_{5/2}$ in $^{199}Hg^+$ (1 064 721 609 899 143 ± 10 Hz) at NIST [6] and for $^2S_{1/2} \rightarrow {}^2D_{3/2}$ in $^{171}Yb^+$ (688 358 979 309 312 ± 6 Hz) at PTB [7], measured with single ions in Paul traps.

2 The Single Ion as a Reference in an Optical Clock

If an atom is held in a trap, several problems that are often encountered in high resolution spectroscopy can be eliminated. One has however to be careful not to perturb the internal level structure with the trap fields, so that the resonance frequency of the free atom can be determined. Ions are very convenient in this respect because they carry electric charge as a "handle", and radiofrequency ion traps (Paul traps [8], cf. Fig. 1) can provide confinement of an ion around the field-free saddle point of an electric quadrupole potential. Using a trap, the interaction time is not limited by the movement of the atom through the finite interaction region, and narrow resonances can be obtained at the limits set by the radiative lifetime of the excited state or by the (technical) linewidth of the interrogating oscillator. Combined with laser cooling, the ion can be brought to the vibrational ground state of the trap potential [9], where the localisation and residual kinetic energy are only determined by quantum limits. Especially for an optical frequency standard the tight confinement in an ion trap is beneficial,

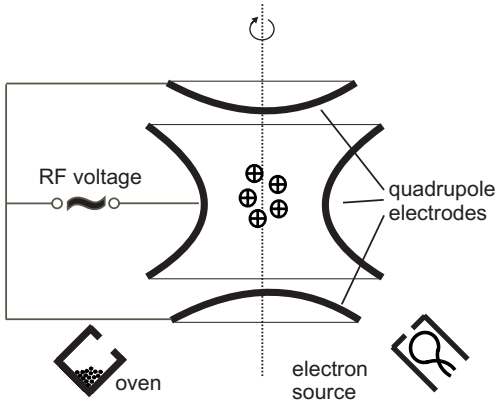


Fig. 1. Schematic of a Paul trap: atoms are evaporated from the oven, ionized by electron impact in the trap and stored in an alternating quadrupole field generated by rotationally symmetric ring and endcap electrodes. For single-ion experiments, the diameter of the ring is of the order of 1 mm.

because it is possible to reach the so-called Lamb–Dicke regime where the oscillation amplitude of the particle is much smaller than the wavelength of the radiation that is used to probe it. In this case, the frequency shifts due to the linear Doppler effect and due to a possible curvature of the phase front of the radiation can be eliminated. The remaining quadratic Doppler shift is usually smaller than a fraction of 10^{-18} of the transition frequency for a laser-cooled ion. If trapped in ultrahigh vacuum the ion will only rarely undergo a collision and it interacts with its environment mainly via relatively well controllable electric fields. These advantages were first pointed out by Dehmelt in the 1970s when he published the proposal of the *mono-ion oscillator* [10] and predicted that it should be possible to reach an accuracy of 10^{-18} with an optical clock based on a dipole-forbidden, narrow-linewidth transition in a single, laser-cooled and trapped ion.

To measure the transition probability on the forbidden optical transition, Dehmelt proposed the *electron shelving* scheme that allows an observation of quantum jumps. The scheme is most conveniently applied if the ion possesses a V-level system, where both a dipole-allowed transition and the forbidden reference transition of the optical clock can be driven with two different laser frequencies from the ground state. The dipole transition is used for laser cooling and the resulting resonance fluorescence can be used for the optical detection of the ion. If the second laser excites the ion to the metastable upper level of the reference transition, the fluorescence disappears and the ion will only light up again after its decay from the metastable state. Every single excitation of the reference transition suppresses the subsequent scattering of a large number of photons on the cooling transition and can thus be detected with practically unity efficiency. To use this kind of double resonance spectroscopy in an optical clock, the two laser radiations have to be applied alternately in time, because the simultaneous presence of cooling laser radiation would lead to a strong light shift of the frequency of the reference transition.

A number of suitable reference transitions with natural linewidths of the order of 1 Hz or below are available in different ions, so that a fractional resolution $\Delta\nu/\nu_0 = 10^{-15}$ is possible. In this case, even with a single ion, a frequency standard can be realized with a relative instability of 10^{-15} in an averaging time of one second, about one order of magnitude better than the best microwave oscillators. Within a few seconds of measurement time, such a clock would allow a frequency measurement at the level of accuracy that the present cesium fountains can only reach after several hours of averaging. Since the instability will decrease inversely proportional to the square root of the averaging time, it will still require a rather lengthy averaging of about 10^6 s, or roughly ten days to reach a final accuracy of 10^{-18}.

Presently, a number of groups is pursuing research along the lines of the mono-ion oscillator proposal with different ions (see [11] for a recent review). High-resolution spectroscopy and precise frequency measurements using femtosecond laser frequency comb generators [12] (see also the contributions by Fischer et al. [13] and by Ye et al. [14] in this book) have been performed recently on the alkali-like Sr^+ [15,16], Hg^+ [17,18], Yb^+ [7,19–21], and the alkaline-earth-

like In^+ [22,23]. Different types of forbidden transitions are investigated in these ions: For the alkali-like ions, higher-order multipole processes lead to the excitation of low-lying metastable ^2D or ^2F levels from the ^2S ground state. In In^+ a transition $^1S_0 \rightarrow {}^3P_0$ exists that connects two levels with total electronic angular momentum $J = 0$ and is made only weakly allowed via the mixing of the electronic and the nuclear spin [23]. The various transitions show different sensitivities to possible systematic frequency shifts and also the technical difficulties associated with trapping and laser cooling of the various elements are diverse.

3 Spectroscopy of the 435.5 nm Clock Transition of $^{171}Yb^+$

In the following, some recent results from the work on $^{171}Yb^+$ at PTB will be presented. This ion is an attractive candidate for optical frequency standards because reference transitions with vanishing low-field linear Zeeman frequency shift are available in a level system with relatively simple hyperfine and magnetic sublevel structure (cf. Fig. 2). In the PTB experiment the electric quadrupole transition $(^2S_{1/2}, F = 0) \rightarrow (^2D_{3/2}, F = 2)$ at 436 nm wavelength serves as the reference transition, with a natural linewidth of 3.1 Hz. The group at NPL [16] is investigating the $(^2S_{1/2}, F = 0) \rightarrow (^2F_{7/2}, F = 3)$ electric octupole transition at 467 nm. The $^2F_{7/2}$ state has a natural lifetime of the order of 10 years, so that the linewidth of this transition can be extremely narrow.

For laser cooling, the low-frequency wing of the quasi-cyclic $(F = 1) \rightarrow (F = 0)$ component of the $^2S_{1/2} \rightarrow {}^2P_{1/2}$ resonance transition is excited, and a static magnetic field of approximately 0.3 mT is applied in order to prevent optical

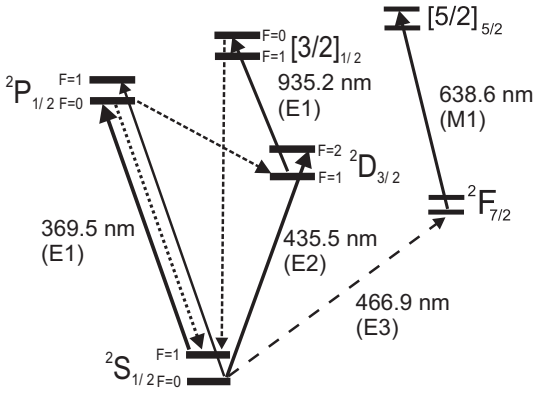


Fig. 2. Level scheme of the $^{171}Yb^+$ ion. The $S - P$ electric dipole (E1) transition is used for laser cooling. The $S - D$ quadrupole (E2) transition or the extremely narrow $S - F$ octupole (E3) transition can be used as reference transitions of the frequency standard. The other lines indicate "repumping transitions" that are driven to ensure cyclic excitation of the ion during the laser cooling phase.

pumping to a nonabsorbing superposition of the magnetic sublevels of the $F = 1$ ground state. The natural linewidth of the resonance transition is 21 MHz, which implies a one-dimensional kinetic temperature of 0.6 mK at the Doppler cooling limit. A weak sideband of the cooling radiation provides hyperfine repumping from the $F = 0$ ground state to the $^2P_{1/2}(F = 1)$ level. At the end of each cooling phase, the hyperfine repumping is switched off in order to prepare the ion in the $F = 0$ ground state. The rapid spontaneous decay from the $^2P_{1/2}$ state to the $(F = 2)$ sublevel of the metastable $^2D_{3/2}$ state that occurs during laser cooling is compensated for by coupling this level to the $[3/2]_{1/2}(F = 1)$ state, from where the ion readily returns to the ground state. The $(F = 2)$ sublevel of the $^2D_{3/2}$ state is not rapidly populated or depleted by the laser cooling excitation. Individual quantum jumps to this state due to excitation of the reference transition can therefore be detected using the electron shelving scheme.

Figure 3 shows four excitation spectra of the $S - D$ electric quadrupole transition on largely different frequency scales from the megahertz to the hertz range to illustrate some particularities of this type of spectroscopy. In Fig. 3a, the carrier at detuning zero is visible together with two sidebands at ± 0.7 MHz detuning. These sidebands are due to the 0.7 MHz radial oscillation of the ion in the trap. Their relative weakness in comparison to the carrier shows that the Lamb–Dicke regime is reached. Prior to the excitation of the reference transition, the magnetic field is reduced to the microtesla range in order to decrease the quadratic Zeeman frequency shift. In Fig. 3b the resolved Zeeman structure of the carrier is shown in a magnetic field of 1.1 μT. At this field strength, the quadratic Zeeman shift of the central $(m_F = 0) \rightarrow (m_F = 0)$ component is only 50 mHz. The spectrum of this component in Fig. 3c shows that it is possible to excite the ion with close to unity efficiency using π-pulses of 10 ms duration, leading to a 900 Hz wide Fourier-limited resonance. A high-resolution spectrum obtained with 90 ms long laser pulses is shown in Fig. 3d. Here one obtains an approximately Fourier-limited linewidth in the range of 10 Hz, or a resolution $\Delta\nu/\nu$ of $1.4 \cdot 10^{-14}$. Since the duration of the probe pulse is longer than the lifetime of the excited state (51 ms), the observed maximum excitation probability is limited by spontaneous decay. In order to operate the system as a frequency standard, both wings of the central resonance of the absorption spectrum are probed alternately, and the probe light frequency is stabilized to the line center according to the difference of the measured excitation probabilities.

4 Absolute Transition Frequency and Frequency Comparison Between Two Ions

Using a femtosecond-laser frequency comb generator [12] the frequency of the 436 nm $^2S_{1/2}(F = 0) \rightarrow {}^2D_{3/2}(F = 2, m_F = 0)$ transition of $^{171}Yb^+$ was measured relative to a cesium fountain microwave frequency standard on three days in December 2000 and January 2001 [7]. A Fourier-limited resonance linewidth of 30 Hz was used, otherwise the optical-excitation conditions were similar to those

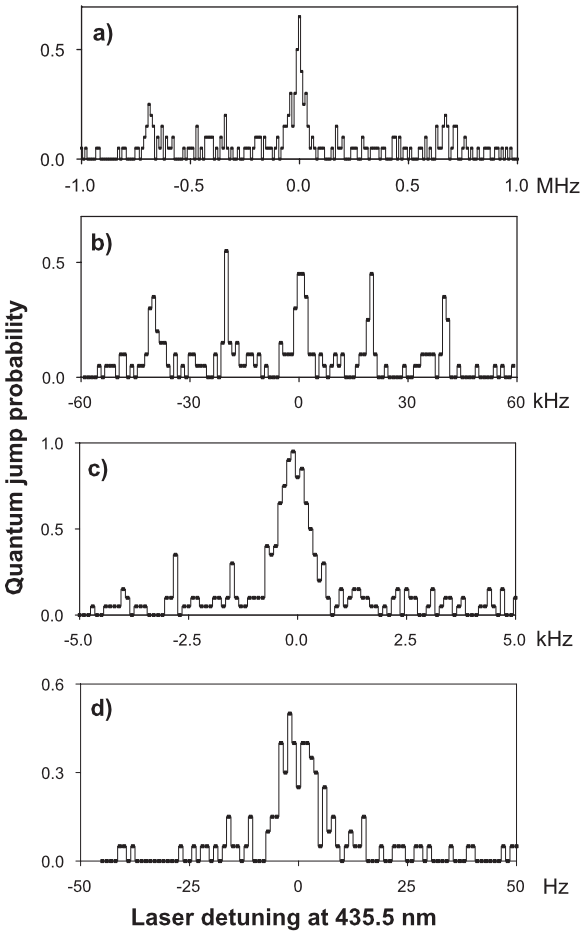


Fig. 3. Excitation spectra of the $S - D$ quadrupole transition of a single laser-cooled $^{171}Yb^{+}$ ion, obtained with the electron shelving method. Note the different frequency scales: (a) central carrier resonance and first order motional sidebands; (b) Zeeman pattern of the carrier in a weak magnetic field with the central $\Delta m_F = 0$ component and the adjacent $\Delta m_F = \pm 1, \ \pm 2$ lines; (c) π-pulse excitation of the $\Delta m_F = 0$ component with 900 Hz resolution; (d) high-resolution scan of the carrier with an approximately Fourier-limited linewidth of 10 Hz. Each data point is the result of 20 cycles of alternating interrogation of the reference and cooling transitions for each laser detuning.

of Fig. 3(d). The measured absolute frequency was 688 358 979 309 312 ± 6 Hz. The total 1σ measurement uncertainty of ±6 Hz consists of approximately equal statistical and systematic contributions. The dominant source of the systematic measurement uncertainty is the so-called quadrupole shift: the shift of the atomic transition frequency due to the interaction of the electric quadrupole moment of the D state with the gradient of the static electric field at trap center. Such

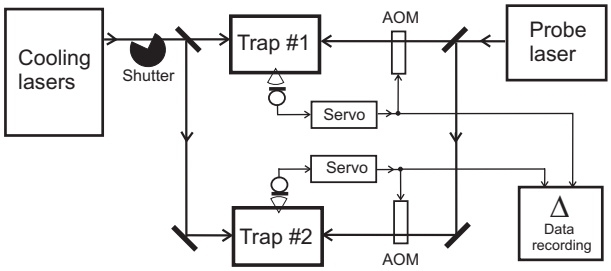


Fig. 4. Experimental setup for the comparison of two $^{171}Yb^{+}$ frequency standards. AOM: acusto-optic modulators, providing independent frequency shifts between the probe laser and the two ion traps.

a field gradient may be produced by patch charges on the trap electrodes. A maximum quadrupole shift of the order of 1 Hz is expected for the $D_{3/2}$ state under the conditions of the experiment. Another non-negligible systematic uncertainty contribution arises from servo errors due to drifts of the probe laser frequency. The uncertainty contributions of other frequency shifting effects are presently negligible: The magnetic field applied during the excitation of the reference transition leads to a quadratic Zeeeman shift of only 50 mHz. Since the trapped ion is cooled to the Doppler limit, the second-order Doppler and Stark effect shifts caused by the trap field are expected to be in the millihertz range.

To enable a quantitative study of these sub-Hertz systematic frequency shifts, we compare the line center frequencies of two $^{171}Yb^{+}$ ions stored in separate traps. The scheme of the frequency comparison experiment is shown in Fig. 4. Both traps use the same cooling laser setup and synchronous timing schemes for cooling, state preparation, and state detection. The beam from the probe laser is split and two separate frequency shift and servo systems are employed to stabilise the probe frequencies to the reference transition line centers of the two ions. The atomic resonance signals were resolved with nearly Fourier-limited linewidths of approximately 30 Hz in both traps. Using two independent digital servo systems, the error signals resulting from the probing of the atomic resonance signals are averaged over typically 20 measurement cycles before the detunings between the probe laser frequency and the probe light beams incident on the traps are corrected. The servo time constants are in the range of 30 s. In order to minimise servo errors due to the drift of the probe laser frequency, a second-order integrating servo algorithm is used. The differences of the detunings imparted on the probe beams are averaged over time intervals of 1 s and recorded. As a measure of the instability, the Allan deviation of this data set is shown in Fig. 5. For averaging times that are longer than the time constants of the servo systems, the so-called quantum projection noise [24] will limit the instability. We have performed a numerical calculation which simulates the effect of quantum projection noise for the realized experimental conditions and include the result as the dashed line in Fig. 5. The observed Allan deviation exceeds the quantum projection noise limit by approximately a factor of two. A possible reason for this

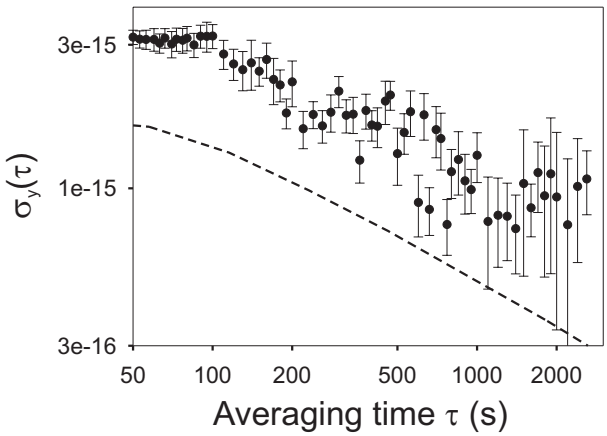


Fig. 5. Allan standard deviation of the frequency differences observed between the two traps, normalized to the optical frequency of 688 THz. The dashed line shows the result of a Monte Carlo simulation of the servo action for the case that the fluctuations of the atomic resonance signals are determined by quantum projection noise.

excess instability are temporal fluctuations of the probe laser emission spectrum which can lead to fluctuating servo errors. The frequency shifts caused by this effect are not necessarily equal for both servo systems because the probe pulse areas by which the two ions were excited were not exactly matched and showed some temporal fluctuations.

The mean frequency difference between the two ions averaged over 8000 s is 0.2 Hz, corresponding to a relative frequency offset of $3 \cdot 10^{-16}$. Since this is smaller than the Allan deviation for long averaging times ($\sigma_y(\tau) \approx 1 \cdot 10^{-15}$ for $\tau \geq 800$ s), the observed offset is not statistically significant. This measurement constitutes, to our knowledge, the most precise experimental test of the identicalness of a specific property of two individual atoms. Thus, this measurement might be considered as a quantitative test of one of the foundations of quantum physics without ensemble averaging. There is of course other evidence for the identity of particles from the successful application of quantum statistics, like for example in systems of degenerate fermions or Bose-Einstein condensates. Stringent experimental tests have been performed with elementary fermions, where violations of the Pauli principle have been searched for (see [25] and references therein).

5 Search for Temporal Variation of the Fine-Structure Constant

The high precision of the single-ion optical frequency standards makes them suitable systems for laboratory searches for drifts of the fundamental constants. Following the analysis of Dzuba, Flambaum et al., relativistic effects in many-electron atoms lead to different sensitivities of the transition frequencies to

Table 1. Sensitivity factor $a = \alpha\,\partial \ln f/\partial\alpha$ for the α-dependence of the transition frequencies of single-ion frequency standards (relative change of frequency f per relative change of α) from calculations by Dzuba et al. for Sr^+, Hg^+ [26], In^+, Tl^+, Ba^+ [27], and Yb^+ [28]. The last column gives the absolute frequency change Δf_{-15} in Hertz for a change of $\Delta\alpha/\alpha = 10^{-15}$.

ion	transition	f [THz]	a	Δf_{-15} [Hz]
Sr^+	$5s\ S_{1/2} \rightarrow 4d\ D_{3/2}$	437	0.389	0.17
Sr^+	$5s\ S_{1/2} \rightarrow 4d\ D_{5/2}$	445	0.428	0.19
Ba^+	$6s\ S_{1/2} \rightarrow 5d\ D_{3/2}$	145	2.41	0.35
Ba^+	$6s\ S_{1/2} \rightarrow 5d\ D_{5/2}$	170	2.11	0.36
Hg^+	$5d^{10}6s\ S_{1/2} \rightarrow 5d^9 6s^2\ D_{3/2}$	1517	-1.74	-2.64
Hg^+	$5d^{10}6s\ S_{1/2} \rightarrow 5d^9 6s^2\ D_{5/2}$	1065	-3.19	-3.40
Yb^+	$4f^{14}6s\ S_{1/2} \rightarrow 4f^{14}5d\ D_{3/2}$	688	0.881	0.61
Yb^+	$4f^{14}6s\ S_{1/2} \rightarrow 4f^{14}5d\ D_{5/2}$	730	0.855	0.62
Yb^+	$4f^{14}6s\ S_{1/2} \rightarrow 4f^{13}6s^2\ F_{7/2}$	642	-5.30	-3.40
In^+	$5s^2\ {}^1S_0 \rightarrow 5s5p\ {}^3P_0$	1268	0.209	0.27
In^+	$5s^2\ {}^1S_0 \rightarrow 5s5p\ {}^3P_2$	1375	0.341	0.47
Tl^+	$6s^2\ {}^1S_0 \rightarrow 6s6p\ {}^3P_0$	1484	0.799	1.19
Tl^+	$6s^2\ {}^1S_0 \rightarrow 6s6p\ {}^3P_2$	1852	1.03	1.91

a change of the fine-structure constant α, with the general trend to find the stronger dependence in the heavier atoms. Table 1 summarizes the relevant results from [26–28] for the transitions that are under study in single-ion frequency standards. While for most atomic transitions the sign of the frequency change is positive for increasing α, the D states of Hg^+ and the F state of Yb^+ present special cases with a negative shift. These states do not arise from the excitation of an s electron to a state of higher angular momentum, but from the excitation of an electron from the filled d and f states respectively to an s level of the next higher shell. The availability of both, positively and negatively shifting transition frequencies could be used to distinguish the signature of a varying fine-structure constant from possible systematic shifts. The Yb^+ ion seems especially promising for this kind of test because in the same system, the $S-D$ and $S-F$ transition frequencies can be compared, which would shift in opposite directions with quite high sensitivity factors a. From the data in Table 1, it can be seen that with the help of at least two different single-ion frequency standards with an accuracy of 1 Hz, a relative change of α of 10^{-15} could be detected with high certainty.

So far no direct frequency comparisons between different single-ion transitions have been reported. The most precise absolute frequency measurements have been performed with reference to local primary cesium clocks, so that a

possible drift of the SI second as realized by these clocks has also to be taken into account. Since the strong interaction contributes to the cesium hyperfine frequency via the nuclear magnetic moment, a relative drift of the latter may even be larger than those of the electronic transition frequency if both time derivatives would be evaluated independently. Under the assumption that the drift of the cesium frequency is linear in time over the duration of the optical frequency comparisons, the drift rate $\partial \ln \alpha / \partial t$ can be obtained from at least *two* measured drift rates of transition frequencies $\partial \ln f / \partial t$ if these have been measured against cesium clocks, and if the sensitivity factors a for the two transitions are sufficiently different. As noted in [29] and in the introduction by Karshenboim and Peik in this book [30], the plot of $\partial \ln f / \partial t$ as a function of a should then yield a straight line where the slope is exclusively determined by $\partial \ln \alpha / \partial t$ and any drifts of the duration of the SI second relative to a common scaling factor like the Rydberg constant would show up in the intercept (cf. Fig. 1 of Chap. 1). Fortunately, the two transitions for which the highest precision frequency measurements are available (Hg^+ $S_{1/2} \rightarrow D_{5/2}$ at 1065 THz and Yb^+ $S_{1/2} \rightarrow D_{3/2}$ at 688 THz) differ in a by 4.1, so that a sensitive determination of the slope is possible.

The mercury ion is investigated as the reference of an optical frequency standard at NIST in Boulder. Laser cooling is performed on the $^2S_{1/2} \rightarrow {}^2P_{1/2}$ transition at 194 nm wavelength and the transition $^2S_{1/2} \rightarrow {}^2D_{5/2}$ at 282 nm (1065 THz) with a natural linewidth of 1.9 Hz serves as the reference. In contrast to the case of Yb^+ (cf. Fig. 2), the branching from the P to the D states is negligible, so that no additional repumping lasers are needed. As in the case of Yb^+, an isotope with nuclear spin 1/2 is used (^{199}Hg) so that a $(F = 0, m_F = 0) \rightarrow (F = 2, m_F = 0)$ component of the reference transition with vanishing linear Zeeman shift is available. In order to reduce the background pressure of neutral mercury, the ion trap is operated at the temperature of liquid helium. The reference transition has been observed with a resolution of 6 Hz [17] and a first frequency measurement with an uncertainty of $9 \cdot 10^{-15}$ was published in 2001 [6]. Since $S - D$ quadrupole transitions are used in both cases, the sensitivities of the Hg^+ frequency standard to systematic shifts [31] are in general comparable to those of Yb^+. Again, the quadrupole shift due to electric field gradients is the systematic effect that is presently most difficult to evalutate, and a 10 Hz estimate for this contribution dominates the uncertainty. Monitoring the frequency with respect to the cesium clock over a period of two years has resulted in a constraint for the fractional variation of f_{Hg}/f_{Cs} at the level of $\pm 7 \cdot 10^{-15}$ 1/yr [18].

The Yb^+ frequency has been remeasured in November 2003 [33], i.e. 34 months after the first measurement [7]. In comparison, the new frequency value is insignificantly lower by an amount of 2 Hz, where the 1σ uncertainty of the new measurement is ± 6 Hz, with similar contributions as in [7]. Consequently, we can deduce a limit on the fractional variation of $f_{\mathrm{Yb}}/f_{\mathrm{Cs}}$ of $(-1.2 \pm 4.4) \cdot 10^{-15}$ 1/yr. The time spans covered by the frequency measurements are MJD 51760 – 52580 for Hg^+ and MJD 51889 – 52947 for Yb^+ (MJD: modified Julian date), so that there is a significant temporal overlap. Any drifts between the two cesium clocks

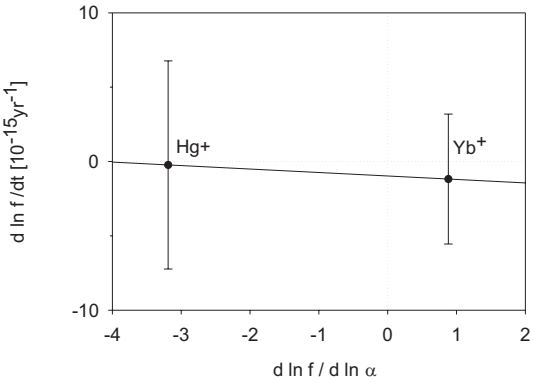


Fig. 6. Relative frequency drift rate versus sensitivity factor a for the reference transitions in the Hg^+ and Yb^+ ions. A significant deviation of the slope of the straight line from zero would indicate a non-zero time derivative of α.

that were used as local references can be excluded at the level of 10^{-15} 1/yr. The time scales of NIST and PTB are compared continuously and the results are documented via the contributions that both laboratories make to the realization of international atomic time TAI [32]. In particular, the cesium fountains of both laboratories have been found to be in agreement to within their evaluated uncertainty of $1.0 \cdot 10^{-15}$ [2].

Combining the 1σ limits on the drift rates of the frequencies of Hg^+ and Yb^+ (cf. Fig. 6) and dividing by the difference between the sensitivity factors for the two transitions we obtain

$$\frac{\partial \ln \alpha}{\partial t} = (-0.2 \pm 2.0) \cdot 10^{-15}\ \mathrm{yr}^{-1} \tag{1}$$

for the present value of the temporal derivative of the fine-structure constant at the confidence level of 1σ. This is so far the most stringent limit on $\dot{\alpha}$ obtained from atomic clock comparisons. This result is based on a largely model-independent analysis [30], not using assumptions on correlations between drift rates of different quantities and not postulating the drift rate of any quantity to be zero.

The intercept of the straight line of optical frequency drift rate with the line $a = 0$ (cf. Fig. 6) can be used to determine the drift rate of the common scaling factor against the cesium hyperfine frequency, i.e. of the Rydberg constant in the SI unit Hertz. The 1σ limit that we can deduce for this quantity is $(-1.0 \pm 3.7) \cdot 10^{-15}$ 1/yr. This quantity is related to the magnetic moment of the Cs nucleus, but, as discussed earlier in this book [30], there is no straightforward relation to fundamental constants.

6 Nuclear Optical Frequency Standard with Th-229

In the context of a laboratory search for temporal variations of the fundamental coupling constants, it would be favorable to have available different types of fre-

quency standards whose operating frequency would be determined in different ways by the four fundamental interactions. A pendulum clock for example may be regarded as a frequency standard that probes the local gravitational force at a rather limited accuracy. The observation of the periods of millisecond pulsars offers a much more promising probe of gravitational physics. Present atomic frequency standards can be categorized in two classes: Those that make use of atomic fine or gross structure transitions (like the single-ion optical frequency standards discussed so far) are predominantely determined by the electromagnetic interaction. The frequencies of those that use hyperfine transitions (e.g. the cesium clock) or molecular vibrations are mainly determined by the electromagnetic interaction but also have important contributions from the strong interaction via nuclear moments and masses. We have recently proposed a system that may be an interesting complement in this discussion [34]: A frequency standard based on a radiative transition in a nucleus would constitute a high-precision oscillator whose frequency is predominantly determined by the strong interaction. The possibility of *optical* nuclear spectroscopy is offered in the Th-229 nucleus that possesses an isomeric state at the unusually low excitation energy of about 3.5 eV above the ground state [35]. The associated transition wavelength is in the range of 350 nm so that the methods of high resolution laser spectroscopy are applicable and an absolute measurement of the nuclear transition frequency is feasible.

Besides the more fundamental interest in monitoring the strength of the strong interaction, nuclear spectroscopy may offer considerable benefits for time and frequency metrology. Apart from motional and collisional frequency shifts that can be extremely well controlled in experiments with single trapped ions, the interaction with ambient electric or magnetic fields usually is the dominant source of systematic uncertainty in frequency standards. The nucleus is much less susceptible to this kind of perturbations than the atomic electron shell: its magnetic dipole moment is typically three orders of magnitude smaller than the magnetic moment of the electron shell, and for the electric quadrupole moment the typical electronic and nuclear scales differ by 8 orders of magnitude. It will often be required and sometimes even advantageous to look at the nucleus in the coupling to the atomic, molecular or crystalline electronic environment. Much of the required formalism is available from the study of hyperfine interactions in spectroscopy of isolated atoms, Mößbauer spectroscopy or nuclear magnetic resonance (NMR). From the viewpoint of atomic physics, the main differences are that the nuclear spin and the associated nuclear moments are not constant in nuclear spectroscopy, and that changes in the interaction between the nuclear charge distribution and the electronic charge density at the nucleus have to be regarded.

In order to illustrate the role of hyperfine interactions in nuclear spectroscopy of an isolated atom, let us consider the Zeeman and Stark shifts of the nuclear transition frequency. In an LS coupling scheme the eigenstates of the coupled electronic and nuclear system are characterised by sets of quantum numbers $|\alpha, I; \beta, L, S, J; F, m_F\rangle$, where I denotes the nuclear spin, L, S, J the orbital, spin and total electronic angular momenta, and F and m_F the total atomic

angular momentum and its orientation. α and β label the involved nuclear and electronic configurations. In the nuclear transition, the nuclear and total angular momentum quantum numbers (α, I, F, m_F) can change, while the purely electronic quantum numbers (β, L, S, J) remain constant. As far as external fields are concerned, the nuclear transition frequency is essentially independent of all mechanisms that produce level shifts depending only on the electronic quantum numbers (β, L, S, J), because these do not change and consequently the upper and the lower state of the transition are affected in the same way. This applies to the scalar part of the quadratic Stark effect, which typically is the dominant mechanism for the shift of electronic transition frequencies due to static electric fields, electromagnetic radiation, and collisions. The observed nuclear transition frequency is however shifted by the so-called hyperfine Stark shift, which depends on F and m_F and is calculated in a higher order of perturbation theory [36]. In the optical frequency range, a relative magnitude of typically 10^{-19} is expected for the hyperfine Stark shift caused by the ≈ 10 V/cm room temperature blackbody radiation field. In order to avoid the influence of the linear Zeeman effect, an electronic state can be chosen such that F is an integer. In this case a Zeeman component $m_F = 0 \rightarrow 0$ is available, that shows only a small quadratic Zeeman effect around zero magnetic field. Further field dependent shifts may arise from the tensor part of the quadratic Stark effect and from the quadrupole interaction between the atomic quadrupole moment and electric field gradients. Both these shifts can be expressed as a product of J-dependent and F-dependent terms and vanish if either $J < 1$ or $F < 1$ [31,36].

From these general considerations it can be seen that for every radiative nuclear transition, an electronic state can be selected which makes the hyperfine coupled nuclear transition frequency immune against the linear Zeeman effect and the quadratic Stark effect as well as the quadrupole shift. For electronic transitions, this combination of advantageous features cannot be obtained. (Even in the case of a hyperfine-induced transition between two levels with $J = 0$ like ${}^1S_0 \rightarrow {}^3P_0$, the electronic polarizabilities will in general be different in the two states, so that the quadratic Stark effect and collisional shifts can not be eliminated.) Since the selected electronic state has to be stable, the choice could be made for the nucleus under question among the ground states of its various positive ions of different charge. In the case of a half integer nuclear spin (like in Th-229), the optimal electronic states are ${}^2S_{1/2}$ or ${}^2P_{1/2}$, and for an integer nuclear spin the states ${}^1S_0$ or ${}^3P_0$ fulfil all criteria. The proposed scheme is readily applicable to low-energy nuclear transitions as available in Th-229 or U-235, while for typical Mößbauer transitions in the 10-100 keV range the competition with internal conversion has to be taken into account, which would alter the charge state of the ion.

In [34] we have shown that the 3.5 eV nuclear transition in Th-229 could be detected in a double resonance method if a single trapped and laser-cooled Th^{3+} ion is used. This system could form the reference in an optical clock of very high accuracy. Even if a calculation of the transition frequency in this very special nucleus with two nearly degenerate ground states may not be possible from first principles, it will be highly interesting to look for any possible drifts between the

reference frequency of this *nuclear* clock and electronic transition frequencies in $^{171}Yb^{+}$, $^{199}Hg^{+}$ or other accurate atomic clocks.

References

1. A. Bauch and H. R. Telle, Rep. Prog. Phys. **65**, 789 (2002).
2. A. Bauch, Meas. Sci. Technol. **14**, 1159 (2003).
3. J. Stenger, H. Schnatz, Chr. Tamm, and H. Telle, Phys. Rev. Lett. **88** 073601 (2002).
4. S. Karshenboim, Can. J. Phys. **78**, 639 (2001).
5. T. J. Quinn, Metrologia **40**, 103 (2003).
6. Th. Udem et al., Phys. Rev. Lett **86**, 4996 (2001).
7. J. Stenger, Chr. Tamm, N. Haverkamp, S. Weyers, and H. Telle, Opt. Lett. **26** 1589 (2001).
8. W. Paul, Rev. Mod. Phys. **62**, 531 (1990).
9. F. Diedrich, J. C. Bergquist, W. M. Itano, and D. J. Wineland, Phys. Rev. Lett. **62**, 403 (1989).
10. H. Dehmelt, IEEE Trans. Instrum. Meas. **31**, 83 (1982).
11. A. A. Madej, J. E. Bernard: 'Single Ion Optical Frequency Standards and Measurement of their Absolute Optical Frequency'. In: *Frequency Measurement and Control: Advanced Techniques and Future Trends, Springer Topics in Applied Research*, ed. by A. N. Luiten (Springer, Berlin, Heidelberg, 2000)
12. J. Reichert, M. Niering, R. Holzwarth, M. Weitz, Th. Udem, T.W. Hänsch, Phys. Rev. Lett. **84**, 3232 (2000); S.A. Diddams, D.J. Jones, J.Ye, S.T. Cundiff, J.L. Hall, J.K. Ranka, R.S. Windeler, R. Holzwarth, Th. Udem, T.W. Hänsch, Phys. Rev. Lett. **84**, 5102 (2000).
13. M. Fischer et al., Precision Spectroscopy of Atomic Hydrogen and Variations of Fundamental Constants, Lect. Notes Phys. **648**, 209–227 (2004)
14. J. Ye et al., Applications of Femtosecond Laser Comb to Nonlinear Molecular Spectroscopy, Lect. Notes Phys. **648**, 275–295 (2004)
15. J. E. Bernard, A. A. Madej, L. Marmet, B. G. Whitford, K. J. Siemsen, S. Cundy, Phys. Rev. Lett. **82**, 3228 (1999).
16. H. S. Margolis et al., Phys. Rev. A **67**, 032501 (2003).
17. R. J. Rafac, B. C. Young, J. A. Beall, W. M. Itano, D. J. Wineland, and J. C. Bergquist, Phys. Rev. Lett. **85**, 2462 (2000).
18. S. Bize et al., Phys. Rev. Lett. **90**, 150802 (2003).
19. Chr. Tamm, D. Engelke, and V. Bühner, Phys. Rev. A **61**, 053405 (2000).
20. Chr. Tamm, T. Schneider, and E. Peik, in: *Proc. of the 6. Symp. on Frequency Standards and Metrology*, Ed.: P. Gill (World Scientific, Singapore, 2002).
21. P. Blythe et al., Phys. Rev. A **67**, 020501(R) (2003).
22. J. v. Zanthier et al., Opt. Lett. **25**, 1729 (2000).
23. Th. Becker, J. von Zanthier, A. Yu. Nevsky, Ch. Schwedes, M. N. Skvortsov, H. Walther, E. Peik, Phys. Rev. A **63**, 051802(R) (2001).
24. W. Itano et al., Phys. Rev. A **47**, 3554 (1993).
25. E. Baron, R. N. Mohapatra, V. L. Teplitz, Phys. Rev. D **59**, 036003 (1999).
26. V. A. Dzuba, V. V. Flambaum, J. K. Webb, Phys. Rev. A **59**, 230 (1999).
27. V. A. Dzuba, V. V. Flambaum, Phys. Rev. A **61**, 034502 (2000).
28. V. A. Dzuba, V. V. Flambaum, M. V. Marchenko, Phys. Rev. A **68**, 022506 (2003).
29. S. G. Karshenboim, eprints physics/0306180 and physics/0311080 (to be published).

30. S.G. Karshenboim, E. Peik, An Introduction to Varying Fundamental Constants, Lect. Notes Phys. **648**, 1–18 (2004)
31. W. M. Itano, J. Res. NIST **105**, 829 (2000).
32. http://www.bipm.org/en/scientific/tai/time_server.html
33. E. Peik, B. Lipphardt, H. Schnatz, T. Schneider, Chr. Tamm, and S. Karshenboim, eprint physics/0402132 (to be published).
34. E. Peik and Chr. Tamm, Europhys. Lett. **61**, 181 (2003).
35. R. G. Helmer, C. W. Reich, Phys. Rev. C **49**, 1845 (1994).
36. J. R. P. Angel, P. G. H. Sandars, Proc. Roy. Soc. A **305**, 125 (1968).

An Optical Frequency Standard Based on the Indium Ion

Petrissa Eckle, Mario Eichenseer, Alexander Y. Nevsky, Christian Schwedes, Joachim von Zanthier, and Herbert Walther

MPI für Quantenoptik and Sektion Physik der LMU München,
Hans-Kopfermann-Str.1, 85748 Garching, Germany

Abstract. A single indium ion stored in a radio frequency trap and laser-cooled to a temperature below 1 mK can serve as an optical frequency standard of exceptionally high accuracy and stability. Probing the $5s^2\,{}^1S_0$ - $5s5p\,{}^3P_0$ transition of $^{115}In^+$ at $\lambda =$ 237 nm with a natural linewidth of 0.8 Hz by a sub-Hertz linewidth laser can lead to an accuracy of 10^{-18} after several days of averaging. So far, a fractional resolution of 1.3×10^{-13} has been obtained, limited by the frequency fluctuations of the laser exciting the ${}^1S_0$ - ${}^3P_0$ resonance. The absolute frequency of the ${}^1S_0$ - ${}^3P_0$ transition was determined to 1 267 402 452 899.92(0.23) kHz ($\Delta\nu/\nu = 1.8\times10^{-13}$), where the measurement uncertainty is dominated by the frequency uncertainty of the He–Ne laser serving as a reference. Further improvements in this measurement together with a comparison with other narrow atomic resonances would allow investigations of variations of fundamental constants in time.

1 Introduction

A single laser cooled ion stored in a radio frequency trap is considered to be one of the most favorable systems for realizing an optical atomic clock [1,2]. With a frequency-stable laser locked to a narrow absorption line of the trapped ion, a stability of $\sigma_y(1\text{ s}) = 10^{-15}$ and an accuracy of a few times 10^{-18} seem possible. So far, high-resolution spectroscopy of narrow optical lines in trapped ions has yielded linewidths below 50 Hz in Hg^+ [3] and Yb^+ [4] and sub-kHz resolution in Ba^+ [5], Sr^+ [6] and Ca^+ [7] and In^+ [8]. By directly locking the probing laser onto a narrow transition in Hg^+ and Yb^+, a stability of $\sigma_y(1\text{ s}) \leq 7\times10^{-15}$ [9] and an accuracy of $\leq 1\times10^{-14}$ were recently obtained in a comparison with a Cesium atomic fountain clock [10,11].

With the exception of In^+, all ions now being investigated for an optical frequency standard are characterized by an alkali-like level scheme where narrow transitions between the ground state with total electron angular momentum $J = 1/2$ and upper states with $J > 1/2$ are used. The quadrupole moment of the excited states, however, makes these transition susceptible to frequency shifts induced by external electric field gradients, which can be of the order of 1 Hz [12,13]. In comparison, In^+ shows an earth-alkali-like electron spectrum. Here, for the clock transition a hyperfine-induced electric dipole transition between the $5s^2\,{}^1S_0$ and the $5s5p\,{}^3P_0$ level at $\lambda = 237$ nm is used, with a natural linewidth

P. Eckle et al., An Optical Frequency Standard Based on the Indium Ion, Lect. Notes Phys. **648**, 263–272 (2004)
http://www.springerlink.com/

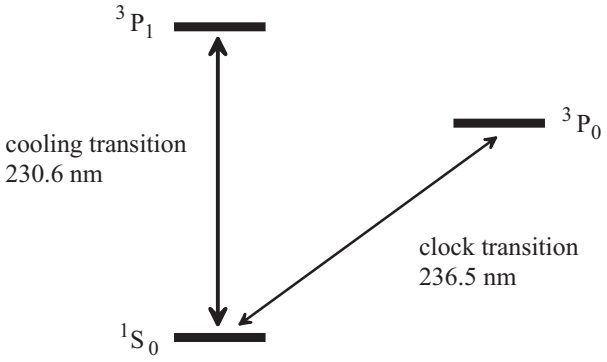


Fig. 1. Level scheme of In^+, showing the three lowest energy levels

of 0.8 Hz (see Fig. 1). For an optical frequency standard this configuration offers several advantages: since the electronic angular momenta in both states vanish, the frequency shifts of this transition due to external magnetic and electric fields are small; the second-order Stark (and also Doppler) shift is at the mHz level at the temperatures to which the ion is cooled in the trap. In particular, since the ion in both levels has no quadrupole moment, the transition is free of the quadrupole shift. The most significant perturbation is expected to result from a small deviation of the g-factor of the $^3P_0$ state from the g-factor of the $^1S_0$ ground state due to hyperfine-induced mixing between the $5s5p$ levels $^3P_1$, $^1P_1$ and $^3P_0$; this leads to a linear (nuclear) anomalous Zeeman shift which was determined to be 224 Hz/G for the $m_F = \pm 1/2 \rightarrow m_F = \pm 1/2$ Zeeman components of $^{115}In^+$ [8]. By controlling external magnetic fields at the level of a few μG, an accuracy of 10^{-18} is obtainable.

Detection and laser cooling of In^+ are performed using the $5s^2\,^1S_0$ - $5s5p\,^3P_1$ intercombination line at 231 nm. The dye laser formerly used for this purpose was recently replaced by a diode-based laser system [14]. The characteristics of this laser will be discussed in more detail below. The $5s^2\,^1S_0$ - $5s5p\,^3P_0$ clock transition is excited by a frequency-quadrupled diode-pumped Nd:YAG laser at 946 nm. This kind of laser has a high intrinsic frequency stability and a long service-free operating time. Excitation of the $^3P_0$ state is detected by probing the ground state population via the fluorescence on the fast $^1S_0$ - $^3P_1$ transition (electron shelving technique) [1]. In high-resolution spectroscopy experiments a resolution of 1.3×10^{-13} has been achieved, where the spectral width was limited by the frequency excursions of the Nd:YAG laser [8]. The frequency stability of the clock laser was strongly improved recently. With a new laser setup based on a quasi-monolithic Nd:YAG ring laser, a laser linewidth < 4 Hz (FWHM) for integration times ≤ 26 s has been obtained [15]. The high frequency-stability of the laser was achieved by locking the laser onto an external reference cavity of high finesse and placing the laser and the reference cavity on an active vibration isolation platform. So far, this laser has not been employed in further high-resolution experiments of the In^+ $^1S_0$ - $^3P_0$ clock transition.

2 Cooling of the Indium Ion

The ion is trapped in a radio frequency trap that is a modification of the original Paul trap and basically consists of only a ring electrode (Paul-Straubel trap [16]). This type of trap is relatively simple to fabricate in miniature size (1 mm diameter of the ring), making it easy to confine the ion to a region in space that is smaller than the optical wavelength (Lamb-Dicke regime). It is also a geometrically open structure that allows good optical access to the trapped ion. The trap is driven by an RF field at $\sim$ 10 MHz of $\sim$ 1000 V amplitude, leading to oscillation frequencies in the time-averaged pseudopotential of about 1.4 MHz in the axial and 0.9 MHz in the radial direction. The ion is detected and laser cooled on the $^1S_0$ – $^3P_1$ intercombination line. With a natural linewidth of 360 kHz [17], this transition is narrow enough to allow resolution of the secular frequencies of the ion in the trap, thus enabling laser sideband cooling of the particle [18]; however, it is also sufficiently broad to record a single ion via its resonance fluorescence and to cool the particle without precooling it on a second, faster transition, even when starting from very high mean vibronic quantum numbers. In the absence of hyperfine splitting of the ground state and with only a weak decay channel from the $^3P_1$ towards the lower-lying metastable $^3P_0$ state (decay rate $\sim$ 35 mHz [17]), a single laser source suffices for laser cooling. With sideband cooling on the $^1S_0$ – $^3P_1$ line we demonstrated reduction of the mean vibrational quantum number of a single In^+ ion from an initial $\langle n \rangle \approx 10^8$ to about $\langle n \rangle = 0.7$, corresponding to a temperature of $\approx$ 60 μK [19]. In the experiments reported here, the laser intensity was higher than in [19] and the detuning to the resonance was decreased to obtain a stronger fluorescence signal. This leads to a slightly increased temperature of about 150 μK, as measured from the strength of the vibrational sidebands in the spectrum of the $^1S_0$ – $^3P_0$ transition.

Since the sidebands of the residual micromotion at 10 MHz induced by stray electric fields can be seen directly in the spectrum, these fields can be compensated with high accuracy using additional DC electrodes mounted close to the trap. In this manner the micromotion of the ion can be reduced to very small values (see Fig. 2). For example, from the spectrum of Fig. 2b, a localization of better than $\lambda/20$ can be deduced, demonstrating that the Lamb-Dicke regime is well attained by this procedure.

3 New Cooling Laser System

In our former setup the ion was detected and cooled on the $^1S_0$ – $^3P_1$ transition by means of a frequency-doubled stilbene-3 ring dye laser at 461 nm, pumped by an Ar^+ laser. This system needs continuous maintenance for laser alignment, change of dye and even replacement of pump laser tubes; therefore, long-term operation as required for an ultra-stable optical clock is not possible with this laser. Furthermore, for tests of fundamental physics such as the search for a time variation of the fine-structure constant by comparing different optical clocks [20–22], a transportable standard would be advantageous. The aim was therefore to

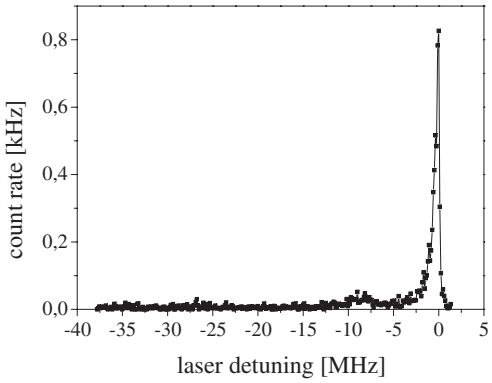


Fig. 2. Spectrum of the In^+ $^1S_0 - {}^3P_1$ cooling transition at 231 nm, after compensation of stray electric fields. The micromotion sidebands are strongly reduced

design a new compact, reliable and transportable laser system allowing permanent operation.

The new cooling laser system [14] consists of an extended-cavity diode laser (ECDL) with 30 mW of output power at 922 nm in Littrow configuration [23,24], where the laser frequency is controlled by the feedback of the first diffraction order of a grating. Due to the high-quality antireflection (AR) coating of the front facet of the diode, the tuning range of the ECDL is rather large, from 880 nm to 960 nm. About 20 per cent of its light is used for frequency stabilizing the laser onto a reference cavity. The main part of the ECDL light is seeded into a tapered amplifier (TA), resulting in a power of $\sim$ 0.5 W at the same frequency. The infrared light is then frequency-doubled a first time in an external enhancement cavity by means of a periodically poled KTP crystal, yielding more than 200 mW at 461 nm. A second external frequency-doubling stage using a BBO crystal generates 1 mW of UV light at 231 nm. Figure 3 shows the setup of the system.

For active frequency stabilization, the laser frequency is locked onto the resonance of a monolithic Fabry-Perot interferometer with a finesse of 11000 (linewidth: 100 kHz) using the Pound-Drever-Hall locking technique. For that purpose the beam is first led in double pass through an acousto-optical modulator to allow for frequency fine tuning and thereafter coupled via a single-mode fibre into the reference resonator. The sidebands, required for the lock, are generated by modulating the ECDL current at 10 MHz. The error signal is applied to the laser current, with a bandwidth of about 500 kHz.

We investigated the frequency fluctuations of the cooling laser in both the free-running and the actively stabilized regimes. The Fourier transform of the frequency noise spectrum allows one to estimate the laser linewidth. For the free-running laser this leads to a value below 500 kHz at 922 nm. In the case of active frequency stabilization, a frequency stability of the laser relative to the reference resonator (quality of lock) of better than 100 Hz is obtained [14]. Contributions of different parasitic effects, such as residual amplitude modulation, spurious

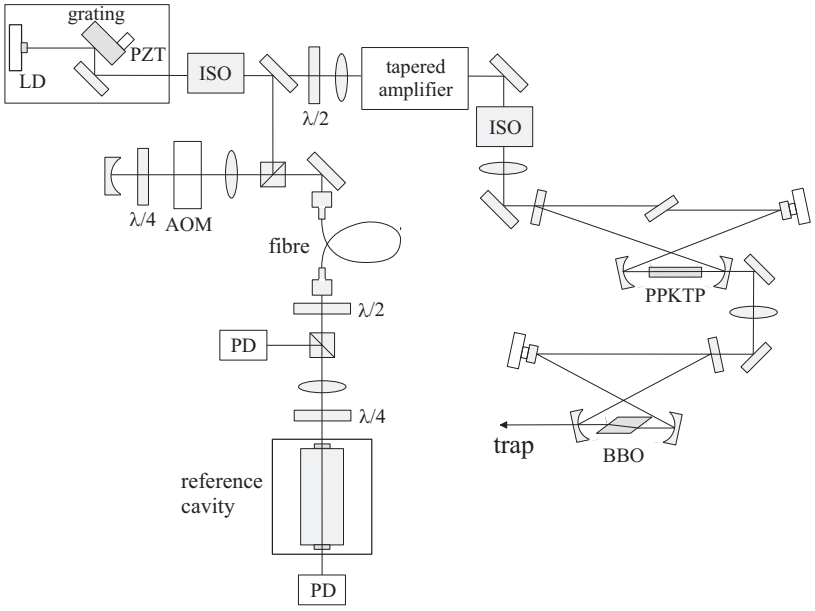


Fig. 3. Illustration of the experimental setup of the new cooling laser system. LD: laser diode, PZT: piezo-actuator, ISO: optical isolator, PD: photodiode, AOM: acousto-optical modulator; for details see text

interferences etc., are estimated to be small at the present level of accuracy; the frequency stability relative to the reference resonator is at present mainly limited by the bandwidth of our servo system.

With the achieved quality of lock, the absolute frequency stability of the cooling laser is mainly determined by the length stability of the reference resonator. The latter depends on the isolation of the reference cavity from environmental perturbations. Our investigations of different cavity supports led us to estimate the absolute laser frequency stability to be smaller than 10 kHz. Decoupling the reference resonator more efficiently from external vibrations, e.g. by active vibration isolation supports, would further reduce the laser linewidth [15]. With improved frequency stabilization schemes at the dekahertz level and below, diode lasers become an attractive laser source also for ultra-high-resolution spectroscopy [4].

Important for the use of an ECDL in series with a TA in high-resolution spectroscopy is the frequency fidelity of the amplifier. For analyzing this property, the output of the TA was heterodyned with the ECDL frequency. For this purpose the frequency of the TA output was shifted with an acousto-optical modulator and the beat note between the TA input and output beams was observed. The result is shown in Fig. 4. The beat note has a width of 100 mHz, limited by the resolution of the spectrum analyzer. This shows that the TA introduces negligible spectral broadening.

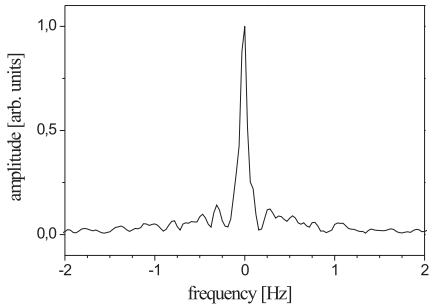


Fig. 4. Beat signal between input and output beams of the TA, showing a linewidth of 100 mHz, limited by the resolution of the measuring system (spectrum analyzer)

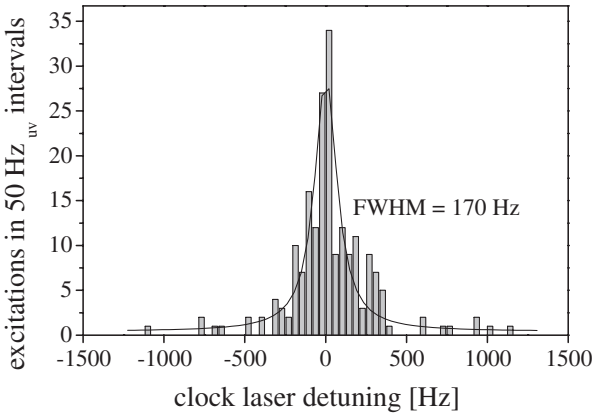


Fig. 5. High-resolution excitation spectrum of the In^+ $^1S_0$ – $^3P_0$ clock transition. The linewidth (FWHM) of the fitted Lorentzian corresponds to 170 Hz

4 High-Resolution Spectroscopy

Spectroscopy of the $^1S_0$ – $^3P_0$ clock transition is performed in optical-optical double resonance [1]. Excitation of the metastable $^3P_0$ state leads to a dark period in the single-ion fluorescence signal on the cooling transition until the level is depopulated via spontaneous decay. In this way absorption of one clock transition photon prevents subsequent scattering of some 10^5 or 10^6 fluorescence photons of the cooling transition. This allows detection of a transition to the metastable state with practically 100 per cent efficiency.

The clock laser system used for the experimental outcome of Fig. 5 consists of a diode-pumped Nd:YAG master laser at 946 nm [25]. The laser is frequency-stabilized to a thermally and acoustically isolated reference cavity of finesse ≈ 60000. A second diode-pumped Nd:YAG slave laser is used for power amplification and efficient intra-cavity second harmonic generation employing a $KNbO_3$ crystal. For transfer of the frequency stability, the light from the master laser

is coupled into the slave laser via injection locking. The blue light of the slave laser is subsequently frequency-doubled by means of a BBO crystal, placed in an external enhancement cavity, to generate the radiation at 237 nm needed to excite the $^1S_0 - {}^3P_0$ transition.

In order to obtain high-resolution spectra of the clock transition any light-shift and broadening of this transition by the cooling laser has to be avoided. The clock laser and the cooling laser beams are therefore applied alternately in time by blocking the beam by means of mechanical shutters. A high-resolution spectrum of the clock transition is shown in Fig. 5. In the experiment the clock laser power was reduced to ≈ 30 nW to avoid saturation broadening. By scanning the laser frequency in steps of 8 Hz, with typically four excitation pulses at each frequency, a spectral linewidth of 170 Hz (FWHM) (fractional resolution $\Delta\nu/\nu = 1.3 \times 10^{-13}$) is derived from a fit with a Lorentzian curve [8]. A spectral window of width 200 Hz contains 50 per cent of all excitations. According to the experimental control of the ion temperature, electromagnetic fields and vacuum conditions, no significant Doppler, Zeeman, Stark or collisional broadening of the ion is expected beyond a level of 1 Hz. The linewidth obtained is essentially determined by the frequency fluctuations of the clock laser; measurement of the frequency stability of the laser using a second independent reference cavity leads to a consistent result.

5 Absolute Frequency Measurements

Realization of an optical frequency standard requires precise frequency determination of the clock transition, at best by comparing it with a primary frequency standard: a cesium atomic clock.

So far, the absolute frequency of the $^{115}In^+$ $^1S_0$ - $^3P_0$ clock transition has been measured twice [26,27]. In a first measurement, by means of a phase-coherent frequency chain, the frequency of the $^1S_0$ - $^3P_0$ resonance was compared with a methane-stabilized He-Ne laser at 3.39 μm [28] and a Nd:YAG laser at 1064 nm whose second harmonic was locked to a hyperfine component of molecular iodine [29]. The transportable He-Ne laser was calibrated against the German atomic cesium time standard at Physikalisch-Technische Bundesanstalt (PTB) with the help of a harmonic frequency chain [30], where the frequency was determined with an uncertainty of $\pm$ 23 Hz [31]; the absolute frequency of the a_{10} hyperfine component of the R(56)32-0 transition in molecular iodine had been measured in [32] with an uncertainty of $\pm$ 40 kHz. A frequency gap in the chain of 1.43 THz at 850 nm was bridged with the help of an optical frequency comb generator [33]. In this way, the frequency of the $^{115}In^+$ clock transition was determined to be 1 267 402 452 914 $\pm$ 41 kHz, corresponding to an accuracy of 3.3 parts in 10^{11} [26] (essentially determined by the uncertainty of the iodine reference). This result represented an improvement in accuracy of more than three orders of magnitude compared with previous measurements of the line.

In a second measurement, the frequency of the indium clock transition was compared with the frequency of the methane-stabilized He-Ne laser at 3.39 μm

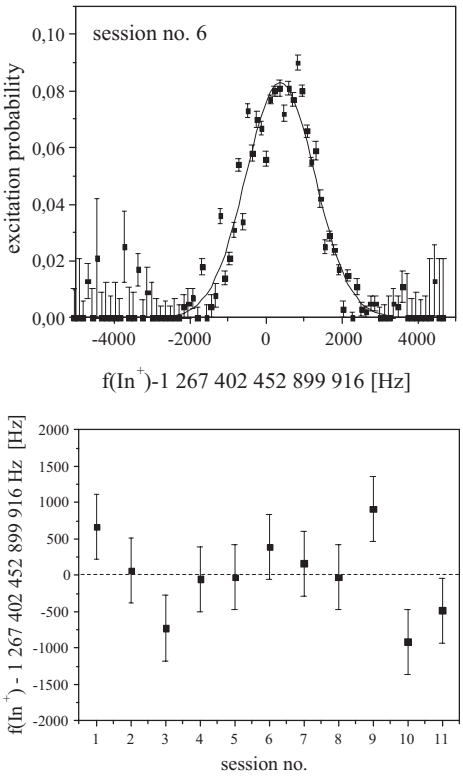


Fig. 6. Top: excitation probability of the $^3P_0$ state as a function of the absolute frequency in a typical measurement session. Bottom: line center of the 11 measurement sessions; the error bars correspond to the uncertainty of the HeNe standard

alone [27]. Here, the He-Ne laser had been calibrated against an atomic cesium fountain clock. A gap of 37 THz in the frequency chain at the fourth harmonic of the He-Ne standard was bridged with the frequency comb of a mode-locked femtosecond laser [34]. In this measurement the frequency of the In^+ clock transition was determined to 1 267 402 452 899.92 ± 0.23 kHz, where the accuracy of 1.8 parts in 10^{13} was mainly limited by the uncertainty of the He-Ne laser [27]. Figure 6 shows the results of the 11 measurement sessions as well as the excitation probability of the In^+ clock transition as a function of the absolute frequency in one particular session. Systematic frequency shifts of the clock transition are negligible at this level of accuracy. The magnetic field dependence is – 636 ± 27 Hz/G for the $|F = 9/2; m_F = 9/2\rangle \rightarrow |\ F' = 9/2; m'_F = 7/2\rangle$ components excited in the experiment [8], where the magnetic field was adjusted to zero with a precision of a few 10 mG. Other systematic frequency shifts such as the quadratic Stark or second-order Doppler shift are orders of magnitude smaller than the Zeeman shift at the temperatures to which the ion is cooled in the trap (T ∼ 150 μK).

Improving the frequency determination of the In^+ $^1S_0$ - $^3P_0$ clock transition and comparing the line with other narrow optical resonances of comparable accuracy using the frequency comb of a mode-locked femtosecond laser would enable the search for possible slow variations of fundamental constants in time as recently discussed in [20–22] (see also the various related contributions in this book). For example, for a relative variation $\Delta\alpha/\alpha$ of the fine-structure constant α of 10^{-15} a frequency shift of 0.26 Hz is predicted for the $^1S_0$ – $^3P_0$ resonance [35]. With the expected accuracy and stability of the In^+ frequency standard much smaller variations of the fine structure could be investigated. To detect the shift, the $^1S_0$ – $^3P_0$ transition could be compared for example with the hydrogen 1S – 2S transition at 122 nm, which has a negligible susceptibility to frequency shifts with $\Delta\alpha$ and in this way could be used as an anchor frequency.

Acknowledgement

We gratefully acknowledge the contributions of our former group members, Th. Becker, E. Peik and M. N. Skvortsov to this work as well as our collaboration with R. Holzwarth, J. Reichert, Th. Udem, Th. Hänsch, P. Pokasov and S. N. Bagayev in the absolute optical frequency measurement.

References

1. H. Dehmelt: IEEE Trans. Instrum. Meas **31**, 83 (1982)
2. For a recent review see: A.A. Madej, J.E. Bernard: 'Single Ion Optical Frequency Standards and Measurement of their Absolute Optical Frequency'. In: Frequency Measurement and Control: Advanced Techniques and Future Trends, Springer Topics in Applied Research, ed. by A.N. Luiten (Springer, Berlin, Heidelberg, 2000) pp. 149-190
3. R.J. Rafac, B.C. Young, J.A. Beall, W.M. Itano, D.J. Wineland, J.C. Bergquist: Phys. Rev. Lett **85**, 2462, (2000)
4. C. Tamm, D. Engelke, V. Bühner: Phys. Rev. A **61**, 053405 (2000)
5. N. Yu, X. Zhao, H. Dehmelt, W. Nagourney: Phys. Rev. A **50**, 2738 (1994)
6. J.E. Bernard, L. Marmet, A. Madej: Opt. Comm. **150**, 170 (1998)
7. H.C. Nägerl, Ch. Roos, D. Leibfried, H. Rohde, G. Thalhammer, J. Eschner, F. Schmidt-Kahler, R. Blatt: Phys. Rev. A **61**, 023405 (2000)
8. Th. Becker, J. von Zanthier, A.Yu. Nevsky, Ch. Schwedes, M.N. Skvortsov, H. Walther, E. Peik: Phys. Rev. A **63**, 051802(R) (2001)
9. S.A. Diddams, T. Udem, J.C. Bergquist, E.A. Curtis, R.E. Drullinger, L. Hollberg, W.M. Itano, W.D. Lee, C.W. Oates, K.R. Vogel, D.J. Wineland: Science **293**, 825 (2001)
10. Th. Udem, S.A. Diddams, K.R. Vogel, C.W. Oates, E.A. Curtis, W.D. Lee, W.M. Itano, R.E. Drullinger, J.C. Bergquist, L. Hollberg: Phys. Rev. Lett. **86**, 4996 (2001)
11. J. Stenger, Ch. Tamm, N. Haverkamp, S. Weyers, H.R. Telle: Opt. Lett. **26**, 1589 (2001)
12. W.M. Itano: J. Res. NIST **105**, 829 (2000)

13. P.J. Blythe, S.A. Webster, K. Hosaka, P. Gill: J. Phys. **B** At. Mol. Opt. **36**, 981 (2003)
14. Ch. Schwedes, E. Peik, J. von Zanthier, A.Yu. Nevsky, H. Walther: Appl. Phys. **B 76**, 143 (2003)
15. A.Yu. Nevsky, M. Eichenseer, J. von Zanthier, H. Walther: Opt. Commun. **210**, 91 (2002)
16. C. Schrama, E. Peik, W.W. Smith, H. Walther: Opt. Comm. **101**, 32 (1993)
17. E. Peik, G. Hollemann, H. Walther: Phys. Rev. A **49**, 402 (1994)
18. D. Wineland, H. Dehmelt: Bull. Am. Phys. Soc. **20**, 637 (1975)
19. E. Peik, J. Abel, Th. Becker, J. von Zanthier, H. Walther: Phys. Rev. A **60**, 439 (1999)
20. V.A. Dzuba, V.V. Flambaum, J.K. Webb: Phys. Rev. Lett. **82**, 888 (1999)
21. J.K. Webb, V.A. Dzuba, V.V. Flambaum, C.W. Churchill, M.J. Drinkwater, J.D. Barrow: Phys. Rev. Lett. **82**, 884 (1999)
22. S.G. Karshenboim, Can. J. Phys. **78**, 639 (2000)
23. C. Wieman and L. Hollberg: Rev. Sci. Instrum. **62**, 1 (1991)
24. L. Ricci, M. Weidemüller, T. Esslinger, A. Hemmerich, C. Zimmermann, V. Vuletic, W. König and T.W. Hänsch: Opt. Comm. **117**, 541 (1995)
25. G. Hollemann, E. Peik, A. Rusch, H. Walther: Opt. Lett. **20**, 1871 (1995)
26. J. von Zanthier, J. Abel, Th. Becker, M. Fries, E. Peik, H. Walther, R. Holzwarth, J. Reichert, Th. Udem, T.W. Hänsch, A.Yu. Nevsky, M.N. Skvortsov, S.N. Bagayev: Opt. Comm. **166**, 57 (1999)
27. J. von Zanthier, Th. Becker, M. Eichenseer, A.Yu. Nevsky, Ch. Schwedes, E. Peik, H. Walther, R. Holzwarth, J. Reichert, Th. Udem, T.W. Hänsch, P.V. Pokasov, M.N. Skvortsov, S.N. Bagayev: Opt. Lett. **25**, 1729 (2000)
28. S.N. Bagayev, A.K. Dmitriyev, P.V. Pokasov: Laser Phys. **7**, 989 (1997)
29. S.N. Bagayev, I.S. Moskalev, A.Yu. Nevsky, M.N. Skvortsov in: Modern Problems of Laser Physics, 1st International Symposium at Novosibirsk, Russia, 1997, ed. by S.N. Bagayev, V.I. Denisov (Siberian Division of Russian Academy of Sciences, Novosibirsk 1997) p. P-062
30. H. Schnatz, B. Lipphardt, J. Helmcke, F. Riehle, G. Zinner: Phys. Rev. Lett. **76**, 18 (1996)
31. Th. Udem, A. Huber, B. Gross, J. Reichert, M. Prevedelli, M. Weitz, T.W. Hänsch: Phys. Rev. Lett. **79** 2646, (1997)
32. P. Jungner, M.L. Eickhoff, S.D. Swartz, J. Ye, J.L. Hall, S. Waltman: Proc. SPIE – Int. Soc. Opt. Eng. **2378** 22, (1995);
    P.A. Jungner, S. Swartz, M. Eickhoff, J. Ye, J.L. Hall, S. Waltman: IEEE Trans. Instr. Meas. **44**, 151 (1995);
    M.L. Eickhoff, J.L. Hall: IEEE Trans. Instr. Meas. **44**, 155 (1995)
33. M. Kourogi, T. Enami, M. Ohtsu: IEEE Photon. Technol. Lett. **6**, 214 (1994)
34. J. Reichert, M. Niering, R. Holzwarth, M. Weitz, Th. Udem, T.W. Hänsch: Phys. Rev. Lett. **84**, 3232 (2000);
    S.A. Diddams, D.J. Jones, J.Ye, S.T. Cundiff, J.L. Hall, J.K. Ranka, R.S. Windeler, R. Holzwarth, Th. Udem, T.W. Hänsch: Phys. Rev. Lett. **84**, 5102 (2000)
35. V.A. Dzuba, V.V. Flambaum: Phys. Rev. A **61**, 034502 (2000)

Applications of Femtosecond Laser Comb to Nonlinear Molecular Spectroscopy

Jun Ye, R. Jason Jones, Lisheng Chen, Kevin W. Holman, and David J. Jones

JILA, National Institute of Standards and Technology and University of Colorado, Boulder, Colorado 80309-0440, USA

Abstract. Recent progress in precision control of pulse repetition rate and carrier-envelope phase of ultrafast lasers has established a strong connection between optical frequency metrology and ultrafast science. A wide range of applications has ensued, including measurement of absolute optical frequencies, precision laser spectroscopy, optical atomic clocks, and optical frequency synthesis in the frequency-domain, along with pulse timing stabilization, coherent synthesis of optical pulses, and phase-sensitive extreme nonlinear optics in the time-domain. In this contribution we discuss the impact of the femtosecond optical frequency comb to molecular spectroscopy. Measurements performed in the frequency-domain provide a global picture of molecular structure at high precision while providing radio frequency clock signals derived from molecular optical standards. Time-domain analysis and experiments give us new possibilities for nonlinear optical spectroscopy and sensitive detections with real-time information.

1 Introduction to Femtosecond Optical Frequency Comb

Precise phase control of femtosecond lasers has become increasingly important as novel applications utilizing the femtosecond laser-based optical comb are developed that require greater levels of precision and higher degrees of control [1]. Improved stability is beneficial for both frequency-domain applications, where the relative phase or "chirp" between comb components is unimportant (e.g. optical frequency metrology), and, perhaps more importantly, time-domain applications where the pulse shape and/or duration is vital, such as in nonlinear optical interactions [2]. For both types of applications, minimizing jitter in the pulse train and noise in the carrier-envelope phase is often critical to achieve the desired level of precision. Phase-stabilized mode-locked femtosecond lasers have played a key role in recent advances in optical frequency measurement [3,4], carrier-envelope phase stabilization [2,5,6], all-optical atomic clocks [7,8], optical frequency synthesizers [9], coherent pulse synthesis [10], and ultra-broad, phase coherent spectral generation [11].

Mode-locked lasers generate short optical pulses by establishing a fixed phase relationship among all of the lasing longitudinal modes. To understand the connection between the time-domain and frequency-domain descriptions of a mode-locked laser and the pulse train that it emits, a key concept is the carrier-envelope phase. This is based on the decomposition of the pulses into an envelope function, $\hat{E}(t)$, that is superimposed on a continuous carrier wave with frequency ω_c, so that the electric field of the pulse is written $E(t) = \hat{E}(t)e^{i\omega_c t}$. The carrier-envelope phase, ϕ_{CE}, is the phase shift between the peak of the envelope and

J. Ye et al., Applications of Femtosecond Laser Comb to Nonlinear Molecular Spectroscopy, Lect. Notes Phys. **648**, 275–295 (2004)
http://www.springerlink.com/

the closest peak of the carrier wave. In any dispersive material, the difference between group and phase velocities will cause ϕ_{CE} to evolve. This group-phase velocity mismatch inside a laser cavity produces a pulse-to-pulse phase shift accumulated over one round-trip as $\Delta\phi_{CE}$.

When ϕ_{CE} is constant, the spectrum of a femtosecond optical comb corresponds to identical pulses emitted by the mode-locked laser. For a single pulse, the spectrum is the Fourier transform of its envelope function and is centered at the optical frequency of its carrier. Generally, for any pulse shape, the frequency width of the spectrum will be inversely proportional to the temporal width of the envelope. For a train of identical pulses, separated by a fixed interval, the spectrum can easily be obtained by a Fourier series expansion, yielding a comb of regularly spaced frequencies, where the comb spacing is inversely proportional to time interval between successive pulses, i.e., the repetition rate (f_{rep}) of the laser. Of course, the Fourier relationship between time and frequency resolution guarantees that any spectrometer with sufficient spectral resolution to distinguish the individual comb line cannot have enough temporal resolution to separate successive pulses. Therefore the successive pulses interfere with each other inside the spectrometer and the comb spectrum occurs because there are certain discrete frequencies at which the interference is constructive. Using the result from Fourier analysis that a shift in time corresponds to a linear phase change with frequency, we can readily see that the constructive interference occurs at $n \times f_{\mathrm{rep}}$, where n is an integer.

When the ϕ_{CE} is evolving with time, such that from pulse to pulse (with a time separation of $T = 1/f_{\mathrm{rep}}$) there is a phase shift of $\Delta\phi_{CE}$, then in the spectral domain a rigid shift will occur for the frequencies at which the pulses add constructively. This shift can be easily determined as $(1/2\pi)\Delta\phi_{CE}/T$. Thus the optical frequencies, ν_n, of the comb lines can be written as

$$\nu_n = nf_{\mathrm{rep}} + f_0 \ , \tag{1}$$

where n is a large integer of order 10^6 that indexes the comb line, and f_0 is the comb offset due to pulse-to-pulse phase shift,

$$f_0 = \frac{1}{2\pi} f_{\mathrm{rep}} \Delta\phi_{CE} \ . \tag{2}$$

The relationship between time and frequency domain pictures is summarized in Fig. 1. The pulse-to-pulse change in the phase for the train of pulses emitted by a mode-locked laser can be expressed in terms of the average phase (v_p) and group (v_g) velocities inside the cavity. Specifically,

$$\Delta\phi_{CE} = \left(\frac{1}{v_g} - \frac{1}{v_p}\right) l_c \omega_c \bmod (2\pi) \ , \tag{3}$$

where l_c is the round-trip length of the laser cavity and ω_c is the "carrier" frequency.

Armed with the understanding of the frequency spectrum of a mode-locked laser, we can now turn to the question of measuring the absolute frequencies of

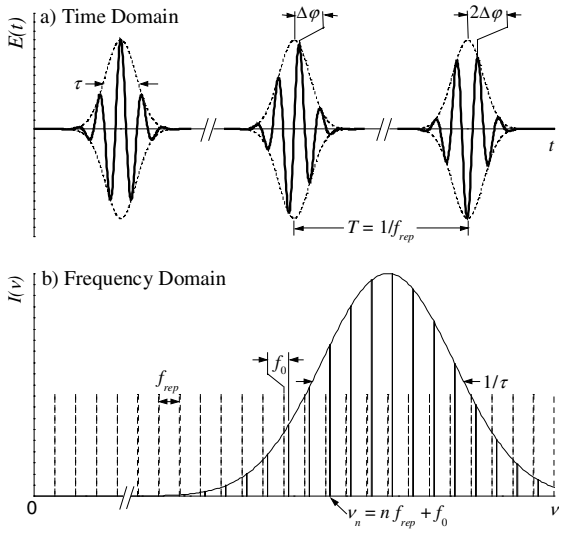


Fig. 1. Summary of the time-frequency correspondence for a pulse train with evolving carrier-envelope phase.

comb lines. For a frequency measurement to be absolute, it must be referenced to the hyperfine transition of ^{133}Cs that defines the second. From (1) we see that determining the absolute optical frequencies of the femtosecond comb requires two radio frequency (RF) measurements, that of $f_{\rm rep}$ and f_0. Measurement of $f_{\rm rep}$ is straightforward; we simply detect the pulse train's repetition rate (from 10's of MHz to several GHz) with a fast photodiode. On the other hand, measurement of f_0 is more involved as the pulse-to-pulse carrier envelope phase shift requires interferometric measurement, whether it is carried out in the time-domain [12] or in the frequency-domain [13]. When the optical spectrum spans an octave in frequency, i.e., the highest frequencies are a factor of 2 larger than the lowest frequencies, then measurement of f_0 is greatly simplified. If we use a second harmonic crystal to frequency double a comb line, with index n, from the low frequency portion of the spectrum, it will have approximately the same frequency as the comb line on the high frequency side of the spectrum with index $2n$. Measuring the heterodyne beat between these yields a difference frequency

$$2\nu_n - \nu_{2n} = 2(nf_{\rm rep} + f_0) - (2nf_{\rm rep} + f_0) = f_0 \ , \tag{4}$$

which is just the offset frequency. Thus an octave spanning spectrum enables a simple measurement of f_0 [5]. Note that an octave spanning spectrum is not required, it is just the simplest. We designate this scheme, as shown in Fig. 2(a), as "self-referencing" as it uses only the output of the mode-locked laser. Self-referencing is not the only means of determining the absolute optical frequencies given an octave spanning spectrum. For example, the absolute optical frequency of a CW laser can be determined if its frequency lies close to comb line n in the low frequency portion of the femtosecond comb spectrum. Then the second

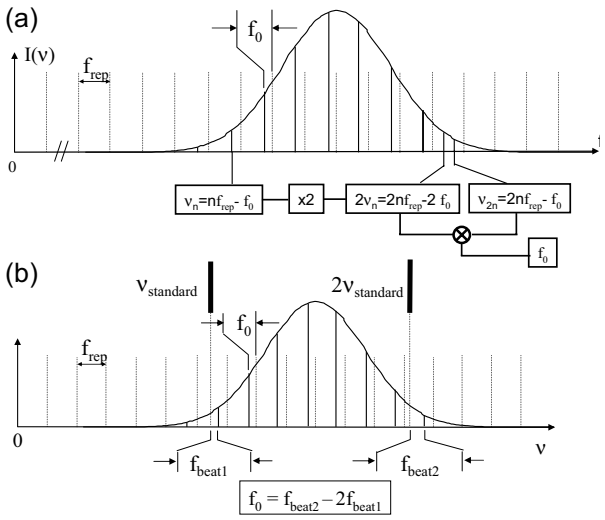


Fig. 2. Two equivalent schemes for measurement of f_0 using an octave-spanning optical frequency comb. In the self-referencing approach shown in (a), frequency doubling and comparison are accomplished with the comb itself. In the second approach shown in (b), the fundamental and the second harmonic of a CW optical frequency standard are used for determining f_0. These two basic schemes are employed for absolute optical frequency measurement and implementation of optical atomic clocks.

harmonic of the CW laser will lie close to the comb line $2n$. Measurement of the heterodyne beat between the CW laser frequency, ν_{cw}, and the frequency of comb line n gives

$$f_{\mathrm{beat1}} = \nu_{cw} - (nf_{\mathrm{rep}} + f_0)$$

and between the second harmonic of the CW laser and comb line $2n$ gives

$$f_{\mathrm{beat2}} = 2\nu_{cw} - (2nf_{\mathrm{rep}} + f_0)\ .$$

Taking the difference,

$$f_{\mathrm{beat2}} - 2f_{\mathrm{beat1}} = 2\nu_{cw} - (2nf_{\mathrm{rep}} + f_0) - (2\nu_{cw} - 2(nf_{\mathrm{rep}} + f_0)) = f_0\ . \qquad (5)$$

This detection scheme is shown in Fig. 2(b) [8].

An octave-bandwidth optical comb is not straightforward to produce. A Fourier-transform limited pulse with a full width at half maximum (FWHM) bandwidth of an octave centered at 800 nm, would only be a single optical cycle in duration. Such short pulses have not been achieved. Fortunately, neither a transform-limited pulse nor a FWHM of an octave is actually needed. The pulse width is unimportant as the measurement and control techniques are purely frequency domain approaches. Experimentally, it has been found that even if the power at the octave spanning points is 40 dB below the peak, it is still possible to observe strong f-to-2f heterodyne beats. Still, the necessary comb bandwidth is larger than that from a typical sub-10-fs mode-locked Ti:sapphire laser. One

approach to produce this sufficient spectral bandwidth is based on self-phase modulation directly inside the Ti:Sapphire crystal itself [14] or inside an additional glass plate located inside the laser cavity with secondary coincident time and space foci [15]. These techniques require carefully designed mirrors and laser cavities. Additional spectral bandwidth can also be obtained by minor changes in the cavity configuration of a high repetition rate laser, although it has not yet yielded sufficient intensity at the octave points for the observation of f-to-2f beats [16]. Another widely adopted approach is to generate the extra comb bandwidth using microstructure fibers that have zero group velocity dispersion (GVD) within the emission spectrum of a Ti:sapphire laser [17]. The large index contrast for waveguiding inside microstructure fibers has two consequences, first the ability to generate a zero in the GVD at visible or near-infrared wavelengths and, secondly, the possibility of using a much smaller core size. This has allowed broadband continuum generation with only nanojoule pulse energies.

For the purpose of using the femtosecond optical comb for absolute optical frequency measurements, it is straightforward to establish the frequency values of all of the comb components. The comb's frequency spacing (f_{rep}) can be phase locked with high precision via detections of higher harmonics of f_{rep} relative to an RF standard. The value of f_0 is determined and controlled using schemes shown in Fig. 2. Control of f_0 requires a servo transducer acting differentially on the intracavity group and phase delays. One common method for adjusting f_0 is to swivel the end mirror in the arm of the laser cavity that contains the prism sequence [18]. An alternative method of controlling f_0 is via modulation of the pump power, with likely contributions from the nonlinear phase, spectral shifts, and the intensity dependence in the group velocity [19]. It is worth noting that a scheme implemented by H. Telle *et al* [20] allows the frequency comb to be free running (without any active stabilization) while making highly precise measurement and/or connections for an optical frequency interval lying within the comb bandwidth.

The dramatic simplification of a complex optical frequency chain to that of a single mode-locked laser has greatly facilitated optical frequency measurement. Another important aspect of this new technology is its high degree of reliability and precision and lack of systematic errors. For example, recent tests have shown that the repetition rate of a mode locked laser equals the mode spacing of the corresponding comb to within the measurement uncertainty of 10^{-16}. The uniformity of the comb mode spacing has also been verified to a level below 10^{-17}, even after spectral broadening in fiber [3]. Comparison between two separate fs comb systems, both linked to a common reference source (microwave or optical), allows one to examine the intrinsic accuracy of a fs comb based frequency measurement system, currently at a level of a few parts in 10^{16} with no measurable systematic effects [21]. Direct comparisons of absolute optical frequency measurements between the femtosecond comb technique and the traditional harmonic frequency chain approach have also produced assuring mutual confirmations at the level of 10^{-12} to 10^{-14} [4,22].

As the measurement precision for optical frequencies is advanced to an ever-higher level, the stability limitation imposed by the available RF standards used

for fs comb stabilization becomes an important issue [22,23]. Instead of operating a fs comb using RF references, it appears to be advantageous to operate the comb by stabilizing it to an optical frequency standard. The fs comb will then in turn produce optically derived stable clock signals in the RF domain, leading to a so-called "optical atomic clock" [7,8,24]. Recent experimental demonstrations support the concept that, in the future, the most stable and accurate frequency standards will be based on optical transitions [25,26]. Stepping down the stability level by one or two orders of magnitude, portable optical frequency standards that offer compact, simple, and less expensive system configurations have also shown competitive performance with (in-)stability near 1×10^{-14} at 1 to 10 s averaging time [27].

To realize an optical atomic clock, an optical comb needs to be stabilized to a pre-selected optical frequency source at a precision level that exceeds the optical standard itself. f_0 can be extracted in a straightforward manner using either schemes shown in Fig. 2. Then f_0 can be stabilized with respect to either f_{rep} or an auxiliary stable RF source. It is worth noting that stabilization of f_0 at a few mHz is more than adequate, as it yields fractional frequency noise of $< 10^{-17}$ for an optical carrier. A heterodyne beat between one of the comb components and the optical standard (ν_{standard}) yields information about fluctuations in f_{rep}. We note for the particular case shown in Fig. 2(b),

$$\begin{aligned} f_{\text{beat2}} - f_{\text{beat1}} &= 2\nu_{\text{standard}} - (2nf_{\text{rep}} + f_0) - (\nu_{\text{standard}} - (nf_{\text{rep}} + f_0)) \\ &= \nu_{\text{standard}} - nf_{\text{rep}} \ , \end{aligned}$$

producing a direct link between the optical frequency ν_{standard} and the radio frequency f_{rep}. After appropriate processing, this error signal is used to stabilize the phase of f_{rep} coherently to ν_{standard}, thereby producing a clock signal output in the RF domain derived from ν_{standard}.

Besides the capability of deriving RF signals from an optical frequency standard, a fs comb can, of course, also be used to transfer the stability of optical standards across vast frequency gaps to other optical spectral regions. Easy access to the resolution and stability offered by optical standards will greatly facilitate the application of frequency metrology to precise spectroscopic investigations. An ideal situation is depicted in Fig. 3 where a high precision optical frequency synthesizer is realized with the help of a stable fs comb linked to an optical frequency standard. For spectroscopy applications, we indeed often desire a single-frequency and reasonably powered optical carrier wave that can be tuned to any desired optical spectral feature of interest. Realization of such an optical frequency synthesizer (analogous to its RF counterpart) will add a tremendously useful tool for modern spectroscopy experiments. One could foresee an array of diode lasers, each covering a successive tuning range of $\sim$ 10 to 20 nanometers that would collectively cover most of the visible spectrum. Each diode laser frequency would be controlled by the stabilized optical comb, and therefore be directly related to the absolute time/frequency standard in a phase coherent fashion, while the setting of the optical frequency would be accomplished via computer control.

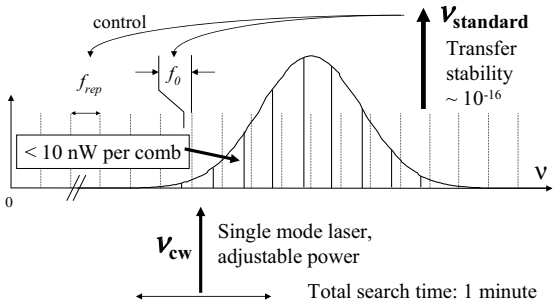


Fig. 3. Random search of and stabilization to the targeted comb position by a single frequency CW laser, whose frequency is continuously guided by the fs comb which is in turn stabilized by an optical frequency standard.

In our preliminary implementation of such an optical frequency synthesizer [9], the fs comb system is referenced by an I_2-based optical frequency standard at 532 nm. A CW diode laser, as well as a CW Ti:sapphire laser, is used to tune through targeted spectral regions (for example, Rb D_1 and D_2 lines at 795 and 780 nm for the diode laser and I_2 hyperfine transitions in the region of 490 - 520 nm) with desired frequency step sizes, while maintaining absolute reference to the stabilized optical comb. A self-adaptive search algorithm first tunes the CW laser to a specified wavelength region with the aid of a wavelength measurement device (100 MHz resolution). A heterodyne beat signal between the laser's frequency and that of a corresponding comb line is then detected and processed. For fine-tuning, an RF source provides a tunable frequency offset for the optical beat. Once the laser frequency tuning exceeds one comb-spacing, we reset the RF offset frequency back to the original value to start the process over again. The laser frequency can thus be tuned smoothly in an "inch-worm" manner along the comb structure. We have demonstrated two fundamental aspects of an optical frequency synthesizer; namely continuous, precise tuning of the optical frequency as well as arbitrary frequency setting on demand. The entire search process takes about a minute.

2 Molecular Spectroscopy Aided by Femtosecond Optical Frequency Comb

Before we delve ourselves into examples of molecular spectroscopy aided by the technology of the precision frequency comb, we would like to discuss briefly the implications of the frequency-domain control of the femtosecond laser to the time-domain experiments. Prior to the development of femtosecond comb technology, mode-locked lasers were used almost exclusively for time-domain experiments. Although the femtosecond comb technology has primarily impacted the frequency-domain applications described earlier, it is having an impact on time-domain experiments and promises to bring about just as dramatic advances in the time-domain as it has in optical frequency metrology and optical clocks.

Indeed, it is fascinating to blur the boundary between traditional CW precision spectroscopy and ultrafast phenomena. The time-domain applications put stringent requirements on the carrier-envelope phase coherence. Stabilization of the "absolute" carrier-envelope phase at a level of tens of milli-radians has been demonstrated and this phase coherence is maintained over an experimental period exceeding many minutes [28], paving the groundwork for synthesizing electric fields with known amplitude and phase at optical frequencies. Working with two independent femtosecond lasers operating at different wavelength regions, we have synchronized the relative timing between the two pulse trains at the femtosecond level [29], and also phase locked the two carrier frequencies, thus establishing phase coherence between the two lasers. By coherently stitching optical bandwidths together, a "synthesized" pulse has been generated [10]. With the same pair of Ti:sapphire mode-locked lasers, we have demonstrated widely tunable femtosecond pulse generation in the mid- and far- IR using difference-frequency-generation [30]. The flexibility of this new experimental approach is evidenced by the capability of rapid and programmable switching and modulation of the wavelength and amplitude of the generated IR pulse. A fully developed capability of producing phase-coherent visible and IR pulses over broad spectral bandwidth, coupled with arbitrary control in amplitude and pulse shape, represents the ultimate instrumentation for coherent control of molecular systems. A pulse train with good carrier-envelope phase coherence is also very promising for experiments that are sensitive to ϕ_{CE}, i.e., the "absolute" pulse phase, not just its pulse-to-pulse change [2]. This can be manifested in "extreme" nonlinear optics experiments, or coherent control.

The capability to precisely control pulse timing (signified by the level of control in f_{rep}) and the pulse-carrier phase allows one to manipulate pulses using novel techniques and achieve unprecedented levels of flexibility and precision, as will be demonstrated in the work on time resolved spectroscopy of molecules. For example, the simultaneous control of timing jitter and carrier-envelope phase can be used to phase coherently superpose a collection of successive pulses from a mode-locked laser. By stabilizing the two degrees of freedom of a pulse train to an optical cavity acting as a coherent delay, constructive interference of sequential pulses will be built up until a cavity dump is enabled to switch out the "amplified" pulse [31]. Such a passive pulse "amplifier", along with the synchronization technique we developed for pulse synthesis, has made a strong impact on the field of nonlinear-optics based spectroscopy and imaging of bio-molecular systems, showing significant improvements in experimental sensitivity and spatial resolution [32,33]. With the enhanced detection sensitivity comes the capability of tracking real time biological dynamics. An ultrafast laser locked to a high stability cavity is also expected to demonstrate extremely low pulse jitter and carrier-envelope phase noise, which will be particularly attractive for time-domain experiments. In addition, we are exploring the use of pulse-cavity interactions to obtain a high sensitivity in intracavity spectroscopy (linear and non-linear) with a wide spectral coverage, as well as to enhance nonlinear interaction strengths for high efficiency nonlinear optical experiments. A more extensive discussion along these lines will be presented in Sects. 4 and 5.

With these new sets of tools in hand, it is appropriate to revisit the topics of precision molecular spectroscopy. It is also interesting to explore spectroscopy in a more broad sense. For example, one can now carry out precision spectroscopy using ultrafast lasers. On the other hand, coherent control of molecular motion can be performed in the spirit of precision measurement. The capability of absolute optical frequency measurements in the visible and IR spectral regions adds a new meaning to the term of precision molecular spectroscopy. Understanding of molecular structure and dynamics often involves detailed spectral analysis over a broad wavelength range. Such a task can now be accomplished with a desired level of accuracy uniformly across all relevant spectral windows, allowing precise investigations of minute changes in the molecular structure over a large dynamic range. For example, absolute frequency measurement of vibration overtone transitions and other related resonances (such as hyperfine splitting) will reveal precise information about the molecular potential energy surface and relevant perturbation effects. We have pursued such a study in iodine molecules, performing high-resolution and high-precision measurement of hyperfine interactions of the first excited electronic state (B) of I_2 over an extensive range of vibrational and rotational quantum numbers towards the dissociation limit. Experimental data demonstrate systematic variations in the hyperfine parameters that confirm calculations based on *ab initio* molecular potential energy curves and electronic wave functions derived from a separated-atomic basis set. We have accurately determined the state-dependent quantitative changes of hyperfine interactions caused by perturbations from other electronic states and identified the respective perturbing states. Our work in I_2 near the dissociation limit is also motivated by the desire to improve cell-based portable optical frequency standards [34]. Indeed, I_2-stabilized lasers have already demonstrated high stability ($< 5 \times 10^{-14}$ at 1 s averaging time) and have served well for optical atomic clocks [8].

3 I_2 Hyperfine Interactions, Optical Frequency Standards and Clocks

The hyperfine structure of I_2 rovibrational levels includes four contributions: nuclear electric quadrupole (eqQ), spin-rotation (C), tensorial spin-spin (d), and scalar spin-spin (δ) interactions. Agreement between experiment and theory using the four-term effective hyperfine Hamiltonian is at the kilohertz level for a few selected transitions. For the first excited electronic state B with the ${}^2P_{3/2}+{}^2P_{1/2}$ dissociation limit, our goal is to perform a systematic high-precision investigation of hyperfine interactions over an extensive range of rovibrational quantum numbers coupled with a large range of internuclear separations. Such a study has allowed us to understand the rovibrational dependence of the hyperfine interactions (as well as the dependence on internuclear distance) based on *ab initio* molecular potential energy curves and the associated electronic wave functions. Careful analysis of various perturbation effects leads to precise determination of molecular structure over a large dynamic range.

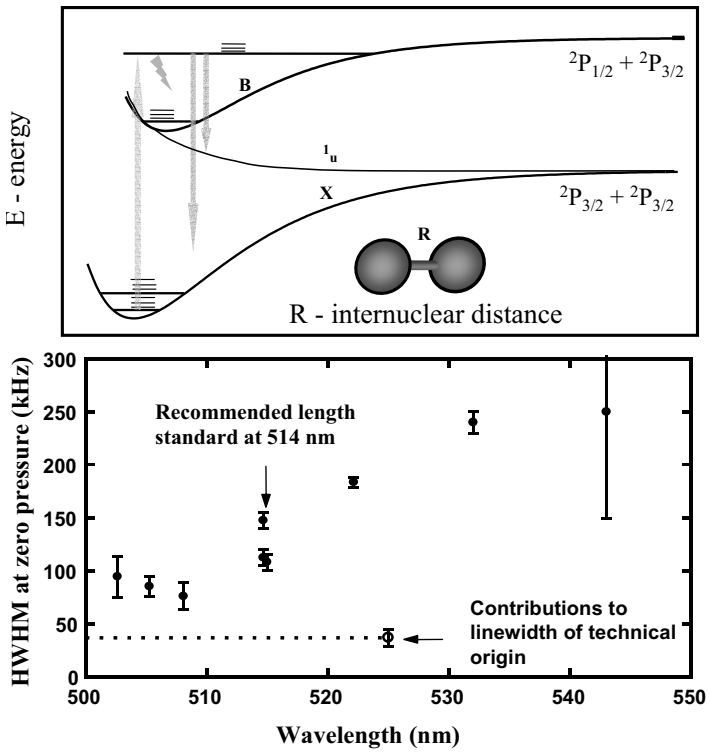


Fig. 4. The ground state and the first excited state of I_2 with their associated dissociation limit. The lower panel shows a narrowing trend of the transition linewidth when the excited state approaches the dissociation limit.

Prior studies have concentrated on a few isolated rovibrational levels for the high vibrational levels $v' = 40$ to 82 in the B state [35–37]. For vibrational levels below $v' = 43$, only functional forms on the state-dependent variations of the hyperfine interactions have been investigated from empirical data [38]. Combining absolute optical frequency metrology with high-resolution and broad wavelength-coverage laser spectroscopy, we have measured ∼80 rovibrational transitions with the upper vibrational levels (from $v' = 42$ up to $v' = 70$) stretching from a closely bonded molecular basis to a separated-atomic basis appropriate for the $^2P_{3/2} + ^2P_{1/2}$ dissociation limit, providing kHz-level line accuracies for most hyperfine components. The study is performed in the wavelength region of 530 to 498 nm. Measurements performed on a large set of rovibrational quantum numbers provide systematic information on state-dependent variations in the hyperfine interactions caused by perturbation from other nearby states. Figure 4 shows a simple schematic of the ground and the first excited electronic states of I_2 and their relevant dissociation limits. The lower panel in Fig. 4 shows a clear trend of linewidth narrowing with decreasing transition wavelength. However, this tendency is complicated by variations in linewidths among different rotational or

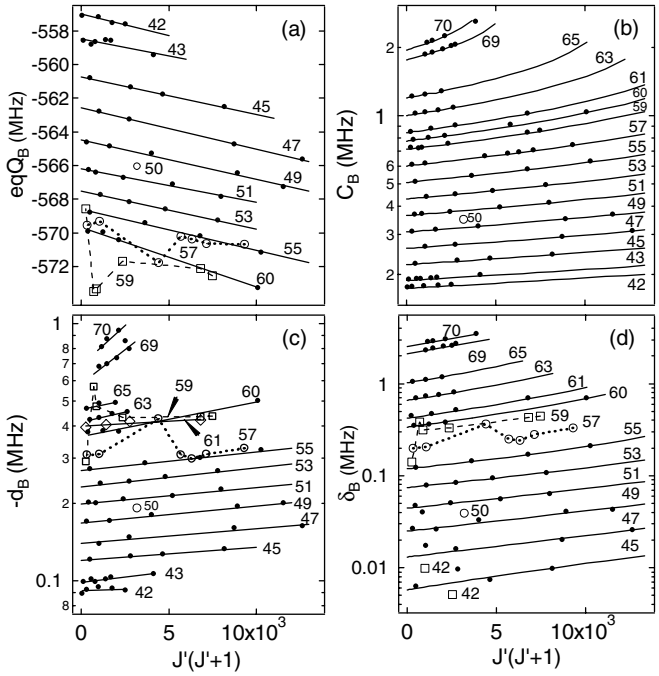


Fig. 5. Rovibrational dependence of the B state hyperfine parameters (a) eqQ_B, (b) C_B, (c) d_B, and (d) δ_B. Note (b), (c), and (d) are semilog plots and the vertical scale of (c) has been inverted. Each solid line is a fit for J-dependence for each vibrational level (v' indicated in the figure). Experimental data in squares and open circles show abnormal variations of eqQ_B, d_B, and δ_B around $v' = 57$ and 59.

hyperfine components when the transitions approach the pre-dissociation region. The initial linewidth narrowing at shorter wavelength may indicate among other interesting effects that the Franck-Condon factor in the transition probability is reduced when the excited state reaches a higher vibration level. As the excited state approaches the dissociation threshold, the limit on life time imposed by predissociation and other effects will need to be taken into consideration.

Figure 5 illustrates systematic rovibrational dependences for all four hyperfine parameters. Each solid line is a fit of experimental data for rotational dependence belonging to a single vibrational level (v'). In general, all hyperfine parameters have monotonic dependence on both rotational and vibrational quantum numbers except for the levels in the vicinity of $v' = 57$ to 59. However, the v-dependence of eqQ_B reverses its trend after $v' = 60$. For the sake of figure clarity, the eqQ_B data for $v' > 60$ are not shown. Another important observation is that for levels of $v' = 57-59$ all hyperfine parameters except for C_B bear abnormal J-dependences due to perturbations from a 1_g state through accidental rotational resonances.

Combining data from this work and the literature [38], investigations of the hyperfine spectra now cover the majority of the vibrational levels ($3 \leq v' \leq 82$)

in the B state. Therefore, it is now possible and useful to explore the global trend of these hyperfine parameters in the B state. Suppressing the rotational dependence, hyperfine parameters as functions of pure vibrational energy $E(v')$ are found to increase rapidly when molecules approach the dissociation limit, which is a result of the increasingly strong perturbations from other high-lying electronic states sharing the same dissociation limit with the B state. While C_B's variation is smooth over the whole range, eqQ_B, d_B, and δ_B all have local irregularities at three positions: $v' = 5$ where the B":1_u state crosses nearby, around $v' = 57$ to 59 (see discussions above), and from $v' = 76$ to 78, due to the same 1_g state [35,37].

To examine these hyperfine parameters in terms of internuclear separation R, the vibrational average of the hyperfine parameters is removed by inverting the expression $O(v', J') = \langle v'_{J'}|O(R)|v'_{J'}\rangle$, where $O(v', J')$ denotes one of the four hyperfine parameters. Figure 6 plots eqQ_B, C_B, d_B, and δ_B against R-centroid evaluated from $\langle v'_{J'}|R|v'_{J'}\rangle$ ($|v'_{J'}\rangle$ properly normalized), along with the corresponding residual errors of the interpolation. In Fig. 6(a), (b), (c), and (d), solid lines are calculated from $\langle v'_{J'}|O(R)|v'_{J'}\rangle$ and symbols are experimental data. Consistent with C_B's smooth variation, the interpolation function $C_B(R)$ has small residual errors (within ± 0.03, relative) for the entire range from $v' = 3$ to 70. On the contrary, the large residual errors in the interpolation of eqQ_B, d_B, and δ_B for $v' \geq 56$ reflect their abnormal variations observed around $v' = 57$ and 59, restricting a reliable interpolation only to levels of $v' < 56$. In the region of $R < 5$ Å, valuable information can be readily extracted from eqQ_B to assist the investigation of I_2's electronic structure. Unlike the other three hyperfine parameters whose major parts originate from perturbations at nearly all possible values of R, a significant part of eqQ_B is due to the interaction between the nuclear quadrupole moment Q and the local electric field gradient $q(R)$ generated by the surrounding charge distribution of a largely B state character. Thus, for $R < 5$ Å, where perturbations from other electronic states are negligible, the vibration-removed interpolation function $eqQ_B(R)$, coupled with *a priori* information on $q(R)$, can be used to determine I_2 nuclear quadrupole moment or serve as a benchmark for molecular *ab initio* calculations of the electronic structure at various values of R.

Precision measurements on $B - X$ hyperfine spectra provide an alternative and yet effective way to investigate the potential energy curves ($PECs$) sharing the same dissociation limit with the B state as well as the associated electronic wave functions. To demonstrate this, we perform calculations of eqQ_B, C_B , d_B, and δ_B based on the available $PECs$ and electronic wave functions derived from a separated-atomic basis set. For both vibrational and rotational dependences, the *ab initio* calculation results agree very well with the experimental data for $v' \geq 42$ (R-centroid ≥ 3.9 Å). In short, we have extended the range of separated-atomic basis calculations from levels near the dissociation limit to low vibrational levels ($v' = 5$) and have found very good agreement with the experimental data on both vibrational and rotational dependences.

Besides these interesting studies in hyperfine structure, the narrow-linewidth I_2 transitions in this wavelength range also provide excellent cell-based optical

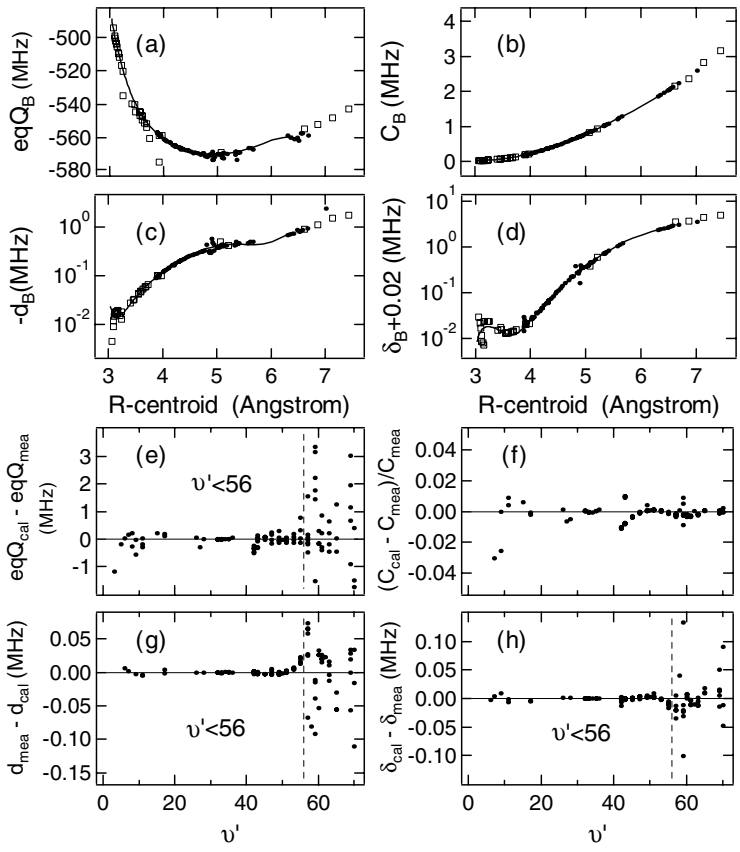


Fig. 6. (a) eqQ_B, (b) C_B, (c) d_B, and (d) δ_B versus R-centroid. Solid lines are calculated from $\langle v'_{J'}|O(R)|v'_{J'}\rangle$. Symbols are experimental data (dots: this work, squares: literature). (e) - (h) show residual errors of the interpolation.

frequency references for laser stabilization. Frequency-doubled Nd:YAG/$^{127}I_2$ at 532 nm has been proved to be one of the best portable optical frequency standards with compact size, reliability, and high stability ($< 5 \times 10^{-14}$ at 1 s). To reach a higher frequency stability, it is useful to explore I_2 transitions at wavelengths below 532 nm, where the natural linewidths decrease at a faster rate than that for the line strengths. We have measured the systematic variation of the I_2 transition linewidths within the range of 532 - 498 nm, with the linewidth decreasing by nearly 6 times when the wavelength is changed from 532 nm to near the dissociation limit [34]. The high S/N results indicate that I_2 transitions in the wavelength range of 532 - 501 nm hold great promise for future development of optical frequency standards, especially with the advent of all solid state Yb:YAG lasers. One exciting candidate is the 514.67 nm standard [39], with a projected stability $< 1 \times 10^{-14}$ at 1 s.

4 Extend Phase-Coherent fs Combs to the Mid-IR Spectral Region

Being able to combine the characteristics of two or more pulsed lasers working at different wavelengths certainly grants a more flexible approach to coherent control. The capability of synchronizing the repetition rates and phase-locking the carrier frequencies of two mode-locked lasers opens many applications. It may be particularly important in the generation of tunable femtosecond sources in other previously unreachable spectral regions. Figure 7 shows the cross-correlation measurement of the two stabilized mode-locked Ti:sapphire lasers using both sum (SFG) and difference frequency generation (DFG). The DFG signal produced by a GaSe crystal can be tuned from 6 micron and onto any longer wavelength regions with a high repetition rate (the same as the original laser's) and a reasonable average power (tens of microwatts). Arbitrary amplitude waveform generation and rapid wavelength switching in these nonlinear signals are simple to implement. The ultimate goal of this work is to make an optical waveform synthesizer that can create an arbitrary optical pulse on demand and use the novel source to study and control molecular motion. For frequency metrology and precision molecular spectroscopy in the IR region, we note that the difference frequency generation approach produces an absolute-frequency calibrated IR comb when the two Ti:sapphire lasers are synchronized and share a common offset frequency f_0.

Another important spectral region is 1.5 μm, where compact, reliable, and efficient mode-locked lasers exist and there are rich families of molecular overtone transitions. Frequency reference grids in this spectral window could also find applications in dense wavelength division multiplexed (DWDM) systems, photonic samplers in high-speed A/D conversion, and distribution of optical frequency standards over optical fiber networks. The synchronization and phase-locking

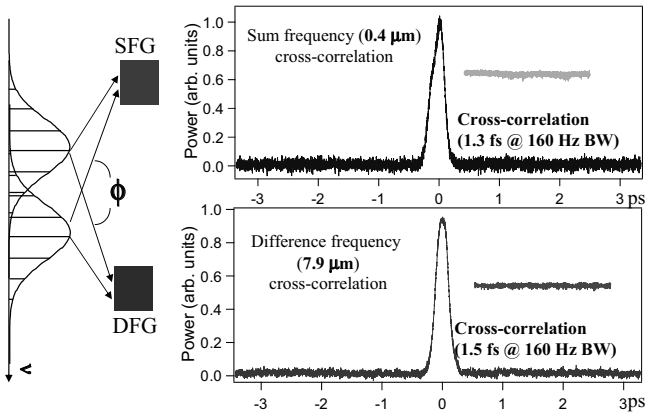


Fig. 7. Simultaneous sum and difference frequency generations from two stabilized femtosecond lasers. Also shown are the amplitude fluctuations at the half power point due to the relative timing jitter between the two parent lasers.

approach for Ti:sapphire lasers can be extended to cover mode-locked lasers at 1.5 μm. We have indeed achieved results of tight synchronization between the repetition rates and coherent phase locking of the optical carriers of the 1.5 μm mode-locked laser sources and a Ti:sapphire-based fs frequency comb, which is used as the clockwork for an optical atomic clock based on a molecular iodine transition.

A phase-coherent link between mode-locked lasers requires two distinct conditions to be met, as shown in Fig. 8(a). The comb spacing of the 1.5 μm source ($f_{\mathrm{rep},1550}$) must be stabilized to the optical clock's fs comb spacing ($f_{\mathrm{rep},775}$). Second, the combs' offset frequencies ($f_{0,775}$ and $f_{0,1550}$) must be phase locked together. This latter step requires spectral overlap between the two combs. The wide bandwidth optical frequency comb generated by the mode-locked fs Ti:sapphire laser is phase locked to a highly stable, iodine-based optical frequency standard. The optical comb of the 1.5 μm source is frequency doubled and compared against the Ti:sapphire comb at a mutually accessible spectral region to generate a heterodyne beat between the two combs.

Under simultaneous control of synchronization and phase locking, Figure 8(b) shows frequency-counting records of the heterodyne beat signal between the laser

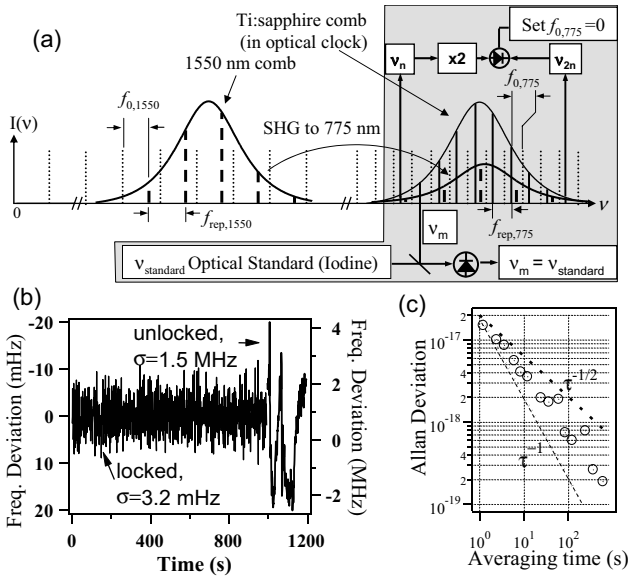


Fig. 8. (a) Schematic diagram of simultaneous synchronization and phase locking between a 1.5 μm mode-locked laser diode and a 775-nm mode-locked Ti:sapphire laser. The shaded area shows the implementation of an optical clock based on a Ti:sapphire fs comb phase-stabilized to an iodine standard. The laser diode's repetition frequency is 8 times that of the Ti:sapphire (not as shown in the figure). (b) The heterodyne beat recorded by a frequency counter at 1 s gate time, under no phase locking (short trace, with respect to the right vertical axis) and phase locking (long trace, with respect to the left vertical axis). (c) The Allan deviation associated with the phase-locked signal.

diode and Ti:sapphire combs. At 1 s gate time, the rms fluctuation ($\sigma_{\rm rms}$) of the heterodyne beat ($2 \times f_{0,1550} - f_{0,775}$) is reduced to 3.2 mHz, in sharp contrast to the 1.5 MHz rms fluctuation when $f_{0,1550}$ is not stabilized. By monitoring the beat error signal produced by the digital phase detector we insure that no cycles have slipped for the phase locked loop over this measurement period. Allan deviation of the stabilized beat frequency record is shown in Fig. 8(c), determined with respect to the 1.5 μm optical carrier frequency [40].

5 Femtosecond Lasers and External Optical Cavities

The combination of ultra-short pulses and optical cavities will open doors for a variety of exciting experiments. This requires the understanding of the intricate pulse-cavity interactions and the subsequent development of techniques to efficiently couple the ultra-short pulses into a high finesse optical cavity and coherently store them in the cavity. An immediate impact is on precision stabilization of ultrafast lasers [41]. Similar to the state-of-art stabilization of CW lasers, a cavity-stabilized ultrafast laser is expected to demonstrate superior short-term stability of both the pulse repetition frequency and the carrier-envelope phase. The improved stability is beneficial in particular for time-domain applications where the signal processing bandwidth is necessarily large. Another attractive application lies in broadband and ultrasensitive spectroscopy. The use of high finesse cavities has played a decisive role for enhancing sensitivity and precision in atomic and molecular spectroscopy. We expect a dramatic advancement in the efficiency of intracavity spectroscopy by exploiting the application of ultrashort pulses. In other words, a high detection sensitivity is achievable uniformly across the broad spectrum of the pulse. Applying cavity-stabilization techniques to femtosecond lasers, the comb structure of the probe laser can be precisely matched to the resonance modes of an empty cavity, allowing an efficient energy coupling for spectroscopic probe. Molecular samples introduced inside the high finesse cavity will have a strong impact on the dispersive properties of the cavity. In fact it is this dispersion-related cavity-pulling effect that will aid our sensitive detection process when we analyze the light transmitted through the cavity. Preliminary data on spectrally resolved, time-domain ring down measurement for intracavity loss over the entire femtosecond laser bandwidth are already quite promising.

To develop sources for ultrafast nonlinear spectroscopy, a properly designed, dispersion-compensated cavity housing a nonlinear crystal will provide efficient nonlinear optical frequency conversion of ultrashort optical pulses at spectral regions where no active gain medium exists. Furthermore, by simultaneously locking two independent mode-locked lasers to the same optical cavity, efficient sum and/or difference frequency generation can be produced over a large range of wavelengths. Under a similar motivation, a passive cavity can be used to explore coherent superposition of ultra short pulses, with cavity stabilization providing the means to phase coherently superpose a collection of successive pulses from a mode-locked laser. The coherently enhanced pulse stored in the cavity can be

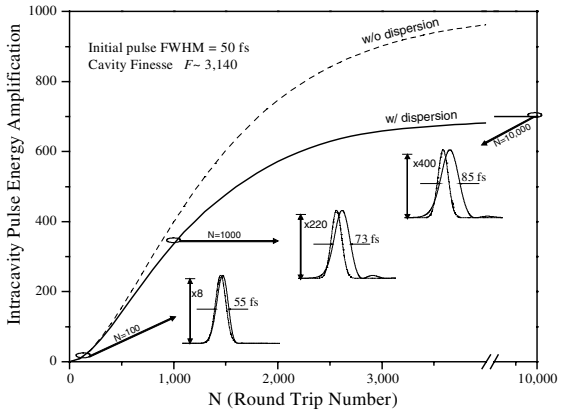


Fig. 9. Coherent evolution of a 50 fs pulse inside the cavity. Dashed line indicates the ideal case of a dispersion free cavity perfectly matched with the incident pulse train, while the solid line shows the effect of cavity dispersion in limiting the amount of energy coupled into the cavity.

switched out using a cavity-dumping element (such as a Bragg cell), resulting in a single phase-coherent amplified pulse. The use of a passive cavity also offers the unique ability to effectively amplify pulses at spectral regions where no suitable gain medium exists, such as for the infrared pulses from difference-frequency mixing or the UV light from harmonic generation. Unlike actively dumped laser systems, the pulse energy is not limited by the saturation of a gain medium or a saturable absorber needed for mode-locking. Instead, the linear response of the passive cavity allows the pulse energy to build up inside the cavity until limited by cavity loss and/or dispersive pulse spreading. Therefore storage and amplification of ultra-short pulses in the femtosecond regime requires precise control of the reflected spectral phase of the resonator mirrors as well as the optical loss of the resonator. While the reflected group delay of the mirrors only change the effective length of the resonator, the group delay dispersion (GDD) and higher-order derivatives of the group delay with respect to frequency affect the pulse shape. The net cavity GDD over the bandwidth of the pulse needs to be minimized in order to maintain the shape of the resonant pulse and allow for the coherent addition of energy from subsequent pulses. Figure 9 illustrates the evolution of a 50 fs pulse inside a cavity with a finesse of 3,140 under the conditions of zero cavity dispersion (dashed curve) and finite dispersion (solid curve). Three representative pulses at different stages of amplification are also shown to illustrate the pulse buildup process. Although the 50 fs pulse is stretched by the dispersive cavity, it is not severely distorted due to its coupling with the incident pulse train. If the incident pulses become too short, the cavity finesse too high, or the laser repetition frequency deviates significantly from the cavity free-spectral-range (FSR) frequency, the intra-cavity pulse may quickly pull apart into several pulses, and the meaning of a single pulse width would be lost.

We have applied the coherent pulse-stacking technique to both picosecond and femtosecond pulses. Initial studies have already demonstrated amplification of picosecond pulses of greater than 30 times at repetition rates of 253 kHz, yielding pulse energies greater than 150 nJ [33]. With significant room left for optimization of the cavity finesse (current value of ~ 350, limited by the cavity input-coupling mirror), we expect that amplifications greater than a hundred times are feasible, bringing pulse energies into the μJ range. While the use of picosecond pulses allows us to separate out complications arising from intra-cavity dispersion, for sub-100 femtosecond pulses, dispersive phase shifts in the cavity mirrors become an important topic. Preliminary results in enhancing low individual pulse energies for ~ 50 fs pulses illustrate the importance of GDD control. The external enhancement cavity incorporated specially designed negative GDD low-loss mirrors to simultaneously compensate for the Bragg cell's 3 mm of fused silica and provide a high finesse. The input-coupling mirror transmission was $\sim 0.8\%$, with a measured cavity finesse of 440. An intracavity energy buildup of 163 is expected, leading to single pulse amplifications of approximately 65 for the current setup given the 40% dumping efficiency of our Bragg cell. The negative GDD mirrors were designed to only partially compensate for the total cavity dispersion. The remaining cavity GDD was estimated at +20 to +30 fs^2. The excess dispersion results in pulse broadening and a non-uniform filtering of the transmitted pulse spectrum. Controlling the intracavity pressure allows fine tuning of the net cavity GDD to zero. Experimental results are in good agreement with independent numerical calculations. An intracavity buildup factor of 120 has already been achieved with pulses under 50 fs duration.

An important application of these advanced pulse control technologies is in the field of nonlinear-optics based spectroscopy and nanoscale imaging. For example, using two tightly synchronized picosecond lasers, we are able to achieve significant improvements in experimental sensitivity and spatial resolutions for coherent anti-Stokes Raman scattering(CARS) microscopy [32]. Vibrational imaging based on CARS spectroscopy is a powerful method for acquisition of chemically selective maps of biological samples [42]. In CARS microscopy, pulsed pump and Stokes beams are focused tightly to a single focal spot in the sample to achieve a high spatial resolution. The third order nonlinear interaction produces a signal photon that is blue-shifted (anti-Stokes signal) with respect to the incident beams. Strong CARS signals are obtained whenever the frequency difference between the pump and Stokes coincides with a Raman-active vibrational mode, which gives rise to the molecule-specific vibrational contrast in the image. Recent studies and technological improvements have demonstrated the exciting capability of CARS microscopy to attain high-resolution vibrational images of unstained living cells.

Practical applications of the CARS microscopy technique require pulsed light sources: optimized peak powers help boost the nonlinear signal. Pulses with temporal widths of 1–2 picoseconds (ps) should be used to match to the vibration bandwidths in order to optimize the CARS signal, with minimized non-resonant background and compromising spectral resolution. An important technical challenge is to achieve tight synchronization between two mode-locked lasers that

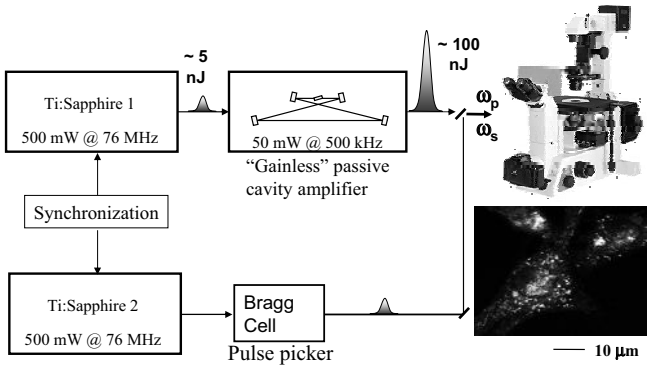


Fig. 10. Schematic setup for coherent anti-Stokes Raman scattering (CARS) microscopy on living cells using two separate picosecond mode-locked lasers. The two lasers are tightly synchronized to minimize detection noise while the pump laser is enhanced in its peak power by a passive optical cavity. High S/N image of distribution of lipids in a live unstained fibroblast cell is achieved via resonant detection of C-H vibration frequency.

produce the frequency difference which matches the vibrational resonance. Another important consideration is that while the repetition rate of the pulse train needs to be low enough to avoid thermal damage to the cell due to a high average power, the peak power of the pulses needs to be reasonably high to aid the nonlinear signal strength. A schematic setup combining CARS spectroscopy and pulse manipulation is shown in Fig. 10. The technologies of pulse synchronization and coherent pulse stacking therefore become ideal tools for carrying out this task of spectroscopy plus microscopy.

Acknowledgments

We thank E. Potma, X.-S. Xie, S. Foreman, I. Thoman, H. Kapteyn, S. Cundiff, T. Fortier, and E. Ippen for fruitful collaborations and J. L. Hall for his interest in our work. We gratefully acknowledge financial support from ONR, NASA, NIST, and NSF. K. Holman is a Hertz Foundation Graduate Fellow. R. J. Jones is a National Research Council Research Associate Fellow.

References

1. S.T. Cundiff, J. Ye: Rev. Mod. Phys. **75**, 325 (2003)
2. A. Baltuska, T. Udem, M. Uiberacker, M. Hentschel, E. Goulielmakis, C. Gohle, R. Holzwarth, V.S. Yakoviev, A. Scrinzi, T.W. Hänsch, F. Krausz: Nature **421**, 611 (2003)
3. T. Udem, J. Reichert, R. Holzwarth, T.W. Hänsch: Phys. Rev. Lett. **82**, 3568 (1999) T. Udem, J. Reichert, R. Holzwarth, T.W. Hänsch: Opt. Lett. **24**, 881 (1999)

4. J. Ye, T.H. Yoon, J.L. Hall, A.A. Madej, J.E. Bernard, K.J. Siemsen, L. Marmet, J.-M. Chartier, A. Chartier: Phys. Rev. Lett. **85**, 3797 (2000)
5. D.J. Jones, S.A. Diddams, J.K. Ranka, A. Stentz, R.S. Windeler, J.L. Hall, S.T. Cundiff: Science **288**, 635 (2000)
6. A. Apolonski, A. Poppe, G. Tempea, Ch. Spielmann, T. Udem, R. Holzwarth, T.W. Hänsch, F. Krausz: Phys. Rev. Lett. **85**, 740 (2000)
7. S.A. Diddams, T. Udem, J.C. Bergquist, E.A. Curtis, R.E. Drullinger, L. Hollberg, W.M. Itano, W.D. Lee, C.W. Oates, K.R. Vogel, D.J. Wineland: Science **293**, 825 (2001)
8. J. Ye, L.-S. Ma, J.L. Hall: Phys. Rev. Lett. **87**, 270801 (2001)
9. J.D. Jost, J.L. Hall, J. Ye: Opt. Express **10**, 515 (2002)
10. R.K. Shelton, L.-S. Ma, H.C. Kapteyn, M.M. Murnane, J.L. Hall, J. Ye: Science **293**, 1286 (2001)
11. A. Baltuska, T. Fuji, T. Kobayashi: Phys. Rev. Lett. **88**, 133901 (2002)
12. L. Xu, C. Spielmann, A. Poppe, T. Brabec, F. Krausz, T.W. Hänsch: Opt. Lett. **21**, 2008 (1996)
13. H.R. Telle, G. Steinmeyer, A.E. Dunlop, J. Stenger, D.H. Sutter, U. Keller: Appl. Phys. B **69**, 327 (1999)
14. T. Fortier, D.J. Jones, S.T. Cundiff, Opt. Lett. **28**, 2198 (2003).
15. R. Ell, U. Morgner, F.X. Kärtner, J.G. Fujimoto, E.P. Ippen, V. Scheuer, G. Angelow, T. Tschudi, M.J. Lederer, A. Boiko, B. Luther-Davies: Opt. Lett. **26**, 373 (2001)
16. A. Bartels, H. Kurz: Opt. Lett. **27**, 1839 (2002)
17. J.K. Ranka, R.S. Windeler, A.J. Stentz: Opt. Lett. **25**, 25 (2000)
18. J. Reichert, R. Holzwarth, T. Udem, T.W. Hänsch: Opt. Comm. **172**, 59 (1999)
19. K.W. Holman, R.J. Jones, A. Marian, S.T. Cundiff, J. Ye: Opt. Lett. **28**, 851 (2003)
20. H.R. Telle, B. Lipphardt, J. Stenger: Appl. Phys. B **74**, 1 (2002)
21. S.A. Diddams, L. Hollberg, L.S. Ma, L. Robertsson: Opt. Lett. **27**, 58 (2002)
22. M. Niering, R. Holzwarth, J. Reichert, P. Pokasov, T. Udem, M. Weitz, T.W. Hänsch, P. Lemonde, G. Santarelli, M. Abgrall, P. Laurent, C. Salomon, A. Clairon: Phys. Rev. Lett. **84**, 5496 (2000)
23. J. Ye, J.L. Hall, S.A. Diddams: Opt. Lett. **25**, 1675 (2000)
24. G. Wilpers, T. Binnewies, C. Degenhardt, U. Sterr, J. Helmcke, F. Riehle: Phys. Rev. Lett. **89**, 230801 (2002)
25. R.J. Rafac, B.C. Young, J.A. Beall, W.M. Itano, D.J. Wineland, J.C. Bergquist: Phys. Rev. Lett. **85**, 2462 (2000)
26. L. Hollberg, C.W. Oates, E.A. Curtis, E.N. Ivanov, S.A. Diddams, T. Udem, H.G. Robinson, J.C. Bergquist, R.J. Rafac, W.M. Itano, R.E. Drullinger, D.J. Wineland: IEEE J. Quant. Electron. **37**, 1502 (2001)
27. J. Ye, L. Robertsson, S. Picard, L.S. Ma, J.L. Hall: IEEE Trans. Instrum. Meas. **48**, 544 (1999)
28. T.M. Fortier, D.J. Jones, J. Ye, S.T. Cundiff, R. S. Windeler: Opt. Lett. **27**, 1436 (2002)
29. R.K. Shelton, S.M. Foreman, L.-S. Ma, J.L. Hall, H.C. Kapteyn, M.M. Murnane, M. Notcutt, J. Ye: Opt. Lett. **27**, 312 (2002)
30. S.M. Foreman, D.J. Jones, J. Ye: Opt. Lett. **28**, 370 (2003)
31. R.J. Jones, J. Ye: Opt. Lett. **27**, 1848 (2002)
32. E.O. Potma, D.J. Jones, J.X. Cheng, X.S. Xie, J. Ye: Opt. Lett. **27**, 1168 (2002)
33. E.O. Potma, C. Evans, X.S. Xie, R.J. Jones, J. Ye: Opt. Lett. **28**, 1835 (2003)

34. W.-Y. Cheng, L. Chen, T.H. Yoon, J.L. Hall, J. Ye: Opt. Lett. **27**, 571 (2002)
35. J. Vigué, M. Broyer, J.C. Lehmann: Phys. Rev. Lett. **42**, 883 (1979)
36. C.J. Borde, G. Camy, B. Decomps, J.-P. Descoubes: J. Phys. (Paris) **42**, 1393 (1981)
37. J.P. Pique, F. Hartmann, S. Churassy, R. Bacis: J. Phys. (Paris) **47**, 191 (1986)
38. B. Bodermann, H. Knöckel, E. Tiemann: Eur. Phys. J. D **19**, 31 (2002)
39. R.J. Jones, W.Y. Cheng, K.W. Holman, L. Chen, J.L. Hall, J. Ye: Appl. Phys. B **74**, 597 (2002)
40. K.W. Holman, D.J. Jones, J. Ye, E.P. Ippen: Opt. Lett. **28**, 2405 (2003)
41. R.J. Jones, J.-C. Diels: Phys. Rev. Lett. **86**, 3288 (2001)
42. A. Volkmer, J.X. Cheng, X.S. Xie: Phys. Rev. Lett. **87**, 023901 (2001)

Ultracold Trapped Molecules: Novel Systems for Tests of the Time-Independence of the Electron-to-Proton Mass Ratio

U. Fröhlich, B. Roth, P. Antonini, C. Lämmerzahl*, A. Wicht, and S. Schiller

Institut für Experimentalphysik, Heinrich-Heine-Universität Düsseldorf,
40225 Düsseldorf, Germany, www.exphy.uni-duesseldorf.de

Abstract. The vibrational and rotational transition frequencies in molecules are functions of the electron-to-nucleon mass ratio. They can therefore be used for experiments searching for a time-dependence of the electron-to-proton and nucleon-to-nucleon mass ratios. We propose to to perform such tests with very high precision using ultracold molecular ions trapped in a radio-frequency trap and sympathetically cooled by atomic ions. The current status of an experiment using Be^+-ions as a coolant medium to cool light diatomic molecular ions is described. An interesting perspective is to perform high-precision spectroscopy on single ultracold molecules. We sketch an approach towards this goal.

1 Introduction

A large experimental effort is currently under way to test the foundation of metric theories of gravity, the Einstein Equivalence Principle (EP) [1]. This effort is motivated in part by the difficulties in unifying the theory of gravity with quantum theory. The EP includes the principle of Local Position Invariance, which states that the fundamental constants of nature, such as the fine structure constant α, the mass ratios of the elementary particles, etc. are independent of time and position.

Tests of the time-independence of the fundamental constants can be performed by laboratory experiments or astronomical observations [2]. In laboratory tests, the frequencies of dissimilar types of oscillators are compared as a function of time, while in astronomical tests, the frequencies of the electromagnetic waves emitted in the far past by oscillators located in distant sources are compared with the current frequency values of the *same* type of oscillators as obtained in the laboratory. Table 1 gives an overview of some microscopic and macroscopic systems that define transition or oscillation frequencies.

Analysis of astronomical observations claims a variation of α on the order of 5 ppm over billions of years [7]. This claim provides increased stimulus for laboratory experiments. These have entered a new era, where use is made of the development of atomic clocks based on ultracold atoms and atomic ions, and new methods of precise comparison of optical and microwave frequencies. An improvement by more than an order of magnitude in the most stringent laboratory limits has already been achieved.

* Present address: ZARM, Universität Bremen, Germany.

U. Fröhlich et al., Ultracold Trapped Molecules: Novel Systems for Tests of the Time-Independence of the Electron-to-Proton Mass Ratio, Lect. Notes Phys. **648**, 297–307 (2004)
http://www.springerlink.com/

Table 1. Dependence of the transition energy or cavity photon energy (in units of the Rydberg energy E_R) of various oscillators on fundamental constants. For the function F, see [3]. The function $G(\alpha)$ is a relativistic correction factor for principal transitions; it reduces to unity in the non-relativistic limit (small-Z atoms) [4]. The dependencies for molecular vibrational and rotational transitions are given for diatomic molecules, where μ is the reduced mass. For vibrational transitions between low-lying levels $H(m_e/\mu) \simeq \sqrt{m_e/\mu}$. The resonance frequency of a cavity is proportional to the size and thus to the interatomic spacing, resulting in $h\nu \propto E_R/\alpha$. Here, propagation of the electromagnetic wave in vacuum is assumed. For the case of monolithic cavities, see [5]. The last entry refers to the weak interaction-induced splitting between molecular levels of mirror molecules (enantiomers) [6] and to the parity non-conservation light shift in atoms. G_F is the Fermi constant, θ_W is the Weinberg angle.

Type	dependence
Hyperfine splitting	$g\,\alpha^2 F(\alpha Z) m_e/m_p$
Fine-structure splitting	α^2
Electronic transitions	$G(\alpha)$
Rotational transitions	m_e/μ
Vibrational transitions	$H(m_e/\mu)$
Cavity frequencies	α^{-1}
Parity violation splitting	G_F , $\sin^2\theta_W$

The best laboratory limits produced thus far constrain certain *combinations* of fundamental constants. For example, the recent high-precision tests concerned a comparison of hyperfine transition frequencies of ultracold atoms that yielded a test of the time-independence of $g_{Cs}/g_{Rb}\,\alpha^{0.44}$ at the level of $7 \cdot 10^{-16}$/yr [8]. A comparison of an optical transition in Hg^+ and the hyperfine frequency of Cs gave a limit of $7 \cdot 10^{-15}$/yr for the combination $g_{Cs}\alpha^6 m_e/m_p$ [9]. Vibrational transition frequencies in room-temperature molecular gases have also been investigated: a comparison of a vibration frequency of methane (CH_4) and the hyperfine frequency of Cs has led to an upper limit on the order 10^{-12}/yr for the time dependence of the ratio [11]. A similar test using the molecule OsO_4 led to a limit $2 \cdot 10^{-13}$/yr [10]. Taking these two results together, the ratio of the two vibrational frequencies is found to be constant at the level of 10^{-12}/yr. This may be interpreted as a limit for the time-dependence of the ratio of the characteristic nuclear masses corresponding to the vibrational modes studied.

From astronomical observations limits for individual fundamental constants have been derived. For example, concerning α the relativistic and spin-orbit energies of various atomic ions in quasars have been compared to laboratory values [4,7]. Concerning m_e/m_p, the ratio of energy differences of ro-vibrational levels in a given electronic state of molecules in distant interstellar clouds can be compared to present-day laboratory values. Based on observation of neutral H_2 absorption lines, constancy within 40 ppm was found [12,13].

2 Molecular Tests of Constancy of Electron-to-Nucleon Mass Ratios

Laboratory tests sensitive to the electron-to-nucleon mass ratio, and insensitive to other fundamental constants to lowest order, can be performed by one of the following methods [14]:

(i) Comparison of a vibrational or rotational molecular frequency and a non-relativistic electronic transition in an atom;
(ii) Comparison of a (ro-)vibrational frequency and a rotational frequency, in the same molecule or in different molecules;
(iii) Comparison of vibrational frequencies of different transitions in the same molecule or in different molecules;
(iv) Comparison of vibrational frequencies of different molecules.

The approach (iii) is based on the fact that the interatomic molecular potential is not harmonic and therefore the energy spacing between adjacent vibrational levels is not constant. The decrease in spacing for a diatomic molecule is itself a function of m_e/μ, where μ is the reduced mass of the oscillating nuclei, and of the particular molecule.

If the comparisons are performed using hydrogen molecules (H_2, H_2^+, H_3^+), the ratio m_e/m_p is accessed directly. If an arbitrary molecule is used, the ratio m_e/μ is accessed. A limit for the time-independence of m_e/m_p can only be given under the assumption that the ratios of nuclear masses are time-independent, i.e. $\mu/m_p = const.$ However, if rotational or vibrational transitions of *different* molecules are compared (approach (iv)) then the ratio of nuclear masses can be probed [15]. An example is the comparison of H_2^+ and D_2^+ or H_2^+ and HD^+. Such tests would probe the time-independence of the strong interaction [16].

Referring to Fig. 1, we can describe the approaches (ii) and (iii) by

$$\frac{\mathrm{d}\,\ln\left(\frac{\nu_a-\nu_b}{\nu_a}\right)}{\mathrm{d}t} = (1-s)\frac{\mathrm{d}\,\ln(m_e/m_p)}{\mathrm{d}t}\ , \tag{1}$$

$$\frac{\mathrm{d}\,\ln\left(\frac{\nu_1}{\nu_2}\right)}{\mathrm{d}t} = (s_1-s_2)\frac{\mathrm{d}\,\ln(m_e/m_p)}{\mathrm{d}t}\ . \tag{2}$$

Here the coefficients s_k are defined by the derivative of the relevant transition frequencies ν_k with respect to the electron-to-proton mass ratio,

$$\frac{m_e/m_p}{\nu_k}\frac{\mathrm{d}\nu_k}{\mathrm{d}(m_e/m_p)} \equiv s_k. \tag{3}$$

The coefficients s_k can be calculated using quantum chemical algorithms. For example, Hilico et al. have performed such calculations for the vibrational levels of the H_2^+ and D_2^+ molecules [17].

To reach an interesting level of sensitivity and open up a direction for future progress, it is certainly necessary to use ultracold molecules. Since the approaches (i-iv) can be applied to essentially any molecule, molecular ions can

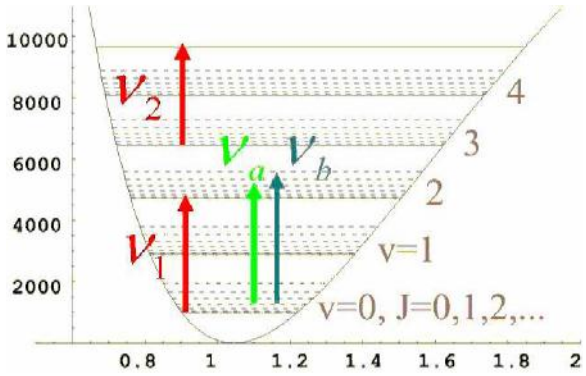


Fig. 1. Principle of the test of the time-independence of m_e/m_p. The relative difference of two ro-vibrational transition frequencies ν_a, ν_b sharing a common level (method (ii)) or the ratio of two vibrational transition frequencies ν_1, ν_2 (method (iii)) is measured over the course of time.

be considered. These are favorable, since they can be (translationally) cooled using sympathetic cooling by laser cooled atomic ions [18]. When crystallized, the molecular ions are in the Lamb-Dicke regime, and the ro-vibrational transitions will exhibit resolved sidebands. The linewidths are expected to be equal to their natural linewidth, on the order of tens of Hz for dipole-allowed vibrational transitions in the ground electronic state, and below 1 Hz for pure rotational transitions in the lowest vibrational state. The quality factors of the transitions are therefore in the range of 10^{11} or larger.

On the experimental side, laser sources for performing fundamental or overtone vibrational or stimulated Raman spectroscopy with such resolution can be implemented. For example, continuous-wave optical parametric oscillators (OPO) currently can provide radiation at wavelengths up to 4 µm with free-running linewidths below 100 kHz. Diode lasers of similar linewidth are available in the telecom wavelength range. Alternatively, difference frequency generation can be used if low power levels are sufficient. The linewidth of the sources can be reduced by frequency-stabilization to cavities. Such laser sources can be used to excite fundamental or overtone vibrational transitions.

The detection of excitation to vibrational or rotational states within the electronic ground state in cold molecular ion ensembles is not straightforward, since the number of molecules will be limited and fluorescence detection is impractical (fluorescence decay rates are small and fluorescence wavelengths lie in the mid-infrared). Destructive detection of excitation is one approach and will be described in the next section. A proposal for nondestructive detection of molecular excitation is presented in the last section.

Rotational transition frequencies can be determined from a frequency difference $\nu_a - \nu_b$ between two ro-vibrational excitations, as shown in Fig. 1, or from the frequency difference of two waves used for a stimulated Raman transition within the same vibrational state, or by direct microwave spectroscopy.

Since the frequencies of the various transitions to be compared will typically be very different (except if method (iv) is used with molecules appropriately selected to have accidental degeneracies), frequency comb techniques will have to be used.

In order to reach interesting sensitivities for time-independence, the evaluation of systematic shifts will be of central importance. The experience gained from work on cold atomic ion frequency standards will certainly be of significance.

3 Sympathetic Cooling of Molecular Ions and Spectroscopy

In our laboratory, two experiments dedicated to translational cooling of molecular ions are under way. The first experiment uses $^{9}Be^{+}$ as a coolant ion, the second uses $^{137}Ba^{+}$. The choice of Beryllium was made in order to be able to efficiently trap light molecular ions, in particular molecular hydrogen ions. The much heavier Barium ions can be used to trap heavier molecular ions, even proteins, provided they are sufficiently highly charged [19]. The issue here is the requirement of stable trapping, which constrains the allowed charge-to-mass ratio of the molecular ions compared to that of the coolant ions.

One of our goals in studying molecular hydrogen ions is to measure the proton and deuteron masses spectroscopically. Besides containing the most fundamental nuclei, the diatomic molecular hydrogen ions H_2^+, D_2^+, HD^+, and HT^+, being relatively simple three-body quantum systems, are also the only ones for which ab-initio theory has the potential of reaching sufficient accuracy in the near future. The Schrödinger equation for such systems can be solved with essentially arbitrary accuracy (10^{-14}) as a function of the masses of the three particles. The required relativistic and QED corrections have so far been worked out to a relative accuracy on the order of 10^{-7}, but with potential for future improvements. Precision measurements on these ions could eventually lead to highly accurate values for m_e/m_p and the nuclear mass ratios m_p/m_d as well as m_p/m_T. Among the above ions, the heteronuclear ion HD^+ is of particular interest from an experimental point of view since dipole-allowed vibrational transitions can be excited. We stress that even at the current stage of theoretical accuracy tests of the time independence of essentially any nuclear-to-electron mass ratio, as opposed to a measurement thereof, can be performed using corresponding molecular ions.

In the Be^{+}-experiment, we use a four-rod linear radio-frequency trap with end caps. The central electrode length is 16 mm, rod diameter is 9.9 mm, rod-to-trap-axis distance is 4.3 mm. The 313 nm cooling radiation is generated using doubly-resonant sum frequency generation (SFG) of a resonantly doubled Nd:YAG laser, and a Ti:Sapphire laser at 760 nm [20]. The Nd:YAG laser is frequency-stabilized to a hyperfine transition in molecular iodine; since the cavity used for SFG is locked to the Nd:YAG laser and the Ti:Sapphire laser is locked to the cavity, the sum frequency wave is then also frequency stabilized. An AOM placed before

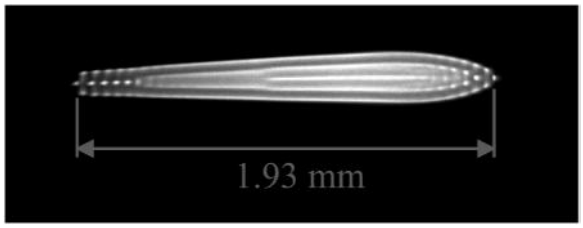


Fig. 2. Stable Beryllium Coulomb crystal. Cooling laser frequency was locked approx. 35 MHz red-detuned from resonance.

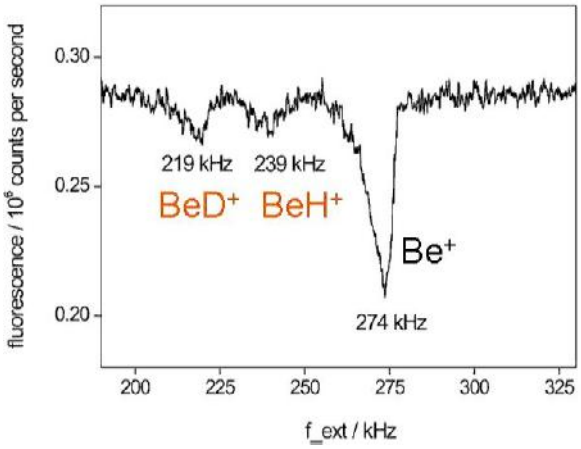


Fig. 3. Secular oscillation mass spectrum showing presence of beryllium hydride and deuteride ions produced by chemical reactions. Shown is fluorescence rate of the Be^+ ions as a function of frequency of the ac voltage applied to a 2 cm wide plate electrode at a distance 9.5 mm from the trap axis. Ac amplitude was 0.5 V.

the iodine stabilization setup allows to shift the stable UV frequency within a range of 340 MHz. UV output powers up to 80 mW were obtained.

Be^+-Coulomb crystals of various sizes and shapes were obtained by varying the trap and loading parameters. Figure 2 is an example of a small crystal.

We have produced molecular ions in situ by leaking HD gas into the UHV chamber for 1–2 min with pressure of $4 \cdot 10^{-10}$ mbar. Chemical reactions between HD and the crystallized Be^+-ions resulted in BeH^+ and BeD^+. Their presence is proven by excitation of the secular oscillation in radial direction. The excitation heats the Be^+ ion ensemble by Coulomb interaction and results in a decrease in Be^+ fluorescence. Figure 3 shows the mass spectrum of the ions contained in the trap. The secular frequencies lie within 1 % of the expected values. The crystal remains stable during excitation, but experiences a small Be^+ ion loss. Beryllium hydride and deuteride are heavier than the coolant ion. It is therefore expected that they are located outside the Be^+-crystal. Indeed, no significant dark regions are visible within the Be^+-crystal. We expect that the molecular

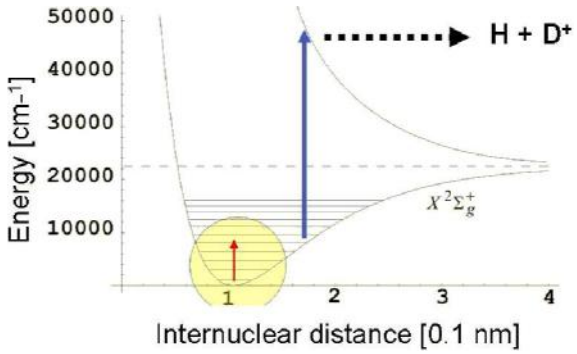


Fig. 4. Detection of molecular excitation by 1+1 REMPI. A vibrational transition can be detected since molecules in an excited vibrational state can be selectively dissociated with UV radiation of appropriate wavelength. The method is also applicable to selectively detect population in any particular rotational state of the lowest vibrational state if first a vibrational excitation laser is used to transfer population from the rotational state to an appropriate vibrational level

ions were sympathetically cooled and crystallized; to prove this statement, we would need to perform molecular spectroscopy or add another fluorescent atomic ion ensemble of greater mass so that the molecular ions would become "visible" as a dark shell between the added atomic ions and the Be^+-crystal [18].

As mentioned above, the detection of population in a long-lived ro-vibrational or rotational state poses a problem. A destructive detection method is possible by means of 1+1 resonance multi-photon ionization (REMPI). We plan to implement this for spectroscopy of HD^+, see Fig. 4.

Since for sympathetically cooled molecular ions in equilibrium with the laser-cooled atomic ions the internal temperature will at most be 300 K (the temperature of the vacuum chamber), in diatomic molecular ions only the lowest vibrational level will be populated initially. A vibrational excitation to be detected will transfer a fraction of these ions into another vibrational level. If a pure rotational transition (e.g. by stimulated Raman transition) is to be detected, it can be followed by a one-photon laser vibrational excitation to an excited vibrational level. The task then is to dissociate molecules preferentially from such an excited vibrational level. At least for HD^+ this is indeed possible, as shown in Fig. 5. A large ratio between the dissociation probabilities from a state $v' \neq 0$ and from $v = 0$ can be achieved for any v' by appropriate choice of dissociation wavelength. A good choice of target vibrational level and dissociation wavelength are $v' = 4$ (corresponding to a $0 \rightarrow 4$ transition wavelength around 1.4 μm) and 266 nm (obtainable by frequency-quadrupling a Nd:YAG laser). The required photodissociation energy densities can be obtained from pulsed or cw lasers. For the $v' = 5$ target level one might be able to use the 313 nm cooling radiation as dissociation light, if it is sufficiently intense. A suitable ion optics and ion counter to extract and detect the dissociation products is required.

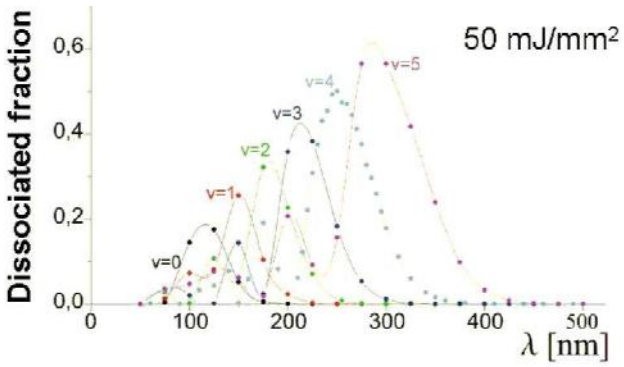


Fig. 5. Theoretical photodissociation probability of HD^+ ions in various vibrational states vs. excitation wavelength for an energy density of 50 mW/mm^2 and 1 s duration. [21,22].

4 Quantum Jump Spectroscopy

Obviously, a destructive detection method is not favourable, because of the need to reload the trap and possibly the need to reestablish a sufficiently similar configuration of the molecular/atomic ion ensemble. Various types of non-destructive detection of molecular excitation can be envisaged. Here we consider one method that is of interest if one seeks to perform spectroscopy on a single molecule. This might be the ultimate goal in precision molecular spectroscopy, since the spatial state of the molecule is then well-defined, and the determination of systematic effects is simpler. The method we propose is related to the concepts of implementing quantum gates [23] and of performing high resolution spectroscopy of atomic ions that do not have laser cooling transitions [24]. Basically, an appropriate atomic ion is used as a monitor of the internal state of the test ion (here a molecule).

Figure 6 sketches a simplified view of the procedure. A single molecular ion and a single coolant atomic ion are prepared in the trap. The atomic ion level scheme must have an accessible long-lived state. This can be an electronic state connected to the ground state by a dipole-forbidden transition, or a hyperfine state in the ground electronic state [24]. Here we consider the first case for simplicity, but the latter case would be relevant if the coolant ion is Be^+. The two-ion crystal is cooled by resolved sideband cooling to the motional ground state of e.g. the axial crystal mode [25]. In our model scheme this requires the P-state linewidth to be smaller than the phonon frequency. However, in practice resolved sideband cooling would be performed on a dipole-forbidden atomic transition, so that this condition can be satisfied. Note that there is no need to focus the sideband cooling radiation selectively onto the atomic ion.

The atomic ion is then transferred to the metastable state (here a D state). Now the internal vibration of the molecule is excited, with the laser tuned to the blue motional sideband. The ion crystal motional oscillation is thereby simultaneously excited. The figure shows a fundamental vibrational excitation ($v, J =$

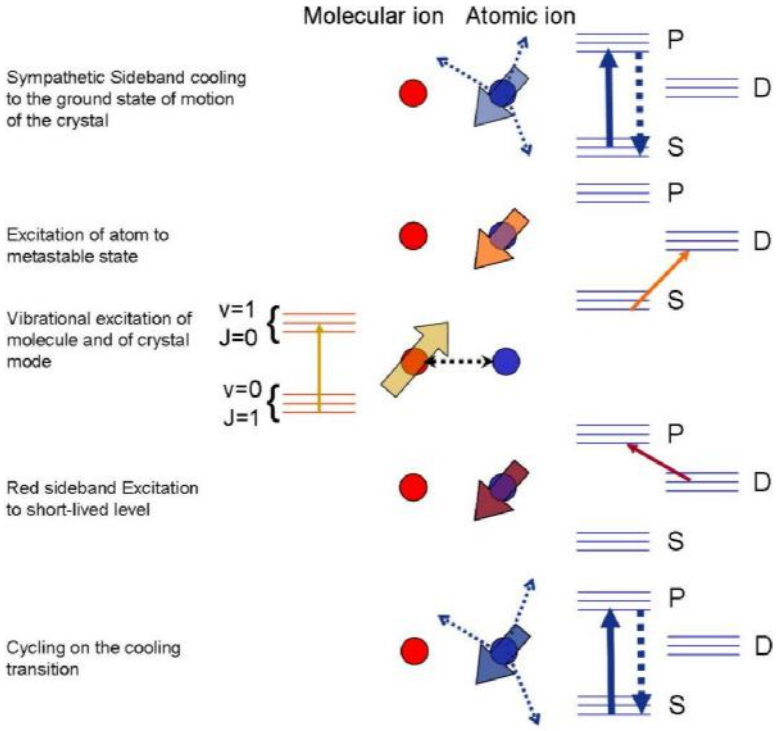


Fig. 6. Schematic of a method for detecting quantum jumps in molecules with high signal-to-noise ratio. The sublevels shown in the level diagrams of molecule (left) and atom (right) are the vibrational states of the crystal oscillation along the trap axis.

$1) \rightarrow (v+1, J=0)$ of the molecule and the excitation of one crystal phonon. The analogue can be done with a pure rotational excitation $(v, J) \rightarrow (v, J \pm 2)$ by using stimulated Raman scattering.

The next step consists in detecting whether a crystal phonon is present or not. To this end, the atomic ion is transferred to one of the levels of the cooling transition, using red-detuned light. This ensures that if no phonon is present, the atomic ion remains in the metastable state.

As a result, the atomic ion has been brought into a level from which the cycling can be driven, resulting in a high fluorescence rate. Thus, if the internal excitation of the molecule has taken place, this results in fluorescence from the atomic ion; the molecular quantum jump has thereby been detected. The fluorescence continues even when the molecule decays back into the ground state. Note that a ro-vibrational molecular transition $(v=0, J=1) \rightarrow (v=1, J=0)$ is a cycling transition, albeit one with a small decay rate, due to the small transition frequency.

For a spectroscopic measurement, the outlined procedure would be repeated for different values of the molecular excitation frequency.

This method, albeit technically difficult, would allow to perform spectroscopy on narrow molecular transitions with similar sensitivity as in the electron shelving method of single atomic ion spectroscopy.

5 Conclusion

In this paper we have pointed out that ultracold molecular ions have the potential for being used for measurements of very high spectroscopic precision, similar to what has already been demonstrated with atomic ions. Vibrational and rotational transitions limited only by the narrow natural linewidth are in principle accessible. Some approaches on how to perform the spectroscopy of molecular transitions in practice have been mentioned.

This opens up fascinating possibilities for the metrology of the electron-to-proton and proton-to-deuteron mass ratios (both absolute determination as well as time-independence tests). Measurements of parity violation, tests of the symmetrization postulate, and of the electron dipole moment in molecules are further perspectives on the use of molecular ions for fundamental physics studies. Of course, on the physical chemistry side, precision molecular structure studies will represent a huge new domain of activity, which is likely to challenge quantum chemical theory significantly.

Acknowledgments

This work has been performed in the framework of DFG Project Schi 431/4-1 and the Gerhard-Hess Program. U.F. was supported by a fellowship of the Düsseldorf Entrepreneurs Foundation. We are grateful to H. Schnitzler, T. Boley, and A. Peters for their essential contributions in the initial phase of the work described in Sect. 3. We thank P. Zoller for mentioning quantum gates in the context of molecular quantum jumps.

References

1. C. Will: *Theory and experiment in gravitational physics*, rev. edn. (Cambridge Univ. Press, Cambridge, 1993)
2. J.P. Uzan, Rev. Mod. Phys. **75**, 403 (2003)
3. J.D. Prestage, R. L. Tjoelker, L. Maleki, Phys. Rev. Lett. **74**, 3511 (1995)
4. V.A. Dzuba, V.V. Flambaum, J.K. Webb, Phys. Rev. A **59**, 230 (1999)
5. C. Braxmaier et al., Phys. Rev. D **64**, 042001 (2001)
6. M. Ziskind et al., Eur. J. Phys. D **20**, 219 (2002)
7. M.T. Murphy, J.K. Webb, V.V. Flambaum, Mon. Not. Roy. Astr. Soc. **365**, 609 (2003)
8. H. Marion et al., Phys. Rev. Lett. **90**, 150801 (2003)
9. S. Bize et al., Phys. Rev. Lett. **90**, 150802 (2003)
10. G.D. Rovera, IEEE Trans. Instrum. Meas. **48**, 571 (1999)
11. Y.S. Domnin et al., Quant. Elec. **26**, 1081 (1996)

12. A.Y. Potekhin et al., Astrophy. J. **505**, 523 (1998)
13. S.A. Levshakov et al., Mon. Not. Roy. Astr. Soc. **333**, 373 (2002)
14. Only molecular tests are considered here. Of course, mass spectrometry in Penning traps is an alternative approach.
15. T. Wiklind and F. Combes, Astron. Astrophys. **328**, 48 (1997)
16. H. Fritzsch, Fundamental Constants and Their Possible Time Dependence, Lect. Notes Phys. **648**, 107–113 (2004)
17. L. Hilico et al., Eur. Phys. J. D **12**, 449 (2000)
18. K. Molhave, M. Drewsen, Phys. Rev. A **62**, 011401 (2000)
19. S. Schiller, C. Lämmerzahl, Phys. Rev. A **68**, 053406 (2003)
20. H. Schnitzler et al., Appl. Opt. **41**, 7000 (2001)
21. M. Tadjeddine, G. Parlant, Mol. Phys. **33**, 1797 (1977)
22. H. Schnitzler, Ph.D. thesis, Univ. Konstanz (2001)
23. I. Cirac, P. Zoller, Phys. Rev. Lett. **74**, 4091 (1995)
24. D.J. Wineland et al.: 'Quantum Computers and Atomic Clocks'. In: *Proc. of the 6th Symp. on Frequency Standards and Metrology*. ed. by P. Gill (World Scientific, New Jersey, 2002) pp. 361–368
25. H. Rhode et al., J. Opt. B **3**, S34 (2001)

35 Years of Testing Relativistic Gravity: Where Do We Go from Here?

Slava G. Turyshev[1], James G. Williams[1], Kenneth Nordtvedt, Jr.[2], Michael Shao[1], and Thomas W. Murphy, Jr.[3]

[1] Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109 USA
[2] Northwest Analysis, 118 Sourdough Ridge Road, Bozeman, MT 59715 USA
[3] Physics Department, University of California, San Diego, CASS-0424, 9500 Gilman Dr., La Jolla, CA 92093, USA

Abstract. This paper addresses the motivation, technology and recent results in the tests of the general theory of relativity in the solar system. We specifically discuss Lunar Laser Ranging (LLR), the only technique available to test the Strong Equivalence Principle (SEP) and presently the most accurate method to test for the constancy of the gravitational constant G. After almost 35 years since beginning of the experiment, LLR is poised to take a dramatic step forward by proceeding from cm to mm range accuracies enabled by the new Apache Point Observatory Lunar Laser-ranging Operation (APOLLO) currently under development in New Mexico. This facility will enable tests of the Weak and Strong Equivalence Principles with a sensitivity approaching 10^{-14}, translating to a test of the SEP violation parameter, η, to a precision of $\sim 3 \times 10^{-5}$. In addition, the v^2/c^2 general relativistic effects would be tested to better than 0.1%, and measurements of the relative change in the gravitational constant, $\dot{G}/G$, would be $\sim 0.1\%$ the inverse age of the universe.

This paper also discusses a new fundamental physics experiment that will test relativistic gravity with an accuracy better than the effects of the second order in the gravitational field strength, $\propto G^2$. The Laser Astrometric Test Of Relativity (LATOR) will not only improve the value of the parameterized post-Newtonian (PPN) γ to unprecedented levels of accuracy of 1 part in 10^8, it will also be able to measure effects of the next post-Newtonian order (c^{-4}) of light deflection resulting from gravity's intrinsic non-linearity, as well as measure a variety of other relativistic effects. LATOR will lead to very robust advances in the tests of fundamental physics: this mission could discover a violation or extension of general relativity, or reveal the presence of an additional long range interaction in the physical law. There are no analogs to the LATOR experiment; it is unique and is a natural culmination of solar system gravity experiments.

1 Introduction

Einstein's general theory of relativity (GR) began with its empirical success in 1915 by explaining the anomalous perihelion precession of Mercury's orbit, using no adjustable theoretical parameters. Shortly thereafter, Eddington's 1919 observations of star lines-of-sight during a solar eclipse confirmed the doubling of the deflection angles predicted by GR as compared to Newtonian and Equivalence Principle arguments. Following these beginnings, the general theory of relativity has been verified at ever-higher accuracy. Thus, microwave ranging to

S.G. Turyshev et al., 35 Years of Testing Relativistic Gravity: Where Do We Go from Here?, Lect. Notes Phys. **648**, 311–330 (2004)
http://www.springerlink.com/

the Viking Lander on Mars yielded an accuracy of ∼0.1% in the tests of GR [1,2]. The astrometric observations of quasars on the solar background performed with Very-Long Baseline Interferometry (VLBI) improved the accuracy of the tests of gravity to ∼ 0.03% [3]. Lunar Laser Ranging (LLR), the continuing legacy of the Apollo program, has provided ∼ 0.01% verification of the general relativity via precision measurements of the lunar orbit [4–8]. Finally, the recent experiments with the Cassini spacecraft have improved the accuracy of the tests to ∼ 0.003% [9–11]. As a result, by now not only is the 'non-relativistic', Newtonian regime well understood, but the first 'post-Newtonian' approximation is well-studied, making general relativity the standard theory of gravity where astrometry and spacecraft navigation are concerned.

The continued inability to merge gravity with quantum mechanics, and recent observations in cosmology indicate that the pure tensor gravity of general relativity needs modification or augmentation. Recent work in scalar-tensor extensions of gravity that are consistent with present cosmological models [12–15] motivate new searches for very small deviations of relativistic gravity in the solar system at levels of 10^{-5} to 10^{-7} of the post-Newtonian effects or essentially to achieve accuracy that enables measurement of the effects of the 2nd order in the gravitational field strength ($\propto G^2$). This will require a several order-of-magnitude improvement in experimental precision from present tests. At the same time, it is well understood that the ability to measure the second order light deflection term would enable one to demonstrate even higher accuracy in measuring the first order effect, which is of the utmost importance for the gravitational theory and is the challenge for the 21st century fundamental physics.

Because of its importance to the tests of gravitational theory, especially to the tests of the Equivalence Principle and search for possible variation of the gravitational constant, we will concentrate on the improvements to these tests expected from LLR in the very near future. We will also discuss the recently proposed LATOR (Laser Astrometric Test Of Relativity) mission [16] that offers a very attractive opportunity to improve fundamental tests of gravitational theory by at least 3 orders-of-magnitude.

LLR is the only technique currently available that allows one to test for a possible Strong Equivalence Principle (SEP) violation as well as providing the best limit on the possible variation of the gravitational constant, G. In the next few months LLR is poised to take a dramatic step forward, enabled both by detector technology and access to a large-aperture astronomical telescope. The Apache Point Observatory Lunar Laser-ranging Operation (APOLLO) is a unique instrument developed specifically to improve accuracies of LLR ranges to retroreflectors on the Moon. The project will exploit a large (3.5 m), high-quality modern astronomical telescope at an excellent site to push LLR into a new regime of multiple return photons per pulse, enabling a determination of the shape of the lunar orbit to a precision of one millimeter [17,18]. As a result, APOLLO will permit improved solutions for parameters describing the Equivalence Principle, relativity theories, and other aspects of gravitation and solar system dynamics. In particular, the Equivalence Principle test would have a sensitivity approaching 10^{-14}, corresponding to a sensitivity for the SEP violation parameter η of ∼

3×10^{-5}; v^2/c^2 general relativistic effects would be tested to better than 0.1%; and measurements of the relative change in the gravitational constant, $\dot{G}/G$, would be $\sim 0.1\%$ the inverse age of the universe.

The LATOR test will be performed in the solar gravity field using optical interferometry between two micro-spacecraft [16]. Precise measurements of the angular position of the spacecraft will be made using a fiber coupled multi-chanelled optical interferometer on the International Space Station (ISS) with a 100 m baseline. The primary objective of the LATOR Mission will be to measure the gravitational deflection of light by the solar gravity to an accuracy of 0.1 picoradians, which corresponds to ∼10 picometers on a 100 m interferometric baseline. In conjunction with laser ranging between the spacecraft and the ISS, LATOR will allow measurements of the gravitational deflection by a factor of 3,000 better than is currently known. In particular, this mission will not only measure the key parameterized post-Newtonian (PPN) γ to unprecedented levels of accuracy of one part in 10^8, it will also measure for the first time the next post-Newtonian order (c^{-4}) of light deflection resulting from gravity's intrinsic non-linearity as well as measure a number of other relativistic effects.

LATOR will lead to very robust advances in the tests of fundamental physics: this mission could discover a violation or extension of general relativity, or reveal the presence of an additional long range interaction in the physical law. By testing gravity to several orders-of-magnitude higher precision, finding a violation of general relativity or discovering a new long range interaction could be one of this era's primary steps forward in fundamental physics. There are no analogs to the LATOR experiment; it is unique and a natural culmination of solar system gravity experiments.

This paper summarizes the science motivation for the precision tests of gravity and focuses on the current and near future techniques and methods that are used to conduct gravity experiments in the solar system. It specifically outlines the methods used in the LLR tests of $\dot{G}$, SEP and other PPN parameters and discusses the order-of-magnitude improvement in these tests that the next-generation of LLR technique enables. The paper also provides an overview for the LATOR experiment including a preliminary mission design.

2 Scientific Motivation

2.1 PPN Parameters and Their Current Limits

Generalizing on a phenomenological parameterization of the gravitational metric tensor field, which Eddington originally developed for a special case, a method called the parameterized post-Newtonian metric has been developed (see [6,19–21]). This method represents the gravity tensor's potentials for slowly moving bodies and weak interbody gravity, and is valid for a broad class of metric theories including general relativity as a unique case. The several parameters in the PPN metric expansion vary from theory to theory, and they are individually associated with various symmetries and invariance properties of the underlying theory. Gravity experiments can be analyzed in terms of the PPN metric, and an

ensemble of experiments will determine the unique value for these parameters, and hence the metric field itself.

The PPN expansion serves as a useful framework to test relativistic gravitation in the context of the LATOR mission. In the special case, when only two PPN parameters (γ, β) are considered, these parameters have clear physical meaning. Parameter γ represents the measure of the curvature of the space-time created by a unit rest mass; parameter β is a measure of the non-linearity of the law of superposition of the gravitational fields in the theory of gravity. GR, which corresponds to $\gamma = \beta = 1$, is thus embedded in a two-dimensional space of theories. The Brans-Dicke theory is the best known of the alternative theories of gravity. It contains, besides the metric tensor, a scalar field and an arbitrary coupling constant ω, which yields the two PPN parameter values $\gamma = (1+\omega)/(2+\omega)$, and $\beta = 1$. More general scalar tensor theories yield values of β different from one [12].

PPN formalism proves to be a versatile method to plan gravitational experiments in the solar system and to analyze the data which is obtained [3,6,20–24]. Different experiments test different combinations of these parameters (for more details, see [21]). The secular trend of Mercury's perihelion, when described in the PPN formalism, depends on another linear combination of the PPN parameters γ and β and the quadrupole coefficient $J_{2\odot}$ of the solar gravity field: $\lambda_\odot = (2 + 2\gamma - \beta)/3 + 0.296 \times J_{2\odot} \times 10^4$. The combination of parameters $\lambda_\odot = 0.9996 \pm 0.0006$, was obtained with the Mercury ranging data [25]. The PPN formalism has also provided a useful framework for testing the violation of the SEP for gravitationally bound bodies. In that formalism, the ratio of passive gravitational mass M_G to inertial mass M_I of the same body is given by $M_G/M_I = 1 + \eta U/(M_0 c^2)$, where M_0 is the rest mass of this body and U is the gravitational self-energy. The SEP violation is quantified by the parameter η, which is expressed in terms of the basic set of PPN parameters by the relation $\eta = 4\beta - \gamma - 3$. Analysis of planetary ranging data recently yielded an independent determination of parameter γ [7,8]: $|\gamma - 1| = 0.0015 \pm 0.0021$; it also gave β with accuracy at the level of $|\beta - 1| = -0.0010 \pm 0.0012$. With LLR finding that Earth and Moon fall toward the Sun at rates equal to 1.5 parts in 10^{13}, even in a conservative scenario where a composition dependence of acceleration rates masks a gravitational self energy dependence, η is constrained to be less than 0.0008 [8]; without such accidental cancelation the η constraint improves to 0.0003. The most precise value for the PPN parameter γ is at present given by Bertotti et al [11] as: $\gamma - 1 = (2.1 \pm 2.3) \times 10^{-5}$, which was obtained from a solar conjunction experiment with the Cassini spacecraft.

We shall now discuss motivations for the precision gravity tests that recently became available from both theory and experiment.

2.2 Motivations for Precision Gravity Experiments

Almost ninety years after general relativity was born, Einstein's theory has survived every test. Such a longevity, along with the absence of any adjustable parameters, does not mean that this theory is absolutely correct, but it serves

to motivate more accurate tests to determine the level of accuracy at which it is violated. A significant number of these tests were conducted over the period of the last 35 years. As an upshot of these efforts, most alternative theories have been put aside; only those theories of gravity flexible enough have survived, the accommodation being provided by free parameters and coupling constants of the theory.

Recently considerable interest has been shown in the physical processes occurring in the strong gravitational field regime. It should be noted that general relativity and some other alternative gravitational theories are in good agreement with the experimental data collected from the relativistic celestial mechanical extremes provided by the relativistic motions in the binary millisecond pulsars. However, many modern theoretical models, which include general relativity as a standard gravity theory, are faced with the problem of the unavoidable appearance of space-time singularities. It is generally suspected that the classical description, provided by general relativity, breaks down in a domain where the curvature is large, and, hence, a proper understanding of such regions requires new physics.

The continued inability to merge gravity with quantum mechanics indicate that the pure tensor gravity of general relativity needs modification or augmentation. The tensor-scalar theories of gravity, where the usual general relativity tensor field coexists with one or several long-range scalar fields, are believed to be the most promising extension of the theoretical foundation of modern gravitational theory. The superstring, many-dimensional Kaluza-Klein, and inflationary cosmology theories have revived interest in the so-called 'dilaton fields', i.e. neutral scalar fields whose background values determine the strength of the coupling constants in the effective four-dimensional theory. The importance of such theories is that they provide a possible route to the quantization of gravity. Although the scalar fields naturally appear in the theory, their inclusion predicts different relativistic corrections to Newtonian motions in gravitating systems. These deviations from GR lead to a violation of the Equivalence Principle (either weak or strong or both), modification of large-scale gravitational phenomena, and generally lead to space and time variation of physical 'constants'. As a result, this progress provides new strong motivation for high precision relativistic gravity tests.

The recent theoretical findings suggest that the present agreement between Einstein's theory and experiment might be naturally compatible with the existence of a scalar contribution to gravity. In particular, Damour and Nordtvedt [12] (see also [13,14] for non-metric versions of this mechanism) have recently found that a scalar-tensor theory of gravity may contain a 'built-in' cosmological attractor mechanism towards GR. A possible scenario for cosmological evolution of the scalar field was given in [12,15]. Their speculation assumes that the parameter $\frac{1}{2}(1-\gamma)$ was of order 1 in the early universe, at the time of inflation, and has evolved to be close to, but not exactly equal to, zero at the present time (Fig. 1 illustrates this mechanism in more detail). The expected deviation from zero may be of the order of the inverse of the redshift of the time of inflation, or somewhere between 1 part per 10^5 and 1 part per 10^7 depending on the total

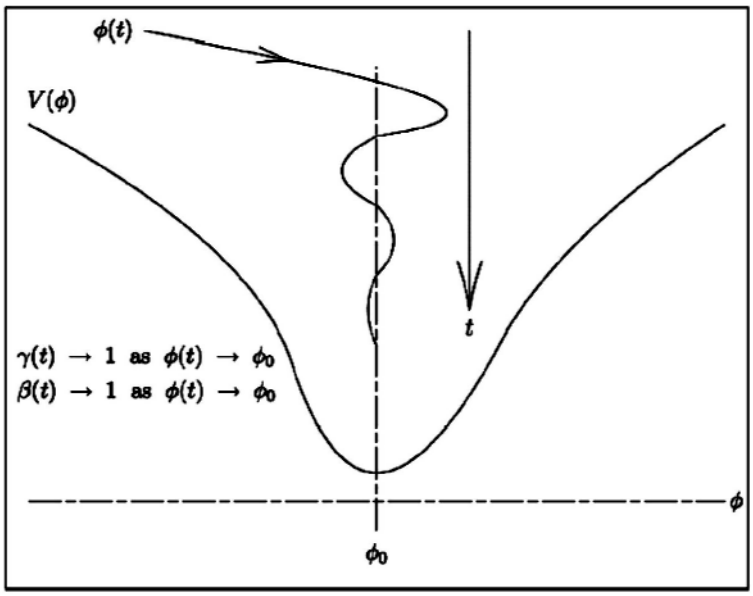


Fig. 1. Typical cosmological dynamics of a background scalar field is shown if that field's coupling function to matter, $V(\phi)$, has an attracting point ϕ_0. The strength of the scalar interaction's coupling to matter is proportional to the derivative (slope) of the coupling function, so it weakens as the attracting point is approached, and both the Eddington parameters γ and β (and all higher structure parameters as well) approach their pure tensor gravity values in this limit. However, a small residual scalar gravity should remain today because this dynamical process is not complete, and that is what experiment seeks to find.

mass density of the universe: $1 - \gamma \sim 7.3 \times 10^{-7}(H_0/\Omega_0^3)^{1/2}$, where Ω_0 is the ratio of the current density to the closure density and H_0 is the Hubble constant in units of 100 km/sec/Mpc. This recent work in scalar-tensor extensions of gravity which are consistent with, indeed often part of, present cosmological models motivates new searches for very small deviations of relativistic gravity in the solar system, at levels of 10^{-5} to 10^{-7} of the post-Newtonian effects.

The theoretical arguments above have been unexpectedly joined by a number of experimental results that motivate more precise gravitational experiments. In particular, there is now multiple evidence indicating that 70% of the critical density of the universe is in the form of a 'negative-pressure' dark energy component; there is no understanding as to its origin and nature. The fact that the expansion of the universe is currently undergoing a period of acceleration now seems inescapable: it is directly measured from the light-curves of several hundred type Ia supernovae [28,29,33], the masses of large-scale structures [27], and independently inferred from observations of CMB (Cosmic Microwave Background) by the WMAP (Wilkinson Microwave Anisotropy Probe) satellite [34] and other CMB experiments [26,35,36]. Cosmic speed-up can be accommodated within general relativity by invoking a mysterious cosmic fluid with large nega-

tive pressure, dubbed dark energy. The simplest possibility for dark energy is a cosmological constant; unfortunately, the smallest estimates for its value are 55 orders-of-magnitude too large (for reviews see [36] and references therein).

Most of the theoretical studies operate in the shadow of the cosmological constant problem, the most embarrassing hierarchy problem in physics. This fact has motivated a host of other possibilities, most of which assume $\Lambda = 0$, with the dynamical dark energy being associated with a new scalar field. The implication of these observations for cosmological models is that a classically evolving scalar field currently dominates the energy density of the universe. Such models have been shown to share the advantages of Λ: compatibility with the spatial flatness predicted inflation; a universe older than the standard Einstein-de Sitter model; and, combined with cold dark matter, predictions for large-scale structure formation in good agreement with data from galaxy surveys. Combined with the fact that scalar field models imprint distinctive signature on CMB anisotropy, they remain currently viable and should be testable in the near future. On the other hand, none of these suggestions is very compelling and most have serious drawbacks. Given the challenge of this problem, a number of authors considered the possibility that cosmic acceleration is not due to some kind of stuff, but rather arises from new gravitational physics (see discussion in [37]). In particular, extensions to general relativity in a low curvature regime were shown to predict an experimentally consistent universe evolution without the need for dark energy. These dynamical models are expected to produce measurable contribution to the parameter γ in experiments conducted in the solar system also at the level of $1-\gamma \sim 10^{-7}-10^{-9}$, thus further motivating the relativistic gravity research. Therefore, the PPN parameter γ may be the only key parameter that holds the answer to most of the questions discussed.

This completely unexpected discovery demonstrates the importance of testing the important ideas about the nature of gravity. We are presently in the 'discovery' phase of this new physics, and while there are many theoretical conjectures as to the origin of a non-zero Λ, it is essential that we exploit every available opportunity to elucidate the physics that is at the root of the observed phenomena. There is also experimental evidence for time-variability in the fine structure constant, α, at the level of $\dot{\alpha}/(\alpha H_0) \sim 10^{-5}$ [30]. This is very similar to time variation in the gravitational constant, which at the post-Newtonian level is expressed as $\dot{G}/(GH_0) \approx \eta = 4\beta-\gamma-3$, thus providing a tantalizing motivation for further tests of the SEP parameter η. A similar conclusion resulted from the recent analysis performed in [31,32,24]. These new findings necessitate the measurements of γ and β in the range from 10^{-6} to 10^{-8} to test the corresponding gravitational scenario, thus requiring new gravitational physics missions.

In summary, there are a number of theoretical reasons to question the validity of GR. Despite the success of modern gauge field theories in describing the electromagnetic, weak, and strong interactions, it is still not understood how gravity should be described at the quantum level. In theories that attempt to include gravity, new long-range forces can arise in addition to the Newtonian inverse-square law. Even at the purely classical level, and assuming the validity of the Equivalence Principle, Einstein's theory does not provide the most general

way to generate the space-time metric. Regardless of whether the cosmological constant should be included, there are also important reasons to consider additional fields, especially scalar fields. Also, the recent accuracy improvement in tests of gravity in the solar system is not sufficient to lead to groundbreaking tests of fundamental physical laws addressed above. This is especially true if the cosmological attractor discovered in [12,15] is more robust, time variation in the fine structure constant would be confirmed in other experiments and various GR extensions would demonstrate feasibility of these methods for cosmology and relativistic gravity.

The new LLR capabilities and the proposed LATOR mission are poised to directly address the challenges discussed above; we shall now discuss these experiments in more details.

3 Lunar Laser Ranging: A Unique Laboratory in Space

3.1 LLR History and Scientific Background

LLR has a distinguished history [14] dating back to the placement of retroreflector arrays on the lunar surface by the Apollo 11 astronauts. Additional reflectors were left by the Apollo 14 and Apollo 15 astronauts, and two French-built reflector arrays were placed on the Moon by the Soviet Luna 17 and Luna 21 missions. Figure 2 shows the weighted RMS residual of laser ranges to these reflector arrays for each year. Early accuracies using the McDonald Observatory's 2.7 m telescope hovered around 25 cm. Equipment improvements decreased the ranging uncertainty to ∼15 cm later in the 1970s. In 1985 the 2.7 m ranging system was replaced with the McDonald Laser Ranging System (MLRS). In the 1980s ranges were also received from Haleakala Observatory on the island of Maui in the Hawaiian chain and the Observatoire de la Cote d'Azur (OCA) in France. Haleakala ceased operations in 1990. A sequence of technical improvements decreased the range uncertainty to the current $\sim$ 2 cm level. The 2.7 m telescope had a greater light gathering capability than the newer smaller aperture systems, but the newer systems fired more frequently and had a much improved range accuracy. The new systems cannot distinguish returning photons against the bright background near full Moon, which the 2.7 m telescope could do, though there are some modern eclipse observations at full moon.

LLR accurately measures the time of flight for a laser pulse fired from an observatory on the Earth, bounced off of a corner cube retroreflector on the Moon, and returned to the observatory. For a general review of LLR see Dickey et al. [14]. A comprehensive paper on tests of gravitational physics is Williams et al. [4]. A recent test of the Equivalence Principle is in Anderson and Williams [8] and other gravitational physics tests are in Williams et al. [38]. An overview of the LLR gravitational physics tests is given by Nordtvedt [42]. Reviews of various tests of relativity, including the contribution by LLR, are given by Will [23].

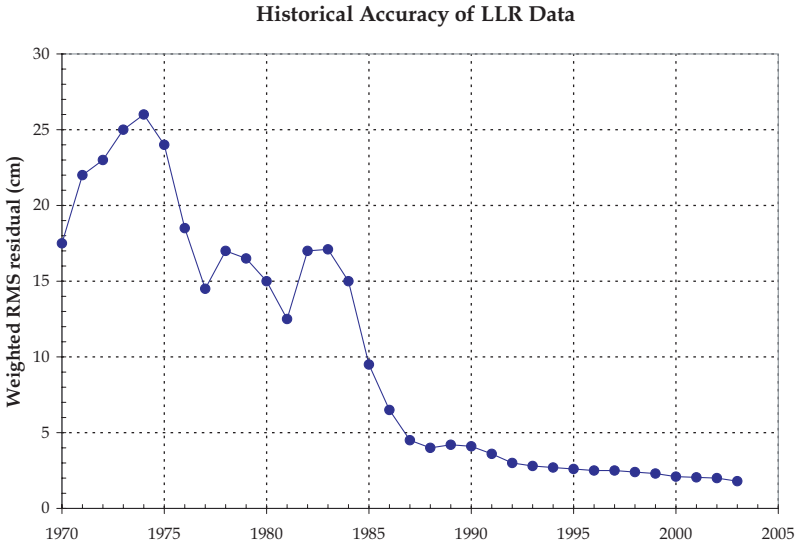


Fig. 2. Historical accuracy of LLR data from 1970 to 2003.

The LLR measurements of the past have contributed to a wide range of scientific investigations [4,8,32], and are today solely responsible for the production of the lunar ephemeris. On the fundamental scientific front, LLR provides the only means for testing the SEP—the statement that *all* forms of mass and energy contribute equivalent quantities of inertial and gravitational mass. In addition, LLR is capable of measuring the time variation of Newton's gravitational constant, G, providing the strongest limit available for the variability of this 'constant'. LLR can also precisely measure the de Sitter precession—effectively a spin-orbit coupling affecting the lunar orbit in the frame co-moving with the Earth-Moon system's motion around the Sun. Finally, current LLR results are consistent with the existence of gravitomagnetism within 0.1% of the predicted level [15,42], thus making the lunar orbit a unique laboratory for gravitational physics where each term in the relativistic equations of motion has been verified to a very high accuracy. Besides the fundamental physics capabilities of LLR, the interior, tidal response, and physical librations (rocking) of the Moon are all probed by LLR, making it a valuable tool for physical selenography [43].

The APOLLO lunar laser-ranging project will yield a one order-of-magnitude improvement in the precision of three important tests of the basic properties of the gravitational interaction. Below we shall discuss some expected results and their significance for fundamental and gravitational physics.

3.2 Equivalence Principle Tests

The Equivalence Principle (EP), the exact correspondence of gravitational and inertial masses, is a central assumption of general relativity and a unique feature of gravitation. It is the equivalence principle that leads to identical accelerations

of compositionally different objects in the same gravitational field, and also allows gravity to be viewed as a geometrical property of spacetime—leading to the general relativistic interpretation of gravitation. EP tests can therefore be viewed in two contexts: tests of the foundations of the Standard Model of Gravity (i.e. general relativity), or as searches for new physics because, as emphasized in [12–15], almost all extensions to the Standard Model of particle physics generically predict new forces that would show up as apparent violations of the EP. Easily the most precise tests of the EP are made by simply comparing the free fall accelerations, a_1 and a_2, of different test bodies, with

$$\frac{\Delta a}{a} \equiv \frac{2(a_1 - a_2)}{(a_1 + a_2)} = \left(\frac{M_G}{M_I}\right)_1 - \left(\frac{M_G}{M_I}\right)_2 , \tag{1}$$

where M_G and M_I represent gravitational and inertial masses of each body. The sensitivity of the EP test is determined by the precision of the differential acceleration measurement divided by the degree to which the test bodies differ (e.g. composition).

The Weak Equivalence Principle. The weak form the EP (the WEP) states that the gravitational properties of strong and electro-weak interactions obey the EP. In this case the relevant test-body differences are their fractional nuclear-binding differences, their neutron-to-proton ratios, their atomic charges, etc. General relativity, as well as other metric theories of gravity, predict that the WEP is exact. However, extensions of the Standard Model of Particle Physics that contain new macroscopic-range quantum fields predict quantum exchange forces that will generically violate the WEP because they couple to generalized 'charges' rather than to mass/energy as does gravity [14]. WEP tests can be conducted with laboratory or astronomical bodies, because the relevant differences are in the test-body compositions.

The Strong Equivalence Principle. The strong form of the EP extends the principle to cover the gravitational properties of gravitational energy itself. In other words it is an assumption about the way that gravity begets gravity, i.e. about the non-linear property of gravitation. Although general relativity assumes that the SEP is exact, alternate metric theories of gravity such as those involving scalar fields, and other extensions of gravity theory, typically violate the SEP [6,15]. For the SEP case, the relevant test body differences are the fractional contributions to their masses by gravitational self-energy. Because of the extreme weakness of gravity, SEP test bodies that differ significantly must have astronomical sizes. Currently the Earth-Moon-Sun system provides the best arena for testing the SEP.

To facilitate investigation of a possible violation of the SEP, the ratio between gravitational and inertial masses, M_G/M_I is expressed in the form

$$\frac{M_G}{M_I} = 1 + \eta \frac{U}{Mc^2} , \tag{2}$$

where U is the gravitational self-energy of the body ($U < 0$), Mc^2 is its total mass-energy, and η is a dimensionless constant. U/Mc^2 is proportional to M, so testing the SEP requires bodies the size of the Moon and planets. For the Earth-Moon system,

$$\frac{U_e}{M_e c^2} - \frac{U_m}{M_m c^2} = -4.45 \times 10^{-10} \, , \tag{3}$$

where the subscripts e and m denote the Earth and Moon, respectively. Therefore, a violation of the SEP would produce an Earth-Moon differential acceleration of $\Delta a/a = -4.45 \times 10^{-10}\eta$.

In general, η is a linear function of seven of the ten Parameterized Post-Newtonian parameters, but considering only β and γ,

$$\eta = 4\beta - \gamma - 3 \, , \tag{4}$$

In general relativity $\eta = 0$. A unit value for η would produce a displacement of the lunar orbit about the Earth [45,46], causing a 13 meter monthly range modulation.

3.3 LLR Tests of the Equivalence Principle

In essence, LLR tests of the EP compare the free-fall accelerations of the Earth and Moon toward the Sun. Lunar laser-ranging measures the time-of-flight of a laser pulse fired from an observatory on the Earth, bounced off of a retroreflector on the Moon, and returned to the observatory [5,24]. If the Equivalence Principle is violated, the lunar orbit will be displaced along the Earth-Sun line, producing a range signature having a 29.53 day synodic period (different from the lunar orbit period of 27 days). Since the first LLR tests of the EP were published in 1976 [2,43,47], the precision of the test has increased by two orders-of-magnitude [4,8,43,26]. (Reviews of contributions to gravitational physics by LLR are given by Nordtvedt [42] and Will [23].)

From the viewpoint of the EP, the Earth and Moon 'test bodies' differ in two significant ways: in composition (the Earth has a massive Fe/Ni core while the Moon has a much smaller core) and in their gravitational self-energies (the Earth is much more massive than the Moon). Therefore, LLR tests the total Equivalence Principle—composition plus self-energy—for the Earth and Moon in the gravitational field of the Sun. Two recent results yield $\Delta a/a$ values of $(-1 \pm 2) \times 10^{-13}$ [38] and $(-0.7 \pm 1.5) \times 10^{-13}$ [42]. The latter corresponds to a 2 ± 4 mm amplitude in range.

The LLR result is a null test so it can be argued that it is unlikely that there would be two compensating violations of the Equivalence Principle—composition and self-energy—that essentially cancel. However, because of the fundamental importance of a good SEP test, laboratory tests of the WEP are used to separate with certainty any composition-dependent and self-energy effects. Recent WEP tests performed at the University of Washington (UW) using laboratory test bodies whose compositions are close to those of the actual Earth and Moon set

upper limits on any composition-dependent Earth-Moon differential acceleration [31,32]. The random and systematic $\Delta a/a$ uncertainties of [31] are 1.4 ± 10^{-13} and 0.2×10^{-13}, respectively. Anderson and Williams [8] used the earlier of these WEP results [32] to limit the SEP parameter $\eta = 0.0002 \pm 0.0008$. If one adopts the more recent WEP test by the UW Eöt-Wash group [31], one gets an η uncertainty of 0.0005. Note that the current intrinsic LLR accuracy, if the WEP were known perfectly, is 0.0003. Therefore, with its 1 mm range accuracy, APOLLO has the capability of determining η to a precision of approximately 3×10^{-5}.

3.4 LLR Tests of Other Gravitational Physics Parameters

In addition to the SEP constraint based on (4), the PPN parameters γ and β affect the orbits of relativistic point masses, and γ also influences time delay [4]. LLR tests this β and γ dependence, as well as geodetic precession, and $\dot{G}/G$. The possibility of a time variation of the constant of gravitation, G, was first considered by Dirac in 1938 on the basis of his large number hypothesis, and later developed by Brans and Dicke in their theory of gravitation (for more details consult [21]). Variation could be related to the expansion of the Universe, in which case $\dot{G}/G = \sigma H_0$, where H_0 is the Hubble constant, and σ is a dimensionless parameter whose value depends on both the gravitational constant and the cosmological model considered. Revival of interest in the Brans-Dicke-like theories, with a variable G, was partially motivated by the appearance of superstring theories where G is considered to be a dynamical quantity [39]). A scale-dependent gravitational constant could mimic the presence of dark matter [40] and could enter discrepancies between the determinations of H_0 at different scales [41]. Williams et al. [38] give uncertainties of 0.004 for β and γ deduced from sensitivity apart from the SEP, and 1.1×10^{-12} yr^{-1} for $\dot{G}/G$ test.

The SEP relates to the non-linearity of gravity (how gravity affects itself), with the PPN parameter β representing the degree of non-linearity. Thus LLR provides the best way to measure β, as suggested by the strong dependence of η on β in (4). The parameter γ has been measured independently via time-delay and gravitational ray-bending techniques. The published Viking [1] and Very Long-Baseline Interferometry (VLBI) [3] uncertainties for γ are 0.002, 0.002, and 0.0022, respectively. Combining the above limits on η from LLR and laboratory WEP tests with the Viking and VLBI results for γ gives $|\beta - 1| < 0.0005$, the limit given by [8]. The uncertainty in β determined in this way is dominated by the uncertainty in γ. Fortunately, a much more accurate result for γ was recently reported by the Cassini experiment [11]; this leads to a significant improvement in the parameter β determination.

In our recent LLR analysis with data to May 2003, the Equivalence Principle was tested at the level of $M_G/M_I = (0.5 \pm 1.4) \times 10^{-13}$, including correction for solar radiation pressure. This result corresponds to the SEP test at the level of $\Delta a/a = (-1.5 \pm 2.0) \times 10^{-13}$ (with a WEP result from [31]) and $\eta = (3.4 \pm 4.5) \times 10^{-4}$ for the SEP violation parameter. Using the Cassini result for γ from [11], the PPN parameter β was measured at the level of $\beta = 1 + (0.9 \pm 1.1) \times 10^{-4}$. The

geodetic precession was tested at the level of $K_{gp} = -0.0035 \pm 0.0066$ and the search for variation in gravitational constant resulted in $\dot{G}/G = (0.46 \pm 1.0) \times 10^{-12}$ yr^{-1}.

Orbital precession depends on β and γ, so their sensitivity depends on the time span of the data. The uncertainty for $\dot{G}/G$ is improving rapidly because its sensitivity depends on the square of the time span. So 1 mm quality data would improve the G rate uncertainty by an order-of-magnitude in ~ 5 yr while γ and geodetic precession would depend on orbital precession time scales: 6.0 yr for argument of perigee, 8.85 yr for longitude of perigee, and 18.6 yr for node.

LLR also has the potential to determine the solar J_2 [38], PPN α_1 [46,48], hunt for influences of dark matter [45,49], and to test the inverse square law at the scale of $\sim$ 20,000 km. A long-range Yukawa interaction has been tested by Müller et al. [50].

3.5 APOLLO Contribution to the Tests of Gravity

The Apache Point Observatory Lunar Laser-ranging Operation is a new LLR effort designed to achieve millimeter range precision and corresponding order-of-magnitude gains in measurements of fundamental physics parameters. The APOLLO project design and leadership responsibilities are shared between the University of California at San Diego and the University of Washington. In addition to the modeling aspects related to this new LLR facility, a brief description of APOLLO and associated expectations is provided here for reference. A more complete description can be found in [17,18].

The overwhelming advantage APOLLO has over current LLR operations is a 3.5 m astronomical quality telescope at a good site. The site in the Sacramento Mountains of southern New Mexico offers high altitude (2780 m) and very good atmospheric 'seeing' and image quality, with a median image resolution of 1.1 arcseconds. Both the image sharpness and large aperture enable the APOLLO instrument to deliver more photons onto the lunar retroreflector and receive more of the photons returning from the reflectors, respectively. Compared to current operations that receive, on average, fewer than 0.01 photons per pulse, APOLLO should be well into the multi-photon regime, with perhaps 5-10 return photons per pulse. With this signal rate, APOLLO will be efficient at finding and tracking the lunar return, yielding hundreds of times more photons in an observation than current operations deliver. In addition to the significant reduction in statistical error ($\sim \sqrt{N}$ reduction), the high signal rate will allow assessment and elimination of systematic errors in a way not currently possible.

The new LLR capabilities introduced by APOLLO offer a unique opportunity to improve the accuracy of a number of fundamental physics tests. Some of them would have a profound effect on our understanding of the evolution of our universe. If G changes at a rate comparable to the reported change in the fine structure constant ($\dot{\alpha}/\alpha \sim 10^{-15}$ yr^{-1}) [27], η would be approximately 10^{-5}. Thus, an order-of-magnitude LLR range improvement would give an uncertainty within reach of the predictions by Damour and Nordtvedt ($\sim 10^{-7} < \eta < 10^{-4}$ [12]), and comparable to the value implied by $\dot{\alpha}$ ($\dot{\alpha}/\alpha \sim 10^{-15}$ yr^{-1} [30]).

The APOLLO project will push LLR into the regime of millimetric range precision which translates to an order-of-magnitude improvement in the determination of fundamental physics parameters. For the Earth and Moon orbiting the Sun, the scale of relativistic effects is set by the ratio $(GM/rc^2) \sim v^2/c^2 \sim 10^{-8}$. Relativistic effects are small compared to Newtonian effects. The Apache Point 1 mm range accuracy corresponds to 3×10^{-12} of the Earth-Moon distance. The resulting LLR tests of gravitational physics would improve by an order-of-magnitude: the Equivalence Principle would give uncertainties approaching 10^{-14}, tests of general relativity effects would be $< 0.1\%$, and estimates of the relative change in the gravitational constant would be 0.1% of the inverse age of the universe. This last number is impressive considering that the expansion rate of the universe is approximately one part in 10^{10} per year.

4 New Test of Relativity: The LATOR Mission

The technology has advanced to the point that one can consider carrying out direct tests in a weak field to second order in the field strength parameter $\propto GM/rc^2$. Although any measured anomalies in first or second order metric gravity potentials will not determine strong field gravity, they would signal that modifications in the strong field domain exist. The converse is perhaps more interesting: if to high precision no anomalies are found in the lowest order metric potentials, and this is reinforced by finding no anomalies at the next order, then it follows that any anomalies in the strong gravity environment are correspondingly quenched. This topic will be the main science goal of the LATOR mission.

4.1 Overview of LATOR

The LATOR experiment would use laser interferometry between two micro-spacecraft (placed in heliocentric orbits, at distances $\sim$ 1 AU from the Sun), whose lines of sight pass close by the Sun, to accurately measure deflection of light in the solar gravity. Another component of the experimental design is a long-baseline ($\sim$ 100 m) multi-channel stellar optical interferometer placed on the International Space Station (ISS). Figure 3 shows the general concept for the LATOR missions including the mission-related geometry, experiment details and required accuracies.

The LATOR mission consists of two low cost micro-spacecraft (the goal is to launch both spacecraft on a single Delta II launch vehicle). with three interferometric links between the craft and a beacon station on the ISS. One of the longest arms of the triangle ($\sim$ 2 AU) passes near the Sun. The two spacecraft are in heliocentric orbits and use lasers to measure the distance between themselves and a beacon station on the ISS. The laser light passes close to the Sun, which causes the light path to be both bent and lengthened. One spacecraft is at the limb of the Sun, the other one is $\sim 1^\circ$ away, as seen from the ISS. Each spacecraft uses laser ranging to measure the distance changes to the other

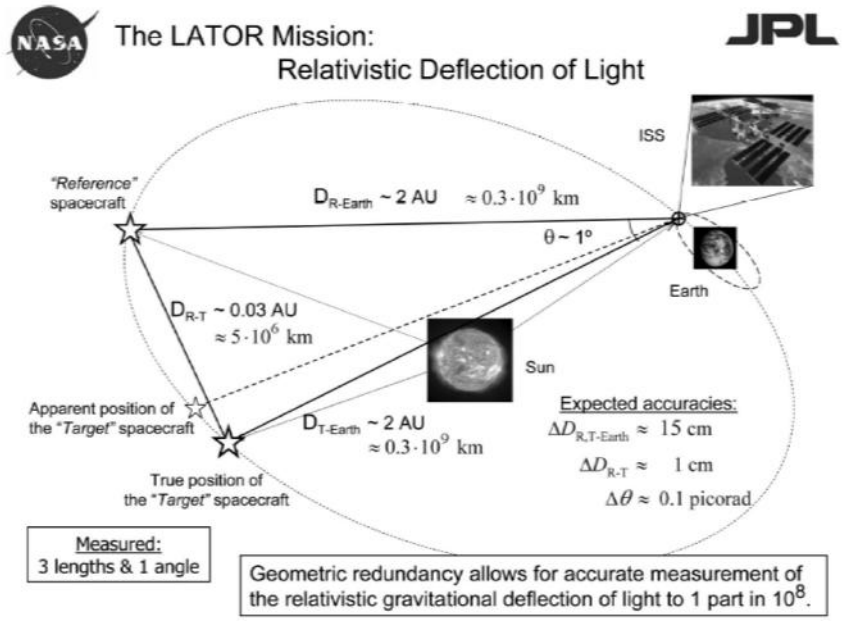


Fig. 3. Geometry of the LATOR experiment to measure deviations from the Euclidean geometry in the solar gravity field.

spacecraft. The spatial interferometer is for measuring the angles between the two spacecraft and for orbit determination purposes.

As evident from Fig. 3, the key element of the LATOR experiment is a redundant geometry optical truss to measure the departure from Euclidean geometry caused by gravity. The triangle in Fig. 3 has three independent arms the lengths of which are monitored with laser metrology. From three measurements one can calculate the Euclidean value for any angle in this triangle. In Euclidean geometry these measurements of the three lengths of the triangle should agree with the angle measured by the interferometer to high accuracy. This geometric redundancy enables LATOR to measure the departure from Euclidean geometry caused by the solar gravity field to a very high accuracy. The difference in the measured angle and its Euclidean value is the non-Euclidean signal. To avoid having to make absolute measurements, the spacecraft are placed in an orbit where their impact parameters, the distance between the beam and the center of the Sun, vary significantly from $10R_\odot$ to $1R_\odot$ over a period of ~ 20 days.

The shortening of the interferometric baseline is achieved solely by going into space to avoid the atmospheric turbulence and Earth's seismic vibrations. On the space station, all vibrations can be made common mode for both ends of the interferometer by coupling them by an external laser truss. This relaxes the constraint on the separation between the spacecraft, allowing it to be as large as a few degrees as seen from the ISS. Additionally, the orbital motion of the ISS pro-

vides variability in the interferometer's baseline projection as needed to resolve the fringe ambiguity of the stable laser light detection by an interferometer.

4.2 The Expected Results from LATOR

The first order effect of light deflection in the solar gravity caused by the solar mass monopole is 1.75 arcseconds (see Table 1 for more details), which corresponds to a delay of ∼0.85 mm on a 100 m baseline. We currently are able to measure with laser interferometry distances with an accuracy (not just precision but accuracy) of < 1 picometer. In principle, the 0.85 mm gravitational delay can be measured with 10^{-9} accuracy versus 10^{-4} available with current techniques. However, we use a conservative estimate for the delay of 10 pm which would produce the measurement of γ to accuracy of 1 part in 10^{-8} (i.e improving the accuracy in determining this parameter by a factor of 30,000) rather than 1 part in 10^{-9}. Note that the Eddington parameter γ, whose value in general relativity is unity, is perhaps the most fundamental PPN parameter, in that $(1-\gamma)$ is a measure, for example, of the fractional strength of the scalar gravity interaction in scalar-tensor theories of gravity. Within perturbation theory for such theories, all other PPN parameters to all relativistic orders collapse to their general relativistic values in proportion to $(1-\gamma)$. Therefore, measurement of the first order light deflection effect at the level of accuracy comparable with the second-order contribution would provide the crucial information separating alternative scalar-tensor theories of gravity from general relativity [12,15].

Where the light deflection by solar gravity is concerned, the magnitude of the first order effect as predicted by GR for the light ray just grazing the limb of the Sun is ~ 1.75 arcsecond (consult Table 1). The effect varies inversely with the impact parameter. The second order term is almost six orders-of-magnitude smaller resulting in ~ 3.5 microarcseconds (μas) light deflection effect, and it falls off inversely as the square of the light ray's impact parameter [51–54]. The relativistic frame-dragging term[1] is ± 0.7 μas, and the contribution of the solar quadrupole moment, J_2, is sized as 0.2 μas (using the value of the solar quadrupole moment $J_2 \simeq 10^{-7}$). The small magnitudes of the effects emphasize the fact that, among the four forces of nature, gravitation is the weakest interaction; it acts at very long distances and controls the large-scale structure of the universe, thus making the precision tests of gravity a very challenging task.

The second order light deflection is approximately 1700 pm and with 10 pm accuracy it could be measured with an accuracy of $\sim 1 \times 10^{-3}$, including first ever measurement of the PPN parameter δ. The frame dragging effect would be measured with $\sim 1 \times 10^{-2}$ accuracy and the solar quadrupole moment (using the theoretical value of the solar quadrupole moment $J_2 \simeq 10^{-7}$) can be modestly measured to 1 part in 20, all with respectable signal to noise ratios.

The laser interferometers use ∼2 W lasers and ∼20 cm optics for transmitting the light between spacecraft. Solid state lasers with single frequency operation

[1] Gravitomagnetic frame dragging is the effect in which both the orientation and trajectory of objects in orbit around a body are altered by the gravity of the body's rotation. It was studied by Lense and Thirring in 1918.

Table 1. Comparable sizes of various light deflection effects in the solar gravity field.

Effect	Analytical Form	Value (μas)	Value (pm)
First Order	$2(1+\gamma)\frac{GM}{c^2R}$	1.75×10^6	8.487×10^8
Second Order	$[(2(1+\gamma)-\beta+\frac{3}{4}\delta)\pi - 2(1+\gamma)^2](\frac{GM}{c^2R})^2$	3.5	1702
Frame-Dragging	$\pm 2(1+\gamma)\frac{GJ}{c^3R^2}$	± 0.7	± 339
Solar Quadrupole	$2(1+\gamma)J_2\frac{GM}{c^2R^3}$	0.2	97

are readily available and are relatively inexpensive. For SNR purposes we assume the lasers are ideal monochromatic sources. For simplicity we assume the lengths being measured are 2AU = 3×10^8 km. The beam spread is 1 μm/20 cm = 5 μrad (1 arcsecond). The beam at the receiver is ∼1,500 km in diameter, a 20 cm receiver will detect 1.71×10^2 photons/sec assuming 50% q.e. detectors. 5 picometer (pm) resolution for a measurement of γ to $\sim 10^{-8}$ is possible with approximately 10 seconds of integration.

As a result, the LATOR experiment will be capable of measuring the angle between the two spacecraft to ~ 0.01 μas, which allows light deflection due to gravitational effects to be measured to one part in 10^8. Measurements with this accuracy will lead to a better understanding of gravitational and relativistic physics. In particular, with LATOR, measurements of the first order gravitational deflection will be improved by a factor of 30,000. LATOR will also be capable of distinguishing between first order ($\propto GM/c^2R$) and second order ($\propto (GM/c^2R)^2$) effects. All effects, including the first and second order deflections, as well as the frame dragging component of gravitational deflection and the quadrupole deflection, will be measured astrometrically.

The LATOR experiment has a number of advantages over techniques which use radio waves to measure gravitational light deflection. Advances in optical communications technology, allow low bandwidth telecommunications with the LATOR spacecraft without having to deploy high gain radio antennae needed to communicate through the solar corona. The use of monochromatic light enables the observation of the spacecraft almost at the limb of the Sun, as seen from the ISS. The use of narrowband filters, coronagraph optics and heterodyne detection will suppress background light to a level where the solar background is no longer the dominant noise source. In addition, the short wavelength allows much more efficient links with smaller apertures, thereby eliminating the need for a deployable antenna. Finally, the use of the ISS will allow conducting the test above the Earth's atmosphere—the major source of astrometric noise for any ground based interferometer. These facts justify LATOR as a space mission.

The LATOR experiment technologically is a very sound concept; all technologies that are needed for its success have been already demonstrated as a part of the JPL's Space Interferometry Mission development. The concept arose from several developments at NASA and JPL that initially enabled optical astrometry and metrology, and also led to developing expertise needed for the precision

gravity experiments. Technology that has become available in the last several years, such as low cost microspacecraft, medium power highly efficient solid state lasers for space applications, and the development of long range interferometric techniques, make the LATOR mission feasible. The LATOR experiment does not need a drag-free system, but uses a geometric redundant optical truss to achieve a very precise determination of the interplanetary distances between the two micro-spacecraft and a beacon station on the ISS. The interest of the approach is to take advantage of the existing space-qualified optical technologies leading to an outstanding performance in a reasonable mission development time. The availability of the space station makes this mission concept realizable in the very near future; the current mission concept calls for a launch as early as in 2009 with a cost of a NASA MIDEX mission.

5 Conclusions

LLR provides the most precise way to test the EP for gravity itself, the best way to test for both non-gravitational long-range fields of dark matter as well as for time variation of Newton's constant. With technology improvements and substantial access to a large-aperture, high-quality telescope, the APOLLO project will take full advantage of the lunar retro-reflectors and will exploit the opportunity provided by the unique Earth-Moon 'laboratory' for fundamental gravitational physics. The expected improvement in the accuracy of LLR tests of gravitational physics expected with the new APOLLO instrument will bring significant new insights to our understanding of the fundamental physics laws that govern the evolution of our universe. The scientific results are very significant which justifies the more than 35 years of history of LLR research and technology development.

The LATOR mission aims to carry out a test of the curvature of the solar system's gravity field with an accuracy better than 1 part in 10^8. In spite of the previous space missions exploiting radio waves for tracking the spacecraft, this mission manifests an actual breakthrough in the relativistic gravity experiments as it allows one to take full advantage of the optical techniques that have recently become available. LATOR will lead to very robust advances in the tests of fundamental physics: this mission could discover a violation or extension of general relativity, or reveal the presence of an additional long range interaction in the physical law. There are no analogs to the LATOR experiment; it is unique and is a natural culmination of solar system gravity experiments.

Acknowledgments

D.H. Boggs participated in the LLR solutions. The work described here was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

1. I.I. Shapiro et al.: JGR **82** 4329 (1977); R.D. Reasenberg et al.: ApJ Lett. **234** L219 (1979)
2. I.I. Shapiro, C.C. Counselman III, & R.W. King: Phys. Rev. Lett. **36**, 555 (1976)
3. D.S. Robertson, W.E. Carter & W.H. Dillinger: Nature **349** 768 (1991); D.E. Lebach et al.: Phys. Rev. Lett. **75** 1439 (1995)
4. J.G. Williams, X.X. Newhall, & J.O. Dickey: Phys. Rev. D **53** 6730 (1996)
5. J.O. Dickey et al: Science **265** 482 (1994)
6. K. Nordtvedt, Jr.: Phys. Rev. **170** 1168 (1968); K. Nordtvedt, Jr.: Phys. Rev. D **43** 10 (1991); K. Nordtvedt, Jr.: CQG **15** 3363 (1998)
7. J.G. Williams et al.: BAAS **33** 836 (2001); J.D. Anderson et al.: BAAS **34** 833 (2002)
8. J.D. Anderson & J.G. Williams: CQG **18** 2447-2456 (2001)
9. L. Iess, G. Giampieri, J.D. Anderson & B. Bertotti: CQG **16** 1487 (1999)
10. J.D. Anderson, E.L. Lau, G. Giampieri: arXiv:gr-qc/0308010
11. B. Bertotti, L. Iess & P. Tortora: Nature **425** 374 (2003)
12. T. Damour & K. Nordtvedt, Jr.: Phys. Rev. Lett. **70** 2217 (1993); T. Damour & K. Nordtvedt, Jr.: Phys. Rev. D **48** 3436 (1993)
13. T. Damour, G.W. Gibbons & G. Gundlach: Phys. Rev. Lett. **64** 123 (1990) T. Damour, & J.H. Taylor: Phys. Rev. D **45** 1840 (1992)
14. T. Damour, & A.M. Polyakov: GRG **26** 1171 (1994) T. Damour, & A.M. Polyakov: Nucl. Phys., B **423** 532 (1994)
15. K. Nordtvedt, Jr.: arXiv:gr-qc/0301024
16. S.G. Turyshev, M. Shao & K. Nordtvedt, Jr.: 'New Experiment to Test General Relativity: The LATOR Mission'. In: *NASA Workshop on fundamental Physics* (Oxnard, CA, June 2003, in press)
17. T.W. Murphy, Jr., et al.: 'The Apache Point Observatory Lunar Laser-Ranging Operation (APOLLO)'. In: *12th International Workshop on Laser Ranging.* (Matera, Italy, November 2000, in press, 2002)
18. T.W. Murphy, Jr., et al.: 'APOLLO: Multiplexed Lunar Laser Ranging,' In: *13th International Workshop on Laser Ranging.* (Washington, D. C., USA, 2002)
19. K. Nordtvedt, Jr.: Phys. Rev. **169** 1014 (1968); Phys. Rev., **169** 1017 (1968)
20. C.M. Will & K. Nordtvedt, Jr.: ApJ **177** 757 (1972)
21. C.M. Will: *Theory and Experiment in Gravitational Physics.* (Cambridge, 1993)
22. K. Nordtvedt, Jr.: CQG **13** A11 (1996); K. Nordtvedt Jr.: Icarus **114** 51 (1995)
23. C.M. Will: "General Relativity at 75: How Right was Einstein?" Science **250** 770 (1990); C.M. Will: *The Confrontation between General Relativity and Experiment*, (2001) [arXiv:grqc/0103036]
24. P.L. Bender et al.: Science **182** 229 (1973)
25. E.V. Pitjeva: Celest. Mech. & Dyn. Astr. **55** 313 (1993)
26. P. de Bernardis et al.: Nature **404** 955 (2000)
27. J.A. Peacock et al.: Nature **410** 169(2001)
28. S. Perlmutter et al.: ApJ **517** 565 (1999)
29. A. Riess et al.: ApJ **116** 1009 (1998)
30. M.T. Murphy et al.: MNRAS **327** 1208 (2001); T. Banks, M. Dine, M.R. Douglas: Phys. Rev. Lett. **88** 131301 (2002)
31. E.G. Adelberger: CQG **18** 2397 (2001)
32. S. Baessler: Phys. Rev. Lett. **83** 3585 (1999)
33. J.L. Tonry et al.: arXiv:astro-ph/0305008

34. C.L. Bennett et al.: arXiv:astro-ph/0302207
35. N.W. Halverson et al.: ApJ **568** 38 (2002); C.B. Netterfield et al.:ApJ **571** 604 (2002)
36. S.M. Carroll: Living Rev. Rel. **4** 1 (2001); P.J. Peebles & B. Ratra: Rev. Mod. Ph. **75** 599 (2003)
37. S.M. Carroll et al.: arXiv:astro-ph/0306438; S.M. Carroll, M. Hoffman, & M. Trodden: arXiv:astro-ph/0301273; G. Dvali & M.S. Turner: arXiv:astro-ph/0301510; K. Freese, & M. Lewis: Phys. Lett. B **540** 1 (2002)
38. J.G. Williams et al.: 'Lunar Laser Tests of Gravitational Physics'. In: *Ninth Marcel Grossmann Meeting*. (World Scientific Publ., eds. V. G. Gurzadyan, R. T. Jantzen, and R. Ruffini, 2002) p. 1797
39. W.J. Marciano: Phys. Rev. Lett. **52** 489 (1984)
40. T. Goldman et al.: Phys. Lett. B **281** 219 (1992)
41. O. Bertolami, J.M. Mouráo, J. Pérez-Mercader: Phys. Lett. B **311** 27 (1993)
42. K. Nordtvedt, Jr.: CQG **16** A101 (1999)
43. J. G. Williams & J.O. Dickey: 'Lunar Geophysics, Geodesy, and Dynamics'. In: *13th International Workshop on Laser Ranging, October 7-11, 2002, Washington, D. C.*, available at `http://cddisa.gsfc.nasa.gov/lw13/lw_proceedings.html`
44. J. Müller & K. Nordtvedt, Jr.: Phys. Rev. D **58** 200 (1998)
45. K. Nordtvedt, Jr., J. Muller & M. Soffel: A&A **293** L73 (1995)
46. T. Damour & D. Vokrouhlicky: Phys. Rev. D **53** 4177 (1996); T. Damour & D. Vokrouhlicky: Phys. Rev. **53** 6740 (1996)
47. J.G. Williams et al.: Phys. Rev. Lett. **36** 551 (1976)
48. J. Müller, K. Nordtvedt, Jr., & D. Vokrouhlicky: Phys. Rev. D **54** R5927 (1996)
49. K. Nordtvedt, Jr.: ApJ **437** 529 (1994)
50. J. Müller et al.: 'Determination of Relativistic Quantities by Analyzing Lunar Laser Ranging Data'. In: *Seventh Marcel Grossmann Meeting*. (World Scientific Publ., eds. R. T. Jantzen, G. M. Keiser, and R. Ruffini, 1996) p. 1517
51. R. Epstein & I. I. Shapiro, Phys. Rev. D **22** 2947 (1980)
52. E. Fishbach & B.S. Freeman: Phys. Rev. D **22** 2950 (1980)
53. G.W. Richter & R.A. Matzner: Phys. Rev. D **26**, 1219 (1982); G.W. Richter & R.A. Matzner: Phys. Rev. D **26**, 2549 (1982); Richter, G. W. & Matzner, R. A., Phys. Rev. D 28, 3007 (1983)
54. K. Nordtvedt, Jr.: ApJ **320** 871 (1987)
55. A. Gerber et al.: "LATOR 2003 Mission Analysis," JPL Advanced Project Design Team (Team X) Report #X-618 (2003)

Search for New Physics with Atomic Clocks

Lute Maleki and John Prestage

Quantum Sciences and Technology Group, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109 USA

Abstract. We will discuss the physical ramifications, and describe an experiment with three high-precision clocks flying to within six solar radii of the sun for a test of a possible variation of the fine structure constant α. Measurement of the drift in ratios between the frequencies generated by each clock will probe for the variation of α. Since the response of each element to a change in α has a specific signature, this measurement will provide specific and unambiguous results. The sensitivity of this experiment to a changing α exceeds the sensitivity of recent tests based on observational astronomy, as well the geophysical bounds on α variations. Thus, the experiment will provide a compelling test of the standard model and the alternative theories.

1 Introduction

Recent developments in both theoretical and observational fronts have fueled a great deal of interest in a search for a variation of the fine structure constant. On the observational side, Webb et al. [1] have found evidence for a cosmological variation of the fine structure constant through an analysis of the absorption lines in galactic halos from quasar-emitted light. Their results indicate that the fractional change in α, averaged over redshift in the range of $0.2 \leq z \leq 3.7$ is $(-0.57 \pm 0.10) \times 10^{-5}$. On the theoretical side, many of the outstanding issues confronting fundamental physics, such as the failure to include gravity in the standard model, and puzzles of cosmology, such as inflation and the apparent accelerated rate of the expansion of the universe, appear to imply the existence of massless, or nearly massless scalar fields. These fields appear as dilaton or moduli in the M-theory, supporting the unification of gravity with other forces, as well as suggesting a possible breakdown of the Equivalence Principle. They also appear as quintessence in models of cosmology aimed at resolving fine tuning and other outstanding problems, such as a nonzero cosmological constant [2]. The scalar fields in these models imply a spatio-temporal variation of constants of nature, such as fine structure and other field coupling constants.

Despite these important developments, at this writing there is no clear consensus amongst researchers regarding the validity of the theoretical predictions, and the observational conclusions are not regarded as inconvertible. The question of if, how, and why the fine structure constant varies remains an open one.

It is clear then that a controlled experiment with sufficient measurement sensitivity beyond the current capabilities will be enormously important in clarifying some of the questions associated with α variations. SpaceTime is a space mission study aimed at providing such an experiment. It is based on flying an

L. Maleki, J. Prestage, Search for New Physics with Atomic Clocks, Lect. Notes Phys. **648**, 331–341 (2004)
http://www.springerlink.com/

instrument based on three clocks that run on ground state hyperfine transitions of three different singly ionized atoms to within six solar radii of the Sun. The "tri-clock" instrument of SpaceTime is capable of testing a variation of α with four orders of magnitude more sensitivity, as compared with the results of quasar observations. As discussed below, the choice of the atomic clocks as the instrument was made to ensure that the results would be conclusive and free of many questions that have confronted previous investigations searching for a varying α.

At this point it is worthwhile to consider some of the consequences of a varying fine structure constant. The fine structure constant has been a point of fascination with physicists since it was introduced, and named, by Sommerfeld in 1916 as a useful constant in spectroscopy; it is a measure of the doublet structure of hydrogen and other atoms with a single valence electron. Sommerfeld also considered α as an indication of an intimate relation between charge and quantum. In the years following Sommerfeld's introduction of α, various physicists, starting with Eddington, have considered the relation between α and other constants of nature. This interest was also fueled by suggestive numerology that relates specific functions of π to the value of α.

The conjecture of varying fundamental constants has also a relatively long history and dates back to Dirac's "Large Number Hypothesis", which was based on the notion that there exists an underlying relationship between constants of nature, as manifested by large numbers, on the order of 10^{39}, that could be obtained by arranging them in various combinations [3]. Other *ad hoc* conjectures similarly have pointed to possible variation of constants, especially the gravitational constant G, through which a variation of α may also arise. These models, nevertheless, were all generally qualitative, and more importantly, lacked any observational support. The picture has changed in the last few years. Since a change in α implies a changing e, the charge of the electron, or c, the speed of light, or Planck's constant, h, through $\alpha = e^2/c\hbar$, several models based on variations of any of these dimensional constants have been devised [4–8]. There is, however, a good bit of controversy regarding the validity of these models, and if their predictions do or do not support [9,10] a violation of the Equivalence Principle, as well.

Atomic clocks have traditionally been used to test the prediction of general relativity. The first such test performed in 1976 by NASA's Gravity Probe A, where the rate of a hydrogen maser clock on a rocket in a sub-orbital trajectory was compared to that of a similar clock on the Earth's surface [11]. This measurement verified the exact prediction of a clock shift by general relativity to a part in 10^4, a precision that still stands unchallenged today. In a recent investigation it was shown that it is also possible to search for a variation in α by comparing the rate of drift of two clocks based on hydrogen and mercury ion [12]. This is because the energy of the hyperfine transition in atoms, which forms the basis for microwave clocks, have an αZ dependence, where Z is the atomic number. This first laboratory attempt to search for a varying α set a limit of $\sim 4 \times 10^{-14}$ per year for its temporal variation. This approach has recently been extended to the comparison of a rubidium and a cesium fountain clock, both based on microwave transitions [13], as well as the comparison of a cesium foun-

tain with an optical mercury ion clock, where an optical transition in the ion was used [14]. These more recent experiments set the limit for a varying α a to be less than about 10^{-15}/yr. This is a less stringent limit than that obtained with an analysis of neutron capture rate applied to a natural thermonuclear reaction that occurred 1.5 billion years ago in Oklo mine, Africa [15], which places the limit on α variation to be less than 5×10^{-17}/yr. SpaceTime's instrument is designed to provide sensitivity to a variation in α at the level of 10^{-20}/yr by searching for any spatial dependence of α.

For alkali atoms, an expression for the hyperfine interval may be obtained, as follows:

$$A_s = \frac{8}{3}\alpha^2 g_I Z \frac{z^2}{n^{*3}} \left(1 - \frac{d\Delta_n}{dn}\right) F(\alpha Z)(1-\delta)(1-\epsilon)\frac{m_e}{m_p} R_\infty c \,. \tag{1}$$

Here, z is the net charge of the ion without the valence electron, and n^* is the effective quantum number with $\Delta n = n - n^*$, δ and ϵ are related to the corrections for finite size of the nucleus. Thus the sensitivity of different clocks, based on atoms of different Z, to a change in the fine structure constant display specific signatures. In particular, the Casimir correction factor, $F(\alpha Z)$, (for the relativistic wave equation of the electron) leads to the differential sensitivity in the alkali microwave hyperfine clock transition frequencies f,

$$f = \alpha^4 \frac{m_e}{m_p} \frac{m_e c^2}{h} F(\alpha Z) \,. \tag{2}$$

It is clear from the above equation that different atomic systems with different Z display different frequency dependencies on a variation of α through the αZ dependent terms. A direct test for a time variation of α can then be devised through a comparison of two clocks, based on two atomic species with different atomic number, Z.

This is a key feature of the SpaceTime instrument that in conjunction with the individual sensitivity of each atomic species to an α variation, can produce clear and unambiguous results.

Since the changing α in all model predictions is mediated by coupling of a scalar field to matter, the fall in the $1/R$ potential near the Sun will allow a direct test of the general relativity, where only the tensor field is allowed, and where the constants are not allowed any variation. This is an important point to consider in clock tests, and other tests searching for an α variation based on a signature of the failure of the equivalence principle (EP). Since EP is currently tested at about the 10^{-12} level [16] with no violations found, any test searching for α variations must have a sensitivity higher than 10^{-12} to EP violation to produce a new result. The expected sensitivity of the differential red shifts as measured by the three clocks that are within six solar radii is at the level of 10^{-13} of the EP, or about six orders of magnitude larger than the GP-A experiment. Thus results of SpaceTime will improve the current state of art in EP violation by an order of magnitude, as well as improving on the results of Webb *et al.* by four orders of magnitude, beyond the capability of all existing and future earth-bound clock experiments.

To improve the measurement sensitivity, our instrument consists of three clocks based on three different atomic species that can be inter-compared for individual signatures. To reduce the influence of systematic errors that can mimic our signal, the three clocks share the same environment. To improve the source of the signal, the tri-clock instrument flies to within six solar radii of the largest body of matter in the solar system, the Sun. Thus the entire experiment is designed to provide a clean and unambiguous result, based on a technology that is proven, and has an outstanding chance for success. Finally, the spinning spacecraft, moving at 300 km/s, or 1/1000 of the speed of light, at its closest approach will test another important question with fundamental underpinning: Is Lorentz symmetry robust, or does it fail at some limit? This question is important since string theory, and theories that extend beyond the Standard Model [17], result in physics without Lorentz and other global symmetries such as CPT.

Beyond this, as mentioned above, a consequence of a changing α is that either c, the speed of light, or e, the charge of the electron, or h, Planck's constant, must change. Theories based on a changing velocity of light have received considerable attention since they solve the outstanding problems in cosmology: the horizon, flatness, cosmological constant, entropy, and homogeneity problems [5]. They nonetheless violate Lorentz invariance. SpaceTime will provide a ten-fold sensitivity for a test of Lorentz invariance, as compared to an earth bound test, due to the order of magnitude smaller orbital speed of earth [18].

2 The Instrument

In the strongly time-dilated spacetime curvature at six solar radii (4.2 Gm), time runs slower than on Earth by about one half microsecond per second. Three atomic clocks based on hyperfine transitions of Hg^+ ($Z = 80$), Cd^+ ($Z = 48$), and Yb^+ ($Z = 70$) are different in their electromagnetic composition (given by the Casimir factor) and will be simultaneously monitored during a solar flyby to determine whether these different clocks will measure the same time interval near the Sun. The atomic clock hardware for the SpaceTime mission is a modification of the linear ion trap frequency standard (LITS) currently being deployed in the Deep Space Network stations worldwide. A laboratory prototype has shown ultra-stable operation in a package far smaller than other clock technologies and represents the state of the art for atomic clocks.

Atomic clocks based on hyperfine transitions and ion traps are the most suitable technology for space applications. This is because of the inherent simplicity of this approach, which does not rely on resonant cavities. In lamp based trapped ion clocks, as in the SpaceTime instrument, the risk associated with the use of lasers in space is eliminated. Ions confined in electromagnetic traps are significantly shielded from environmental perturbations such as collisions with the walls or each other. The relatively large hyperfine splitting of singly ionized systems also reduce their sensitivity to ambient magnetic fields, as compared with atoms with smaller hyperfine frequencies.

The classical ion trap consisting of a three-electrode structure made with hyperbolic electrodes confines charged particles of particular charge to mass

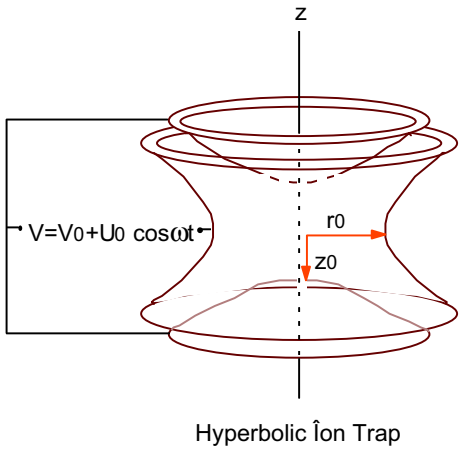


Fig. 1. Ion Trap

ratios based on the applied dc and rf potentials (see Fig. 1). In this geometry, ions are confined in a spherical region as a result of the applied ponderomotive forces.

A geometry based on linear electrodes, first introduced at JPL for clock applications, improves the clock stability by providing a geometry whereby the temperature (kinetic energy) of the ions resulting from the micro-motion in the trap is reduced [19]. This configuration was further refined at JPL [20] to put the ability to move the charged particles from one region of space to another, to separate the ion preparation region from the region where the microwave field produced by a local oscillator (LO) interacts with the clock transition of ions (Fig. 2). By separating these regions it is possible to significantly reduce the requirement of magnetic shielding which must protect the ions undergoing interaction with the microwave field. Higher pole traps are also employed in order to further reduce ion density space-charge related ion-heating. This is key to the reduction of the size and weight of the clock, parameters that are particularly important for space instruments.

The instrument for this mission is composed of three ion trap clocks in a package where much of the hardware is common to all of the clocks. Because some of the clock systematic frequency perturbations will be common to all three clocks and will have a characteristic signature that can be identified and removed from the difference of the clock frequencies, relative stabilities to 10^{-16} in the inter-comparison can be reached. The local oscillator (LO) will simultaneously interrogate each of the three clock transitions thereby removing LO noise in the inter-comparison, and greatly improving short-term clock noise so that 10^{-16} resolution in the difference in clock rates can be obtained within the 15-hour close encounter. Because ion-trap-based clocks are relatively immune to temperature and magnetic field changes, a simple, robust electronics package is sufficient for ultra-stable operation.

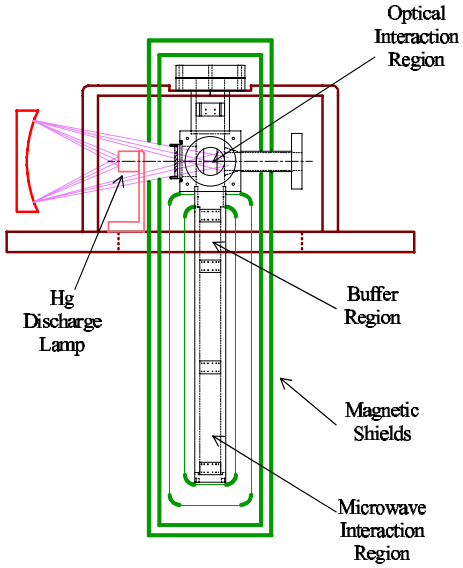


Fig. 2. Schematic of Linear Ion Trap

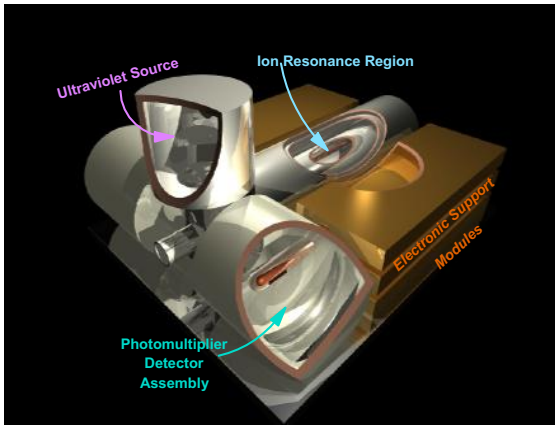


Fig. 3. Tri-clock Instrument

The basic architecture of the "tri-clock" instrument is three LITE (Linear Ion Trap Extended) units, each operating with a single element Hg^+, Cd^+ or Yb^+, and will be packaged into one housing with many shared components for mass reduction. This configuration is shown in Fig. 3. Each separate clock is based upon a linear multi-pole trap [20]. For optical state-selection, ions are trapped around the rf quadrupole electric field node along the centerline where they are prevented from escaping by dc fields applied at each end. By applying dc positive bias to all trap rods in one region along the length of the trap, ions can be excluded from that region and 'transported' into another section where

the rods are at dc ground. Ions can thus be moved from one end of the trap to the other. This allows the optical state selection and interrogation to be carried out in an unshielded region while the much more critical clock hyperfine resonance is probed in a small, well shielded region, away from magnetic optical components and openings in the shields for light entry and exit. The ion-number (space-charge) induced frequency pulling is reduced by more than a factor of 20 in the multi-pole arrangement as compared to the linear quadrupole [21,20,22,23].

The three traps will be operated with a common rf voltage source so that related trapping forces confine the three different ion species. In this way small variations in the trapping strength will affect each ion cloud in a characteristic manner that can be readily identified. Another unique feature of this clock comparison is the use of the ultra-stable local oscillator. Space-qualified quartz oscillators achieve short-term stabilities of 10^{-13} over tens of seconds averaging intervals. This will limit a conventional high performance atomic clock to about 10^{-13} at 1 second averaging time, falling from there as $\tau^{-1/2}$ where τ is the averaging interval in seconds. For the clock comparison at the near-solar flyby, the largest change in gravitational potential occurs over a 15-hour period, i.e., 54 000 s. This LO-limited performance gives 4×10^{-16} at 15 hours and falls short of the design goal. We have demonstrated atomic clock performance at $(2-3) \times 10^{-14}$ at one second but LO noise degrades the performance for a single operating atomic clock. For a comparison between two or more clocks, however, a single LO can be used to interrogate all clock transitions simultaneously, and the LO noise will be common. This common noise in individual atomic line-center measurements will not be present in the differences of these and we can recover the $(2-3) \times 10^{-14}/\sqrt{\tau}$ and reach the 10^{-16} stability level in 15 hours averaging.

The tri-clock measurement offers a suppression of other common mode frequency shifts of the three atomic transitions. The suppression of systematic frequency pulling can also be applied to variations of the solar magnetic field along the spacecraft trajectory. This approach will save mass and power in magnetic shielding. A set of four layers of magnetic shields will enclose the clock resonance tube. An additional layer will house the final package. Since the unshielded Hg^+ atom sensitivity is about 2×10^{-13}/mG (at an operating point of 50 mG), 20×10^{-13}/mG for Yb^+, and 15×10^{-13}/mG for Cd^+, a shielding factor of 10^7 is required to reduce a 1-G solar field variation during the spacecraft flyby to below one part in 10^{16} relative clock stability. A 1-G field variation might be expected during the solar flyby. This level of shielding is very difficult to achieve within the mass and power budget.

The differential response of the three clocks to a common field variation has a characteristic signature that will identify this systematic shift and will enable its removal in post analysis. The magnetic sensitivity of the three hyperfine levels is well understood in the atomic physics of the clock transitions. The change of the clock frequency as the operating field changes by δH_0 is given by $\delta y \equiv \delta f/f_0 = 2\beta H_0 \delta H_0$ where the constant β describes the field sensitivity of each of the three clock transitions. The atoms with a smaller hyperfine splitting f_0 shift more. Note that this behavior is very different from the sensitivity to a

change in α as given in [12]. In that paper it is shown that the atoms with larger atomic number Z shift more with a change in α than the low Z atoms.

The two simultaneous equations for the variation of the difference frequencies are

$$\delta y_{AB} = \Big(L(Z_A) - L(Z_B)\Big)\frac{\delta\alpha}{\alpha} + \left(1 - \frac{\beta_B f_A}{\beta_A f_B}\right)\frac{2\beta_B}{f_A}\frac{H_0\delta H}{S}\ ,$$
$$\delta y_{AC} = \Big(L(Z_A) - L(Z_C)\Big)\frac{\delta\alpha}{\alpha} + \left(1 - \frac{\beta_C f_A}{\beta_A f_C}\right)\frac{2\beta_B}{f_A}\frac{H_0\delta H}{S}\ . \qquad (3)$$

We have taken the variation of the clock transitions with operating field, H_0, to be given by $f = f_0 + \beta H_0^2$ and the shielding factor for external fields to be S, i.e., $\delta H_0 = \delta H/S$. δH is the variation of the solar magnetic field along the spacecraft trajectory. The α sensitivities, $L(Z)$, are found in Fig. 1 of [12].

For the hyperfine clock transitions in Hg, Cd, and Yb, these equations can be inverted to solve for $\delta\alpha/\alpha$ and $(2\beta_A/\nu_A)H_0\delta H/S$ along the trajectory of the near-Sun flyby. Thus, even with imperfect magnetic shielding and the accompanying clock frequency pulling, an unambiguous variation of α could be extracted.

2.1 Temperature Induced Frequency Shifts

Ambient temperature changes of the clocks can cause spurious frequency pulling δy_{AB} and δy_{AC} and must be completely removed to the 10^{-16} level. Unlike magnetic sensitivities, which can be to a large extent understood as incomplete shielding of the atomic transition, temperature-induced frequency shifts are more difficult to predict from first principles. The only definitive measurement of temperature sensitivity must be carried out with a fully assembled and operating system. The differential sensitivity coefficients to be used in separating any observed effect from a temperature induced δy_{AB} and δy_{AC}, must be generated in-situ. Once these sensitivities are measured, we can use the two return data channels to distinguish temperature effects from any observed violations.

Some temperature effects have very clear signatures, completely distinguishable from any a variation along the spacecraft trajectory. For example, ion temperature variations will lead to clock frequency changes via second-order Doppler shifts, by an amount proportional to $-kT/mc^2$ where T is the ion temperature and m is the ion mass. Any temperature change, δT, common to all three ionic species will shift the three clock frequencies by an amount inversely proportional to their mass. This will allow this systematic frequency offset to be removed as in the magnetic case above. For these shifts,

$$\delta y_{AB} = \Big(L(Z_A) - L(Z_B)\Big)\frac{\delta\alpha}{\alpha} + \left(1 - \frac{m_A}{m_B}\right)\frac{k\delta T}{m_A c^2}\ ,$$
$$\delta y_{AC} = \Big(L(Z_A) - L(Z_C)\Big)\frac{\delta\alpha}{\alpha} + \left(1 - \frac{m_A}{m_C}\right)\frac{k\delta T}{m_A c^2}\ , \qquad (4)$$

showing that these temperature variations can be separated from the variations that come from a non-zero $\delta\alpha/\alpha$ along the solar flyby trajectory. We have assumed no mass dependent heating, δT, which will almost certainly be present.

However, a pre-launch ground measurement will be carried out to catalog differential frequency shifts vs rf trap level, buffer gas pressure, etc.

2.2 Mission Design

The only economical technique to get sufficient change in velocity to fly near the Sun is to go via Jupiter. This is because the angular momentum associated with the orbiting earth must be lost so the spacecraft will fall to the Sun in a reasonable length of time. Thus, SpaceTime will launch in a direct transfer orbit to Jupiter and then a fast trajectory to the Sun. A kick stage is integrated with the spacecraft on a "spin table" that spins the entire integrated package during the launch. The spinning spacecraft does not have to be despun following injection, as with a typical three-axis stabilized spacecraft. This eliminates the mass and reliability penalties of a despin hardware.

Figure 4 illustrates the entire interplanetary trajectory to the Sun including the first leg after injection. The time tics are 50-day intervals. Approaching Jupiter, a precision orbit determination is completed using only radio tracking data, and a precise final aiming maneuver is completed. The gravity assist flyby is used to: 1) reduce (almost canceling) the trajectory angular momentum, allowing the spacecraft to fall into a 6-R_S perihelion, 2) rotate the plane of the heliocentric orbit to a final inclination of 90.0 degrees and 3) establish the time of perihelion to produce a quadrature trajectory geometry (Sun-spacecraft-Earth angle = 90.0 degrees) at perihelion. This latter condition is fundamental to the spacecraft architecture, which always has the shield pointed at the Sun and the high gain antenna (HGA) pointed at Earth.

Following the Jupiter flyby, the spacecraft is on its final trajectory toward the perihelion. The perihelion flyby trajectory is shown in Fig. 4 from $P-24$ to $P+24$ hr. This is the prime data acquisition period for the mission. The view in Fig. 4 is from Earth illustrating the effects of the quadrature trajectory geometry by the schematic drawings of the spacecraft. The spacecraft is a spinning drum with the direction of its spin axis toward the Earth (out of the page). The thermal shield for the spacecraft, as the spacecraft spins, maintains its orientation toward the Sun at all times protecting the sensitive elements from the extreme thermal

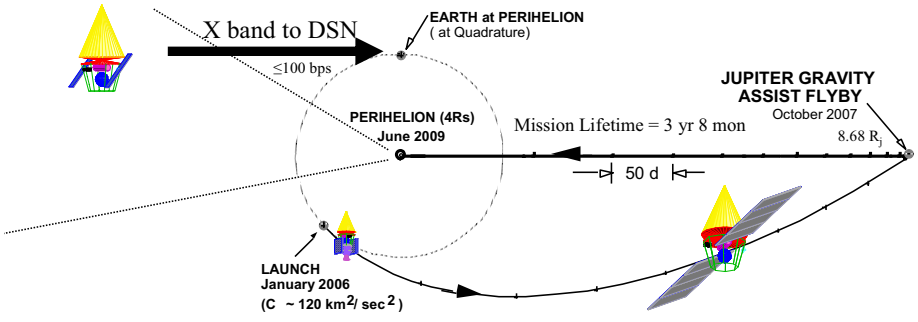


Fig. 4. Spacecraft trajectory

environment. This is a passive attitude control technique that simplifies the control of the spacecraft and allows a very robust design in this otherwise hostile environment.

It is interesting to point out that the most challenging aspect of the mission, affecting orbital trajectory and the number of passes (single) by the Sun is the power requirements. Because of the extreme heat encountered near the Sun, solar panels, even those designed for high temperature, cannot be used. Instead, a bank of batteries must provide the needed power to the spacecraft systems, and the instrument. The mass associated with the batteries ultimately limits the choices of a trajectory with a given launch vehicle, as well as the size of the spacecraft and associated systems. This ironic limitation (shortage of power while so near the Sun) is the major design issue that affects virtually all aspects of the mission.

3 Conclusion

We have briefly discussed a mission design study based on the inter-comparison of the oscillation frequencies of three atomic clocks based on three different species of singly ionized atoms. By flying this instrument to within six solar radii of the Sun it is possible to search for a variation of fine structure constant to a level that is not accessible to earth-based instruments. At this point two other points regarding this approach are worth noting. First, one may ask the question why the choice of atomic clocks, as opposed to other instruments (see a description of LATOR mission in this volume [24]). As briefly mentioned above, and discussed elsewhere in this volume (see, for example, the paper by Flambaum *et al.*) the detail of theories that predict a temporal or spatial variation in fine structure constant, such as M-theory or theories based on varying c or e, are rather tentative. Experimental tests of these theories based on a search for varying α then must produce direct and unambiguous results to be most valuable. The three-clock comparison discussed here is indeed such an approach. As discussed above, each atomic clock will drift in a specific manner with varying α and inter-comparison of these variations assures that an observed signal produces a clear result. Secondly, the technology of atomic clocks is well developed, and a space test based on clocks has an inherently large probability of success.

Acknowledgements

This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. The SpaceTime science team in this study consisted of Eric Adelberger (University of Washington), John Armstrong (JPL), Thibault Damour (Institut des Hautes Etudes Scientifiques), Kenneth Johnston (US Naval Observatory), Alan Kostelecky (Indiana University), Claus Lämmerzahl (Heinrich-Heine-Universität Duesseldorf), Lute Maleki (PI, JPL), Kenneth Nordtvedt (Montana State University) and John Prestage (JPL).

References

1. J.K. Webb et al., Phys. Rev. Lett. **87**, 091301 (2001).
2. T. Damour, A. M. Polyakov, Nucl. Phys. B. **423**, 532 (1994).
3. P.A.M. Dirac, Nature **139**, 323 (1937).
4. J.D. Bekenstein, Phys. Rev. D **25**, 1527 (1982).
5. J.D. Barrow, J. Magueijo, astro-ph/9811072.
6. H. Sandvik, J.D. Barrow, J. Magueijo, astro-ph/0107512.
7. J.W. Moffat, astro-ph/01093501.
8. D. Youm, hep-th/0108237
9. J.D. Bekenstein, Phys. Rev. D **66**, 123514 (2002).
10. T. Damour, Astro. Phys. Space Sci. **283**, 445 (2003).
11. R. Vessot et al., Phys. Rev. Lett. **45**, 2081 (1980).
12. J.D. Prestage, R.L. Tjoelker, L. Maleki, Phys. Rev. Lett. **74**, 3511 (1995).
13. H. Marion et al. Phys. Rev. Lett. **90**, 150801 (2003)
14. S. Bize et al. Phys. Rev. Lett. **90**, 150802 (2003)
15. T. Damour, F. Dyson, Nucl. Phys. B. **480**, 37 (1996).
16. E.G. Adelberger, Class. Quant. Grav. **18**, 2397 (2001).
17. V.A. Kostelecky, *CPT and Lorentz Symmetry*. World Scientific, Singapore, 1999.
18. R. Bluhm, V.A. Kostelecky, C. Lane, N. Russell, Phys. Rev. Lett. **88**, 090801 (2002)
19. J.D. Prestage, G. J. Dick and L. Maleki, J. Appl. Phys. **66**, 1013 (1989).
20. J.D. Prestage, R.L. Tjoelker and L. Maleki, Topics Appl. Phys. **79**, 195 (2001), Frequency Measurement and Control, ed. A. N. Luiten (Springer, Heidelberg, 2001).
21. J.D. Prestage, R.L. Tjoelker, L. Maleki, Proc. 2000 IEEE Freq. Control Symp., p. 459 (2000).
22. J.D. Prestage, S. Chung, E. Burt, L. Maleki, and R.L. Tjoelker, Proc. 2002 IEEE Freq. Control Symp., p. 706 (2000).
23. R.L. Tjoelker, J.D. Prestage, P. K. Koppang, and T. B. Swanson, Proc. 2003 IEEE Freq. Control Symp., p. 1066 (2003).
24. S.G. Turyshev et al., 35 Years of Testing Relativistic Gravity: Where Do We Go from Here?, Lect. Notes Phys. **648**, 311–330 (2004)

Index