Time and Matter

LaTeX

Proceedings of the 2nd International Conference on

TIME AND MATTER

26 – 31 August 2007
Bled, Slovenia

Edited by

Martin O'Loughlin
Samo Stanič
Darko Veberič

University of Nova Gorica Press

Proceedings of the 2nd International Conference on TIME AND MATTER
26 – 31 August 2007, Bled, Slovenia

Edited by: prof.dr. Martin O'Loughlin, prof.dr. Samo Stanič and doc.dr. Darko Veberič

Cover design: Eva Kosel

Typeset by: doc.dr. Darko Veberič

Published by: University of Nova Gorica Press, P.O. Box 301, Vipavska 13, SI-5001 Nova Gorica, Slovenia

Publication year: 2008

Printed by: Tiskarna Pleško d.o.o., Slovenia, 150 copies

ISBN 978-961-6311-48-9

CIP - KatAložni zapis o publikaciji
Narodna in univerzitetna knjižnica, Ljubljana

530.1(063)(082)

INTERNATIONAL Conference on Time and Matter (2 ; 2007 ; Bled)
Proceedings of the 2nd International Conference on Time and Matter, 26 – 31 August 2007, Bled, Slovenia / edited by Martin O'Loughlin, Samo Stanič, Darko Veberič. – Nova Gorica : University, 2008

ISBN 978-961-6311-48-9
1. O'Loughlin, Martin John
237509376

International Advisory Committee:

Danilo Zavrtanik (University of Nova Gorica), Chair
Ikaros Bigi (University of Notre Dame)
Martin Faessler (LMU Munich)
Don Howard (University of Notre Dame)
Cecilia Jarlskog (Lund University)
Hermann Nicolai (MPI für Gravitationsphysik)
Leonard Susskind (Stanford University)
Gabriele Veneziano (Collège de France)
Julius Wess (LMU Munich)
Bruce Winstein (University of Chicago)
Hans Dieter Zeh (University of Heidelberg)

Local Organizing Committee:

Samo Stanič (University of Nova Gorica), Chair
Andrej Filipčič (Jožef Stefan Institute)
Biagio Forte (University of Nova Gorica)
Matej Horvat (Jožef Stefan Institute)
Martin O'Loughlin (University of Nova Gorica)
Darko Veberič (University of Nova Gorica)
Serguei Vorobiov (University of Nova Gorica)
Danilo Zavrtanik (University of Nova Gorica)
Marko Zavrtanik (Jožef Stefan Institute)

Organiser:

Laboratory for astroparticle physics, University of Nova Gorica
Vipavska 13, POB 301, SI-5001 Nova Gorica, Slovenia
Email: `tam@p-ng.si` Phone: +386 5 3315 237 Fax: +386 5 3315 385

Previous conferences:

Venice, Italy, 11 – 17 August 2002
Proceedings: World Scientific Publishing Company (April 30, 2006)
ISBN-10: 9812566341

The conference was sponsored by

Mobitel d.d.
Ljubljana

Zavarovalnica Triglav d.d.
Ljubljana

Institut Français Charles Nodier
Ljubljana

Zveza bank
Celovec

and supported by

Slovenian Research Agency (ARRS)

Preface

The primary aim of the second Time and Matter conference, held at Lake Bled, Slovenia in August 2007 was to provide a meeting place for ideas from various fields of natural sciences involving the most fundamental concepts in nature, namely those of time and matter. The discussions focused on different phenomena related to time and matter on various scales of magnitude, both micro and macroscopic, with the prospect of obtaining a better grasp of the "big picture" and of the interconnections between these specific phenomena.

Time is considered to be one of the most fundamental concepts in physics. The definition of the unit of time itself – the oscillation rate of atomic clocks – exceeds those of all other basic units by orders of magnitude in precision, with the precision continually increasing. Furthermore, all other quantities in physics and their units can be related to time and its unit using only physical constants. In science and in philosophy time together with space has traditionally been regarded as an *a priori* condition of perception, an independent frame or coordinate system where events take place. The two major legacies of 20th century physics – the theory of relativity and quantum mechanics – have, however, taught us to view space and time as being intertwined with the phenomena they are hosting. The concept of an absolute space and time was abandoned and an increasing doubt has been cast on the notion that four space-time dimensions really suffice to describe the physical universe. It was realized that space-time does not merely represent a stage independent of the drama that is being played by the material objects. Matter actually creates and deforms the space and time in which it moves.

The expansion of the Universe has been discovered and mapped out with increasing accuracy showing that there was a beginning of space and time, popularly named the Big Bang. The ultimate fate of the Universe – perpetual expansion or reversal leading to a Big Crunch – hangs precariously in the balance and depends on the amount of matter contained in the Universe. There is an increasing amount of evidence for the abundant presence of “dark matter” and for a cosmological constant possibly arising from the presence of “dark energy”. Recent data also suggests that the cosmic microwave background radiation exhibits tiny inhomogeneities, which in a theory of an inflating Universe could lead to the grand structures that we see through a telescope today.

An essential ingredient in the study of time and matter is the understanding of the role of discrete symmetries in nature. Violations of discrete symmetries, the reversals of time, charge and parity and all their combinations have indeed been observed in particle physics and major resources are being invested into further experimental studies of these fundamental phenomena. We are witnessing a fascinating development; branches of the physical sciences that had apparently evolved in completely different directions over the centuries suddenly find themselves looking at problems that are intimately connected. One example is provided by elementary particle physics, which explores the tiniest domains in space and time, and cosmology, which deals with the largest of such entities. The relation between violation of microscopic time reversal symmetry and the baryon number of the Universe, between CP violation and dark matter of the Universe are wondrous examples of such developments.

On the human scale, the violation of macroscopic time reversal invariance has been known for a long time and is the essence of the second law of thermodynamics. The new approaches to the concept of time as a complex entity now lead us to raise questions and address problems that before would have been considered outside the realm of scientific jurisdiction. Is the arrow of time an intrinsic property of the Universe or does it depend on dynamics, for example undergoing a change when one evolves from a Big Bang to a Big Crunch? Quantum mechanics has added even more texture to the notion of time; the usual interpretation combines a continuous time evolution as controlled by differential equations versus the sudden impact of the “so-called” collapse of the wave-function. The quantum Zeno effect and the Einstein-Podolski-Rosen correlations challenge our notions of reality and locality. Our understanding is also tested and maybe even challenged by some phenomena observed in tunneling transitions. The next step, namely, quantizing space and time itself, raises the question of whether a space-time lattice represents the real world at its fundamental level rather than being just a technical device to overcome some mathematical challenges of computation.

The discussions on these issues at "Time and Matter 2007" were grouped into six sections:

Section I: *Measuring Time;* the latest precision measurements of time using cold atom clocks and accurate single ion optical clocks, the management of universal time and other applications of very precise clocks

Section II: *Causality and Signal Propagation;* superluminal signal propagation, interaction between gravitational and electromagnetic radiation

Section III: *Coherence, De-coherence and Entanglement;* entanglement role in studies of *CP* violation at B factories and the tests of Einstein-Podolski-Rosen correlations, decoherence measurements in fullerene interferometry

Section IV: *CP and T Violation; CP* and *T* violation measurements, experimental tests of *CPT* symmetry in the neutral kaon system, *CP* violation in B meson decays, *CP* violation measurements at Large Hadron Collider

Section V: *Quantum Gravity;* canonical quantum gravity, problem of time in quantum gravity, gravitational limitations on space and time measurements and fundamental loss of coherence in quantum theory, emergent space-time

Section VI: *Big Bang Evolution and Structure Formation of the Universe;* origin of time and its error, cosmological constant, new matter

which are presented in the following chapters of these proceedings.

On behalf of both the Local Organizing Committee and the International Advisory Committee, I would like to express our thanks to all the speakers for their interesting lectures and their written contributions. The next "Time and Matter" conference is planned to take place in 2009.

Samo Stanič
Local Organizing Committee, Chair

Contents

Section I: Measuring Time

latest precision measurements of time using cold atom clocks and accurate single ion optical clocks
management of universal time and other applications of very precise clocks

TIME AND MATTER 2007

Time and Matter
2007 Bled, Slovenia

Stable and Accurate Single-ion Optical Clocks

J.C. BERGQUIST[1*], A. BRUSCH[1], S.A. DIDDAMS[1], T.M. FORTIER[1], T.P. HEAVNER[1], L. HOLLBERG[1], D.B. HUME[1], S.R. JEFFERTS[1], L. LORINI[1,2], T.E. PARKER[1], T. ROSENBAND[1], J.E. STALNAKER[1] AND D.J. WINELAND[1]

[1] *Time and Frequency Division, National Institute of Standards and Technology, Boulder, CO 80305, USA*

[2] *IEN, Str. delle Cacce 91, 10135 Torino, Italy*

Abstract: In recent years, several groups throughout the world have initiated research toward the development and systematic evaluation of frequency and time standards based on narrow optical transitions in laser-cooled atomic systems. In this paper we discuss some of the key ingredients to the make-up and operation of single atom, optical clocks and why they offer higher stability and accuracy than the best clocks of today. We also present some of the results obtained at NIST through comparative studies of the $^{199}Hg^+$ single-ion optical clock, the $^{27}Al^+$ single-ion optical clock and the Cs fountain, primary frequency standard (NIST-F1). The frequencies of the clocks are compared with each other using an octave-spanning optical frequency comb, which is tightly phase locked to one of the clock lasers. The most recent frequency comparison between the Hg^+ optical clock and NIST-F1 shows an uncertainty of about 9×10^{-16} limited by the integration time, and recent measurements of the frequency ratio between the Al^+ and Hg^+ standards show an overall uncertainty of several parts in 10^{-16}. The extremely precise measurements of the frequency ratios of these clocks over time have begun to offer more stringent limits on any temporal variation of the fine structure constant α as well as other tests of general relativity.
Tests of the temporal stability of the fine structure constant α are possible with both the Hg^+/Cs and the Hg^+/Al^+ frequency comparisons. From Hg^+/Cs measurements, temporal variation of α is estimated to be lower than 1.3×10^{-16} yr^{-1}, assuming stability of the other fundamental constants involved. This limit is determined from the historical series of frequency comparisons of these two standards spanning more than five years. From the measurements of the frequency ratios

* berky@boulder.nist.gov

of various optical clocks it is possible to directly estimate any present-day temporal variation of α without constraints on other constants. Preliminary data from the measurements of the Hg^+/Al^+ frequency ratio spanning a period of several months indicate a more stringent limit on the time variation of α is possible.
Results from Hg^+/Cs frequency comparisons can also be used to test the postulate of Local Position Invariance (LPI). LPI states that atomic clocks experience the same fractional frequency shift when they move through the same change in gravitational potential. The test presented here uses the natural variation of gravitational potential given by the earths revolution about the sun to set limits on possible violations of LPI.

Mercury ion frequency standard

The first proposal to use the 282 nm transition from the ground $5d^{10}6s\ ^2S_{1/2}$ state to the metastable $5d^96s^2\ ^2D_{5/2}$ state of Hg^+ in an optical frequency was made by Bender *et al.* [1]. The metastable state has a natural lifetime of around 90 ms [2, 3, 4, 5], giving this transition a lifetime limited Q of around 6×10^{14}. Work began at NIST on mercury ion optical frequency standards several years later. The transition was first observed by Doppler-free two-photon absorption of a cloud of trapped $^{198}Hg^+$ ions [2]. The transition was later observed in a single trapped $^{198}Hg^+$ ion by single-photon electric-quadrupole absorption [6]. Doppler broadening was eliminated in this case by confinement of the ion to less than the wavelength of the radiation [7]. The observed linewidth of about 30 kHz was due to the laser linewidth and to the magnetic field instability. Further work resulted in narrowing the frequency width of the laser to less than 1 Hz [8]. Line broadening due the magnetic field was reduced by using the $(F = 0)$ to $(F = 2, m_F = 0)$ hyperfine-Zeeman component in $^{199}Hg^+$, which has only a quadratic Zeeman shift. With these improvements, the 282 nm resonance was observed with a linewidth as low as 6.7 Hz [9]. The frequency of the laser was servo-locked to the atomic resonance so that the apparatus functioned as a frequency standard. With the development of the self-referenced femtosecond laser frequency comb [10, 11, 12], it became possible to compare the frequency of the Hg^+-stabilized laser to microwave or other optical frequency standards [13, 14, 15, 16, 17, 18].

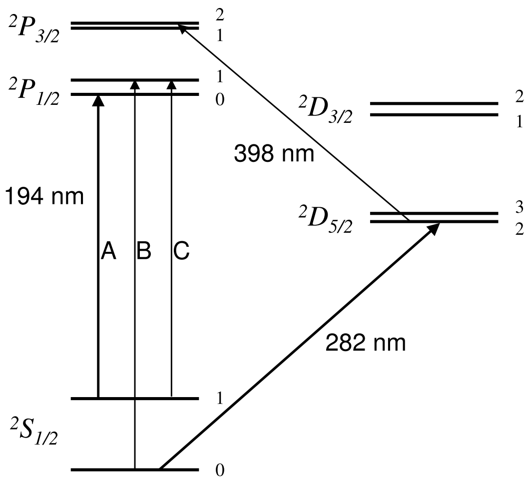


Figure 1: Energy levels of $^{199}Hg^+$. Numbers to the right of the hyperfine energy levels are the values of F, the total angular momentum quantum number. Transitions induced by the lasers are indicated by arrows. For clarity, the energy differences between hyperfine levels are expanded relative to the electronic energy differences.

State preparation and measurement

The basic methods used for laser cooling, state preparation, and detection of the clock transition have been described previously [9], but some additional laser beams are now used to reduce the dead time in the measurement cycle and thereby improve the frequency stability. The energy levels of $^{199}Hg^+$ which are relevant to the operation of the frequency standard are shown in Fig. 1. The 194 nm $5d^{10}6s\ ^2S_{1/2}$ to $5d^{10}6p\ ^2P_{1/2}$ transition is used for Doppler laser cooling and fluorescence detection. The main laser cooling beam is tuned to the $(F = 1)$ to $(F = 0)$ component, labeled A in Fig. 1. To a first approximation, this is a cycling transition, since decay from $(F = 0)$ to $(F = 0)$ is forbidden. However, if the magnetic field is low, the ion can be trapped in a non-absorbing dark state. This can be prevented by applying a large magnetic field, which is undesirable for a frequency standard, or by polarization modulation of the laser [19]. In earlier work, two beams having different propagation directions but with the same frequency were used. The polarization of one beam was modulated between right and left circular polarization [20]. More recently, trapping in a dark state has been prevented by irradiating the ion with three non-collinear beams with frequencies differing by several megahertz [16]. This has enabled reducing the magnetic field to around 8 μT. To prevent trapping of the ion in the ground $(F = 0)$ state, due to off-resonant optical pumping through the ${}^2P_{1/2}$ $(F = 1)$ state, a weak laser beam tuned to the $(F = 0)$ to $(F = 1)$ component, labeled B in Fig. 1, is introduced.

In order to determine whether the ${}^2S_{1/2}$ to ${}^2D_{5/2}$ 282 nm clock transition has been driven, 194 nm radiation at the A and B frequencies is applied. If 194 nm fluorescence is observed, then the transition did *not* occur. If the transition *did* occur, then no fluorescence is observed, and it is necessary to wait 90 ms on the average for the ion to decay back to the ${}^2S_{1/2}$ state before attempting to drive the 282 nm transition again. In the current setup, a laser tuned to the 398 nm ${}^2D_{5/2}(F = 2)$ to ${}^2P_{3/2}(F = 2)$ transition is used to empty the ${}^2D_{5/2}$ state. This reduces the average dead time for the measurement cycle. The ${}^2P_{3/2}$ state decays to the ${}^2S_{1/2}$ state 350 times more frequently than to the ${}^2D_{5/2}$ state [21], so there is very little possibility that the ion will return to the ${}^2D_{5/2}(F = 3)$, which is not emptied by the laser.

Prior to driving the 282 nm clock transition, the ion must be prepared in the ${}^2S_{1/2}$ $(F = 0)$ hyperfine state. Previously, this was done by shutting off the $(F = 0)$ to $(F = 1)$ 194 nm radiation (component B) while leaving on the $(F = 1)$ to $(F = 0)$ 194 nm radiation (component A). This introduced some dead time into the measurement cycle, since around 20 ms had to be allowed for the ion to be pumped into the $(F = 0)$ ground state by off-resonant excitation of the ${}^2P_{1/2}$ $(F = 1)$. In the current setup, a 194 nm laser tuned to the $(F = 1)$ to $(F = 1)$ frequency (component C) is introduced to quickly drive the ion from the ${}^2S_{1/2}$ $(F = 1)$ state to the ${}^2S_{1/2}$ $(F = 0)$ state, thereby reducing the dead time.

Laser frequency servo

In order to lock the frequency of the clock laser to the center of the 282 nm transition, the probability for driving the clock transition is measured for frequencies slightly above and slightly below the estimated resonance frequency. If the transition probability for excitation on the high-frequency side is denoted by $P(H)$ and for excitation on the low-frequency side by $P(L)$, then a measurement result of $P(H) > P(L)$ indicates that the laser frequency is too low, and vice versa. The signal-to-noise ratio is fundamentally limited by quantum projection noise, due to the fact that the atom is found to be in one state or the other when measured, rather than in some superposition [22]. This means that several measurements must be averaged in order to reliably determine the difference between the laser and the clock transition. Two possible sources of frequency error of the locked laser are drift of the signal amplitude and drift of the resonant frequency of the Fabry-Pérot cavity to which the laser is locked. The algorithm for the laser frequency servo attempts to address both issues.

Amplitude drift cancelation

Drift of the signal amplitude might be caused by drift in the laser intensity at the position of the ion. If the intensity drifts downward in time, and if the servo error signal is derived from a high-frequency measurement followed by a low-frequency measurement (HL) then the servo algorithm would cause the laser frequency to be set too low. If the signal drift is linear with time, then this can be compensated by following the (HL) sequence with a (LH) sequence and averaging the results. That is, a linear signal drift is compensated by deriving the error signal from a $(HLLH)$ or $(LHHL)$ sequence of measurements. It has been shown that this method can be generalized to compensate for drifts having arbitrary polynomial time dependence [23]. The error signal for the Hg^+ frequency servo is derived from the $(HLLHLHHL)$ sequence, which compensates for linear or quadratic signal drifts. Since several measurements must be made in order to reduce the amount of projection noise, there is essentially no cost to using this sequence.

Cavity drift compensation

The clock laser is stabilized to a vibrationally-isolated Fabry-Pérot cavity [8]. While this results in sub-hertz laser linewidths, long-term drifts of the locked laser frequency of around 1 Hz/s are observed, and the drift rate can change significantly during an experiment. If a simple servo algorithm is used, then the drift will lead to a frequency error, dependent on the servo gain, because the servo will never quite catch up to the atomic resonance frequency. In general, local-oscillator frequency drifts of this sort are compensated by introducing another stage of integration into the servo response function.

In the Hg^+ frequency servo, the additional integration stage is implemented by introducing a chirped radiofrequency oscillator, whose frequency is added to that of the clock laser with an acousto-optic modulator (AOM). The chirped oscillator consists of a frequency synthesizer whose frequency can be changed, under computer control, while maintaining phase continuity. If the drift rate of the cavity does not match the rate of frequency change of the chirped frequency synthesizer, then the frequency corrections made by the servo algorithm will tend to be all in the same direction, and the frequency of the locked laser (relative to the cavity resonance frequency) will change linearly with time. Periodically, the computer controlling the frequency servo performs a least-squares fit to the record of frequency corrections and modifies the rate of frequency change of the chirped frequency synthesizer in order to decrease the mismatch.

Averaging of the quadrupole shift

Previously, the uncertainty of the frequency of the Hg^+ optical frequency standard was dominated by the uncertainty of the electric quadrupole shift [14]. This shift comes about because the electronic charge density of the $^2D_{5/2}$ state has an electric quadrupole moment, which leads to an energy shift if a static electric field gradient is present. Although no static electric field gradient is applied deliberately, small, uncontrolled electric field gradients might be present, and would be difficult to detect.

The $^2D_{5/2}$ electric quadrupole moment Θ was measured by observing the shift of the clock frequency on applying an electric field gradient [24]. It was found that $\Theta = (-0.510 \pm 0.018)\, ea_0^2$, where e is the elementary charge, and a_0 is the Bohr radius. When no electric field gradient was deliberately applied, it was found that the fractional shift of the clock frequency was less than 10^{-16}. A multi-configuration Dirac-Hartree-Fock calculation gave the result Θ = -0.564 ea_0^2, which disagrees with the experiment by about 10% [25]. Recently, Θ has been calculated by the Fock-state unitary coupled-cluster theory to be -0.517 ea_0^2, in good agreement with the experiment [26].

It is actually not necessary to know the value of the electric quadrupole moment in order to eliminate its effect on the frequency standard. At least two methods are available for canceling the quadrupole shift, both of which make use of the symmetries of the electric quadrupole interaction, but do not depend on its magnitude. First, the shift is zero when averaged over any three mutually perpendicular quantization axes [27]. Second, the shift vanishes when an average over m_F components is done [28].

The Hg^+ frequency standard makes use of the first of the two methods in order to eliminate the quadrupole shift. The orientation of the static magnetic field is switched among three mutually orthogonal directions, so that an equal amount of time is spent at each orientation [16]. The remaining

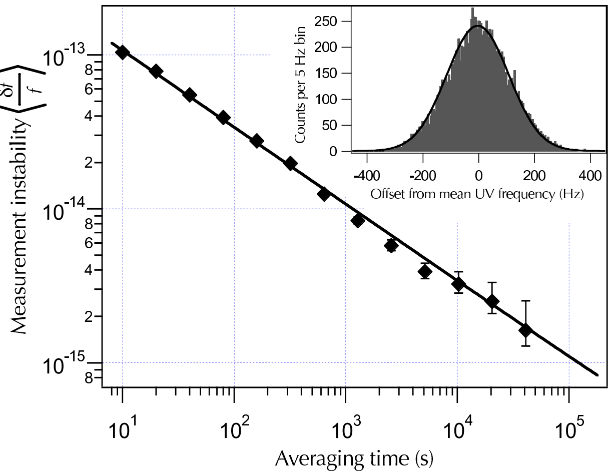


Figure 2: Instability of the ratio of the Hg^+ clock frequency, relative to the cesium frequency standard. The quantity plotted is the Total Deviation. The inset is a histogram of the frequency values and a fitted Gaussian function.

fractional frequency uncertainty of about 10^{-17} is due to the uncertainty of the magnetic field orientations.

Other systematic uncertainties

The total fractional systematic uncertainty of the Hg^+ clock frequency is 3.2×10^{-17}. The various contributions to the systematic uncertainty have been discussed previously [16]. Some of these contributions have been reduced in recent work. Here we discuss the most important contributions in general terms. The fractional second-order Doppler shift due to thermal motion is less than 10^{-17}, because the ion is laser cooled to near the Doppler cooling limit. The fractional second-order Doppler shift due to rf micromotion is also than 10^{-17}, because the stray electric fields that lead to excess micromotion are compensated using a rf-phase-sensitive fluorescence detection method [29]. The static magnetic field is periodically measured by interrupting the frequency servo and observing the resonance line of a first-order magnetic-field dependent Zeeman component of the $^2S_{1/2}(F=0)$ to $^2D_{5/2}(F=2)$ line. The uncertainty of the fractional shift due to variations in the static magnetic field is less than 10^{-17}. The blackbody radiation shift is negligible because the trap is operated at liquid helium temperature. An AC Zeeman shift due to unbalanced rf currents in the trap electrodes has been considered. This fractional shift is estimated to be less than 3×10^{-17}.

Hg^+ optical to Cs microwave frequency comparison

Since the SI second is based on the frequency of the cesium ground-state hyperfine transition, making an absolute frequency measurement at the highest level of accuracy requires a primary cesium frequency standard as a reference. The frequency of a laser frequency-locked to the Hg^+ clock frequency can be compared to a microwave frequency by using a self-referenced femtosecond laser frequency comb [10, 11, 12]. The NIST-F1 cesium atomic fountain has a fractional frequency uncertainty of around 4×10^{-16} [30]. Comparisons have been made over several years between NIST-F1 and the Hg^+ optical frequency standard [14, 15, 16, 17, 18]. The results of a series of measurements of the Hg^+ frequency, referenced to NIST-F1, are shown in Fig. 2 [16]. The Total Deviation is plotted as a function of the averaging time. The Total Deviation is similar to the better-known Allan deviation, but is a better predictor of long-term fractional frequency instability [31]. The inset is a histogram of the optical frequency measurements, together with a fitted Gaussian function. A more recent measurement gives the value of the Hg^+ frequency $f(\mathrm{Hg}^+)$ = 1 064 721 609 899 145.30 $\pm$ 0.69 Hz. The fractional frequency of 6.5×10^{-16} is within a factor of 1.5 of the uncertainty of the NIST-F1 frequency standard [18].

Aluminum ion frequency standard

The possibility of using the transitions from the ground $^1S_0$ state to the metastable $^3P_0$ state in $^{27}\mathrm{Al}^+$ and other group IIIA ions for an optical frequency standard was first pointed out by Dehmelt [32, 33]. These transitions have extremely high Qs and have the additional advantage of not having an electric quadrupole shift, since a $J = 0$ state has zero electric quadrupole moment.

The energy levels of $^{27}\mathrm{Al}^+$ which are relevant to the operation of the frequency standard are shown in Fig. 3. While the strongly-allowed 167 nm $3s^2\ {}^1S_0$ to $3s3p\ {}^1P_1$ transition would be useful for laser cooling and state detection, narrowband, tunable lasers are not available at that wavelength.

To get around this problem, Wineland proposed to simultaneously trap an auxiliary ion, which could be laser cooled and optically detected at a more

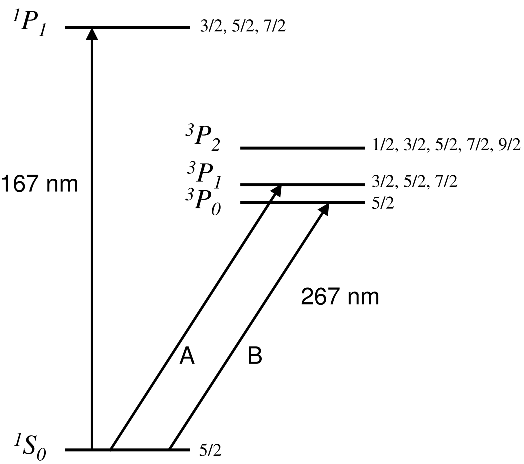


Figure 3: Energy levels of $^{27}Al^+$. Numbers to the right of the energy levels are the values of F, the total angular momentum quantum number. The energy separations between hyperfine energy levels are not shown. The strong transition at 167 nm is not directly driven, for lack of a tunable laser at that wavelength. Transitions near 267 nm induced by the lasers are are labeled A and B. The energy differences between the ^3P_J fine-structure levels are expanded for clarity.

convenient wavelength [34]. Since the two ions are coupled through the Coulomb interaction, the "clock" ion (e. g. $^{27}Al^+$) is also cooled. Further, the superposition state of the clock ion can be transferred to the auxiliary ion, making use of the fact that they share a vibrational degree of freedom. That is, if the state of the clock ion is $(\alpha|S\rangle + \beta|P\rangle)$, the state of the auxiliary ion becomes $(\alpha|1\rangle + \beta|2\rangle)$, where $|1\rangle$ and $|2\rangle$ are two of the hyperfine ground state sublevels. This makes it possible to detect whether the clock ion has been driven to the metastable state by observing the fluorescence of the auxiliary ion.

The basic methods were demonstrated at NIST with a $^{27}Al^+$ clock ion and a $^9Be^+$ auxiliary ion [35]. The $^{27}Al^+$ $^1S_0$ to $^3P_1$ transition (A in Fig. 3) was used for this demonstration. This transition has a natural linewidth of about 500 Hz, so it is not the best choice for an optical frequency standard [36]. More recently, the $^{27}Al^+$ $^1S_0$ to $^3P_0$ transition (B in Fig. 3) has been observed [37]. This transition is the basis for the $^{27}Al^+$ optical frequency standard. Detection of the $^1S_0$ to $^3P_0$ transition depends on the mapping of the $^1S_0$ to $^3P_1$ superposition state to a superposition state of the ^{9}Be$^+$ ion. Details of the methods used to frequency-lock a laser to the ^{9}Be$^{+1}S_0$ to $^3P_0$ resonance have been published [37].

Atomic system

In a nonrelativistic approximation, the $3s3p\ ^3P_1$ state does not decay to the ground state, while the $3s3p\ ^1P_1$ decay is fully allowed. However, spin-orbit and other relativistic interactions mix the $3s3p\ ^3P_1$ and $3s3p\ ^1P_1$ states and both states to decay. In the absence of hyperfine interaction, the $3s3p\ ^3P_0$ state does not decay by any single-photon process. However, the hyperfine interaction mixes the $3s3p\ ^3P_0$ state with other states that do decay to the ground state, mainly the $3s3p\ ^1P_1$ and $3s3p\ ^3P_1$ states. This makes the $3s3p\ ^3P_0$ to $3s^2\ ^1S_0$ decay weakly allowed [38]. The lifetime of the $^3P_0$ state was measured to be 20.6 ± 1.4 s, so the transition has a natural Q of 1.45×10^{17}.

Since the nuclear spin of ^{27}Al, the only stable isotope of aluminum, has spin $I = 5/2$, both the $^1S_0$ and the $^3P_0$ state have total angular momentum $F = 5/2$. Therefore, each m_F component of the $^1S_0$ and $^3P_0$ states has a linear Zeeman shift. In the absence of hyperfine-induced mixing between J-states, the g-factors of the ground and excited states would be nearly equal to each other and to the nuclear g-factor. This would make the frequencies of the $^1S_0(F = 5/2, m_F)$ to $^3P_0(F = 5/2, m_F)$ transitions nearly independent of magnetic field. In fact there is a large shift of the g-factor of the $^3P_0$ state compared to that of the $^1S_0$ state that has recently been measured [37]. It was found that $g(^3P_0)$ = -0.001 976 86(21) and $g(^1S_0)$ = -0.000 792 48(14), where the g-factors are defined in terms of the Bohr magneton, and the numbers in parentheses are the uncertainties in units of the least significant digits. The shift of $g(^3P_0)$ relative to $g(^1S_0)$ is due to the hyperfine interaction mixing the $^3P_0$ with other other J-states, mainly $3s3p\ ^3P_1$. This type of g-factor shift was first observed in the $6s6p\ ^3P_0$ states of ^{199}Hg and ^{201}Hg [39].

Comparison of atomic calculations with experiment

Recently, some calculations of diagonal and off-diagonal hyperfine constants of ^{27}Al$^+$ have been carried out [40]. The `GRASP92` set of programs [41] was used to generate the atomic state functions for the $\{^1P_1, ^3P_2, ^3P_1, ^3P_0\}$ set of states by the multi-configuration Dirac-Hartree-Fock method, and the `HFS92` program [42] was used to calculate the mag-

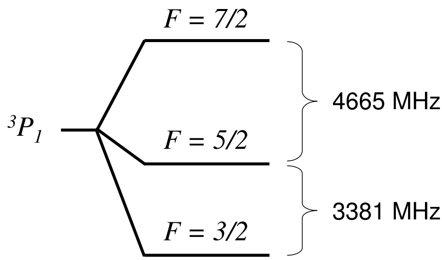


Figure 4: Hyperfine structure of the $6s6p\ ^3P_1$ state of $^{27}\mathrm{Al}^+$.

netic dipole (A) and electric quadrupole (B) hyperfine constants. Values for the nuclear magnetic moment $\mu_I = 3.64150687(65)\mu_N$ and electric quadrupole moment $Q = 0.1466(10) \times 10^{-24}\, e\, \mathrm{cm}^2$ were assumed [43, 44]. The results are listed in Table 1.

These calculations can be connected with experiment in at least three different ways:

First, the hyperfine separations of the $^3P_1$ state have been measured at NIST (see Fig. 4). The separations were measured to be 4664.903(1) MHz and 3380.688(1) MHz [45]. The uncertainties are mainly due to the uncertainty of the quadratic Zeeman shift, since the measurements were made at nonzero magnetic field. Extraction of the diagonal A and B coefficients is not straightforward, because the second-order magnetic dipole energy is comparable to the first-order electric quadrupole energy. One way to test the calculation is to compare the predicted separations, calculated to second order in perturbation theory, with the observed separations. The predicted separations are 4693 MHz and 3400 MHz, which agree with experiment to less than 2%.

Second, we can use the off-diagonal A coefficients which mix the $^3P_0$ state with the $^1P_1$ and $^3P_1$ states to calculate the radiative decay rate of the $^3P_0$ state [46]. This calculation also requires the radiative lifetimes of the $^3P_1$ state and the $^1P_1$ state. We take the former from experiment [36] and the latter from the NIST database [47]. The result is 22.7 s, compared to the experimental result of 20.6 ± 1.6 s [37].

Third, we can use the same off-diagonal A coefficients to calculate the difference between the g-factors of the $^3P_0$ and $^1S_0$ states [46]. The result is $g(^3P_0)$ - $g(^1S_0) = -1.181 \times 10^{-3}$, compared to the experimental result $-1.118437(8) \times 10^{-3}$ [37].

Table 1: Calculated values of the diagonal and off-diagonal hyperfine A and B constants within the $3s3p\ ^{1,3}P_J$ set of states.

Γ	Γ'	$A(\Gamma,\Gamma')$ (MHz)	$B(\Gamma,\Gamma')$ (MHz)
$^3P_2$	$^3P_2$	1149	31.42
$^3P_2$	$^3P_1$	-539	7.84
$^3P_2$	$^1P_1$	845	-0.05
$^3P_2$	$^3P_0$	0	13.57
$^3P_1$	$^3P_1$	1348	-15.62
$^3P_1$	$^1P_1$	1571	0.18
$^3P_1$	$^3P_0$	-1320	0
$^1P_1$	$^3P_0$	-1045	0

Systematic uncertainties

The total fractional systematic uncertainty of the Al^+ clock frequency is 3.7×10^{-17}. The various contributions are discussed in detail elsewhere [37]. The greatest part of the systematic uncertainty is due to the second-order Doppler shift, from both the thermal motion and the micromotion. The linear Zeeman shift is cancelled by alternately observing the $^1S_0(F = 5/2, m_F = +5/2)$ to $^3P_0(F = 5/2, m_F = +5/2)$ and the $^1S_0(F = 5/2, m_F = -5/2)$ to $^3P_0(F = 5/2, m_F = -5/2)$ transitions and averaging the frequencies. The quadratic Zeeman shift has been measured, and it contributes less than 10^{-18} to the fractional frequency uncertainty.

The blackbody radiation shift is unusually small for an optical frequency standard because of a fortuitous cancelation between the $^1S_0$ and $^3P_0$ quadratic Stark shifts [48]. The fractional frequency shift at 300K is $8(5) \times 10^{-18}$. At the normal operating temperature of the frequency standard, it is $12(5) \times 10^{-18}$.

While to first order, the $^3P_0$ state has zero electric quadrupole moment, the hyperfine interaction mixes this state with states having $J = 1, 2$, such as the nearest $^3P_1$ and $^3P_2$ states. This results in an electronic quadrupole moment of about $-1.2 \times 10^{-5}\ ea_0^2$ [40]. This quadrupole can interact with the electric field gradients due to the ion trap and to the presence of the

$^9Be^+$ ion, resulting in a fractional frequency shift which is less than 10^{-18} and can be neglected.

Hg^+ - Al^+ frequency comparisons

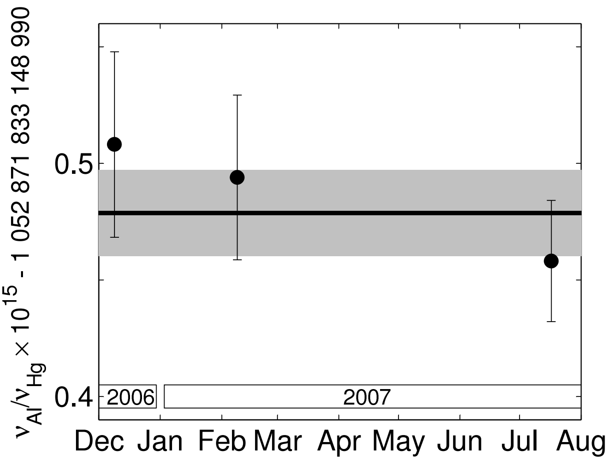


Figure 5: Measurements of the ratio of the Al^+ and Hg^+ clock frequencies as a function of the date the measurements were made.

The Hg^+ and Al^+ frequency standards have been operated simultaneously. The frequency of one mode or "tooth" of the self-referenced femtosecond laser frequency comb can be phase-locked to the frequency of one standard. The frequency of the heterodyne beat-note of the other frequency standard with the nearest comb tooth is measured. The measurement can be converted to the frequency ratio of the two standards and does not depend on the accuracy of any microwave frequency standards used as references. Figure 5 is a plot of measurements of the ratio of the frequency of the Al^+ standard to that of the Hg^+ standard. The reproducibility of the ratio is seen to be better than 1×10^{-16}, which is better than the accuracy of the primary Cs frequency standard. Figure 6 shows the fractional frequency instability of the ratio as a function of averaging time.

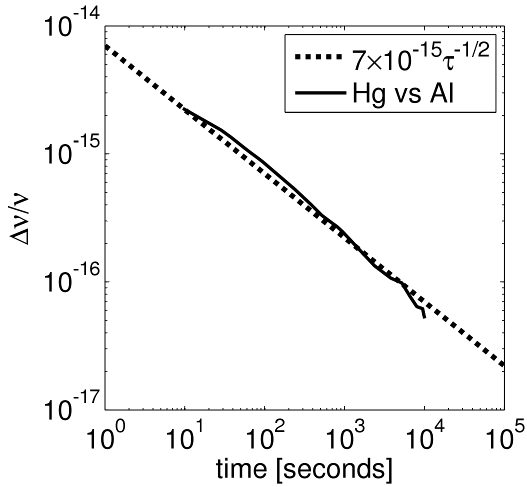


Figure 6: Instability of the ratio of the Al^+ and Hg^+ clock frequencies. The quantity plotted is the Allan deviation, the square root of the Allan variance.

Acknowledgments

This work was partially supported by the Office of Naval Research. This work was performed by an agency of the U. S. government and is not subject to U. S. copyright.

References

[1] P.L. Bender, J.L. Hall, R.H. Garstang, F.M.J. Pichanick, W.W. Smith, R.L. Barger and J.B. West, Bull. Am. Phys. Soc. **21** (1976) 599.

[2] J.C. Bergquist, D.J. Wineland, W.M. Itano, H. Hemmati, H.-U. Daniel and G. Leuchs, Phys. Rev. Lett. **55** (1985) 1567.

[3] J.C. Bergquist, R.G. Hulet, W.M. Itano and D.J. Wineland, Phys. Rev. Lett. **57** (1986) 1699.

[4] W.M. Itano, J.C. Bergquist, R.G. Hulet and D.J. Wineland, Phys. Rev. Lett. **59** (1987) 2732.

[5] A.G. Calamai and C.E. Johnson, Phys. Rev. A **42** (1990) 5425.

[6] J.C. Bergquist, W.M. Itano and D.J. Wineland, Phys. Rev. A **36** (1987) 428.

[7] R.H. Dicke, Phys. Rev. **89** (1953) 472.

[8] B.C. Young, F.C. Cruz, W.M. Itano and J.C. Bergquist, Phys. Rev. Lett. **82** (1999) 3799.

[9] R.J. Rafac, B.C. Young, J.A. Beall, W.M. Itano, D.J. Wineland and J.C. Bergquist, Phys. Rev. Lett. **85** (2000) 2462.
[10] S.A. Diddams, D.J. Jones, J. Ye, S.T. Cundiff, J.L. Hall, J.K. Ranka, R.S. Windeler, R. Holzwarth, T. Udem and T.W. Hänsch, Phys. Rev. Lett. **84** (2000) 5102.
[11] D.J. Jones, S.A. Diddams, J.K. Ranka, A. Stentz, R.S. Windeler, J.L. Hall, and S.T. Cundiff, Science **288** (2000) 635.
[12] R. Holzwarth, T. Udem, T.W. Hänsch, J.C. Knight, W.J. Wadsworth and P.S.J. Russell, Phys. Rev. Lett. **85** (2000) 2264.
[13] 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 and D.J. Wineland, Science **293** (2001) 825.
[14] 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** (2001) 4996.
[15] 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 and J.C. Bergquist, Phys. Rev. Lett. **90** (2003) 150802.
[16] W.H. Oskay, S.A. Diddams, E.A. Donley, T.M. Fortier, T.P. Heavner, L. Hollberg, W.M. Itano, S.R. Jefferts, M.J. Delaney, K. Kim, F. Levi, T.E. Parker and J.C. Bergquist, Phys. Rev. Lett. **97** (2006) 020801.
[17] T.M. Fortier, N. Ashby, J.C. Bergquist, M.J. Delaney, S.A. Diddams, T.P. Heavner, L. Hollberg, W.M. Itano, S.R. Jefferts, K. Kim, F. Levi, L. Lorini, W.H. Oskay, T.E. Parker, J. Shirley and J.E. Stalnaker, Phys. Rev. Lett. **98** (2007) 070801.
[18] J.E. Stalnaker, S.A. Diddams, T.M. Fortier, L. Hollberg, J.C. Bergquist, W.M. Itano, M.J. Delaney, L. Lorini, W.H. Oskay, T.P. Heavner, S.R. Jefferts, F. Levi, T.E. Parker and J. Shirley, Appl. Phys. B (in press).
[19] D.J. Berkeland and M.G. Boshier, Phys. Rev. A **65** (2002) 033413.
[20] D.J. Berkeland, J.D. Miller, J.C. Bergquist, W.M. Itano and D.J. Wineland, Phys. Rev. Lett. **80** (1998) 2089.
[21] D.H. Crandall, R.A. Phaneuf and G.H. Dunn, Phys. Rev. A **11** (1975) 1223.
[22] W.M. Itano, J.C. Bergquist, J.J. Bollinger, J.M. Gilligan, D.J. Heinzen, F.L. Moore, M.G. Raizen and D.J. Wineland, Phys. Rev. A **47** (1993) 3554.
[23] G.E. Harrison, M.A. Player and P.G.H. Sandars, J. Phys. E **4** (1971) 750.
[24] W.H. Oskay, W.M. Itano and J.C. Bergquist, Phys. Rev. Lett. **94** (2005) 163001.
[25] W.M. Itano, Phys. Rev. A **73** (2006) 022510.
[26] C. Sur and R.K. Chaudhuri, Phys. Rev. A (in press), [arχiv:0707.3587v2].
[27] W.M. Itano, J. Research National Institute of Standards and Technology **105** (2000) 829.
[28] P. Dubé, A.A. Madej, J.E. Bernard, L. Marmet, J.-S. Boulanger and S. Cundy, Phys. Rev. Lett. **95** (2005) 033001.
[29] D.J. Berkeland, J.D. Miller, J.C. Bergquist, W.M. Itano and D.J. Wineland, JȦppl. Phys. **83** (1998) 5025.
[30] T.P. Heavner, S.R. Jefferts, E.A. Donley, J.H. Shirley and T.E. Parker, Metrologia **42** (2005) 411.
[31] C.A. Greenhall, D.A. Howe and D.B. Percival, IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46** (1999) 1183.
[32] H. Dehmelt, Bull. Am. Phys. Soc. **20** (1975) 60.

[33] H.G. Dehmelt, IEEE Trans. Instrum. Meas. **31** (1982) 83.
[34] D.J. Wineland, J.C. Bergquist, J.J. Bollinger, R.E. Drullinger and W.M. Itano, Proceedings of the 6th Symposium on Frequency Standards and Metrology, ed. P. Gill (2002) 361–368 (World Scientific, Singapore).
[35] P.O. Schmidt, T. Rosenband, C. Langer, W.M. Itano, J.C. Bergquist, and D.J. Wineland, Science **309** (2005) 749.
[36] E. Träbert, A. Wolf, J. Linkemann and X. Tordoir, J. Phys. B **32** (1999) 537.
[37] T. Rosenband, P.O. Schmidt, D.B. Hume, W.M. Itano, T.M. Fortier, J.E. Stalnaker, K. Kim, S.A. Diddams, J.C.J. Koelemeij, J.C. Bergquist and D.J. Wineland, Phys. Rev. Lett. **98** (2007) 220801.
[38] T. Brage, P.G. Judge, A. Aboussaïd, M.R. Godefroid, P. Jönsson, A. Ynnerman, C.F. Fischer and D.S. Leckrone, Astrophys. J. **500** (1998) 507.
[39] B. Lahaye and J. Margerie, J. Phys. (Paris) **36** (1975) 943.
[40] W.M. Itano (unpublished calculation).
[41] F.A. Parpia, C.F. Fischer and I.P. Grant, Comput. Phys. Commun. **94** (1996) 249.
[42] P. Jönsson, F.A. Parpia and C.F. Fischer, Comput. Phys. Commun. **96** (1996) 301.
[43] P. Raghavan, At. Data Nucl. Data Tables **42** (1989) 189.
[44] P. Pyykkö, Mol. Phys. **99** (2001) 1617.
[45] T. Rosenband *et al.* (unpublished work).
[46] E. Peik, G. Hollemann and H. Walther, Phys. Rev. A **49** (1994) 402.
[47] "NIST Atomic Spectra Database," `http://physics.nist.gov/PhysRefData/ASD/index.html`.
[48] T. Rosenband, W.M. Itano, P.O. Schmidt, D.B. Hume, J.C.J. Koelemeij, J.C. Bergquist and D.J. Wineland, Proc. 20th European Time and Frequency Forum (2006) 289–292, [arχiv:physics/0611125v2].

Time and Matter
2007 Bled, Slovenia

Optical Clocks with Trapped Ions and the Search for Temporal Variations of Fundamental Constants

E. PEIK*, B. LIPPHARDT, H. SCHNATZ, CHR. TAMM, S. WEYERS AND R. WYNANDS
Physikalisch-Technische Bundesanstalt
Bundesallee 100, D-38116 Braunschweig, Germany

Abstract: The techniques of trapping and laser cooling of ions have allowed to perform laser spectroscopy of forbidden transitions with a resolution of a few hertz. These systems will be used as optical atomic clocks that offer higher stability and greater accuracy than the best primary cesium clocks available today. At PTB we have built an optical clock based on a single trapped ytterbium ion and have shown that the frequencies realized in two independent ion traps agree to within a few parts in 10^{16}. An interesting question from fundamental physics that can be investigated with optical clocks of this precision is the search for possible temporal variations of fundamental constants, based on comparisons between different transition frequencies over time. Currently, we can infer an upper limit for the relative change of the fine structure constant of $4 \cdot 10^{-16}$ per year.

Introduction

Application of the methods of laser cooling and trapping has led to significant improvements in the precision of atomic clocks and frequency standards over the last years [1, 2]: Primary cesium clocks based on atomic fountains [3, 4] realize the unit of time, the SI second, with a relative uncertainty below 10^{-15}. Research towards *optical* clocks with trapped ions and atoms [5, 2] has resulted in several systems that approach the accuracy of the primary cesium clocks or show an even higher reproducibility. A significant advantage of an optical clock lies in the high frequency of the oscillator, that allows one to perform precision frequency measurements in a short averaging time. The lowest instabilities that have been demon-

* `ekkehard.peik@ptb.de`

strated are a few parts in 10^{15} in 1 s only, improving further like the inverse square root of the averaging time.

If an atom or ion is held in a trap [6], several problems that are encountered in atomic frequency standards can be eliminated: 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 linewidth of the interrogating oscillator. Combined with laser cooling, the ion can be brought to the vibrational ground state of the trap potential [7], 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, 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* [8] 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 detect the excitation on the forbidden optical transition, Dehmelt's *electron shelving* scheme is used: 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 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.

A number of suitable reference transitions with natural linewidths of the order of 1 Hz or below are available in different ions and several groups pursue research along the lines of the mono-ion oscillator proposal (see [5, 9] for recent reviews). High-resolution spectroscopy and precise frequency measurements using femtosecond laser frequency comb generators [10] have been performed on the $S \rightarrow D$ electric quadrupole transitions in the alkali-like ions Sr^+ [11], Hg^+ [12, 13, 14, 15] and Yb^+ [16, 17, 18, 19].

Measurements of the Yb^+ optical frequency standard at 688 THz

The $^{171}Yb^+$ ion is attractive as an optical frequency standard because it offers narrow reference transitions with small systematic frequency shift. At PTB, the electric quadrupole transition $(^2S_{1/2}, F = 0) \rightarrow (^2D_{3/2}, F = 2)$ at 436 nm wavelength (688 THz frequency) with a natural linewidth of 3.1 Hz is investigated [17, 19, 20, 21]. A single ytterbium ion is trapped in a miniature Paul trap and is laser-cooled to a sub-millikelvin temperature by exciting 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 at 370 nm (cf. Fig. 1). The reference transition is probed by the frequency doubled radiation from a diode laser emitting at 871 nm. The short term frequency stability of this laser is derived from a temperature-stabilized and seismically isolated high-finesse reference cavity.

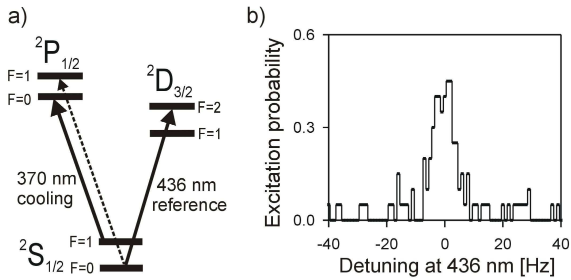


Figure 1: a) Simplified level scheme of the $^{171}Yb^+$ ion. b) Excitation spectrum of the $^2S_{1/2}(F = 0, m_F = 0) \rightarrow {}^2D_{3/2}(F = 2, m_F = 0)$ transition, obtained with 20 probe cycles for each value of the detuning. The linewidth of 10 Hz is approximately at the Fourier limit for the employed probe pulses of 90 ms duration.

Figure 1b shows a high-resolution excitation spectrum obtained with 90 ms long laser pulses, leading to an approximately Fourier-limited linewidth 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 resonance are probed alternately, and the probe light frequency is stabilized to the line center according to the difference of the measured excitation probabilities [21].

For quantitative studies of systematic frequency shifts we have compared the line center frequencies of two $^{171}\mathrm{Yb}^{+}$ ions stored in separate traps [20]. In the absence of external perturbations we found a mean frequency difference between the two trapped ions of 0.26(42) Hz, corresponding to a relative difference of $3.8(6.1)\cdot 10^{-16}$. This is comparable to the agreement found in the most accurate comparisons between cesium fountain clocks. The dominant source of systematic uncertainty in the $^{171}\mathrm{Yb}^{+}$ optical frequency standard is the so-called quadrupole shift of the atomic transition frequency. It is due to the interaction of the electric quadrupole moment of the $D_{3/2}$ state with the gradient of static electric patch fields in the trap. The frequency difference between the two systems shows a relative instability (Allan deviation) of $\sigma_y(1000\,\mathrm{s}) = 5\cdot 10^{-16}$, limited essentially by quantum projection noise [21], i.e. by the fact that the signal is derived from single ions.

The absolute frequency of the $^{171}\mathrm{Yb}^{+}$ standard at 688 THz was measured relative to PTB's cesium fountain clock CSF1 [22, 23]. The link between optical and microwave frequencies is established with a femtosecond-laser frequency comb generator [10]. The $^{171}\mathrm{Yb}^{+}$ frequency was measured first in December 2000 [16]. Fig. 2 shows the results of all frequency measurements performed up to June 2006, showing a steady decrease of the measurement uncertainty and excellent overall consistency. Technical improvements over the years were made in the resolution of the ionic resonance, in the control of trapping conditions, and in the frequency comb generator setup [24]. Between July 2005 and June 2006 five absolute frequency measurements with continuous averaging times of up to 36 h were performed. The obtained statistical uncertainties (about $6\cdot 10^{-16}$ after two days) are dominated by the white frequency noise of the cesium fountain. The relative systematic uncertainty contribution of the cesium reference of $1.2\cdot 10^{-15}$ is based on a reevaluation of CSF1 that is applicable to the 2005/06 measurements. The systematic uncertainty contribution for the Yb^{+} standard of $1.5\cdot 10^{-15}$ is largely dominated by an estimate of 1 Hz for the stray-field induced quadrupole shift.

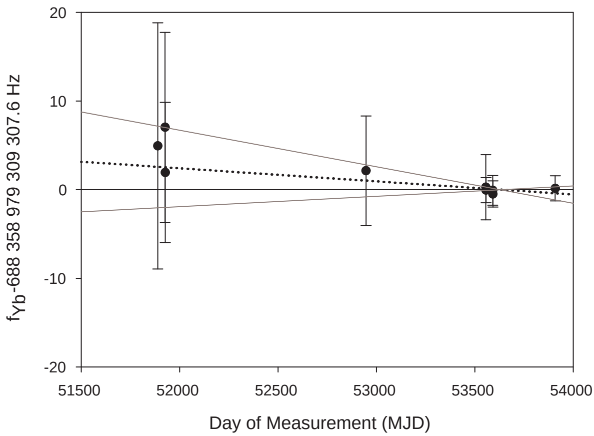


Figure 2: Results of absolute frequency measurements of the Yb^+ optical frequency standard at 688 THz, plotted as a function of measurement date (MJD: Modified Julian Date), covering the period from December 2000 to June 2006. The dotted straight line is the result of a weighted linear regression through the data points. The gray lines indicate the $\pm 1\sigma$ range for the slope as determined from the regression. The intercepts of the gray lines are chosen such that they cross in the point determined by the weighted averages of frequency and measurement date.

The weighted average over these measurements gives the present result for the frequency of the ${}^{171}\mathrm{Yb}^+\ {}^2S_{1/2}(F=0) \rightarrow {}^2D_{3/2}(F=2)$ transition:

$$f_{\mathrm{Yb}} = 688\,358\,979\,309\,307.6(1.4)\,\mathrm{Hz} \tag{1}$$

with a total relative uncertainty of $2.0 \cdot 10^{-15}$. This frequency value refers to a measurement at room temperature (296 K). Extrapolating to zero temperature (as it was done for the cesium clock) the frequency would have to be corrected for the AC Stark shift from blackbody radiation, i.e. increased by $0.37(5)$ Hz.

From a weighted linear regression of the sequence of frequency measurements over time (cf. Fig. 2) we obtain a slope of $(-0.54 \pm 0.97)\,\mathrm{Hz/yr}$, corresponding to a value for the fractional temporal variation of the frequency ratio $\mathrm{d}\ln(f_{\mathrm{Yb}}/f_{\mathrm{Cs}})/\mathrm{d}t = (-0.78 \pm 1.40) \cdot 10^{-15}\,\mathrm{yr}^{-1}$, consistent with zero. This result shows the consistency between the optical clock and the cesium clock. In addition, it can be used to obtain limits on possible temporal variations of the fundamental constants underlying the atomic transition frequencies.

Search for Variations of the Fine Structure Constant

Over the past few years there has been great interest in the possibility that the fundamental constants of nature might show temporal variations over cosmological time scales [25, 26, 27, 28, 29]. Such an effect – as incompatible as it seems with the present foundations of physics – appears quite naturally in the attempt to find a unified theory of the fundamental interactions. An active search for an indication of variable constants is pursued mainly in two areas: observational astrophysics and laboratory experiments with atomic frequency standards. The two most important quantities under test are Sommerfeld's fine structure constant $\alpha = e^2/(4\pi\epsilon_0\hbar c)$ and the proton-to-electron mass ratio $\mu = m_p/m_e$. While α is the coupling constant of the electromagnetic interaction, μ also depends on the strength of the strong force and on the quark masses via the proton mass. It is important that these quantities are dimensionless numbers so that results can be interpreted independently from the conventions of a specific system of units. Quite conveniently, both constants appear prominently in atomic and molecular transition energies: α in atomic fine structure splittings and other relativistic contributions, and μ in molecular vibration and rotation frequencies as well as in hyperfine structure.

A multitude of data on variations of α has been obtained from astrophysical observations but the present picture that is obtained is not completely consistent: Evidence for a variation of α has been derived from a shift of wavelengths of metal ion absorption lines produced by interstellar clouds in the light from distant quasars [30]. These observations suggest that about 10 billion years ago (redshift range $0.5 < z < 3.5$), the value of α was smaller than today by $\Delta\alpha/\alpha = (-0.543 \pm 0.116) \cdot 10^{-5}$, representing 4.7σ evidence for a varying α [31]. Assuming a linear increase of α with time, this would correspond to a drift rate $\mathrm{d}\ln\alpha/\mathrm{d}t = (6.40 \pm 1.35) \cdot 10^{-16}\,\mathrm{yr}^{-1}$ [31]. Other evaluations of spectra obtained with a different telescope and using other selections of quasar absorption systems reach similar or even higher sensitivity for $\Delta\alpha/\alpha$ but are consistent with $\Delta\alpha = 0$ for all look-back times [32, 33]. It should be kept in mind that the uncertainties cited here are usually the result of averaging over a large ensemble of data points with much higher individual statistical uncertainties. The significance of the results that detect non-zero variations therefore hinges on a complete understanding of all systematic effects that may introduce bias or correlations in the data and that may not be easy to detect.

The approach of using precision laboratory experiments with frequency standards to search for variations of the fundamental constants [34] has the advantage that the relevant parameters are under the control of the experimenter, permitting a detailed investigation of possible systematic effects. The obvious disadvantage is that the limited duration of the experiments of typically a few years only is not well adapted to search for evolution on a cosmological time scale. Precision data are available for the ground state hyperfine frequency of ^{87}Rb [35, 4], the $1S \rightarrow 2S$ two-photon transition in atomic hydrogen [36], the transition $^2S_{1/2} \rightarrow {}^2D_{5/2}$ at 1065 THz in Hg^+ [37, 14, 15], and the $^2S_{1/2} \rightarrow {}^2D_{3/2}$ transition at 688 THz in $^{171}Yb^+$ (this work). Significant changes of α would produce a clear signature in these experiments, because they would influence the transition frequencies differently [28]. Atomic gross structure scales with the Rydberg constant R_∞, hyperfine structure with the product of $\alpha^2 R_\infty$ and the nuclear magnetic moment. In addition, the fine structure constant α appears in relativistic contributions to the level energies that increase with the square of the nuclear charge and are consequently much more important for heavy atoms [38]. The sensitivity of a specific transition frequency to changes of α can be derived from *ab initio* relativistic atomic structure calculations [39, 40].

For the analysis we use a simple parametrization that includes only a minimum of assumptions [28]. The electronic transition frequency is expressed as

$$f = \text{const} \cdot Ry \cdot F(\alpha) \tag{2}$$

where $Ry = m_e e^4/(8\epsilon_0 h^3) \simeq 3.2898 \cdot 10^{15}\,\text{Hz}$ is the Rydberg frequency, appearing as the common atomic scaling factor. $F(\alpha)$ is a dimensionless function of α that takes relativistic contributions to the level energies into account [38, 39]. The numerical constant in front depends only on integer or half-integer quantum numbers characterizing the atomic structure and is independent of time. The relative temporal derivative of the frequency f can be written as:

$$\frac{\mathrm{d}\ln f}{\mathrm{d}t} = \frac{\mathrm{d}\ln Ry}{\mathrm{d}t} + A\frac{\mathrm{d}\ln \alpha}{\mathrm{d}t} \quad \text{with } A \equiv \frac{\mathrm{d}\ln F}{\mathrm{d}\ln \alpha}. \tag{3}$$

The first term $\mathrm{d}\ln Ry/\mathrm{d}t$ represents a variation that would be common to all measurements of electronic transition frequencies: a change of the numerical value of the Rydberg frequency. This statement refers to a dimensional quantity (frequency) in an SI unit (hertz). Since the SI second is fixed to the hyperfine structure of the ^{133}Cs nucleus, possible variations of the quantity Ry are not purely electromagnetic in origin – as one would expect for the abstract physical quantity "Rydberg constant" – but are also related to the strong interaction. This is a complication that arises because dimensional quantities unavoidably depend on the system of units that is used to measure them. The second term in Eq. 3 is specific to the atomic transition under study. The sensitivity factor A for small changes of α has been calculated for the relevant transitions by Dzuba, Flambaum *et al.* [39, 40].

To obtain a limit on $\mathrm{d}\ln\alpha/\mathrm{d}t$ we combine the data from $^{171}\mathrm{Yb}^+$ with results on the constancy of the transition frequency in $^{199}\mathrm{Hg}^+$, measured at NIST in Boulder [13, 14, 15]. The transition $5d^{10}6s\ ^2S_{1/2} \rightarrow 5d^9 6s^2\ ^2D_{5/2}$ of the mercury ion at 282 nm (1065 THz) with a natural linewidth of 1.9 Hz is studied. A sequence of measurements over the period from 2001 to 2006 has resulted in a constraint on the fractional variation of the frequency ratio $\mathrm{d}\ln(f_{\mathrm{Hg}}/f_{\mathrm{Cs}})/\mathrm{d}t = (0.37 \pm 0.39)\cdot 10^{-15}\,\mathrm{yr}^{-1}$ [15].

The sensitivities of the ytterbium and mercury transition frequencies to changes of α are quite different [39, 40]: $A_{\mathrm{Yb}} = 0.88$ and $A_{\mathrm{Hg}} = -3.19$. The change of sign of A between the two transitions reflects the fact that in Yb^+ a 6s-electron is excited to the empty 5d-shell, while in Hg^+ a hole is created in the filled 5d-shell when the electron is excited to 6s. The two measured drift rates together with Eq. 3 can now be used to calculate:

$$\frac{\mathrm{d}\ln\alpha}{\mathrm{d}t} = (-0.28 \pm 0.36)\cdot 10^{-15}\,\mathrm{yr}^{-1}, \tag{4}$$

$$\frac{\mathrm{d}\ln Ry}{\mathrm{d}t} = (-0.53 \pm 1.10)\cdot 10^{-15}\,\mathrm{yr}^{-1}. \tag{5}$$

We conclude that to within an uncertainty of 4 parts in 10^{16} per year, the fine structure constant *is constant* in the present epoch. This result can be qualified as model-independent because it is only the simple expression Eq. 2 and the *ab initio* atomic structure calculations of the A-values that have been used in the interpretation of the experimental data. This limit is slightly more stringent than that reported in [15] because the more recent data for the $^{171}\mathrm{Yb}^+$ frequency have been used here. The constancy of Ry implies that there is no temporal drift between atomic gross structure frequencies and the cesium hyperfine splitting. This is of importance for metrology because it means that a cesium clock and an optical atomic clock can be regarded as fundamentally equivalent.

A very precise result has been published on the constancy of the ratio of ground state hyperfine frequencies in $^{87}\mathrm{Rb}$ with respect to $^{133}\mathrm{Cs}$ [4]: $\mathrm{d}\ln(f_{\mathrm{Cs}}/f_{\mathrm{Rb}})/\mathrm{d}t = (0.05 \pm 0.53)\cdot 10^{-15}\,\mathrm{yr}^{-1}$. Since this frequency ratio involves the nuclear magnetic moments, it can be combined with the above results to derive a limit on a variation of the proton-to-electron mass ratio [41, 42].

Though over a very different time scale, the laboratory experiments have now reached the same sensitivity for $d \ln \alpha / dt$ as the analysis of quasar absorption spectra, if a linear time evolution of α is assumed. The laboratory search will soon become more sensitive, as the precision of frequency standards continues to improve. In addition, a larger variety of systems is now being investigated so that it will be possible to perform tests of the consistency if a first observation of a variation in the laboratory is reported.

Acknowledgments

We thank S. Karshenboim for many inspiring discussions. This work was partly supported by Deutsche Forschungsgemeinschaft in SFB 407 and by grant RFP1-06-08 from the Foundational Questions Institute (`fqxi.org`).

References

[1] A. Bauch and H.R. Telle, Rep. Prog. Phys. **65** (2002) 789.
[2] S.A. Diddams, J.C. Bergquist, S.R. Jefferts and C.W. Oates, Science **306** (2004) 1318.
[3] R. Wynands and S. Weyers, Metrologia **42** (2005) S64.
[4] S. Bize *et al.*, J. Phys. B: At. Mol. Opt. Phys. **38** (2005) S449.
[5] P. Gill *et al.*, Meas. Sci. Tech. **14** (2003) 1174.
[6] W. Paul, Rev. Mod. Phys. **62** (1990) 531.
[7] F. Diedrich, J.C. Bergquist, W.M. Itano and D.J. Wineland, Phys. Rev. Lett. **62** (1989) 403.
[8] H. Dehmelt, IEEE Trans. Instrum. Meas. **31** (1982) 83.
[9] 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).
[10] Th. Udem, R. Holzwarth, T.W. Hänsch, Nature **416** (2002) 233.
[11] H.S. Margolis *et al.*, Science **306** (2004) 1355.
[12] R.J. Rafac, B.C. Young, J.A. Beall, W.M. Itano, D.J. Wineland, and J.C. Bergquist, Phys. Rev. Lett. **85** (2000) 2462.
[13] S. Bize *et al.*, Phys. Rev. Lett. **90** (2003).
[14] W.H. Oskay *et al.*, Phys. Rev. Lett. **97** (2006) 020801.
[15] T.M. Fortier *et al.*, Phys. Rev. Lett. **98** (2007) 070801.
[16] J. Stenger, Chr. Tamm, N. Haverkamp, S. Weyers and H. Telle, Opt. Lett. **26** (2001) 1589.
[17] Chr. Tamm, D. Engelke and V. Bühner, Phys. Rev. A **61** (2000) 053405.
[18] E. Peik, B. Lipphardt, H. Schnatz, T. Schneider, Chr. Tamm and S.G. Karshenboim, Phys. Rev. Lett. **93** (2004) 170801.

[19] Chr. Tamm, B. Lipphardt, H. Schnatz, R. Wynands, S. Weyers, T. Schneider and E. Peik, IEEE Trans. Instrum. Meas. **56** (2007) 601.
[20] T. Schneider, E. Peik and Chr. Tamm, Phys. Rev. Lett. **94** (2005) 230801.
[21] E. Peik, T. Schneider and Chr. Tamm, J. Phys. B: At. Mol. Opt. Phys. **39** (2006) 145.
[22] S. Weyers, U. Hübner, B. Fischer, R. Schröder, Chr. Tamm and A. Bauch, Metrologia **38** (2001) 343.
[23] S. Weyers, A. Bauch, R. Schröder and Chr. Tamm, Proc. 6th Symposium on Frequency Standards and Metrology (World Scientific, 2002) 64–71.
[24] F. Adler, K. Moutzouris, A. Leitensdorfer, H. Schnatz, B. Lipphardt, G. Grosche and F. Tauser, Opt. Expr. **12** (2004) 5872.
[25] J.-P. Uzan, Rev. Mod. Phys. **75** (2003) 403.
[26] *Astrophysics, Clocks and Fundamental Constants*, Lecture Notes in Physics Vol. 648, eds. S.G. Karshenboim and E. Peik (Springer, Heidelberg, 2004).
[27] J.D. Barrow, Phil. Trans. Roy. Soc. **363** (2005) 2139.
[28] S.G. Karshenboim, Gen. Rel. Grav. **38** (2006) 159, [arχiv:physics/0311080].
[29] K.A. Bronnikov and S.A. Kononogov, Metrologia **43** (2006) R1.
[30] J.K. Webb *et al.*, Phys. Rev. Lett. **87** (2001) 091301.
[31] M.T. Murphy, J.K. Webb and V.V. Flambaum, Mon. Not. R. Astron. Soc. **345** (2003) 609.
[32] R. Quast, D. Reimers and S.A. Levshakov, Astr. Astrophys. **415** (2004) L7.
[33] R. Srianand, H. Chand, P. Petitjean and B. Aracil, Phys. Rev. Lett. **92** (2004) 121302.
[34] S.G. Karshenboim, Can. J. Phys. **78** (2001) 639.
[35] H. Marion *et al.*, Phys. Rev. Lett. **90** (2003) 150801.
[36] M. Fischer *et al.*, Phys. Rev. Lett. **92** (2004) 230802.
[37] Th. Udem *et al.*, Phys. Rev. Lett. **86** (2001) 4996.
[38] J.D. Prestage, R.L. Tjoelker and L. Maleki, Phys. Rev. Lett. **74** (1995) 3511.
[39] V.A. Dzuba, V.V. Flambaum, J.K. Webb, Phys. Rev. A **59** (1999) 230.
[40] V.A. Dzuba, V.V. Flambaum and M.V. Marchenko, Phys. Rev. A **68** (2003) 022506.
[41] S.G. Karshenboim, V.V. Flambaum, E. Peik, in *Handbook of Atomic, Molecular and Optical Physics*, ed. G. Drake, 2. edition (Springer, New York, 2005) 455-463, [arχiv:physics/0410074].
[42] E. Peik, B. Lipphardt, H. Schnatz, Chr. Tamm, S. Weyers and R. Wynands, [arχiv:physics/0611088].

Section II: Causality and Signal Propagation

superluminal signal propagation
interaction between gravitational and electromagnetic radiation

Time and Matter
2007 Bled, Slovenia

Generation and Detection of Gravitational Waves at Microwave Frequencies by Means of a Superconducting Two-body System

R.Y. CHIAO*
School of Natural Sciences and School of Engineering, University of California, P.O. Box 2039, Merced, CA 95344, U.S.A.

Abstract: The 2-body system of a superconducting sphere levitated in the magnetic field generated by a persistent current in a superconducting ring, can possibly convert gravitational waves into electromagnetic waves, and vice versa. Faraday's law of induction implies that the time-varying distance between the sphere and the ring caused by the tidal force of an incident gravitational wave induces time-varying electrical currents, which are the source of an electromagnetic wave at the same frequency as the incident gravitational wave. At sufficiently low temperatures, the internal degrees of freedom of the superconductors are frozen out because of the superconducting energy gap, and only external degrees of freedom, which are coupled to the radiation fields, remain. Hence this wave-conversion process is loss-free and therefore efficient, and by time-reversal symmetry, so is the reverse process. A Hertz-like experiment at microwave frequencies should therefore be practical to perform. This would open up observations of the gravitational-wave analog of the Cosmic Microwave Background from the extremely early Big Bang, and also communications directly through the interior of the Earth.

Consider the configuration of a superconducting sphere levitated above a superconducting ring, as shown in Figure 1 [this was suggested by Clive Rowe, personal communication]. For levitation to occur, it is required that the downwards force of Earth's gravity $\mathbf{F}_{\mathrm{GR}}(0)$

on sphere be exactly balanced by the upwards electromagnetic force $\mathbf{F}_{\mathrm{EM}}(0)$

on the sphere due to the complete expulsion of the **B** field in the Meissner effect [1], so that

$$\mathbf{F}_{\mathrm{EM}}^{(0)} = -\mathbf{F}_{\mathrm{GR}}^{(0)}$$

* rchiao@merced.edu

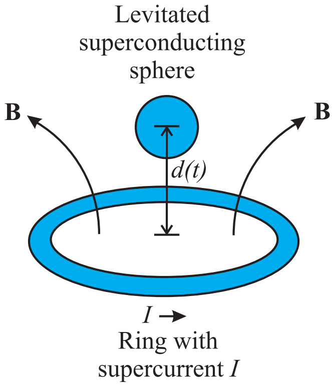


Figure 1: The DC magnetic **B** field generated by a superconducting ring with a persistent current I levitates a superconducting sphere. When a (+) polarized gravitational wave propagates radially inwards and converges upon the center of mass of this two-body system, the distance between the sphere and the ring $d(t)$ changes periodically with time, for example, at a microwave frequency, with small-amplitude excursions around an average distance of on the order of the microwave wavelength. This superconducting two-body configuration, viewed as a gravitational-wave antenna, can possibly efficiently convert this incident gravitational (GR) wave into an electromagnetic (EM) wave, and vice versa. Under the operation of time reversal, the current I and the magnetic field **B** of the ring in this Figure are reversed in direction, in order to achieve the time-reversed process of converting the EM wave back into the GR wave.

Note that the zero superscripts denote forces that are evaluated in the zeroth order of perturbation theory, i.e. in the absence of any perturbing radiation fields.

Now let a single mode of a weak, linearly polarized gravitational wave [2] at a microwave frequency (such as from the Big Bang), with one of its polarization axes along the vertical axis, be incident upon this two-body system. For simplicity, let us assume that the mass of the ring is much larger than that of the sphere, so that the reduced mass of the two-body system becomes simply that of the sphere, and the center of mass of the system approximately coincides with the center of the ring. The gravitational wave then exerts a time-varying tidal force upon the sphere relative to the ring. According to a distant inertial observer, the space between the sphere and the ring is being periodically squeezed and stretched with time, so that the distance $d(t)$ is a sinusoidal function of time.

The mechanism for the wave conversion is the following: In the presence of the DC $\mathbf{B}$ field of the superconducting ring, Faraday's law of induction implies that the time-varying distance $d(t)$ between the sphere and the ring causes induced electrical currents to circulate azimuthally around the bottom pole of the sphere at a microwave frequency, and similarly, induced countercurrents around the ring. These time-varying induced currents are a source of electromagnetic radiation at the same frequency as that of the incident gravitational radiation. Thus a gravitational (GR) wave can in principle be converted into an electromagnetic (EM) wave [3]. Lenz's law implies that there exists an instantaneous EM force $\mathbf{F}_{\mathrm{EM}}$ acting on the sphere which opposes the instantaneous GR wave tidal force $\mathbf{F}_{\mathrm{GR}}$ acting on the sphere, such that under certain circumstances (to be discussed below) quasi-static mechanical equilibrium holds, i.e.,

$$\mathbf{F}_{\mathrm{EM}} = -\mathbf{F}_{\mathrm{GR}}.$$

Note that this equation implies an equality of the magnitudes of the electromagnetic and gravitational forces on the sphere.

Let us now examine whether or not there can exist circumstances under which a complete conversion of incoming GR wave power into outgoing EM wave power in Figure 1 can in principle occur. Under these circumstances, energy conservation demands that, in terms of quantities as measured by the distant inertial observer, the power absorbed from GR wave is equal to the power emitted into EM wave,

$$\langle \mathbf{F}_{\mathrm{GR}} \cdot \mathbf{v}_{\mathrm{rad}} \rangle = -\langle \mathbf{F}_{\mathrm{EM}} \cdot \mathbf{v}_{\mathrm{rad}} \rangle \tag{1}$$

The angular brackets denote the time average over one period of the radiation fields, and $\mathbf{v}_{\mathrm{rad}}$ is the instantaneous radiation-damping velocity of the sphere undergoing rigid-body motion relative to the ring, which is parallel to, and in phase with, the instantaneous force $\mathbf{F}_{\mathrm{GR}}$, as seen by the distant observer. These circumstances can occur in a regime in which the electromagnetic radiation damping of the motion of the superconducting two-body system is the dominant damping mechanism as compared to all other loss mechanisms.

The superconducting sphere and ring can both undergo rigid-body motion in response to the gravitational wave, because the quantum adiabatic theorem [3, 4] applies to all perturbations due to any kind of radiation, whenever the frequency of these perturbations is below the BCS gap frequency [1]. Superconductors stay adiabatically, hence rigidly, in the BCS ground state of the system, because no internal excitations are allowed within the BCS energy gap [1, 3, 5].

Also, there must be little energy dissipation, such as into heat, other than that due to radiation damping or conversion. This requirement is satisfied by the strictly zero resistance of the superconductors, which follows from the fact that no dissipative excitations into heat due to ohmic processes (or phonon generation to be discussed later) are allowed within the BCS energy gap, when the frequency of the radiation is lower than the BCS gap frequency [1, 3, 5].

Thus we see that a complete, loss-free conversion from GR wave energy to EM wave energy can indeed occur, provided that all internal degrees of freedom of the superconducting two-body system are frozen at low temperatures, so that only its external degrees of freedom remain. For these external degrees of freedom, in the case when electromagnetic radiation damping is dominant, there then can exist a quasi-static mechanical equilibrium at each instant of time, such that the instantaneous forces obey the equality $\mathbf{F}_{\text{GR}} = -\mathbf{F}_{\text{EM}}$, just like in the case of a linear induction motor (or generator). Then the power in a single, incoming GR wave mode can be completely converted into the power in a single, outgoing EM wave mode.

Under the operation of time reversal, the single, outgoing EM wave mode becomes a single, incoming EM wave mode incident upon the superconducting two-body system. This EM wave mode is then back-converted into an outgoing GR wave mode. This reverse (or reciprocal) process has the same complete, loss-free conversion efficiency as the forward process by time-reversal symmetry [3, 6]. Energy conservation demands that, the power absorbed from EM wave is equal to the power emitted into GR wave,

$$\langle \mathbf{F}_{\text{EM}} \cdot \mathbf{v}_{\text{rad}} \rangle = -\langle \mathbf{F}_{\text{GR}} \cdot \mathbf{v}_{\text{rad}} \rangle \tag{2}$$

But this is satisfied by Lenz's law and the equality $\mathbf{F}_{\text{GR}} = -\mathbf{F}_{\text{EM}}$ that is also valid for the reverse process. Here the force $\mathbf{F}_{\text{EM}}$ is squeezing and stretching periodically the space between the sphere and the ring, so that the sphere is acting upon space, not moving through space, in a time-reversal of the forward process.

Under the action of the gravitational wave, the space between the sphere and the ring is undergoing small amplitude, anisotropic, periodic time-varying strains h_{ij} of the metric tensor [2], as if the space were an elastic medium. According to the Equivalence Principle [2], in both the forward and the reverse processes, the sphere is not accelerating at all with respect to a tiny, local inertial observer located at its center, so that the sphere is not moving, i.e., accelerating, through space. Hence the objection that its inertia limits the amount of emitted GR radiation to extremely small values is invalid; see section 12 of [3]a. Similarly, the objection that the motion of the sphere through a flat-space background must be relativistic (i.e., with velocities close to c) in order to generate any appreciable amount of GR waves is also invalid.

Therefore we propose here to perform a Hertz-like experiment at 12 GHz, in which the EM-to-GR wave-conversion process becomes the source of GR waves, and the GR-to-EM wave-conversion process becomes the receiver of GR waves. Faraday cages consisting of normal metals at room temperature (the metallic casings of the two well-separated liquid helium storage dewars to be described below) prevent the transmission of EM waves, so that only GR waves, which can easily pass through all ordinary, classical matter, such as the normal (i.e., dissipative) metals of which standard, room-temperature Faraday cages are composed, are transmitted between the two halves of the apparatus that serve as the source and the receiver, respectively [3]. Such an experiment should be practical to perform using standard microwave sources and receivers, since the scattering cross-sections and the wave conversion efficiencies of the two-body superconducting system such as the one depicted in Figure 1, when viewed as a gravitational-wave antenna, should be large enough to be experimentally interesting [3]; see also below].

We plan to use lead (Pb) as the type I superconducting material for both the sphere and the ring. The sphere could consist of a light, low-density material (e.g., Styrofoam) coated on its surface with a Pb film which is much thicker than the penetration depth [5]. The ring could consist of a flat copper gasket electroplated with a thick coating of Pb. The central hole of the gasket could have a diameter of 34 mm, which should levitate the superconducting sphere above the hole of the ring at a distance of around 6 mm, i.e., approximately a quarter of a microwave wavelength at 12 GHz where we plan to work (12 GHz is over an order of magnitude below the BCS gap frequency of Pb, so that the losses at this frequency are negligible [3]). A persistent current in the ring sufficient to levitate the sphere could be induced by means of a nearby permanent magnet, which would then be removed after the ring has been cooled to become superconducting below the transition temperature of Pb at 7.2 K [5]. We plan also to experiment with other two-body geometries, such as with a pair of superconducting coaxial rings with opposing trapped magnetic flux [suggested by S. Minter], as realizations of the superconducting two-body system. We have chosen to use a type I superconductor instead of a type II superconductor, in order to avoid microwave losses arising from the Abrikosov vortex-motion degrees of freedom.

We propose to perform the first experiments in a standard 4 K liquid helium storage dewar using a dewar insert, which consists of a 50 mm diameter cylindrical structure with a detachable sample holder that is capable of containing a specific sphere-ring configuration (or other superconducting two-body geometries) at its bottom end. Such a dewar insert could be inserted directly into the liquid helium inside a storage dewar, in order to explore the parameters for the levitation of the sphere by the ring, by adjusting the dimensions of the sphere-ring configuration. Optical access and viewing could be made possible by means of glass prisms placed inside the sample can near the ring. Alternatively, a Hall probe sensor could detect the onset of the Meissner effect and levitation of the sphere. These dewar-insert experiments would enable us to test the computer simulations we are presently conducting for the levitation of type I superconducting spheres by various magnetic field configurations.

Two of these dewar inserts, each with a sphere-ring configuration (or some other superconducting two-body configuration) immersed deep inside two well separated 4 K liquid helium dewars, with the associated microwave electronics, would then allow us to perform the Hertz-like experiment. In preparation for this experiment, we plan to determine at room temperature the microwave antenna design parameters of the sphere-ring and the two-coaxial-ring configurations (and other possible configurations) using a network analyzer. We plan to use an appropriately designed magnetic loop antenna to couple to the TE_{01} microwave cylindrical mode of a superconducting ring for transferring energy from the GR wave mode to the EM wave mode, and vice versa.

We have already built and tested microwave sources and receivers at 12 GHz [6]. We have used a frequency-doubled 6 GHz microwave-cavity oscillator as the source, with an output power level at 12 GHz of around 10 mW, and have measured that a standard, commercial satellite-dish KU band receiver [Precision model PMJ-LNB KU gold label series from Astrotel Communications Corp.] has a noise figure of 0.6 dB at 12 GHz, so that our receiver system Noise-Equivalent-Power (NEP) sensitivity will be on the order of 10^{-25} W/Hz$^{1/2}$. This will allow us to achieve a very good signal-to-noise ratio for the Hertz-like experiment [3], assuming that our estimates for the cross-sections and wave-conversion efficiencies for the sphere-ring system are of the correct orders of magnitude.

If we should be successful in the Hertz-like experiment, one of the first tests to see if we truly have GR rather than EM coupling between the transmitter (or source) and receiver (or detector) halves of the apparatus, is to tilt the transmitter part of the apparatus by $+22.5°$ with respect to the vertical around the line of sight joining the transmitter and receiver, and to tilt the receiver part of the apparatus by $-22.5°$ with respect to the line of sight. The signal should be extinguished at the resulting $45°$ relative orientation between the two halves of the apparatus, and not at $90°$, as would the case for EM waves. This would be a clear signature that we have successfully generated and detected GR waves rather than EM waves [this was suggested by Kirk Wegter-McNelly, personal communication].

Another of our initial goals if we should be successful in detecting a signal in the Hertz-like experiment would be to directly measure the speed of gravitational waves for the first time. This can be done either by changing the distance between the two experimental dewars containing the two sphere-ring (or other) configurations, and then measuring the resulting change in phase of the received signal relative to that of the source signal, or by modulating the carrier signal, and measuring the delay in the arrival time of the demodulated signal. We will thereby be able to directly check experimentally in the laboratory for the first time the theoretical prediction by Einstein that the speed of these waves is identical to that of light in vacuum $c = 3.00 \cdot 10^8$ m/s [2] , to within $\pm 1\%$ accuracy.

It may be objected that sound waves will be generated as a dissipative mechanism in the proposed experiment. However, sound waves have a speed that is typically five orders of magnitude smaller than c. If Einstein were correct, then the typical generated sound wavelength of 250 nm or smaller would be much less than the gravitational wavelength of 25 mm, so that the typical sound wave would be badly mismatched to the incident GR wave. The generation of sound waves (or phonons) within the superconductor at microwave frequencies inside the BCS gap by the incident gravitational wave would therefore be forbidden due to the Mössbauer-like (or zero-phonon) effect described in section 9 of [3]a; see also [3]b. Hence the sphere and the ring would undergo rigid-body motion in response to the incident gravitational wave.

It may also be objected that the microwave-frequency electrical currents induced through Faraday's law in the GR to EM wave conversion process may be too feeble and too difficult to detect. However, we are not proposing to detect these currents, but rather to detect the power transmitted from the GR wave to the EM wave, and vice versa. With an antenna design which takes into account impedance-matching and cross-section considerations properly, the power transmitted from the source to the receiver should be readily measurable.

To better understand the cross-section estimate for the above proposed experiment, it is useful to introduce a weak-field representation of the linearized Einstein equations for the gravito-electric field $\mathbf{E}_G \equiv \mathbf{g}$ (i.e., the gravitational acceleration of a test mass), and for the gravito-magnetic field $\mathbf{B}_G$ (i.e., the far-field, time-varying analog of the Lense-Thirring field), as observed in the coordinate system of a distant inertial observer. This representation takes the form of the Maxwell-like equations [7]

$$\nabla \cdot \mathbf{E}_G = -\frac{\rho_G}{\varepsilon_G} \tag{3}$$

$$\nabla \times \mathbf{E}_G = -\frac{\partial \mathbf{B}_G}{\partial t} \tag{4}$$

$$\nabla \cdot \mathbf{B}_G = 0 \tag{5}$$

$$\nabla \times \mathbf{B}_G = \mu_G \left(-\mathbf{j}_G + \varepsilon_G \frac{\partial \mathbf{E}_G}{\partial t} \right) \tag{6}$$

where ρ_G is the mass density and $\mathbf{j}_G$ is the mass current density of slowly-moving matter in the source, and the gravitational analog of the electric permittivity of free space ε_0 is

$$\varepsilon_G = \frac{1}{4\pi G} = 1.19{\cdot}10^9 \text{ units} \tag{7}$$

where G is Newton's constant, and the gravitational analog of the magnetic permeability μ_0 of free space is

$$\mu_G = \frac{4\pi G}{c^2} = 9.31{\cdot}10^{-27} \text{ units} \tag{8}$$

where c is the speed of light.

The fields $\mathbf{E}_G \equiv \mathbf{g}$ and $\mathbf{B}_G$ are measurable quantities that obey the gravitational analog of the Lorentz force law [7]

$$\mathbf{F} = m(\mathbf{E}_G + 4\mathbf{v} \times \mathbf{B}_G) \tag{9}$$

where m is the mass of a test particle, and $\mathbf{v}$ is its velocity, as seen by the distant inertial observer.

These Maxwell-like equations are linear, so that the fields obey the superposition principle not only in the vacuum outside of the source, but also in the matter inside the source, provided that the field strengths are sufficiently weak and the matter is sufficiently slowly moving, so that there exists a regime of a linear response of the matter to the applied fields. The resulting optics for gravitational waves is therefore linear, just like the linear optics for electromagnetic waves.

Therefore any argument involving the characteristic power scale [2]

$$P_0 = \frac{c^5}{G} = 3.6 \cdot 10^{52}\,\mathrm{W} \tag{10}$$

is irrelevant in the context of linear optical devices such as linear mirrors, since it should always be possible to reduce the intensity of the incident GR waves so that there exists a linear regime of response of such devices to these sufficiently weak GR wave amplitudes. Then there cannot exist any characteristic scale of power in any linear response of such devices, such as that given by Equation (10).

Also, it is important to emphasize that the matter generating the GW waves in such linear devices as the two-body superconducting system described above, needs not be moving relativistically with velocities close to the speed of light c relative to each other, in order to generate any appreciable amount of GR waves. By the time-reversal symmetry argument given above, it is clear that the generation of GR waves by the two-body superconducting system is not due to the matter moving through space, but rather is due to the matter acting upon space directly. Hence it is irrelevant whether such matter is moving close to c through space or not.

In the case of the vacuum, where ρ_G and $\mathbf{j}_G$ both vanish, these equations lead to wave propagation at a wave speed exactly equal to the speed of light

$$c = \frac{1}{\sqrt{\varepsilon_G \mu_G}} = 3 \cdot 10^8\,\text{SI units.} \tag{11}$$

A plane wave solution of the Maxwell-like equations possesses the gravitational characteristic impedance of free space given by [8, 3]

$$Z_G = \sqrt{\frac{\mu_G}{\varepsilon_G}} = 2.79 \cdot 10^{-18}\,\text{SI units,} \tag{12}$$

which is the analog of the electromagnetic characteristic impedance of free space

$$Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} = 377\,\text{Ohms.} \tag{13}$$

The gravitational characteristic impedance of free space Z_G, like Z_0, plays a central role in all radiation problems, such as in a comparison of the radiation resistance of gravitational-wave antennas to the value of this impedance, in order to estimate the coupling efficiency of these antennas to free space. The numerical value of Z_G is extremely small, but the impedance of all material objects must be "impedance matched" to this extremely small quantity before significant power can be transferred efficiently from gravitational waves to these objects, or vice versa.

In contrast to the electromagnetic-wave case, in the gravitational-wave case, all ordinary, classical matter, such as Weber bars, possesses impedances much larger than that of the gravitational characteristic impedance of free space Z_G. It is therefore extremely difficult to impedance-match gravitational waves to any ordinary, classical matter. As a consequence, it is a general rule that all ordinary, classical matter, such as a Weber bar, is essentially completely transparent to these waves.

However, phase-coherent, loss-free quantum matter, such as superconductors, which can possess strictly zero dissipation due to the presence of the BCS energy gap, can be exceptions to this general rule. Experimental evidence for this "quantum dissipationlessness" is the fact that persistent currents in annular superconducting rings have been observed to last much longer than the age of the Universe [1]. One important consequence of the zero-dissipation property of a superconductor is that a mirror-like reflection of sufficiently low-frequency gravitational waves can occur at a planar interface between the vacuum and the superconductor. Note that by "low frequency" we mean frequencies much less than the BCS gap frequency.

In the electromagnetic case, the reflection coefficient R of a wave being transmitted down a transmission line with impedance Z, which is terminated by a resistance R which vanishes like that of a superconductor, is given by

$$R_G = \left| \frac{Z - R}{Z + R} \right|^2 \to 100\% \quad \text{as} \quad R \to 0. \tag{14}$$

This implies that a mirror-like reflection of the wave occurs from a superconductor, when it is used as a short-circuit termination of the transmission line.

Similarly, the reflection coefficient RG of a sufficiently low-frequency gravitational wave from a superconductor-vacuum interface is given by

$$R_G = \left| \frac{Z_G - R_G}{Z_G + R_G} \right|^2 \to 100\% \quad \text{as} \quad R_G \to 0. \tag{15}$$

where the gravitational characteristic impedance of free space Z_G is given by Equation (12), and R_G is the gravitational analog of the resistance R of the superconductor. This implies that a mirror-like reflection of a sufficiently low-frequency gravitational wave could in principle occur from the surface of the superconductor.

Therefore mirrors for gravitational waves can in principle exist. Curved, parabolic mirrors can focus these waves, and Newtonian telescopes for gravitational radiation can in principle be constructed. In the case of scattering of gravitational waves from superconducting bodies, the above mirror-like-reflection condition implies hard-wall boundary conditions at the surfaces of these superconducting bodies, so that the scattering cross-section of these waves from large superconducting bodies can in principle be geometric, e.g., hard-sphere, in size. For example, for a superconductor in the form of a sphere of radius which is much larger than the wavelength, the cross-section for GR wave scattering by the sphere is given by

$$\sigma_{\text{hardsphere}} = 2\pi a^2, \tag{16}$$

where a is the radius of the superconductor.

It may be objected that any kind of reflection of gravitational waves from matter, including any kind of partial reflection, would constitute a kind of "anti-gravity" effect. Now it is certainly true that only positive masses exist in nature, and therefore that the screening of longitudinal gravito-electric fields, like the Earth's gravitational field, including its partial screening by any kind of matter, including superconductors as has been falsely claimed in the so-called "Podkletnov effect", is impossible. However, both positive and negative mass currents can exist in nature, and therefore the screening of transverse gravito-magnetic fields, like those of a gravitational wave, including the partial screening by the reflection of these waves from matter, should indeed be possible. The strength of the reflection can in principle depend on the details of the nature of the matter.

A closely related objection is that, according to one interpretation of the Equivalence Principle, the response of all kinds of matter, whether classical or quantum mechanical in nature, should be universally the same to all gravitational fields. Locally, all dropped objects undergo exactly the same free-fall motion, which is independent of their mass, composition, or thermodynamic state. Specifically, the response of all matter to gravitational fields, such as the Earth's, is independent of whether the thermodynamic state of the matter is classical or quantum mechanical in nature.

Thus this interpretation of the Equivalence Principle would forbid any difference between the linear response of incoherent classical matter, such as Weber bars, and the linear response of coherent quantum mechanical matter, such as superconductors, to any kind of gravitational field. However, whilst this interpretation is valid for the local response of any kind of matter, whether coherent or incoherent, to all gravito-electric fields, like that of all dropped, freely-falling objects in their response to the gravitational field

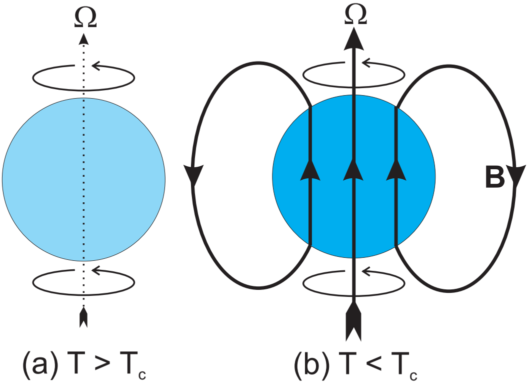


Figure 2: Comparison of a rotating normal metal with a rotating superconductor. In (a), the normal metal is a neutral sphere consisting of a rotating superconducting metal at a temperature above its transition temperature T_c. In (b), this body is slowly cooled down below T_c, whilst it is rotating. Now there is the onset of a magnetic field **B** proportional to its rate of rotation Ω with respect to the distant "fixed stars".

of the Earth, it is not valid for the nonlocal response of matter, whether coherent or incoherent, to gravito-magnetic fields, such as to the Lense-Thirring field arising from the distant matter of the Universe (i.e., from the distant "fixed stars").

Empirical evidence for this latter fact lies in the difference between the response of normal metals and the response of superconductors which are rotating with respect to the distant "fixed stars" [9]. Consider an experiment in which there is a spherical body rotating at a fixed angular speed Ω, which consists of a normal metal that is a superconductor above its transition temperature T_c. For all temperatures $T > T_c$, no magnetic field is produced by this neutral, rotating body.

Now consider what happens when this rotating body is slowly cooled, whilst it is undergoing rotation with respect to the distant "fixed stars," down below its transition temperature T_c. Now for all temperatures $T < T_c$, there exists a magnetic field B produced by this rotating superconducting body in the "London moment" effect. The London moment arises from the constructive quantum interference of the Cooper pairs of electrons in the superconductor near its surface after one round trip around the perimeter of the body, as seen by the distant inertial observer. This quantum interference effect implies that the Cooper pairs everywhere near the surface of the body in its lowest energy state will come to a complete halt with respect to the distant "fixed stars". The nuclei near the surface, however, continue to rotate with respect to the distant "fixed stars". Thus this differential motion of the Cooper pairs with respect to the nuclei near the surface produces a surface electrical current. It is this current that produces the magnetic field B which is sketched in Figure 2(b).

However, according to the above interpretation of the Equivalence Principle, there cannot be any difference in the response of the rotating body above or below its transition temperature to any kind of gravitational field, including the Lense-Thirring field which arises from the distant "fixed stars". Specifically, the response of the rotating body to the Lense-Thirring field arising from these distant "fixed stars" should be independent of the thermodynamic state of this body, and in particular, it should be independent of whether the rotating body is composed of incoherent, classical matter, or coherent, quantum mechanical matter. This interpretation of the Equivalence Principle is contradicted by experiments that demonstrate the existence of the London moment [9].

Let us now generalize the above considerations to the case of the time-varying, transverse fields of gravitational radiation, when the frequency of the radiation is much lower than the BCS gap frequency, so that the quantum adiabatic theorem holds. Let us first consider the case of a normal metal, which consists of a superconducting metal above its transition temperature T_c. Let this neutral metallic body have the shape of a planar slab with a thickness of half a wavelength of the incident gravitational plane wave, which is propagating towards the slab at normal incidence along the x axis from the left, as depicted in Figure 3. The entrance face of the slab is at $x = 0$, and the exit face of the slab is at $x = \lambda/2$.

The linear response of a normal electron above the superconducting transition temperature Tc near the surface of the metal to the tidal field g of the incident wave is represented by the local, instantaneous velocity vector v, which would be the free-fall velocity of the normal electron, assuming the absence of any dissipation in the metal. (All measurable quantities are those that are being measured by a distant inertial observer represented by the eye in Figure 3). Since the normal electrons undergo local free fall together with the nearby nuclei (neglecting for the moment the weak restoring forces arising from the ionic lattice), no electrical currents are produced in the linear response of this neutral body to the local, instantaneous accelerations due to gravity g arising from tidal forces of the gravitational wave.

Let us now choose two rectangular loops, one above the center of mass (c.m.) of the system, and one below it, as shown in Figure 3, in order to

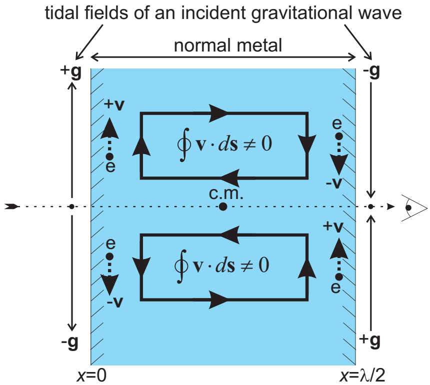


Figure 3: A slab of normal metal half a wavelength thick, which is a superconducting metal above its transition temperature T_c, in its linear response to the tidal fields of a normally incident gravitational plane wave. The center of mass of the slab is represented by the dot below the label "c.m." The eye on the right represents a distant inertial observer. However, below T_c, the upper and lower electrons form a Cooper pair in a Bohm singlet state, that is an entangled state which violates Bell's theorem and is nonlocal. The electrons local free-fall motion is thus suppressed.

evaluate the circulation of the normal electrons around these loops in their linear response to the incident wave. It is obvious upon inspection of Figure 3 that these two closed loop integrals of the velocity v of the normal electron, as seen by the distant inertial observer, obey the inequality

$$\oint \mathbf{v} \cdot \mathrm{d}\mathbf{s} \neq 0, \tag{17}$$

so that the circulation of the normal electrons around these loops does not vanish. This is because there is a reversal of the sign of the tidal forces as the wave propagates from the entrance face of the slab at $x = 0$ to the exit face at $x = \lambda/2$, and therefore there is also a reversal of the sign of the normal electron velocity in their linear response to these tidal forces.

Next, let us slowly cool this system down below the superconducting transition temperature T_c whilst the incident gravitational wave is still present. The normal electrons now become Cooper pairs, which possess quantum phase coherence over long distances. If the Cooper pairs near the surface of the metal in Figure 3 were to undergo free fall with exactly the same instantaneous, local velocity $\mathbf{v}$ in their linear response to the tidal fields g of the incident gravitational wave, as the instantaneous, local free-fall velocity $\mathbf{v}$ of the normal electrons in the normal state of the metal (as would be demanded by the above erroneous interpretation of the Equivalence Principle), then their macroscopic wavefunction (or the complex order parameter of Ginzburg and Landau) would no longer by single valued after one round trip around the closed loop, and the constructive quantum interference for these Cooper pairs after a round trip would be impossible. This is because the instantaneous velocity $\mathbf{v}$ is related to the phase ϕ of the Cooper pair macroscopic wavefunction (or of the complex order parameter) by

$$\mathbf{v} = \frac{\hbar}{m} \nabla \phi, \tag{18}$$

where $\hbar$ is Planck's constant$/2\pi$ and m is the mass of a Cooper pair.

However, experiments verifying the existence of the London moment [9] demonstrate that the Cooper pairs do in fact always possess a single-valued macroscopic wavefunction (or complex order parameter), as determined by the observations in the inertial frame of the distant observer, i.e., that of the distant "fixed stars". Hence it must be the case that below the temperature T_c, the circulation of the Cooper pairs vanishes identically, i.e.,

$$\oint \mathbf{v} \cdot d\mathbf{s} = 0, \tag{19}$$

in the case of the lowest energy state of the Cooper pairs. This implies that their local, instantaneous velocity vanishes everywhere along the left and the right vertical segments at $x = 0$ and $x = \lambda/2$ of the arbitrarily chosen closed loops in Figure 3, so that in the limit of arbitrarily skinny loops, it follows that

$$\mathbf{v} = 0 \tag{20}$$

everywhere near the surface of the superconductor. Therefore it follows that the time-dependent, transverse gravito-electric field

$$\mathbf{E}_G \equiv \mathbf{g} = \left(\frac{\partial \mathbf{v}}{\partial t} \right)_{\text{near surface}} = 0 \tag{21}$$

also vanishes everywhere near the surface of the superconductor. This also implies hard-wall boundary conditions that lead to a mirror-like reflection of gravitational waves. (See also footnotes 11 and 12 of [3a].)

The above conclusions also follow from the fact that the entire superconducting slab stays adiabatically in the BCS ground state of the system with a single value of the global phase ϕ of the macroscopic wavefunction (or of the complex order parameter) inside the entire superconducting body, i.e.,

$$\phi = \text{constant everywhere.} \tag{22}$$

This is true even in the presence of the weak, local perturbations due to the incident gravitational wave, provided that the wave frequency is well below the BCS gap frequency, so that the wave cannot cause any quantum transitions out of the BCS ground state.

The Cooper pairs in the BCS ground state are in entangled states of momentum and spin, which are nonlocal, so it is impossible to know whether it is the upper or the lower electron in Figure 3 that is moving upwards or downwards in response to the tidal g fields. The large number of these phase-coherent electrons in the system (on the order of Avogadro's number) in the case of superconductors, compensates for the usual weakness of the interaction between gravitation radiation and matter, and leads to a superradiant, mirror-like reflection [3b].

Therefore, just as in the case of the London moment below the transition temperature T_c, the Cooper pairs everywhere near the entrance and exit faces of the superconducting slab must come to a complete halt with respect to the distant inertial observer (i.e., with respect to the distant "fixed stars") below Tc. The vanishing of v everywhere near the surfaces of the slab must be independent of the size of the small amplitude of a sufficiently weak gravitational wave, in the regime of the linear response of the superconductor to the wave. This is the behavior of an extremely rigid material in its linear response to the gravitational wave, so that, once again, one concludes that a mirror-like reflection of a sufficiently weak incident gravitational wave should occur at the planar superconductor-vacuum interface. The resulting scattering cross-section for large superconducting bodies should therefore once again be of the order of magnitude of the geometric cross-section given by Equation (16). These counter-intuitive predictions will be tested in the above proposed experiment.

It is to be emphasized that throughout the above discussion, all measurable quantities in the above equations are those which are being observed and measured by the distant inertial observer. The coordinate system being used in these measurements is the one set up by means of light signals sent to and from this distant observer, and the time coordinate being used is the one set up by means of this observer's clock [11].

This experiment could lead to important applications in science and engineering. In science, it would open up the possibility of gravitational-wave astronomy at microwave frequencies. One important problem to explore would be observations of the analog of the Cosmic Microwave Background (CMB) in gravitational radiation. Since the Universe is much more transparent to gravitational waves than to electromagnetic waves, such observations would allow a much more penetrating look into the extremely early Big Bang towards the Planck scale of time, than the presently well-studied CMB. Different cosmological models of the very early Universe give widely differing predictions of the spectrum of this penetrating radiation, so that by measurements of the spectrum, one could tell which model, if any, is close to the truth [10]. The anisotropy in this radiation would also be very important to observe.

In engineering, it could open up the possibility of intercontinental communications by means of microwave-frequency gravitational waves directly through the interior of the Earth, which is transparent to such waves. This would eliminate the need of communications satellites, and would allow an economical means of communication with people deep underground or underwater in submarines in the oceans. Such a new direction of gravitational-wave engineering could aptly be called "gravity radio" [3].

Acknowledgments

This work was supported in part by the CTNS-STARS program. I would like to thank Prof. Thomas Kibble for his helpful comments.

References

[1] M. Tinkham, *Introduction to Superconductivity*, 2nd edition (Dover Books on Physics, New York, 2004).

[2] C.W. Misner, K.S. Thorne, J.A. Wheeler, *Gravitation* (Freeman, San Francisco, 1972).

[3] R.Y. Chiao, *New directions for gravity-wave physics via Millikan oil drops*, to be published in Visions of Discovery (a Volume in honor of C.H. Townes) edited by R.Y. Chiao, W.D. Phillips, A.J. Leggett, M.L. Cohen and C.L. Harper Jr. (Cambridge University Press, Cambridge, 2007), [arχiv:gr-qc/0610146v16]; R.Y. Chiao, *The Interface between Quantum Mechanics and General Relativity*, Lamb Medal Lecture, J. Mod. Optics **53** (2007) 2349, [arχiv:quant-ph/0601193v7].

[4] L. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1955).

[5] C. Kittel, *Introduction to Solid State Physics*, 8th edition (John Wiley & Sons, New York, 2005).

[6] R.Y. Chiao and W.J. Fitelson, *Time and Matter in the Interaction between Gravity and Quantum Fluids: Are There Macroscopic Quantum Transducers between Gravitational and Electromagnetic Waves?*, Proceedings of the Time and Matter Conference, Venice, Italy, 11-17 August 2002, eds. Ikaros Bigi and Martin Faessler (World Scientific, Singapore, 2006), p. 85, [arχiv:gr-qc/0303089].

[7] R.M. Wald, *General Relativity* (University of Chicago Press, Chicago, 1984), section 4.4, see also footnote 6 of [3a].

[8] C. Kiefer and C. Weber, *On the Interaction of Mesoscopic Quantum Systems with Gravity*, Annalen der Physik (Leipzig) **14** (2005) 253.
[9] J. Tate, B. Cabrera, S.B. Felch and J.T. Anderson, *Precise Determination of the Cooper-Pair Mass*, Phys. Rev. Lett. **62** (1989) 845; *Determination of the Cooper-Pair Mass in Niobium*, Phys. Rev. B **42** (1990) 7885.
[10] R.Y. Chiao, *Proposed Observations of Gravity Waves from the Early Universe via "Millikan Oil Drops"*, Proceedings of the NASA Conference Quantum to Cosmos, [arχiv:gr-qc/0606118], to be published in Int. J. Mod. Phys. D.
[11] J.L. Synge, *Relativity: The General Theory* (North-Holland Publishing Company, Amsterdam, 1960).

TIME AND MATTER 2007

Time and Matter
2007 Bled, Slovenia

Search for Frame-Dragging-Like Signals Close to Spinning Superconductors

M. TAJMAR*, F. PLESESCU, B. SEIFERT, R. SCHNITZER
AND I. VASILJEVICH
Space Propulsion and Advanced Concepts, Austrian Research Centers GmbH – ARC, A-2444 Seibersdorf, Austria

Abstract: High-resolution accelerometer and laser gyroscope measurements were performed in the vicinity of spinning rings at cryogenic temperatures. After passing a critical temperature, which does not coincide with the material's superconducting temperature, the angular acceleration and angular velocity applied to the rotating ring could be seen on the sensors although they are mechanically decoupled. A parity violation was observed for the laser gyroscope measurements such that the effect was greatly pronounced in the clockwise-direction only. The experiments seem to compare well with recent independent tests obtained by the Canterbury Ring Laser Group and the Gravity-Probe B satellite. All systematic effects analyzed so far are at least 3 orders of magnitude below the observed phenomenon. The available experimental data indicates that the fields scale similar to classical frame-dragging fields. A number of theories that predicted large frame-dragging fields around spinning superconductors can be ruled out by up to 4 orders of magnitude.

Introduction

Gravity is the weakest of all four fundamental forces; its strength is astonishingly 40 orders of magnitude smaller compared to electromagnetism. Since Einstein's general relativity theory from 1915, we know that gravity is not only responsible for the attraction between masses but that it is also linked to a number of other effects such as bending of light or slowing down of clocks in the vicinity of large masses. One particularly interesting aspect of gravity is the so-called Thirring-Lense or Frame-Dragging effect: A rotating mass should drag space-time around it, affecting for example

* martin.tajmar@arcs.ac.at

the orbit of satellites around the Earth. However, the effect is so small that it required the analysis of 11 years of LAGEOS satellite orbit data to confirm Einstein's prediction within $\pm 10\%$ [3]. Presently, NASA's Gravity-Probe B satellite is aiming at measuring the Thirring-Lense effect of the Earth to an accuracy better than 1% [7]. Therefore, apart from Newton's mass attraction, gravitational effects are believed to be only accessible via astronomy or satellite experiments but not in a laboratory environment.

That assumption was recently challenged [19, 20, 5, 22], proposing that a large frame-dragging field could be responsible for a reported anomaly of the Cooper-pair mass found in Niobium superconductors. A spinning superconductor produces a magnetic field as the Cooper-pairs lag behind the lattice within the penetration depth. This magnetic field, also called London moment, only depends on the angular velocity of the superconductor and the mass-to-charge ratio of the Cooper-pairs. By measuring precisely the angular velocity and the magnetic field, it is possible to derive the Cooper-pair mass with a very high precision, knowing that it consists of two elementary charges. The most accurate measurement to date was performed at Stanford in 1989 using a Niobium superconductor with a very surprising result: the Cooper-pair mass derived was larger by 84 ppm compared to twice the free-electron mass [25, 26]. The theoretical analysis predicted a Cooper-pair mass 8 ppm smaller than the free-electron mass including relativistic corrections. Several efforts were published in the last 15 years that investigated additional correction factors [16, 12]. However, the discrepancy between measurement and theoretical prediction remained unsolved.

Within the classical framework, frame-dragging is independent of the state (normal or coherent) of the test mass. Over the last years, several theoretical approaches were developed that propose significantly amplified non-classical frame-dragging fields for superconductors with respect to normal matter [23, 1, 6, 4]. Since 2003, an experimental program was established at the Austrian Research Centers (ARC) to search for such frame-dragging anomalies in the vicinity of spinning masses down to cryogenic temperatures using laser gyroscopes and accelerometers [21, 24]. This paper will give an overview of our experimental setup and the results obtained so far and will compare it with other recent tests performed by the Canterbury Ring Laser Group and the results from the Gravity Probe-B satellite.

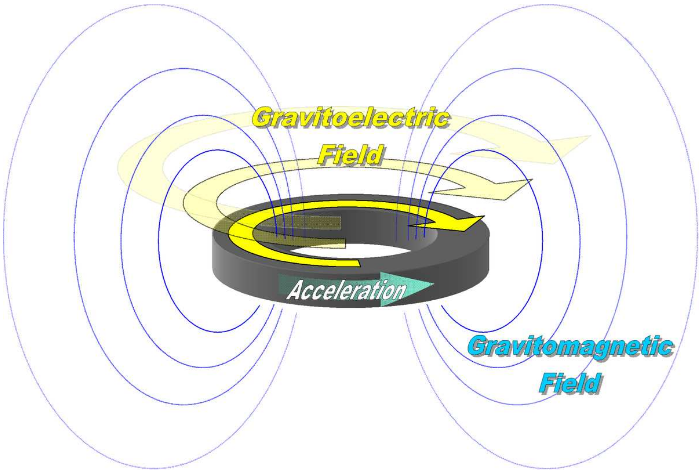
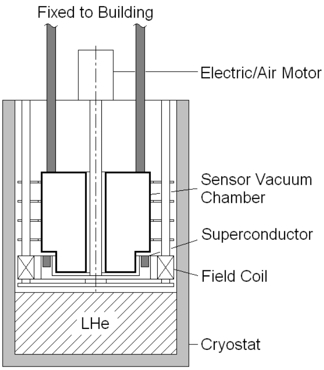


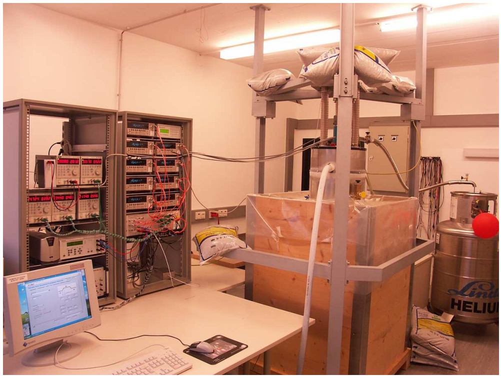
Figure 1: Gravitomagnetic and gravitoelectric field generated by a rotating and angularly accelerated superconductor.

Experimental setup

The core of our setup is a rotating ring inside a large cryostat. Several rings have been used so far including Niobium and Aluminum to test a classical low-temperature superconductor as well as a non-superconductor material for reference purposes (outer diameter of 150 mm, wall thickness of 6 mm and a height of 15 mm), and a YBCO high-temperature superconductor (outer diameter of 160 mm and wall thickness of 15 mm). The ring can be rotated using a brushless servo motor or a pneumatic air motor to minimize any electromagnetic influence. According to the theoretical concepts, a frame-dragging-like field should be produced directly proportional to the superconductor's angular velocity. Another aspect of Einstein's theory is that a time-varying frame-dragging field should give rise to non-Newtonian gravitational fields, also called accelerational frame-dragging. Therefore, any angular acceleration of the superconductor should produce a gravitational field along the ring's surface. A short illustration of the expected fields around the rotating superconductor is shown in Fig. 1. Laser gyroscopes and low-noise accelerometers can be used to detect those frame-dragging fields if they are rigidly fixed to avoid any mechanical movement.



(a) Schematic setup.



(b) Facility in the laboratory at the Austrian Research Centers.

Figure 2: Experimental setup.

The experimental setup is illustrated in Fig. 2. The motor and the superconductor assembly are mounted on top and inside a liquid helium cryostat respectively, which is stabilized in a 1.5 t box of sand to damp mechanical vibrations induced from the rotating superconductor. The accelerometers and gyros are mounted inside an evacuated chamber made out of stainless steel, which acts as a Faraday cage and is directly connected by three solid shafts to a large structure made out of steel that is fixed to the building floor and the ceiling. The sensors inside this chamber are thermally isolated from the cryogenic environment due to the evacuation of the sensor chamber and additional MLI isolation covering the inside chamber walls. Only flexible tubes along the shafts and electric wires from the sensor chamber to the upper flange establish a weak mechanical link between the sensor chamber and the cryostat. This system enables a very good mechanical de-coupling of the cryostat with the rotating superconductor and the sensors even at high rotational speeds. A minimum distance of at least 5 mm is maintained between the sensor chamber and all rotating parts such as the motor axis or the rotating sample holder. In order to obtain a reliable temperature measurement, a calibrated silicon diode (DT-670B-SD from Lakeshore) was installed directly inside each rotating ring. A miniature collector ring on top of the motor shaft enabled the correct readout even during high speed rotation. Two temperature fixpoints enabled a temperature calibration during each run: the liquid helium temperature of 4.2 K and the evaluation of the critical temperature of the superconductor using the field coil. When the superconductor was cooled down, the field coil was switched on with a field below the critical field strength from the superconductors used. A Honeywell SS495A1 solid state Hall-sensor was installed inside the sensor chamber. Initially, the superconductor acted as a magnetic shield. But when the superconductor passed T_c, the magnetic field from the coil was recorded on the Hall sensor. At this point in time, the temperature read-out from the silicon diode must then correspond to the critical temperature.

The sample rings are glued inside an aluminum sample holder using STYCAST cryogenic epoxy. The bottom plate of the sample holder is made out of stainless steel as well as the rotating axis. In general, only non-magnetic materials were used throughout the facility, only the bearings are partly made out of steel with a non-negligible magnetic permeability.

The sensor-chamber can be equipped with accelerometers, gyros, temperature sensors and highly sensitive magnetic field sensors (based on the Honeywell HMC 1001 with 0.1 nT resolution). All sensors are mounted on the same rigid mechanical structure fixed to the upper flange. Each sensor level (In-Ring, Above-Ring and Reference) is temperature controlled to 25 °C to obtain a high sensor bias stability. Trade-offs between different sensors can be found in [21].

Accelerator measurements

After a survey on commercially available accelerometers, we decided to use the Colibrys Si-Flex SF1500S due to its low noise of $300\,\mathrm{ng}/\sqrt{\mathrm{Hz}}$, small size and low sensitivity to magnetic fields which we evaluated as $5\cdot 10^{-4}\,\mathrm{g/T}$ (expressing the acceleration in the unit g of the Earth's standard acceleration). Using only the air motor, the magnetic fields inside the sensor chamber were always below $1\,\mu\mathrm{T}$ at maximum speed (originating from the rotating bearings).

Therefore any magnetic influence is less than 1 ng and thus well below the sensors noise level. The biggest systematic effect found is the so-called vibration rectification, a well known effect common to all MEMS accelerometers [2]. Due to nonlinearities in the response of the pendulum, an anomalous DC offset appears when the sensor is exposed to vibration although the time average of the vibration of zero. Such vibrations are present due to the acoustic noise from the bearings as well as from the helium evaporation. Fortunately, the vibration rectification always leads to negative DC offsets independent of the angular speed orientation. By alternating between clockwise and counter-clockwise rotations and subtracting both signals, the anomalous DC offset can be eliminated and any real signal remains. A further improvement is to mount several sensors along the rings surface ("curl configuration") as shown in Fig. 3. This allows to further reduce sensor offsets and to increase the number of measurements. The accelerometer setup is described in more detail in [21]. The accelerometers were read out using Keithley 2182 Nanovoltmeters with a measurement rate of 10 Hz.

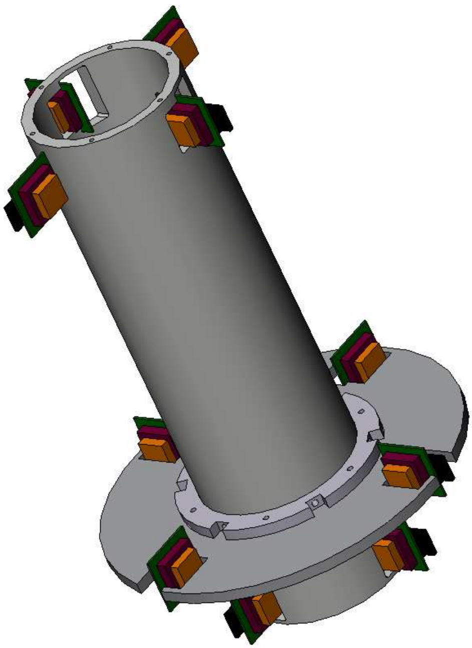
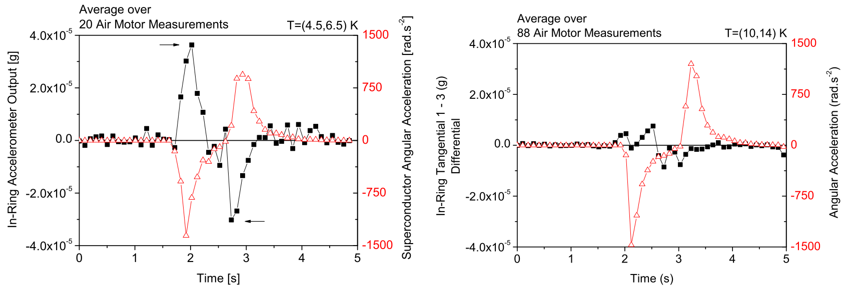


Figure 3: Accelerometer insert for sensor chamber in curl configuration.

Many tests were carried out in the time frame from 2003–2006 to reduce the noise level on the sensors and to obtain an optimum mechanical decoupling between the sensor chamber and the rotating parts of the facility [21, 24]. At the end we achieved a ground noise level on the sensors of a few μg_{rms} while the ring was at rest and $20\,\mu g_{rms}$ when the ring was rotating. The noise level did not steadily increase with rotational speed but strongly increased above a speed of 350 rad/s (a resonance peak appeared at a speed of 400 rad/s). In the final analysis, all sensor signals above this speed were therefore damped by a factor of 5. Using signal averaging over many profiles, the accuracy could be even further reduced.



(a) Superconducting.　(b) Normal conducting.

Figure 4: Signal averaged in-ring sensor data (■) vs. applied angular acceleration (△).

The high accuracy tests were carried out using the Niobium ring. After the reduction of all vibration offsets, a signal above the noise level still remained when the ring was cooled down close to LHe temperature. An example is shown in Fig. 4 with the signal averaged plots from the in-ring position for the acceleration field and the applied angular acceleration to the ring. The difference between the temperature range in which the Niobium is superconducting ($g/\dot{\omega} = -2.26 \pm 0.3 \cdot 10^{-8}\,\mathrm{g\,rad^{-1}\,s^2}$) and normal conducting ($g/\dot{\omega} = -1.24 \pm 1 \cdot 10^{-9}\,\mathrm{g\,rad^{-1}\,s^2}$) is clearly visible. Also the correlation between measured acceleration and applied acceleration is good (0.78) but for the first peak only, the second sensor peak seems to precede the applied acceleration for 0.2 s and is less correlated. The signals were obtained in a differential configuration, i.e. the signals from the reference position were subtracted from the in-ring values. That is done to remove any mechanical artifacts, such as a real sensor chamber movement, and to reduce the noise level.

If an angular acceleration of 1500 rad/s is applied to the superconducting ring, the tangential accelerometers show a counter-reaction of about 30 μg. What is the origin of these signals? The most important systematic error could still be the vibration rectification as also the helium evaporation is greatly contributing to the noise environment below 10 K. If the vibration sensitivity varies over the sensors, then our subtraction strategy in the curl configuration and also the subtraction of the reference position could maybe lead to false signals. Therefore, it was decided to further investigate this phenomenon using laser gyroscopes which are much less sensitive to vibration.

Laser gyroscope setup

Gyro measurements

Our requirements for the laser gyroscope include a low random angle walk (RAW), good bias stability and a high resolution, as well as a small size to fit in our sensor chamber. We therefore selected the KVH DSP-3000 fiber optic gyroscope that also features a digital output, which is much less affected by the electromagnetic environment compared to analog signals. It has a RAW of $2 \cdot 10^{-5}$ rad/s and a resolution of $1 \cdot 10^{-7}$ rad. Laser gyros are sensitive to magnetic fields due to the Faraday effect. We found values ranging from 2 to 0.5 rad/s T depending on the gyro's axis. Since we measured maximum magnetic fields during rotation in the order of hundreds of nT, we decided to put each laser gyro into a μ-metal shielding box to further reduce magnetic influence. This reduced the maximum sensitivity to 0.04 rad/s T. At maximum speed, the magnetic influence is therefore below the gyro's resolution. However, due to the magnetic influence, the field coil must be off during the measurements as this can otherwise introduce sensor offsets when the superconducting ring passes T_c. This actually caused wrong signals in our first reported preliminary data [24]. However, as these values were used to estimate the mechanical artifacts, the overestimated values gave a correct upper limit.

The gyro setup is illustrated in Fig. 5 showing the sensors inside the open vacuum chamber. Four gyros are mounted in three positions and two are mounted above the rotating ring (one showing up and one down to investigate any offset problems similar to the accelerometers). The axial distance from the top ring surface to the middle of each gyro position is 45.8 mm, 92.8 mm and 220.8 mm respectively. All sensors are mounted at a radial distance of 53.75 mm.

Due to the reduced vibration sensitivity, the following analysis was done without subtracting between alternating speed orientations or automatic subtraction between signals close to the spinning rings and the reference sensors. The gyro outputs for all three positions (reference, middle and above) using Niobium, Aluminum and YBCO samples with respect to the applied angular velocity of the rings is shown in Figs. 6 and 7. All profiles were recorded at an average temperature between 4 and 6 Kelvin. For reference purposes, we also removed the sample ring holder with the bottom plate but leaving the axis, bearings and motor assembly unchanged. The result is shown in Fig. 7(b). The above position (LG 3–4) was obtained by subtracting the two above-ring gyros from each other due to their alternating orientations. That should eliminate any signal offsets if present.

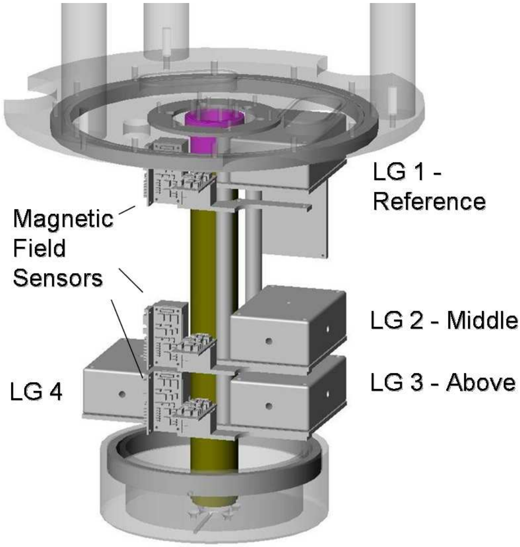
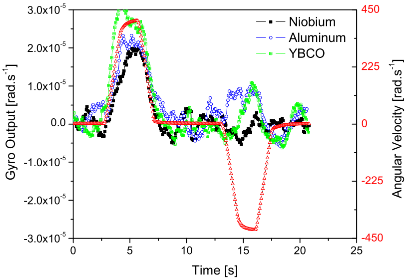


Figure 5: Laser gyroscope setup – open sensor chamber (outer chamber wall removed for illustration) and rotating ring with sample holder on bottom.

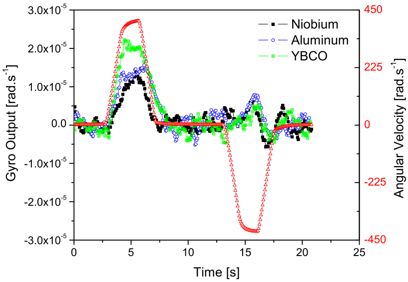
During each experimental test, more than 1000 profiles could be measured. Therefore, we were able to usually average over more than 20 profiles in every 2 K interval thus increasing statistical confidence.

The results are very surprising. First, the gyro outputs show a parity violation. Indeed, the gyro follows the applied angular velocity, but only if the ring is rotated in clockwise orientation. In order to check for any signal processing systematic, we cooled the YBCO sample down to 4.2 K and performed 40 successive clockwise rotations – and the gyro effect could always be measured. Then we cooled down again to 4.2 K and performed 40 successive counter-clockwise rotations – but any effect was an order of magnitude reduced compared to the clockwise rotation effect.

Second, YBCO gave the strongest signals while Niobium and Aluminum had similar responses to the applied angular velocity. Aluminum has a T_c of 1.2 K and is therefore not superconductive at liquid helium temperatures. Therefore, the sample holder was made out of Aluminum as no signal contribution was expected from that material (the same applies to the stainless steel plate at the bottom). The difference of the YBCO sample with respect to the other ones is that the YBCO ring has a larger wall

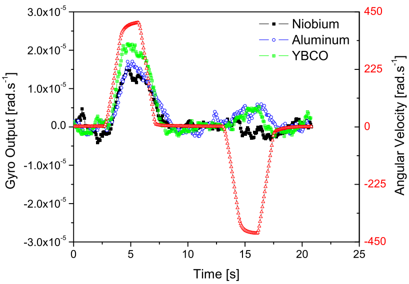


(a) LG 1 (reference).

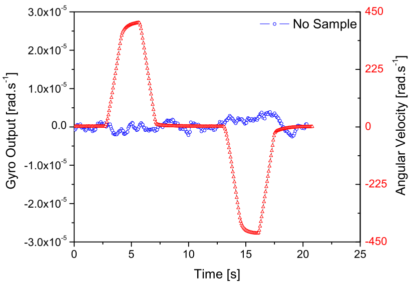


(b) LG 2 (middle).

Figure 6: Laser gyro output for Niobium, Aluminum and YBCO vs. applied angular velocity (△) between a temperature of 4 to 6 K.



(a) LG 3-4 (above).



(b) LG 3-4 (above).

Figure 7: Laser gyro output for Niobium, Aluminum and YBCO (a) and without sample holder (b) vs. applied angular velocity (△) between a temperature of 4 to 6 K.

thickness (15 mm versus 6 mm) and that the outer diameter of the sample holder is also slightly larger (165 mm versus 160 mm). As the effect vanishes when the sample holder is removed, the observed gyro responses must be related to the sample holder and its sample rings.

Third, the effect does not decay as one would expect from a dipolar field distribution. The reference position in fact even gives the highest signal responses while the above and middle positions have similar values. Of course, we have to keep in mind that the laser gyro averages only the z-component of any present field over a large 89×48 mm area (more specifically, the fiber coil has an elliptic shape) and it is therefore impossible to derive a field distribution using these measurements. Nevertheless, one would have expected that the field decays over distance if the source is the spinning ring. It is important to note that for the case of the above-ring gyros, if the gyro's orientation is flipped, also the effect sign flips and the effect is only present for the same clockwise-rotating of the spinning ring. The measurement also tells us that the effect is at least rotationally symmetric.

And fourth, the gyro responses do not correlate with the accelerometer measurements if one assumes the standard induction law. Any parity violation was not visible as the accelerometer measurements were always done subtracting the alternating speed profiles from each other in order to eliminate the anomalous DC offsets from the vibration rectification. From the induction law, the Niobium signals should be higher by a factor of about 100. Assuming the validity of the gyroscope measurements, that means that either the accelerometer measurements are vibration artifacts or the standard induction law does not apply. That could be explained by the breakdown of the usual weak-field approximation used to derive the Maxwell-like equations out of general relativity. Also, new theoretical concepts actually predict that the acceleration-induced effect is stronger by about two orders of magnitude compared to the gyro response as observed in our measurements [4]. However, the measurements so far already rule out our initial theoretical approach that modeled the effect as proportional to the ratio between the matter densities in the coherent state with respect to the lattice [23].

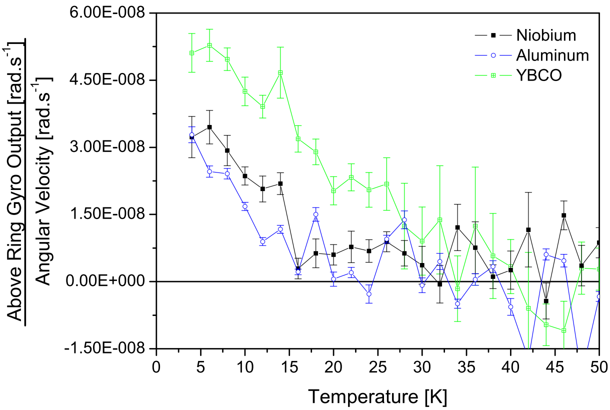


Figure 8: Variation of above ring gyro output vs. angular velocity with temperature (for clockwise rotation).

The ratio of the above gyro versus angular velocity response for clockwise rotation with respect to temperature for Niobium, Aluminum and YBCO is shown in Fig. 8. It is interesting to see that the effect occurs below a critical temperature that does not coincide with the superconducting temperature of the materials. For Niobium and Aluminum, the critical temperature is about 16 K whereas for YBCO it is close to 32 K. Apart from the value at 4 K, the Niobium curve is above Aluminum by about $1{\cdot}10^{-8}$. The YBCO curve is quite constant below 32 K until the Aluminum critical temperature at 16 K. It seems like the effect of Aluminum adds then to the YBCO curve. So either Aluminum does indeed produce an effect and therefore contributes to Nb and YBCO due to the Al sample holder material, or Aluminum provides a reference case due to the vibration environment that leads to gyro sensor offsets. If the Aluminum part is removed from the curves, than YBCO has a more or less constant coupling factor of $2.2{\cdot}10^{-8}$ below 32 K and Niobium $1.6{\cdot}10^{-8}$ below 16 K.

Systematic effects

Electromagnetic fields

The sensors are inside an evacuated vacuum chamber, which is grounded and acts as a Faraday cage. The supply voltages for the gyros are generated and stabilized inside the vacuum chamber. Moreover, the gyro output is digital, which greatly eliminates possible electromagnetic interference of the signals along their transmission to the computer. From the magnetic sensitivity, an induced coupling factor of $4{\cdot}10^{-11}$ has been measured, which is 3 orders of magnitude below the observed effect. In addition, gyro measurements were done with and without μ-metal shielding of the gyros with very similar results. Therefore, the explanation of the observed effects due to electromagnetic effects is very unlikely.

Pressure effect

The rotation of the ring causes a strong evaporation of the liquid helium. This pressure increase can maybe tilt or turn the vacuum chamber and cause sensor offsets. We investigated the pressure increase of the helium gas in the cryostat by mounting a Keller PA-23 pressure transmitter on the top of the facility. After filling up of the facility with liquid helium, the first profile caused a peak pressure increase of 100 mbar as some liquid was evaporated due to the stirring of the rotating ring. All successive profiles showed no change in the pressure within the sensor resolution of

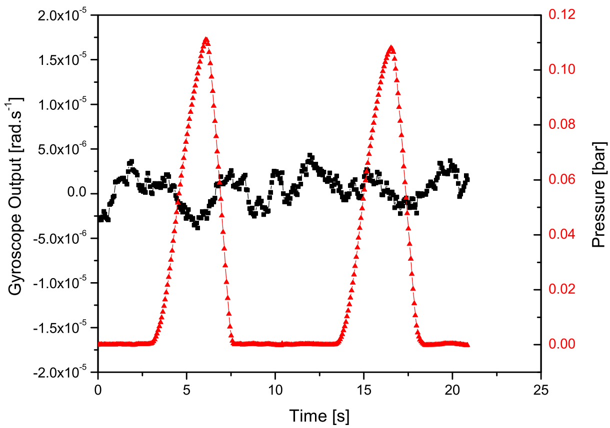


Figure 9: Gyro output vs. pressure.

3 mbar. As a worst-case scenario, we simulated the pressure increase using compressed air that was connected to the liquid helium line at room temperature. Instead of running the motor, the pressure was increased during each profile until the sensor measured the 100 mbar. The result is shown in Fig. 9. No gyro response other than the usual noise is seen during the pressure increase. Therefore, the pressure increase due to liquid helium expansion cannot lead to the observed effects and especially no parity violation.

Vibration offsets

Although the manufacturer noted no vibration sensitivity, we performed a dedicated test by putting the gyro on a table next to a shaker table. We were indeed able to produce a vibration offset but only using very large amplitudes at a frequency of 60 Hz, which is similar to the frequencies of the air motor at maximum speed as shown in Fig. 10. This offset was always negative and it was sensitive to the orientation of the sensor axis on the shaker table. We found out that the vibration offset was only present when the gyro's axis was pointing to the $Z+$ or $X+$ direction ($Z+$ gave the strongest response). The difference between the two directions is also evident in Fig. 10. Pointing the gyro's in the Y direction showed no vibration offsets other than noise. The magnitude of the vibration offset was directly related to the amplitude of the vibration. When the amplitude was reduced

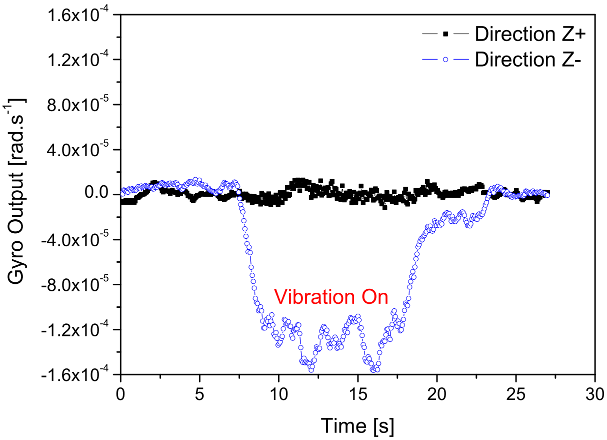


Figure 10: Gyro output vs. vibration.

so that the noise of the gyro's output did not increase during the vibration, the vibration offset vanished below detectability ($< 2 \cdot 10^{-6}$ rad/s).

During all gyro measurements, no increase in the noise level of the gyros was detected during the ring's rotation similar to the case as described above. The measured gyro's response to the ring's rotation was positive and not negative as the vibration offsets. The same signal with alternating sign was also measured by flipping the gyro's orientation axis as in the case of LG3 and LG4 which would not be possible with the vibration offsets. Also the difference between the YBCO and Nb/Al rings as well as the parity violation cannot be explained. Therefore, the explanation of our gyro results as vibration offsets is not likely – but it is the most important error source that has to be further analyzed since the real noise environment is more complex than the signals produced from the shaker table.

Sensor tilting

Due to the helium gas flow, the sensor chamber might be tilted. Since the gyro measures the Earth's rotation, tilting of the sensors can induce a false signal due to the different offset from the Earth's measurement. Assuming a tilting angle α, the offset can be expressed as

$$\text{Offset} = \text{Offset}_{\text{Earth}} \left[\sin(\text{latitude}) - \sin(\text{latitude} + \alpha) \right]. \qquad (1)$$

Therefore, in order to get offsets in the gyro's signal similar to the observed effects, a tilting angle of at least 20° is necessary with OffsetEarth $= 73 \cdot 10^{-6}$ rad/s and a latitude of 48°. It is impossible that the sensor chamber tilted by as much as 20° during each profile. This artifact can be therefore ruled out.

Mechanical friction

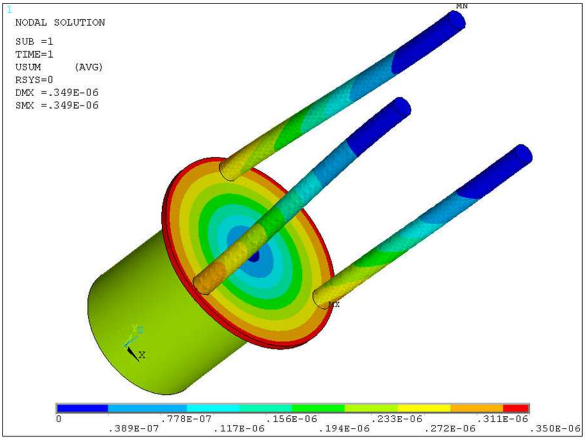


Figure 11: Finite-element analysis of sensor chamber and connecting rods drilling.

During the rotation of the ring, there is a 5 mm gap with helium gas between the sample holder and the sensor vacuum chamber. Since the viscosity of helium gas at 5 K is an order of magnitude below the viscosity of air at room temperature [14], mechanical friction from the rotating helium gas cannot explain the observed effects as the effect only occurs when passing through a critical cryogenic temperature. Nevertheless, we shall examine the order of magnitude from such friction effects. We therefore built a finite element model of the sensor chamber with its connecting rods using ANSYS as shown in Fig. 11. The force on the bottom of the sensor chamber from the rotating gas can be calculated using Stoke's law as

$$F \cong \eta \frac{Av}{\Delta x}, \tag{2}$$

where η is the viscosity, A the area, v the velocity of the gas and Δx the gap between the rotating ring and the sensor chamber. Using our geometry we obtain a maximum friction force on the bottom of the sensor chamber of about 1 mN at maximum speed. Using the finite element model, this force leads to a drilling of the sensor chamber assembly of $5{\cdot}10^{-8}$ rad. The upper limit false coupling factor from friction is therefore $5{\cdot}10^{-11}$, which is three orders of magnitude below the observed effects similar to the magnetic sensitivity.

Discussion

The effects and the analysis so far lead to the following possible interpretations:

1. The effects are real: A frame-dragging-like signal was detected from spinning rings at cryogenic temperatures. The effect occurs below a critical temperature which does not coincide with the superconducting temperature of the rings. The strength of the effect depends on the material of the ring. The coupling factor of the observed effect with respect to the applied angular velocity is in the range of 3 to $5{\cdot}10^{-8}$. We observed a parity violation, such that the effect is dominant in the clockwise rotation (when looking from above) in our laboratory setup. No classical or systematic explanation has been found so far for the observed effects. The field expansion is not clear at the moment and needs further investigation.

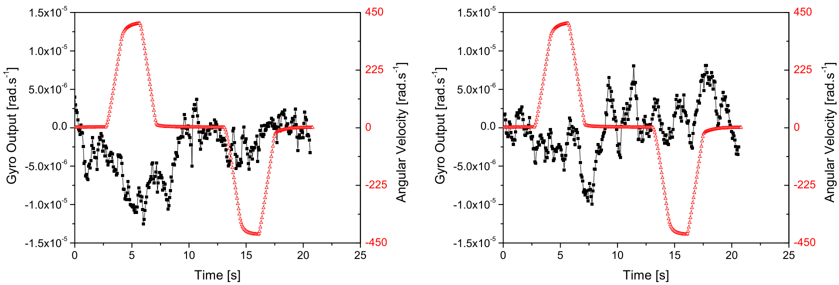
2. The reference signal has to be subtracted from the measurements due to sensor chamber movements: The coupling factor is then reduced to about $1.3{\cdot}10^{-8}$. Now it is less clear if the effect's origin could be indeed related to superconductivity because Nb still shows a signal but Al does not as shown in Fig. 12 (they have the same sample holder dimensions). Nevertheless, the critical temperature for the effect is different than the superconducting critical temperature. Also in this case, parity violation is observed.

3. Aluminum is the reference case which has to be subtracted from the measurements: The vibration environment from the helium gas expansion causes sensor offsets. Since aluminum is not a superconductor, this material can be considered the reference case which has to be subtracted from the results. As discussed above together with Fig. 8, the coupling factor for Nb is then reduced to $1.6{\cdot}10^{-8}$ and for

YBCO to $2.2{\cdot}10^{-8}$. The effect is now indeed related to superconductivity, however as in the cases discussed above, the critical temperature does not coincide with the superconducting temperature. We still observe a parity violation.

4. All signals are false due to systematic effects: Of course this may still be a possibility. The strongest indication is that the signals do not decay over the three positions measured. As this looks like giving room to facility artifacts, it could be also possible that the field is propagating along the spinning motor axis. All systematic effects analyzed so far contribute to less than 3 orders of magnitude to the observed effects. Only vibration effects may still contribute to the gyro output but they seem unlikely to explain all different aspects of the effect such as parity violation and the dependence on the ring material.

Since the systematic effects analyzed so far are below the laser gyro measurement accuracy and therefore cannot account for our measurements, we suggest that the first interpretation is correct. That in turn puts severe limits on theoretical models that were proposed to predict frame-dragging fields generated by superconductors as the phenomenon that we observe is apparently not related to superconductivity and shows a parity violation.



(a) Niobium. (b) Aluminum.

Figure 12: Above laser gyro output minus reference (LG3–4 − LG1) vs. applied angular velocity (△) between a temperature of 4 to 6 K.

Comparison with other experiments

In order to determine if our results are indeed genuine or facility artifacts, it is important to compare them with other experiments and to check for consistency. Fortunately, at least two similar experiments are available that can be used for such a comparison: a recent test of a spinning lead superconductor close to the world's largest ring laser gyroscope (Canterbury Ring Laser Group) and the Gravity-Probe B satellite using spinning superconducting gyroscopes to detect Earth's frame dragging field. Due to differences between the experiments, models are necessary for such a comparison to extrapolate our results to the other setups. In addition to the different models that were already proposed [23, 6, 4], we will follow a phenomenological approach assuming that the fields detected are an amplification of classical frame-dragging fields without linking it to superconductivity. After passing a critical temperature, the classical frame-dragging fields B_{g0} are enhanced using the simple expression

$$B_g = \gamma B_{g0}. \tag{3}$$

According the our measurements so far, the enhancement factor γ (or frame-dragging relative permeability factor) is temperature and material dependent, and shows a parity violation. For the comparison with our experiments and the Canterbury setup, the classical frame-dragging field B_{g0} at the center of a ring and a disc is given by,

$$B_{g0} = \frac{4\mathrm{G}}{\mathrm{c}^2}\frac{m}{R_o + R_i}\omega, \tag{4}$$

where m is the spinning mass, R_o the outer radius, R_i the inner radius (= zero for the case of a disc), and ω the angular velocity. Using our measurements, we get $\gamma \cong 1.2$ to $1.7 \cdot 10^{18}$ for the various material combinations at a temperature of 5 K (we assumed that the field measured at the gyro's location is similar to the one in the center of the spinning ring).

For the comparison with the Gravity-Probe B data, we have to calculate the classical frame-dragging field for a spinning shell or sphere with radius R along the central axis z, given by

$$B_{g0} = \begin{cases} \frac{2G}{c^2}\frac{I}{z^3}\omega & z > R, \\ \frac{5G}{c^2}\frac{I}{R^3}\omega & z = 0 \quad \text{(sphere)}, \\ \frac{2G}{c^2}\frac{I}{z^3}\omega & z = 0 \quad \text{(shell)}, \end{cases} \tag{5}$$

where I is the moment of inertia.

It is important to note that the classical fields scale with mass and geometry of the spinning source. If the effect would be related to a superconductive-like phenomenon, the field strengths are expected to be independent of mass and geometry of the spinning source similar to the London moment-magnetic field produced by a rotating superconductor that is only proportional to the angular velocity and the charge-to-mass ratio of the Cooper-pair.

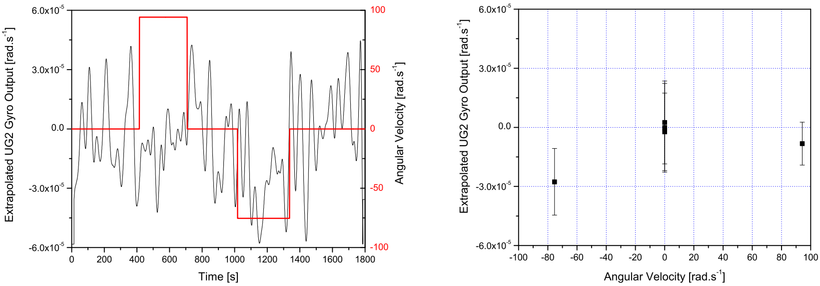
Comparison with Canterbury Ring Laser Group experiment

Recently, an independent experimental test was carried out, where a lead disc at liquid helium temperature was spinning close to the world's most precise ring laser gyro UG2 from the Canterbury Ring Laser Group [9]. Contrary to our setup, the gyro here is operated outside of the cryostat facility due to its large dimensions of 21×39.7 m. This should greatly reduce any vibration offsets associated with the evaporating helium gas. Fig. 13a shows the gyro's response to the speed of the spinning superconductor[1]. Here too, we see that the gyro reacts if the superconductor is rotated. Again, we note a parity violation as the gyro response is greater for the counter-clockwise rotation – which is the opposite direction as in our experiments. Since this experiment was carried out in the southern hemisphere and our experiments in the northern hemisphere, a first hint at the origin of the parity effect could be the Earth's rotation. A similar parity anomaly that may be related to our effect was reported on gyro weight and free-fall experiments in a Japanese laboratory [10, 11], which showed the same parity direction as in our laboratory and was also located on the northern hemisphere. However, these claims could not be verified in a number of replication attempts up to now ([17], and references therein).

As the UG2 gyro is large compared to the actual rotating superconductor, a field distribution has to be assumed in order to compare our results with the UG2 results. The Canterbury group used a dipolar distribution for the interpretation of their experimental data. As computed in [9], the UG2 gyro output was multiplied by a factor of $1.7 \cdot 10^6$ to get an estimate of the frame-dragging-like signal strength in the vicinity of the rotating superconductor. Fig. 13(b) shows the gyro response corrected by the dipolar field distribution and the applied angular velocities. The coupling factor (Gyro Output)/ω computed for the counter-clockwise direction is $3.8 \pm 3 \cdot 10^{-7}$. It is possible to improve the statistics with additional filter. By applying a 200 pt digital moving average filter, the counter clockwise direction coupling factor yields $3.8 \pm 1.3 \cdot 10^{-7}$. This is nearly one order of magnitude above our measurements in the close vicinity of the spinning supercon-

[1] The gyro data from [9] was converted into rad/s and the offset at zero angular velocity was removed for consistency with the present paper.

ductor. One major difference is the disc shape in comparison with our ring superconductors.



(a) UG2 gyro response (black) and applied angular velocity of lead of superconductor (red).

(b) UG2 gyro response vs. applied angular velocity.

Figure 13: Experimental results from the response of the UG2 ring laser gyro to the rotation of a lead superconductor – extrapolated to the spinning source and normalized (data taken from [9]).

Indeed, by computing the enhancement factor γ from the angular momentum of the spinning disc and the UG2 gyro output, we get $\gamma \cong 1.9 \cdot 10^{18}$ for the counter-clockwise rotation, a value very close to the ones in our setup. This suggests that the fields observed are scaling like classical frame-dragging fields. A superconducting-like phenomenon such as a gravitomagnetic London moment would give the same results in the Canterbury and in our setup.

Comparison with Gravity-Probe B Results

Since Gravity-Probe B (GP-B) measured with unprecedented accuracy the precession of superconducting gyros in a polar orbit around Earth, one would expect to see the effect of an enhanced frame-dragging-like field in their setup as well. The experiment consists of four gyros, which are equally spaced and aligned along the roll axis of the satellite towards the guide star IM Pegasi [13]. Gyro 1 and 3 are rotating in one direction and Gyro 2 and 4 in the other direction so that the poles are always facing each other. Note that for the gravitational case, similar poles attract contrary to magnetic fields where opposite poles attract [27]. If the spinning gyros now produce a frame-dragging-like field, a restoring torque would appear proportional to the misalignment of the gyro's axis with the spacecraft axis towards the guide star. This in turn will cause additional precession of the gyros. We can express this drift in function of the misalignment angle Ψ as [8, 18],

$$\Omega = \frac{B_g}{2} \sin \Psi, \tag{6}$$

where B_g is the frame-dragging-like field at the location from one gyro caused by the others. As a first approximation, we use a dipolar field expansion with a gyro separation of 75 mm and a gyro diameter of 38 mm. The final speed of the four gyros was measured to be 79.4, 61.8, 82.1 and 64.9 Hz respectively. The gyros are made out of fused quartz spheres coated with a 1.25 μm layer of Niobium. An anomalous torque proportional to the misalignment angle was indeed seen in the GP-B experiment with drift rates of a couple of arcsec/day/deg. The anomalous torque anomaly is presently modeled as an electrostatic patch effect due to a variation of the electric potential along the gyro's surface. Without distinguishing between a patch effect or frame-dragging origin of this effect, we can at least express an upper value for any non-classical frame-dragging field generated by the rotating superconducting Nb shells. Using the average torque of all four gyros, the upper-limit coupling factor $B_g/\omega \cong 1 \cdot 10^{-9}$ at the center of the spinning superconductor. This is more than an order of magnitude smaller compared to our setup.

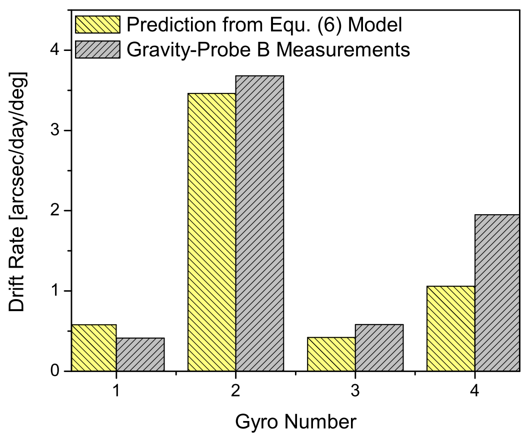


Figure 14: GP-B gyro misalignment torque drifts and spin-spin approximation from Equ. 6 - data taken from [15] with linear fit up to 1° and γ_{SiO_2}=0.6·10^{18}.

By using the angular momentum of the gyro, we can estimate an upper limit for the enhancement factor by comparing the fields to the observed anomalous torques. Since the gyro has only a Nb layer, the SiO_2 will dominate the angular momentum by orders of magnitude. By taking an enhancement factor γ for SiO_2 that is half the value for Nb, we get a fairly good match with the observed anomalous torques as shown in Fig. 8 (here we assumed again a parity violation similar to the other experiments). This suggests that our effect might be an alternative explanation for the anomalous torques observed on Gravity-Probe B. Since we did not specifically measure SiO_2 in our setup, we can set this enhancement factor as a possible upper limit for this material at 4 K. Since Nb contributes very little to the angular momentum, the upper limit enhancement factor is about a factor of 500 larger compared to the values from our experiment. That would still leave enough room for a possible patch effect assuming that SiO_2 does not contribute to the frame-dragging effect. Further experiments are necessary to clarify this point.

Summary and conclusion

High-resolution accelerometer and laser gyroscope measurements were performed in the vicinity of spinning rings at cryogenic temperatures. After passing a critical temperature, which does not coincide with the material's superconducting temperature, the angular acceleration and the angular velocity applied to the rotating ring could be seen on the sensors although they are mechanically de-coupled. A parity violation was observed for the laser gyroscope measurements such that the effect was greatly pronounced in the clockwise-direction only.

Tab. 1 summarizes our laser gyroscope experiments and compares them with an independent test performed by the Canterbury Ring Laser Group and the Gravity-Probe B anomalous torque measurements. The data is given as a simple coupling factor with respect to the applied angular velocity (as predicted by superconducting-like models) as well as an enhancement factor amplifying the classical frame-dragging field of the spinning source. We see that the effect scales well with an enhancement factor of $\cong 1 \cdot 10^{18}$ throughout the different experimental setups. Apart from the parity violation, this suggests that the effect behaves similar to a classical frame-dragging field but greatly amplified. It does not show the signature of a superconductivity-like phenomenon, that would lead to similar B_g/ω coupling factors between the different setups.

Table 1: Comparison of all experimental data for a frame-dragging-like field at $T = 4\,\mathrm{K}$ at the center of the spinning source.

Experiment	Material	Coupling $(B_g/\omega) \times 10^8$		Enhancement $\gamma \times 10^{18}$	
		CW	CCW	CW	CCW
Tajmar *et al.*	YBCO+Al	5.3±0.2	−1.2±0.1	1.7±0.1	−0.4±0.1
	Nb+Al	3.2±0.5	−0.4±0.3	1.2±0.2	−0.1±0.1
	Al	3.8±0.3	−0.7±0.3	1.7±0.1	−0.3±0.1
Graham *et al.*	Pb	−5.3±8.5	37.7±13.2	0.5±0.8	1.9±0.7
GP-B*	SiO_2	< 0.1		< 0.6	
	Nb	< 0.1		< 500	

* Gravity-Probe B upper limit, average over all gyros

The results can also be used to compare with different theoretical models that have been proposed predicting large frame-dragging fields around rotating superconductors. Apart from the parity violation and the non-superconductor critical temperatures observed in the experiments, especially the Gravity-Probe B data rules out all present models by up to 4 orders of magnitude. The experimental data also rules out our initial Cooper-pair mass anomaly hypothesis by 5 orders of magnitude [19, 20].

Table 2: Comparison of existing theoretical models with experimental limits at $T = 4\,\mathrm{K}$ at the center of the spinning source.

Theory	Coupling $(B_g/\omega) \times 10^8$		
	Tajmar config. (Nb)	Graham config.	GP-B config.
Tajmar and de Matos	395	332	395
Dröscher and Hauser	130	130	130
de Matos and Beck	1.6	1.6	1.6
Experimental results*	3.2 ± 0.5	37.7 ± 13.2	< 0.1

* We neglect the parity violation and use the maximum value from the CW or CCW direction

The gyro responses do not correlate with the accelerometer measurements because they are lower by a factor of 100 if one assumes the standard induction law. It is not clear at the moment if systematic effects such as vibration rectification are responsible for the accelerometer mismatch or if this discrepancy between accelerometer- and gyro measurements is correct as this was recently theoretically predicted [4].

All laser gyro systematic effects modeled and analyzed so far show that facility artifacts from mechanical friction, magnetic fields or vibration effects are at least 3 orders of magnitude below the high-resolution gyro measurements. Although vibration offsets might still be present in our data, which will be investigated in further testing, all data and analysis suggests that the observed effects are real.

Nomenclature

A = surface area of sample holder (m^2)
B_g = frame-dragging field (rad/s)
c = speed of light ($= 3 \cdot 10^8$ m/s)
γ = enhancement or relative frame-dragging permeability factor
G = gravitational constant ($= 6.67 \times 10^{-11}\, m^3/kg\, s^2$)
g = gravitational field (in unit of Earth standard acceleration $= 9.81\, m/s^2$)
I = angular momentum ($kg\, m^2/s$)
m = mass (kg)
η = viscosity (Pa s)
$\dot{\omega}$ = angular acceleration (rad/s^2)
ω = angular velocity (rad/s)
Ω = gyro precession (rad/s)
Ψ = gyro misalignment angle (°)
R = radius (m)
T = temperature (K)
T_c = critical superconducting temperature (K)
v = velocity of helium gas (m/s)
Δx = gap between sample holder and sensor chamber (m)
z = distance along central spinning axis (m)

Acknowledgements

This research program is funded by the Austrian Research Centers GmbH – ARC. Part of this research was sponsored by the European Space Agency under GSP Contract 17890/03/F/KE and by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant number FA8655-03-1-3075. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. We would like to thank R.Y. Chiao, R. Packard, C.J. de Matos and J. Overduin for many stimulating discussions.

References

[1] R.Y. Chiao, *Millikan oil drops as quantum transducers between electromagnetic and gravitational radiation* (2007), [arχiv:gr-qc/0702100].

[2] L.A. Christel, M. Bernstein, R. Craddock and K. Peterson, *Vibration Rectification in Silicon Machined Accelerometers*, IEEE Conference Proceedings (1991).

[3] I. Ciufolini and E.C. Pavlis, *A confirmation of the general relativistic prediction of the lense-thirring effect*, Nature **431** (2004) 958–960.

[4] C.J. de Matos and C. Beck, *Possible measurable effects of dark energy in rotating superconductors*, (2007), [arχiv:0707.1797v1].

[5] C.J. de Matos and M. Tajmar, *Gravitomagnetic london moment and the graviton mass inside a superconductor*, Physica C **432** (2005) 167–172.

[6] W. Dröscher and J.Hauser, *Advanced propulsion systems from artificial acceleration fields*, (2007).

[7] F. Everitt, *First results from gravity probe b*, (2007), `http://einstein.stanford.edu/content/aps_posters/APS_talk_Everitt.pdf`.

[8] R.L. Forward, *General relativity for the experimentalist*, in Proceedings of the IRE (1961).

[9] R.D. Graham, R.B. Hurst, R.J. Thirkettle, C.H. Rowe and B.H. Butler, *Experiment to detect frame-dragging in a lead superconductor*, (2007), `http://www2.phys.canterbury.ac.nz/~physrin/papers/SuperFrameDragging2007.pdf`.

[10] H. Hayasaka and S. Tageuchi, *Anomalous weight reduction on a gyroscope's right rotations around the vertical axis on the earth*, Phys. Rev. Lett. **63** (1989) 2701.

[11] H. Hayasaka, H. Tanaka, T. Hashida, T. Chubachi and T. Sugiyama, *Possibility for the existence of anti-gravity: Evidance from a free-fall experiment using a spinning gyro*, Speculations in Science and Technology **20** (1997) 173.

[12] Y. Jiang and M. Liu, *Rotating superconductors and the london moment: Thermodynamics versus microscopics*, Phys. Rev. B **63** (2001) 184506.

[13] G.M. Keiser, S. Buchman, W. Bencze and D.B. DeBra, *The expected performance of the gravity probe b electrically suspended gyroscopes as differential accelerometers*, in AIP Conference Proceedings (1998).

[14] W.E. Keller, *Calculation of the viscosity of gaseous he_3 and he_4 at low temperatures*, Phys. Rev. **105** (1957) 41–45.

[15] D.K. Gill and S. Buchman, *Evidence for patch effect forces on the gravity probe b gyroscopes*, (2007), `http://einstein.stanford.edu/content/aps_posters/EvidenceForPatchEffectForces.pdf`.

[16] M. Liu, *Null result for violation of the equivalence principle with free-fall rotating gyroscopes*, Phys. Rev. Lett. **81** (1998) 3223–3226.

[17] J. Luo, Y.X. Nie, Y.X. Zhang and Z.B. Zhou, *Null result for violation of the equivalence principle with free-fall rotating gyroscopes*, Phys. Rev. D **65** (2002)

42005.
[18] B. Muhlfelder, private comm. (2007).
[19] M. Tajmar and C.J. de Matos, *Gravitomagnetic field of a rotating superconductor and of a rotating superfluid*, Physica C **385** (2003) 551–554.
[20] M. Tajmar and C.J. de Matos, *Extended analysis of gravitomagnetic fields in rotating superconductors and superfluids*, Physica C **420** (2005) 56–60.
[21] M. Tajmar and C.J. de Matos, *Gravitomagnetic fields in rotating superconductors to solve tate's cooper pair mass anomaly* Physica C **420** (2006) 56–60.
[22] M. Tajmar and C.J. de Matos, *Gravitomagnetic fields in rotating superconductors to solve tate's cooper pair mass anomaly*, in AIP Conference Proceedings **813** Issue 1 (2006).
[23] M. Tajmar and C.J. de Matos, *Local photon and graviton mass and its consequences*, (2006), [arχiv:gr-qc/0603032].
[24] M. Tajmar, F. Plesescu, B. Seifert and K. Marhold, *Measurement of gravitomagnetic and acceleration fields around rotating superconductors*, in AIP Conference Proceedings **880** Issue 1 (2007).
[25] J. Tate, B. Cabrera, S.B. Felch and J.T. Anderson, *Precise determination of the cooper-pair mass*, Phys. Rev. Lett. **62** (1989) 845–848.
[26] J. Tate, B. Cabrera, S.B. Felch and J.T. Anderson, *Determination of the cooper-pair mass in niobium*, Phys. Rev. B **42** (1990) 7885–7893.
[27] R. Wald, *Gravitational spin interaction*, Phys. Rev. D **6** (1972) 406–413.

Section III: Coherence, De-coherence and Entanglement

entanglement role in studies of *CP* violation at B factories
tests of Einstein-Podolski-Rosen correlations
decoherence measurements in fullerene interferometry

Time and Matter
2007 Bled, Slovenia

Measurement of EPR-type Flavour Entanglement in $\Upsilon(4S) \rightarrow B^0\overline{B}^0$ Decays

AURELIO BAY* FOR THE BELLE COLLABORATION
Ecole Polytechnique Fédérale Lausanne

Abstract: The neutral B-meson pair produced at the $\Upsilon(4S)$ should exhibit a non-local correlation of the type discussed by Einstein, Podolski and Rosen. The time-dependent flavour asymmetry of the B mesons decaying into flavour eigenstates is used to test such a correlation. The asymmetry obtained from semileptonic B^0 decays is in agreement with the prediction from quantum mechanics and far away from the predictions of the local realistic models tested.

The observation of partial decoherence in EPR systems could be a signal for New Physics: we have tested for such effects, and found our results are consistent with no decoherence.

Introduction

The concept of entangled states (i. e. states which cannot be represented as product states of their parts) was born in the '30s in the midst of several conceptual difficulties with Quantum Mechanics (QM). In 1935 Einstein, Podolski and Rosen (EPR) wrote a paper which was an effort to clarify the conceptual basis of QM and arrived at the conclusion that QM could not be considered a "complete" theory [1]. EPR considered a pair of particles produced by the same interaction, subsequently freely propagating in space but still linked by momentum conservation. EPR found a contradiction when realism and locality are applied to the predictions of QM on a couple of non-commuting observables (position and momentum, in their paper). The conceptual problem is better understood considering the 1951 variant by David Bohm using spin correlations [2]. In the EPR-Bohm experiment

* aurelio.bay@epfl.ch

the two-particle singlet state can be written as:

$$|\psi\rangle = \frac{1}{\sqrt{2}}[|\uparrow\rangle_1 \otimes |\downarrow\rangle_2 - |\downarrow\rangle_1 \otimes |\uparrow\rangle_2] \quad (1)$$

where $|\uparrow\rangle_j$ ($|\downarrow\rangle_j$) describes the spin state of j^{th} particle (j=1,2) with spin up (down) respectively. Measurement of the spin on one particle, undetermined prior to the measurement, will "collapse" the wave function to one of the eigenstates and therefore predicts with certainty the outcome of the spin measurement on the second particle without actually doing any measurement. The important point is that the spin of the second particle in a given direction is defined by the choice of the polarizer orientation on the first particle. The orientation can be chosen at the "last moment", just prior to the arrival of the particle, and cannot be communicated to the second particle system unless superluminal signals are invoked. We should conclude that in a way or another the second particle carries the information needed to behave correctly for any possible choices of the measurement in the system of the first particle. Indeed, following EPR, one can define "elements of reality" for spin in S_x and S_y direction for the second particle, determined from the spin measurements done on the first particle. But according to QM the observables S_x and S_y do not commute and therefore cannot have definite values at the same time. Following the EPR-Bohm conclusion the description of reality given by QM is incomplete. This points to the need of extra information, "hidden variables" (HV) for instance, to complement QM. In 1964 J. S. Bell found a general scheme to test QM against HV theories: he showed that a certain inequality which is always satisfied by all local hidden variable models, can instead be violated by QM [3]. Following the demonstration by J. Clauser, M. Horne, A. Shimony and R.Holt (CHSH), the correlation function for the measurements of the two particles with spin analyzers with orientation a and b is given by

$$E(a,b) = \int_\Gamma A(a,\lambda)B(b,\lambda)\rho(\lambda)\,d\lambda. \quad (2)$$

$A(a,\lambda)$, $B(b,\lambda)$ are the results of the measurements (+1 or –1, corresponding for instance to spin up and down in an experiment performed on spin 1/2 particles), and the HV are represented by the symbol λ. The HV follow a normalized probability distribution $\rho(\lambda)$. CHSH show that the inequality

$$|E(a,b) - E(a,c)| + |E(b',b) + E(b',c)| \leq 2 \quad (3)$$

is always satisfied by any local realistic theory featuring HV, while QM can violate it for some particular values of the analyzer orientations a, b, c,

b'. The key point of the demonstration is the presence of products of the kind $B(b,\lambda)B(c,\lambda)$, representing the measurement of the same event by the same apparatus, but with different orientations. The additional information carried by λ is used to infer the result of such classical measurement, incompatible with QM when orthogonal observables are measured.

Several Bell-CHSH experiments have been performed by measuring the linear polarization of pairs of photons produced in a correlated state.

For this kind of experiment an optimized choice of angles brings to the following Bell-CHSH inequality:

$$|S(\phi)| = |3E(\phi) - E(3\phi)| \leq 2 \tag{4}$$

in which only two orientations of the analyzers need to be considered. In this case QM predicts $E(\phi) = cos(2\phi)$ for the two photons, giving a maximal value $S_{max} = 2\sqrt{2}$, when $\phi = \pi/8$. Note that in the case of a spin 1/2 system, we would have obtained $E(\phi) = cos(\phi)$, the maximum for S occurring at $\phi = \pi/4$.

In the experiment of A. Aspect, P. Grangier and G. Roger [4] the photons with correlated linear polarization are produced from a ^{40}Ca source. The atom excited by a laser undergoes a cascade $(J = 1) \to (J = 0) \to (J = 1)$. A notable feature of this experiment is the usage of two two-channel polarizers allowing true dichotomic polarization measurements. On the other hand a normalization is introduced to account for the limited detector efficiency

$$E(\phi) = \frac{(R_{++} + R_{--}) - (R_{-+} + R_{-+})}{(R_{++} + R_{--}) + (R_{-+} + R_{-+})}(\phi) \tag{5}$$

where the $R_{\pm\pm}$ are the four coincidence rates for relative orientation ϕ of the two polarizers. For the optimal ϕ value, the results are in good agreement with QM, and violate LR by many standard deviations.

T. D. Lee and C. N. Yang suggested to use neutral kaon pairs as an entangled EPR system, the kaon flavour (the Strangeness S=1 or S=–1) playing the role of spin up and down in a 1/2 spin system. Strangeness-entangled pairs can be produced from the decay of a ϕ, for instance. The two experiments CPLear (at CERN) and KLOE (Frascati) have studied correlations in K^0-$\overline{K}^0$ pairs, and they obtained results in agreement with QM predictions [5, 17].

In this paper we present a study of EPR correlation in the flavour (the beauty +1 or –1) of neutral B-meson pairs from $\Upsilon(4S)$ decays. The system is described by a wavefunction analogous to (1) [6, 7]:

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left[\left|B^0\right\rangle_1 \otimes \left|\overline{B}^0\right\rangle_2 - \left|\overline{B}^0\right\rangle_1 \otimes \left|B^0\right\rangle_2\right]. \quad (6)$$

An important difference with the photon experiment resides in the fact that the meson's flavour is known only when it decays and only if it decays into "flavour-specific" modes (like one of the semi-leptonic channels, in which the flavour can be inferred from the electric charge of the lepton). From 6 we deduce that decays occurring at the same proper time are fully correlated: the flavour-specific decay of one meson fixes the (previously undetermined) flavour (B^0 or $\overline{B}^0$) of the other meson. From 6 we can also deduce the time-dependent rate for decay into two flavour-specific states for opposite flavour (OF, $B^0\overline{B}^0$) and same flavour (SF, B^0B^0 or $\overline{B}^0\overline{B}^0$) decays, and the corresponding time-dependent asymmetry (see Fig 1, curve QM):

$$R_{\substack{OF\\SF}}(\Delta t) = e^{-\Delta t/\tau_{B^0}}/(4\tau_{B^0})\{1 \pm \cos(\Delta m_d \Delta t)\}, \quad (7)$$

$$A_{\text{QM}}(\Delta t) \equiv \frac{R_{\text{OF}} - R_{\text{SF}}}{R_{\text{OF}} + R_{\text{SF}}}(\Delta t) = \cos(\Delta m_d \Delta t) \quad (8)$$

$\Delta t \equiv |t_1 - t_2|$ is the proper-time difference of the decays, and Δm_d the mass difference between the two B^0-$\overline{B}^0$ mass eigenstates. (We have assumed a lifetime difference $\Delta\Gamma_d = 0$ and neglected the $O(10^{-4})$ effects of CP violation in mixing.) The fact that the asymmetry depends only on Δt, and not on the absolute decay (measurement) times, t_1 and/or t_2, is a manifestation of EPR-type entanglement at a distance. It must be noticed that experimentally it is very difficult to measure the absolute times t_1 and t_2, hence only Δt is available.

Recently the question arose about the possibility to perform a Bell test with neutral K or B mesons. Several authors [7, 8] have suggested that the oscillation of a neutral mesons meson during a time Δt plays a role analogous to the choice of the angular orientation of a spin analyzer, while in a passive mode. If this is correct, then the time dependent asymmetry $A(\Delta t)$ (Eq. 8) can be thought as a B-meson version of the correlation $E(\phi)$, Eq. 5, with $\Delta m_d \Delta t \equiv \phi$. Like in the Aspect *et al.* experiment, the denominator of the expression for $A(\Delta t)$ accounts for the inefficiencies of the apparatus and, in addition, for the fast rate reduction due to the short B^0 lifetime. Following Ref. [7], the Bell-CHSH test can be performed considering an expression similar to Eq. 4. In this spin 1/2 system the maximal violation should appear for $\phi = \pi/4$ ($\Delta t \approx 2.6$ ps). The results presented in Ref. [9] are in agreement with QM and 3σ above the LR limit of 2. This approach was criticized by [10, 11]. These authors consider the Bell tests unaccessible due to the rapid decrease in time of the B-meson amplitudes, and because of the passive character of the flavour measurement.

Here we propose to demonstrate the existence of an intrinsic problem by exhibiting a model HV based capable of violating the proposed inequality. This is found to be the case for the Spontaneous and immediate Disentanglement model (SD), in which the B-meson pair separates into a B^0 and $\overline{B}^0$ with well-defined flavour immediately after the $\Upsilon(4S)$ decay, which then evolve independently [12].

In the SD model, the time-dependent asymmetry is

$$\begin{aligned} A_{\rm SD}(t_1, t_2) &= \cos(\Delta m_d t_1)\cos(\Delta m_d t_2) \\ &= \frac{1}{2}[\cos(\Delta m_d(t_1+t_2)) + \cos(\Delta m_d \Delta t)]. \end{aligned} \tag{9}$$

Note the additional $t_1 + t_2$ dependence, which can be removed by integrating the OF and SF functions for fixed values of Δt. The result is represented on Fig. 1 by the curve SD. From this result we can compute $|S|$ as function of Δt. It is found that this HV model violates LR limit of two, for $\Delta t \approx 1.3$ ps and $\Delta t \approx 4.2$. This proves that we are not in the presence of a genuine Bell-CHSH test and we are brought to abandon this approach.

To probe the non-local behaviour of the B^0 pair we can pragmatically limit ourselves to verify that, first, QM reproduces the experimental asymmetry, and, second, this is not the case for any other "reasonable" HV-based model. Within the definition of "reasonable" we include the capability to reproduce the B^0-$\overline{B}^0$ oscillation behaviour for each boson taken individually, after the $\Upsilon(4S)$ decay. In conclusion, we have chosen to compare our results with the predictions of QM and two other models. We stress the fact that to keep open the possibility of testing more models we also provide a fully corrected experimental time-dependent asymmetry, i. e. the background is subtracted and the detector effects corrected by a deconvolution method.

Our first candidate is the SD model seen before. The second is the local realistic model by Pompili and Selleri (PS) [13]. In PS, each B transports flavour information (B^0 or $\overline{B}^0$), and mass (corresponding to the heavy and light B_H, B_L eigenstates). There are thus four basic states: B^0_H, B^0_L, $\overline{B}^0_H$, $\overline{B}^0_L$. The model imposes mass and flavour anti-correlations at equal times $\Delta t = 0$; mass values are stable, but the system is programmed to allow random simultaneous jumps in flavour within the pair. The model is also required to reproduce the QM predictions for uncorrelated B-decays. No other assumptions are made: the result is an upper and a lower bound for

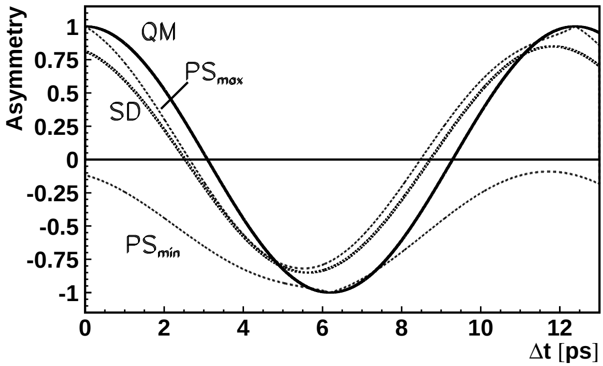


Figure 1: Time-dependent asymmetry predicted by (QM) quantum mechanics and (SD) spontaneous and immediate disentanglement of the B-pair, and (PS_{min} to PS_{max}) the range of asymmetries allowed by the Pompili and Selleri model. $\Delta m_d = 0.507\,\mathrm{ps}^{-1}$ is assumed.

the asymmetry. For instance, the upper bound is

$$\begin{aligned} \mathrm{A}_{\mathrm{PS}}^{\mathrm{max}}(t_1, t_2) = 1 - |\{1 \quad &- \cos(\Delta m_d \Delta t)\} \cos(\Delta m_d t_{\mathrm{min}}) \\ &+ \sin(\Delta m_d \Delta t) \sin(\Delta m_d t_{\mathrm{min}})|. \end{aligned} \tag{10}$$

The additional $t_{\mathrm{min}} = \min(t_1, t_2)$ dependence is again removed by integrating the OF and SF functions for fixed values of Δt. A similar function is given for$\mathrm{A}_{\mathrm{PS}}^{\mathrm{min}}$. We obtain the curves $\mathrm{PS}_{\mathrm{max}}$ and $\mathrm{PS}_{\mathrm{min}}$ shown in Fig. 1.

Finally, assuming QM as the correct model, we can consider hypothetical effects which can disturb the propagation of the entangled wave function [14, 15], and affect the time-dependent asymmetry. For instance, decoherence can originate from the "interaction" with a foamy space-time. The time evolution of the system state represented by a density matrix $\rho(t)$ can be described by a modified Liouville equation

$$\frac{d\rho}{dt} = i[\rho, H] + D[\rho], \tag{11}$$

where $D[\rho]$ represents a dissipative term. In the most simple approach a single "decoherence parameter" $\lambda \leq 0$ is introduced, $D[\rho] \sim \lambda$. From Ref. [16] the order of λ cannot be larger than $E^2/m_P l$, with E the typical energy scale of the mass system, emphasizing the interest to consider systems with the largest possible energy. In this framework Eq. 7 becomes

$$R_{\substack{OF\\SF}} = \frac{e^{-\Delta t \tau_{B^0}}}{4\tau_{B^0}}(1 \pm (1 - e^{\lambda t_{min}}) \cos(\Delta m_d \Delta t)). \tag{12}$$

The effect on $A(\delta t)$ is a modulation of the amplitude. Not having access to t_{min}, a simplified parametrisation of the global effect has been adopted. Making the hypothesis of a partial disentanglement into mass or flavour eigenstates the asymmetry becomes, respectively

$$A = (1 - \zeta_{B_H B_L}) A_{\mathrm{QM}}, \text{ and} \tag{13}$$

$$A = (1 - \zeta_{B^0\overline{B}^0}) A_{\mathrm{QM}} + \zeta_{B^0\overline{B}^0} A_{\mathrm{SD}} \tag{14}$$

(Eq. (13) corresponds to formula 3.5 in Ref. [17], for $\Delta\Gamma = 0$).

Data analysis

To determine the asymmetry we use 152×10^6 $B\overline{B}$ pairs collected by the Belle detector at the $\Upsilon(4S)$ resonance at the KEKB asymmetric-energy (3.5 GeV on 8.0 GeV) e^+e^- collider [18].

The Belle detector [19] is a large-solid-angle spectrometer consisting of a silicon vertex detector (SVD), central drift chamber (CDC), aerogel Cherenkov counters (ACC), time-of-flight counters (TOF) and a CsI(Tl) electromagnetic calorimeter (ECL) inside a 1.5T superconducting solenoid. The flux return is instrumented to detect K_L^0 and identify muons (KLM).

The $\Upsilon(4S)$ is produced with $\beta\gamma = 0.425$ close to the z axis. As the B momentum is low in the $\Upsilon(4S)$ center-of-mass system (CMS), Δt can be determined from the z-displacement of B-decay vertices: $\Delta t \approx \Delta z / \beta\gamma c$. The SVD provides Δz with a precision of about 100 μm.

The event selection for this study (see Ref. [20] for details) was optimized for theoretical model discrimination. To enable direct comparison of the result with different models, we subtract both background and mistagged-flavour events from the data, and then correct for detector effects by deconvolution. The flavour of one neutral B was obtained by reconstructing the decay $B^0 \to D^{*-}\ell^+\nu$, with $D^{*-} \to \overline{D}^0\pi_s^-$ and $\overline{D}^0 \to K^+\pi^-(\pi^0)$ or $K^+\pi^-\pi^+\pi^-$ (charge-conjugate modes are included throughout this paper). The D^0 candidates must have a reconstructed mass compatible with the known value. A D^* is formed by constraining a D^0 and a slow pion to a common vertex. We require a mass difference $M_{\mathrm{diff}} = M_{Kn\pi\pi_s} - M_{Kn\pi} \in [144.4, 146.4]$ MeV$/c^2$ (Fig. 2, left), and CMS momentum $p^*_{D^*} < 2.6$ GeV$/c$, consistent with B-decay.

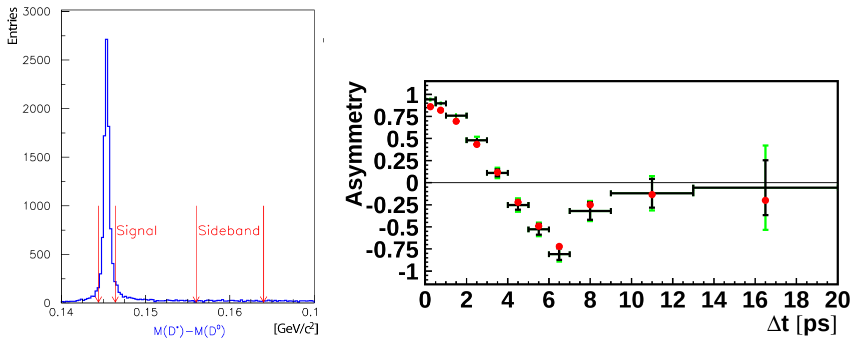


Figure 2: Left: M_{diff} distribution. Right: asymmetries before (red dots) and after (crosses) the corrections for the background and wrong flavour events. Statistical (black) and total errors (green) are superimposed.

The CMS angle between the D^* and lepton needs to be greater than $90°$. From the relation $M_\nu^2 = (E_B^* - E_{D^*\ell}^*)^2 - |\vec{p}_B^{\,*}|^2 - |\vec{p}_{D^*\ell}^{\,*}|^2 + 2|\vec{p}_B^{\,*}||\vec{p}_{D^*\ell}^{\,*}|\cos(\theta_{B,D^*\ell})$, where $\theta_{B,D^*\ell}$ is the angle between $\vec{p}_B^{\,*}$ and $\vec{p}_{D^*\ell}^{\,*}$, we can reconstruct $\cos(\theta_{B,D^*\ell})$ by assuming a vanishing neutrino mass. We require $|\cos(\theta_{B,D^*\ell})| < 1.1$. The neutral B decay position is determined by fitting the lepton track and D^0 trajectory to a vertex, constrained to lie in the e^+e^- interaction region. The remaining tracks are used to determine the second B decay vertex and flavour(see Ref. [21]; in this analysis we use only the high purity leptonic tags).

In total 8565 events are selected (6718 OF, 1847 SF). To compensate for the rapid fall in event rate with Δt, the time-dependent distributions are histogrammed in 11 variable-size bins (see Table 1). The raw asymmetry is shown in Fig. 2, right (dots). Background subtraction is then performed bin-by-bin; systematic errors are likewise determined by estimating variations in the OF and SF distributions, and calculating the effect on the asymmetry.

A GEANT-based Monte Carlo (MC) sample was analysed with identical criteria, and used for consistency checks, background estimates and subtraction, and to build deconvolution matrices.

Four types of background events have been considered: $e^+e^- \to q\bar{q}$ continuum, fake D^*, wrong D^*–lepton combinations, and $B^+ \to \overline{D}^{**0}\ell\nu$ events. Off-resonance data (8.3 fb^{-1}) were used to estimate the continuum background, which was found to be negligible. Fake D^0 reconstruction and misassigned slow pions producing a fake D^* background were estimated

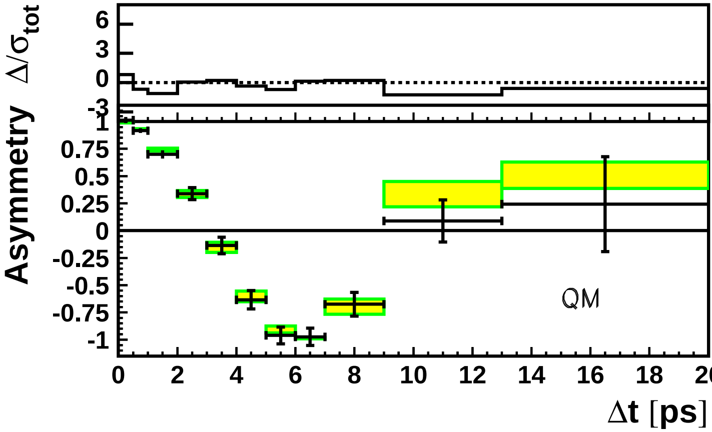
Table 1: Time-dependent asymmetry in Δt bins, corrected for experimental effects, with total uncertainties.

bin	window [ps]	A and total error	bin	window [ps]	A and total error
1	0.0 – 0.5	1.013 ± 0.028	7	5.0 – 6.0	-0.961 ± 0.077
2	0.5 – 1.0	0.916 ± 0.022	8	6.0 – 7.0	-0.974 ± 0.080
3	1.0 – 2.0	0.699 ± 0.038	9	7.0 – 9.0	-0.675 ± 0.109
4	2.0 – 3.0	0.339 ± 0.056	10	9.0 – 13.0	0.089 ± 0.193
5	3.0 – 4.0	-0.136 ± 0.075	11	13.0 – 20.0	0.243 ± 0.435
6	4.0 – 5.0	-0.634 ± 0.084			

from the sideband in M_{diff} (Fig. 2, left). The contamination from wrong D^*–lepton combinations was obtained by a reverse lepton momentum method, the validity of which was confirmed by MC studies. A fit of the $\cos(\theta_{B,D^*\ell})$ distribution allows the extraction of the D^{**-} component. The MC is then used to compute the fraction from charged B mesons which must be subtracted (as it has no mixing).

After correction for wrong flavour assignments (an event fraction of 0.015 ± 0.005) using OF and SF distributions from wrongly-tagged MC events, we obtain the time-dependent asymmetry shown in Fig. 2, right (crosses).

Remaining experimental effects (e.g. resolution in Δt, selection efficiency) are corrected by a deconvolution procedure based on the singular value decomposition method described in Ref. [22]. 11×11 response matrices are built separately for SF and OF events, using MC $D^*\ell\nu$ events indexed by generated and reconstructed Δt values. The procedure has been optimised, and its associated systematic errors inferred by a toy Monte Carlo where sets of several hundred simulated experiments are generated with data and MC samples identical in size to those of the real experiment, but assuming the three theoretical models. We test the consistency of the method applied to our data by fitting the B^0 decay time distribution (summing OF and SF samples), leaving the B^0 lifetime as a free parameter. We obtain 1.532 ± 0.017(stat) ps, consistent with the world average [23]. We have also repeated the deconvolution procedure using a subset of events with better vertex fit quality, and hence more precise Δt values: consistent results are obtained. The final results are shown in Table 1 and Fig. 3.



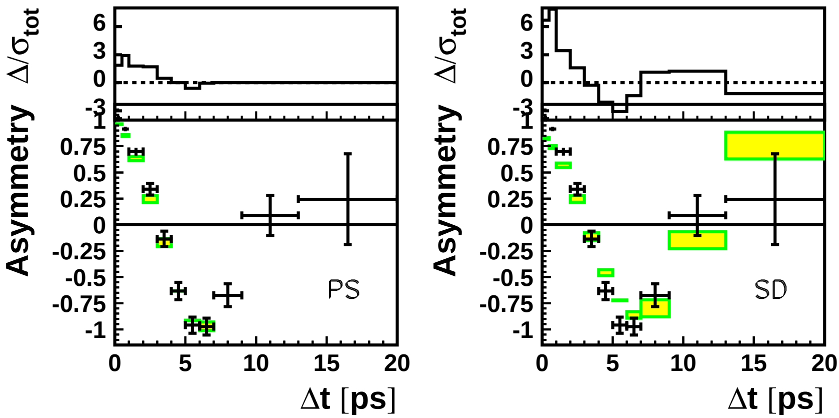


Figure 3: The bottom part of each plot represents the time-dependent flavour asymmetry (crosses) and the results of weighted least-squares fits to the model (the rectangles, showing $\pm 1\sigma$ errors on Δm_d). The top part show the differences $\Delta \equiv A_{\rm data} - A_{\rm model}$ in each bin, divided by the total experimental error $\sigma_{\rm tot}$. In the case of the PS model (bottom left) bins where $\mathrm{A}_{\rm PS}^{\rm min} < A_{\rm data} < \mathrm{A}_{\rm PS}^{\rm max}$ have been assigned a null deviation: see the text.

Comparison with the theoretical models

The model testing is done by a least-square fit to $A(\Delta t)$, leaving Δm_d free, but taking the world-average Δm_d into account. To avoid bias, we discard BaBar and Belle measurements, which assume QM correlations: this yields [24] $\langle \Delta m_d \rangle = (0.496 \pm 0.014)\,\text{ps}^{-1}$.

Our data is in agreement with the prediction of QM: we obtain $\Delta m_d = 0.501 \pm 0.009$ ps^{-1} with $\chi^2 = 5.2$ for 11 dof (see Fig. 3). SD is rejected by $\chi^2 = 174$ ($\Delta m_d = 0.419 \pm 0.008$). To fit PS we have used the closest boundary to our data $\mathrm{A}_{\mathrm{PS}}^{\mathrm{max}}$, Eq. (11), or $\mathrm{A}_{\mathrm{PS}}^{\mathrm{min}}$, but assumed a null deviation for data falling inside the boundaries. We obtain $\chi^2 = 31.3$ ($\Delta m_d = 0.447 \pm 0.010$ ps^{-1}): the data favour QM over PS at the 5.1σ level.

As noted above, CP violation in mixing can be neglected. Introducing a lifetime difference $\frac{\Delta\Gamma_d}{\Gamma_d} = 0.009 \pm 0.037$ [24] has a negligible effect on the fit.

Decoherence studies

We have examined the possibility of a partial loss of coherence just after the decay of the $\Upsilon(4S)$ resonance.

The fraction of events with disentangled B^0 and a $\overline{B}^0$ can be estimated by fitting our asymmetry with the mixture of Eq. (14), leaving $\zeta_{B^0\overline{B}^0}$ free. The fit finds $\zeta_{B^0\overline{B}^0} = 0.029 \pm 0.057$, consistent with no decoherence.

The second possibility considered is a partial decoherence into mass eigenstate, for which we expect a reduction in the amplitude of $A(\Delta t)$, as given by Eq. (13). The result of a fit gives a value of $\zeta_{B_H B_L} = 0.004 \pm 0.017$ (preliminary), also compatible with zero.

Conclusion

The neutral B-meson pair produced at the $\Upsilon(4S)$ should exhibit a non-local correlation of the EPR type. Ideally a Bell test should be performed to check that hidden variables are not active in the system. On the other hand we have seen that there is little hope to perform such a test. *Ultima ratio* we have decided to compare our data to QM and to two local realistic models: the model of Pompili and Selleri and a model with spontaneous and immediate disentanglement in which definite-flavour B^0 and $\overline{B}^0$ evolve independently.

We have measured with the Belle apparatus neutral B pairs from $\Upsilon(4S)$ decay. We have determined the time-dependent asymmetry due to flavour oscillations. The distribution has been corrected for experimental effects: the background has been subtracted and the resolution effects corrected by a deconvolution model, in such a way that the resulting distribution can be directly compared to theoretical models. We have found that QM reproduces our results well, while the two other models are strongly disfavoured.

The observation of partial decoherence in EPR systems could be a signal for New Physics. We have studied a possible disentanglement into mass or flavour eigenstates. We have found that our data is consistent with a null fraction of events with a loss of entanglement.

References

[1] A. Einstein, B. Podolski and N. Rosen, Phys. Rev. **47** (1935) 777.
[2] D. Bohm, *Quantum Theory* (Prentice Hall, Englewood Cliffs, NJ, 1951), pp. 614–622.
[3] J.S. Bell, Physics **1** (1964) 195.
[4] A. Aspect, P. Grangier and G. Roger, Phys. Rev. Lett. **47** (1981) 460; A. Aspect, J. Dalibard and G. Roger, Phys. Rev. Lett. **49** (2) (1982) 91-94; A more recent experiment is described in M.A. Rawe *et al.*, Nature **409** (2001) 791.
[5] A. Apostolakis *et al.* (CPLEAR Collaboration), Phys. Lett. B **422** (1998) 339; F. Ambrosino *et al.* (KLOE Collaboration), Phys. Lett. B **642** (2006) 315.
[6] A. Datta and D. Home, Phys. Lett. A **119** (1986) 3.
[7] N. Gisin and A. Go, Am. J. Phys. **69** (2001) 264.
[8] A. Bramon and M. Nowakowski, Phys. Rev. Lett. **83** (1999) 1.
[9] A. Go, Journal of Modern Optics **51** (2004) 991.
[10] R.A. Bertlmann, A. Bramon, G. Garbarino and B.C. Hiesmayr, Phys. Lett. A **332** (2004) 355.
[11] A. Bramon, R. Escribano and G. Garbarino, J. Mod. Opt. **52** (2005) 1681.
[12] The model is inspired by W.H. Furry, Phys. Rev. **49** (1936) 393.
[13] A. Pompili and F. Selleri, Eur. Phys. J. C **14** (2000) 469.
[14] R. Omnès, Rev. Mod. Phys. **64** (1992) 339.
[15] R.A. Bertlmann, Lect. Notes Phys. **689** (2006) 1 and references within, in particular Ref. 42; R.A. Bertlmann and W. Grimus, Phys. Rev. D **64** (2001) 056004.
[16] J. Ellis *et al.*, Phys. Rev. D **53** (1996) 3846.
[17] R.A. Bertlmann, W. Grimus and B.C. Hiesmayr, Phys. Rev. D **60** (1999) 114032.
[18] S. Kurokawa and E. Kikutani, Nucl. Instr. Meth. A **499** (2003) 1 and other papers in this volume.
[19] A. Abashian *et al.* (Belle Collaboration), Nucl. Instr. Meth. A **479** (2002) 117.
[20] A. Go, A. Bay *et al.* (BELLE Collaboration), Phys. Rev. Lett. **99** (2007) 131802, [arχiv:quant-ph/0702267] and references within.

[21] K. Abe *et al.* (Belle Collaboration), Phys. Rev. Lett. **87** (2002) 091802; Phys. Rev. D **66** (2002) 32007; H. Kakuno *et al.*, Nucl. Instr. Meth. A **533** (2004) 516.
[22] A. Höcker and V. Kartvelishvili, Nucl. Instr. Meth. A **372** (1996) 469.
[23] Particle Data Group, W.-M. Yao *et al.*, J. Phys. G **33** (2006) 1.
[24] E. Barberio *et al.* (Heavy Flavour Averaging Group), [arχiv:hep-ex/0603003].

Section IV: *CP* and *T* Violation

CP **and** ***T*** **violation measurements**
experimental tests of ***CPT*** **symmetry in the neutral kaon system**
CP **violation in B meson decays**
CP **violation measurements at Large Hadron Collider**

Time and Matter
2007 Bled, Slovenia

CP and *T* Violation with *K* Mesons

M.S. SOZZI*
University of Pisa and INFN, Largo Pontecorvo 3, 56126 Pisa, Italy

Abstract: Recent measurements of charge-conjugation parity and time-reversal symmetry violations in the kaon system are reviewed and discussed, and ongoing and future activities in the field are briefly described.

Introduction

Since its discovery in 1964 [1] the investigation of *CP* violation was always high on the agenda of fundamental research in particle physics, although it is only in the last decade that the unceasing experimental efforts actually led to the gathering of qualitatively new pieces of information. Still, however, the deep meaning and origin of this subtle effect remains unclear, although the recognition of the importance of its rôle in Nature grew steadily.

This paper will try to give a broad and shallow, but fairly complete, overview of the investigations performed on *CP* violation (and its close counterpart, *T* violation) using the system of *K* mesons, which is uniquely attractive for this purpose. In the spirit of the Time and Matter conference, the review is addressed to interested non-specialists; the interested reader can find many more details e.g. in [2].

Kaons: some history

K mesons (kaons) are the lightest form of "flavoured" matter, as they contain a quark or an anti-quark from one of the heavier families which are not present in ordinary matter, namely the strange quark. For this very reason they played a central role in shaping what we now know as the Standard

* marco.sozzi@cern.ch

Model (SM) of particle physics: one should just recall the Cabibbo angle, the fall of parity, the prediction of the existence of the charm quark and of course CP violation. Arguably, no other particle extended our understanding of the structure of Nature at the fundamental level more than the kaon, as it opened the field of flavour physics.

The tiny violation of the combined parity and charge-conjugation symmetries (CP violation) first manifested itself in a not-so-straightforward way as the evidence of two-pion decay of the long-lived neutral K meson, a physical state which was known to decay copiously to the three-pion state of opposite CP-parity.

The importance of CP violation needs not being recalled here: let us just note that it is the tiniest observed violation of a discrete symmetry and, as a required ingredient in any (not-too-weird) cosmological baryogenesis recipe, it is deeply related to the observed matter-antimatter asymmetry of the Universe, and (through the "sacred" CPT theorem) to microscopic time reversal violation. The involvement of charge conjugation implies that the study of CP symmetry is only accessible to high-energy physics experiments.

The importance of the kaon system for CP violation does not stem only from it being the one in which such phenomenon was originally discovered, but also from the fact that it displays all the features of flavour physics, and in particular all the (known) kinds of CP violation (of the same magnitude as in other systems, e.g. B mesons), and moreover it is a rather simple system, easily accessible from an experimental point of view; in this sense kaons can be considered as the *minimal flavour laboratory*. Their main downside is related to their theoretical understanding: as it is often the case at low (with respect to the QCD scale) energies, the link from the observed properties to the fundamental parameters of the underlying theory is marred from nightmarish QCD difficulties in the theoretical treatments, which so far resulted in precise computations not being available for kaons.

All possible ways in which CP and T violation can manifest themselves can be and have been investigated with kaons: the existence of physical states which are not CP eigenstates (the aforementioned decay of the long-lived state to $K_L \to 2\pi$, showing it has a $CP = +1$ component), the transition between states with different CP eigenvalues ($K_2 \to 2\pi$, direct CP violation), the time asymmetry of virtual transition probablilities $P(K^0 \to \overline{K}^0) \neq P(\overline{K}^0 \to K^0)$, the differences in the behaviour of CP-conjugate states (charge asymmetries in $K^{\pm}$ decays), and the search for non-zero T-odd quantities (transverse lepton polarization in $K_{\ell 3}$ decays). In the following all the above signatures will be briefly discussed.

Progress in the investigation of CP violation showed a rather long hiatus, since after the initial discovery not much qualitatively new information was obtained for about 36 years, and the phenomenon remained quite elusive and moreover apparently confined to the neutral kaon system, despite a rather aggressive worldwide experimental program. Such a situation allowed the survival of the super-weak *ansatz* [3], according to which the only known manifestation of CP violation could be attributed to a tiny new kind of interaction, which would be enhanced by the sensitivity linked to the small mass difference between the neutral K states, while remaining practically unobservable anywhere else. The big question was therefore whether CP violation was indeed an intrinsic property of weak interactions (as suggested by the CKM picture in the SM) or rather a more exotic phenomenon. The key in answering this question laid in the search for CP violation in the decay amplitudes, or *direct* CP violation, i.e. an effect which could not possibly be attributed to a property of the decaying system itself (the neutral K), in contrast to the known *indirect* CP violation, parameterized by the quantity

$$|\epsilon| \sim \left| \frac{A(K_L \to \pi\pi)}{A(K_S \to \pi\pi)} \right| \simeq 2{\cdot}10^{-3} \tag{1}$$

and related to the virtual $K^0 - \overline{K}^0$ oscillations. Only the latter effect could be accounted for by a super-weak model, while the CKM scheme predicts the existence of both (although not quantitatively). This second manifestation of CP violation, which turned out to be much smaller than the first for the K system, in general requires the presence of two interfering amplitudes with interacting hadrons in the final state.

The final clarification of this long-standing puzzle concerning the existence of direct CP violation in Nature had to wait 1999, when the first results from the latest round of dedicated neutral kaon experiments were announced; eventually both the E832 (KTeV) experiment at Fermilab [4] and the NA48 experiment at CERN [5] proved that direct CP violation exists, confirming a (disputed) earlier indication of the NA31 CERN experiment [6]. The above experiments showed that CP violation is indeed present in the decay amplitudes of the long-lived neutral kaon to two-pion final states, as quantified by the small but non-zero parameter ϵ',

$$\epsilon' \simeq \frac{1}{3} \left[\frac{A(K_L \to \pi^+\pi^-)}{A(K_S \to \pi^+\pi^-)} - \frac{A(K_L \to \pi^0\pi^0)}{A(K_S \to \pi^0\pi^0)} \right]. \tag{2}$$

Averaging the most precise results on $\mathrm{Re}(\epsilon'/\epsilon)$ one obtains

$$\mathrm{Re}(\epsilon'/\epsilon) = (16.3 \pm 2.3)\cdot 10^{-4}. \tag{3}$$

A graphical depiction of the present data is shown in figure 1; the interested reader can find more details on this whole story for example in [7].

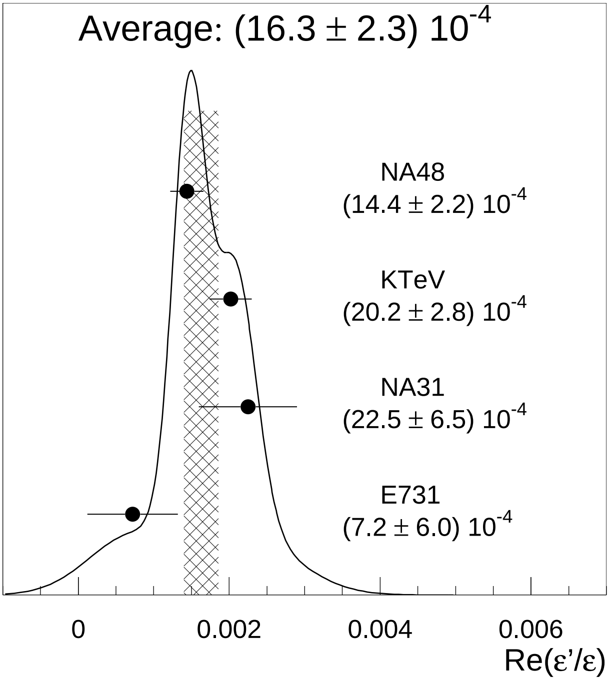


Figure 1: Ideogram of recent $\mathrm{Re}(\epsilon'/\epsilon)$ measurements. The curve shows the (unnormalized) probability distribution according to the PDG procedure [27]. The quoted error is inflated to reflect the unsatisfactory statistical consistency of the results.

Direct *CP* violation actually represents the most straightforward matter-antimatter asymmetry effect, and when the above result is rewritten as a partial decay width difference both its physical meaning and its numerical significance are clearer,

$$\frac{\Gamma(K^0 \to \pi^+\pi^-) - \Gamma(\overline{K}^0 \to \pi^+\pi^-)}{\Gamma(K^0 \to \pi^+\pi^-) + \Gamma(\overline{K}^0 \to \pi^+\pi^-)} = (5.04 \pm 0.82)\cdot 10^{-6}. \tag{4}$$

While the main importance of the above result is expressed by the fact that $\epsilon' \neq 0$ (with a significance which at present exceeds 7 standard deviations), regardless of its exact value, one should not oversee the fact that this parameter is now measured at the $\sim$ 15% level, and improvements on the precision are expected when the final result from the full KTeV statistics will be available.

Right after the above results appeared, a steady flow of CP-violating measurements in the heavier system of neutral B mesons started emerging from B-factories [8] [9]: the main advantage of such system with respect to K is that among many more available decay modes some can be found which are better tractable theoretically, so that the experimental results can be used to extract information on the parameters of the underlying theory, namely the angles of the CKM unitarity triangles.

On the contrary, the theoretical difficulties (linked to strong interaction effects) in the computation of the ϵ and ϵ' parameters are formidable, and despite the good accuracy of the experimental results such parameters cannot yet be used as quantitative constraints on the Standard Model in a precise way, although progress is expected to come (since quite some years, actually) from lattice QCD simulation. What can be fairly said today concerning the comparison of ϵ' with the Standard Model is that the measured value is fully consistent with the theory, which indeed predicts such an effect to exist within an order of magnitude.

One point of interest in this respect is linked to the hypothesis of "antigravity", that is a possible difference in the gravitational interaction of matter and anti-matter, which periodically resurfaces in physics. The $K^0 - \overline{K}^0$ coupled system is a very precise interferometric system on which such hypothesis could be tested, as any difference in the gravitational coupling of the two particles would perturb the oscillation pattern; indeed the smallness of the experimental asymmetry was used to set limits on antigravity [10], but the same argument was later turned around to propose that the measured effect usually ascribed to CP violation might indeed be an (anti-)gravitational effect. The measurement of direct CP violation, not arising in particle-antiparticle oscillations, breaks the above connection, sending back the antigravity hypothesis to a limbo.

Another experimental approach for kaon experiments should be mentioned at this point, which is analogous to that of B meson factories: the KLOE experiment at the DAΦNE e^+e^- collider in Frascati, working from 1999 to 2006, produced entangled $K^0\overline{K}^0$ (and K^+K^-) pairs by running at the centre of mass energy of the ϕ resonance (1020 MeV). In this way several interesting measurements can be made with a technique which is quite complementary to the one used at hadron colliders. While a lack of luminosity was an obstacle to high-statistics measurements, and the original aim of measuring direct *CP* violation was not reached, KLOE provided many other interesting measurements, including the first evidence for QM interference in the kaon system.

An entangled-pair experiment as KLOE might in principle offer an alternative way of measuring direct *CP* violation parameters: by studying the correlated decay probability of the pair into two different hadronic states ($\pi^+\pi^-$ and $\pi^0\pi^0$) as a function of their time difference, both the real and imaginary part of ϵ'/ϵ (the latter quantity being linked to direct *CPT* violation) could be extracted [11], although much higher statistics is required than is available at KLOE.

More *CP* violation with neutral kaons

More traditional ways of measuring *CP* violation also saw a renaissance in recent years, as a byproduct of the high-statistics experiments discussed above. The $K_L \to \pi^+\pi^-$ decay amplitude was re-measured at KLOE, KTeV and NA48, reaching the 0.5% accuracy level, using different experimental approaches.

The radiative decays $K \to \pi\pi\gamma$ offer a different way of probing *CP*-violating asymmetries. While for the most part dominated by *Bremsstrahlung* which (being an EM effect) is not expected to introduce any new feature, such decays also exhibit photon emission from the (weak) decay vertex, which can be accessed experimentally at the high end of the centre-of-mass photon energy distribution. Such "direct emission" contributions were measured both in $K_L \to \pi^+\pi^-\gamma$ [12] and in $K^\pm \to \pi^\pm\pi^0\gamma$, where recently NA48/2 also observed the interference term between the two terms [13], a potential source for observing *CP* violation effects, estimated at the $\mathcal{O}(10^{-4})$ level. It should also be mentioned that more than 10 years ago the E731 experiment at FNAL studied $K_L - K_S$ interference in radiative decays behind a regenerator, obtaining a result fully consistent (at modest accuracy) with the known *CP* violation in $\pi\pi$ decays, thus showing no new sources of asymmetry [14].

Another more recent use of radiative decays is linked to the detection of the rare decay $K_L \to \pi^+\pi^-e^+e^-$ in the '90s by KTeV and NA48 (BR $\simeq 3 \cdot 10^{-7}$). Internal photon conversion allows the polarization analysis of the $\pi^+\pi^-\gamma$ decay, as the lepton plane orientation is correlated to the photon helicity; in this way a rather large asymmetry between the $\pi^+\pi^-$ and e^+e^- planes is expected to originate from *CP* violation (as polarization states are no longer summed over). Such asymmetry was indeed measured to be $\simeq$ 13%, consistent with indirect CP violation in $\pi\pi$ decay [15, 16] (see fig. 2).

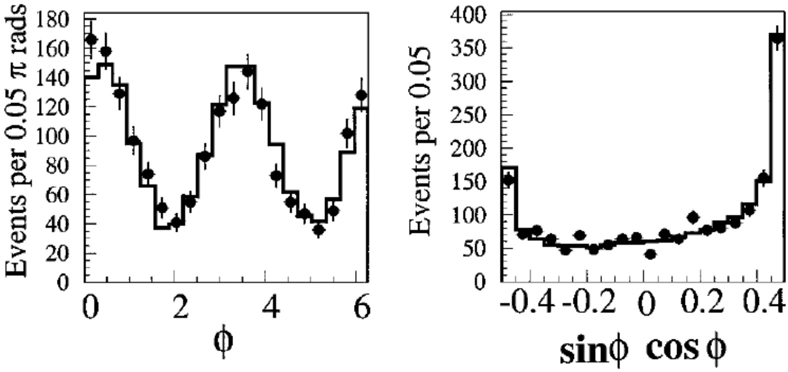


Figure 2: Distributions of the angle ϕ between the lepton and dipion planes in $K_L \to \pi^+\pi^-e^+e^-$ from KTeV.

A similar kind of measurement was also performed in the purely leptonic decays $K_L, K_S \to e^+e^-e^+e^-$ [17, 18], where no significant CP violation was detected, as expected.

Precise measurements of the charge asymmetry of semi-leptonic decay rates with several hundred millions K_L decays were obtained both by KTeV [19] and NA48 [20]. The results give

$$\delta_L^{(e)} \equiv \frac{\Gamma(K_L \to \pi^- l^+ \nu) - \Gamma(K_L \to \pi^+ l^- \overline{\nu})}{\Gamma(K_L \to \pi^- l^+ \nu) + \Gamma(K_L \to \pi^+ l^- \overline{\nu})} = (3.34 \pm 0.07)\cdot 10^{-3}, \qquad (5)$$

which, assuming CPT conservation, provides a measurement of the (indirect) CP violation parameter ϵ which is fully consistent (and more precise) that the one obtained from hadronic decays (no direct CP violation is possible in this decay mode due to lack of interfering amplitudes). The corresponding quantity for the $K_L \to \pi\mu\nu$ decay is measured to be fully consistent with the above, as expected from $\mu - e$ universality.

A new result was obtained by the KLOE experiment, which detected for the first time the semi-leptonic decays of K_S and measured their (indirect CP-violating) charge asymmetry in the $\pi e \nu$ mode to be [22],

$$\delta_S^{(e)} = (1.5 \pm 9.6 \pm 2.9)\cdot 10^{-3}. \qquad (6)$$

The equality of the above asymmetry with the precisely-measured one of K_L ($\delta_L(e)$) provides in principle a test of CPT violation, which is however still far from being statistically significant.

Besides the usual CP violation in $K_L \to 2\pi$ decays, searches were performed also in $K_S \to 3\pi$ decays (K_S would be expected to be a pure $CP = +1$ state if CP were conserved, while 3π states are dominantly $CP = -1$), where any effect is expected to be much smaller due to the large $K_S \to 2\pi$ decay rate. Interference of K_S and K_L decays into 3π final states was searched for at hadron machines (NA48/1 experiment) [24], while by exploiting pure tagged K_S decays the best current limit was set by KLOE [25],

$$\mathrm{BR}(K_S \to 3\pi^0) < 1.2 \cdot 10^{-7} \quad (90\%\,\mathrm{CL}), \tag{7}$$

still far from the expected figure $\approx 10^{-9}$.

Unitarity imposes a relation linking the (indirect) $CP-$ and $CPT-$ violating parameters to all the physical decay amplitudes: for the kaon system, with a limited number of decay modes, such Bell-Steinberger relation [21] is useful to obtain information on the symmetry-violating parameters.

The analysis of the consequences of the above relation is periodically repeated as new data becomes available; using the last experimental input this gives [26],

$$\mathrm{Re}(\epsilon) = (160.2 \pm 1.3) \cdot 10^{-5}, \qquad \mathrm{Im}(\delta) = (1.2 \pm 1.9) \cdot 10^{-5}, \tag{8}$$

where δ is the phenomenological parameter describing (indirect) CPT violation, linked to differences in mass and decay widths between K^0 and $\overline{K}^0$.

In this respect an older CERN experiment should also be mentioned, which provided a wealth of measurements on the kaon system with a different technique, akin to that used to study CP violation in B mesons at hadronic colliders. CPLEAR [23] exploited low-energy $\overline{p}p$ collisions to produce tagged $K^0, \overline{K}^0$ mesons and to study their time evolution. While unable to pursue the initial goal of (what else?) measuring direct CP violation, from 1990 to 1996 this experiment measured several quantities, some of them unique, greatly contributing to the knowledge of the strange flavour sector.

CP violation with charged kaons

Charged particles are free from mixing effects, and any CP asymmetry there would be necessarily of the direct kind. Again, theoretical predictions of direct CP-violating effects are quite difficult, and only order of magnitude estimates are usually available; the smallness of strong rescattering phases leads in general to tiny values for the asymmetries in the SM (below the 10^{-4} level).

While several asymmetry measurements were performed with charged kaons in the past, all with null results, until recently a large gap still remained between the ballpark values expected in the Standard Model and the experimental results, which allowed for possible large enhancements, actually predicted for some parameter values of new physics models, such as SUSY.

A great improvement in the accuracy of the measurement was achieved by the NA48/2 experiment, which studied the most copious charged kaon decay modes in which CP violation might be expected, namely $K^\pm \to \pi^\pm\pi^+\pi^-$ and $K^\pm \to \pi^\pm\pi^0\pi^0$ (with BR in the percent range).

Since decay rate asymmetries are expected to be suppressed and are also more difficult to measure experimentally, the experiment focused on the search for $K^\pm$ differences in the shapes of Dalitz plots, which for such 3-body decays are parameterized by slope parameters with respect to two Lorentz-invariant kinematical variables [27]. Control of systematics is of paramount importance for such kind of measurement, and NA48/2 exploited a unique configuration with two simultaneous, superimposed, narrow momentum spectra ($60 \pm 3\,\mathrm{GeV}/c$) K^+ and K^- beams (see fig. 3), and extracted asymmetries on the linear slope parameters g for the dependence of the decay rate on the kinematical variable corresponding to the CM energy of the odd-sign pion. These were obtained from the measured ratios of normalized kinematical distributions, equalizing acceptance differences mostly linked to the presence of an analyzing magnetic field by frequently reversing its polarity. In this way a robust measurement was obtained in two years of data-taking (2003-04), in which first-order instrumental asymmetries were eliminated and systematics from second-order effects were kept low enough to match the statistical precision obtained by collecting more than $3 \cdot 10^9$ $K^\pm$ decays.

The use of quadruple ratios of decay distributions, in which configurations with opposite magnetic fields orientations entered, allowed the cancellation of beam-related differential effects and global time instabilities, only leaving a sensitivity to small residual time differences of left-right instrumental asymmetries on short (few hours) timescale, whose effect was bounded by measurements of the possible corresponding sources and Monte Carlo estimates of the sensitivities.

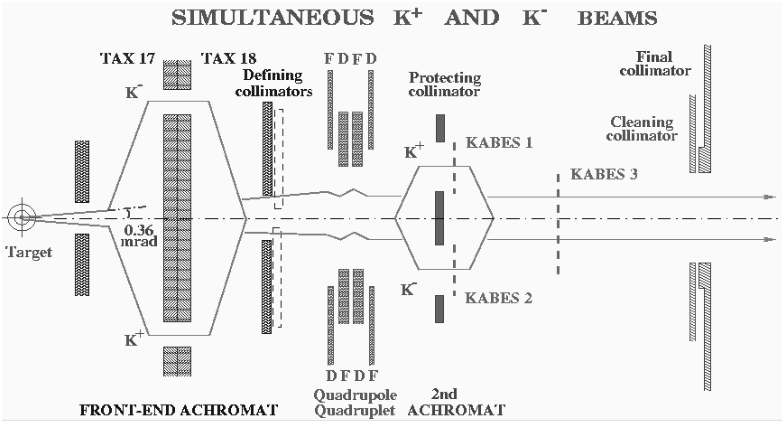


Figure 3: Schematic drawing of the beam configuration for the NA48/2 charged kaon experiment.

While the measurement for the decay mode with three charged pions only involved the magnetic spectrometer, that for the mode with two π^0 used only the electro-magnetic calorimeter, thus resulting in a completely indipendent measurement which, despite the lower branching ratio and acceptance, could match the former in precision thanks to the more favourable kinematical configuration.

The statistically-dominated final results from the experiment are consistent with no direct CP violation effects [28],

$$\frac{g_+ - g_-}{g_+ + g_-}(K^\pm \to \pi^\pm\pi^+\pi^-) = (-1.5 \pm 2.2)\cdot 10^{-4}, \tag{9}$$

$$\frac{g_+ - g_-}{g_+ + g_-}(K^\pm \to \pi^\pm\pi^0\pi^0) = (1.8 \pm 1.8)\cdot 10^{-4}, \tag{10}$$

and represent a ten-fold improvement on earlier results (see fig. 4), closing the gap for possible non-SM enhancements of direct CP-violation in these modes.

CPT conservation, or *T* violation

Since the early measurements of CP violation, tests were performed to check whether such phenomenon was accompanied by time-reversal vi-

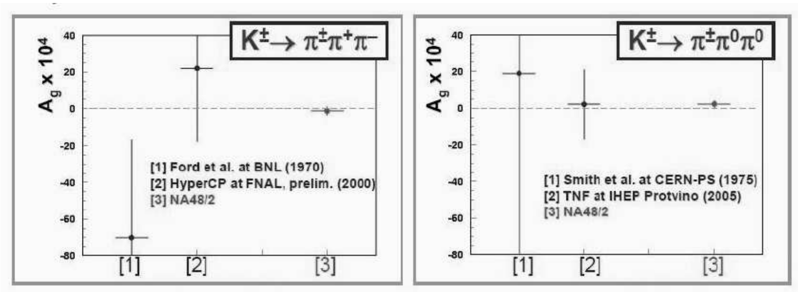


Figure 4: Results on Dalitz plot slope asymmetries in $K^{\pm} \to 3\pi$ decays.

olation – as CPT conservation would require – or CPT symmetry was actually violated. First indications that the former scenario is favoured came from the measurement of the phase difference of the $K_L, K_S \to 2\pi$ amplitudes in interference experiments: since CPT symmetry requires such phase difference to be close to a value dictated by the mass and decay width differences of the two physical states, which turns out to be close to $\pi/2$, while in the opposite case of T conservation (and CPT violation) such difference should be $\approx \pi$. The data, including the most recent measurements by KTeV [4] confirm that the usual indirect CP violation is not accompanied by (indirect) CPT violation,

$$\phi_{+-} = 43.4 \pm 0.7^\circ, \qquad \phi_{00} = 43.7 \pm 0.8^\circ, \tag{11}$$

but rather by T violation.

A direct test of time reversal symmetry was proposed long ago by P. Kabir, namely the comparison of the forward and backward probabilities of $K^0 - \overline{K}^0$ virtual transitions,

$$A_T = \frac{\Gamma(\overline{K}^0 \to K^0) - \Gamma(K^0 \to \overline{K}^0)}{\Gamma(\overline{K}^0 \to K^0) + \Gamma(K^0 \to \overline{K}^0)}. \tag{12}$$

Exploiting semi-leptonic decays of neutral kaons, the CPLEAR experiment could actually perform the first (and so far sole) measurement of time-reversal violation [29],

$$A_T = (6.6 \pm 1.3 \pm 1.0) \cdot 10^{-3}, \tag{13}$$

which is fully consistent with the value expected from the known mixing (indirect) CP-violation parameter ϵ, assuming CPT conservation.

Other approaches have been used in searching for T violation effects, such as the measurement of T-odd correlations in multi-body decays. In a 3-body final state a T-odd quantity involving spin can be built; the classic example, with a long history, is the lepton polarization transverse to the decay plane in $K \to \pi\ell\nu$ decays (ℓ being a lepton),

$$P_T^{(\ell)} = \frac{\boldsymbol{p}_\pi \times \boldsymbol{p}_\ell \cdot \boldsymbol{S}_\ell}{|\boldsymbol{p}_\pi \times \boldsymbol{p}_\ell||\boldsymbol{S}_\ell|}, \tag{14}$$

where $\boldsymbol{p}$ are momenta and $\boldsymbol{S}$ is spin. The above parameter can be non-zero due to the interference of two decay amplitudes with different phases (CP violation), which are only sizable in the case $\ell = \mu$. Experiments in the '70s reached the final-state interaction limit for K_L decays, while experiments continued on $K^+ \to \pi^0\mu^+\nu$ decay: E246 at KEK, an elegant third-generation precision experiment (see fig. 5), recently published its final result [30],

$$P_T^{(\mu)}(K^+ \to \pi^0\mu^+\nu) = (-1.7 \pm 2.3 \pm 1.1)\cdot 10^{-3}, \tag{15}$$

based on the analysis of 12 million decays of stopped K^+.

A planned experiment at the new Japanese proton facility J-PARC is expected to push further this technique to achieve a tenfold improvement in accuracy, approaching the final-state interaction limit.

T-odd quantities involving only momenta can be formed in 4-body decays, an example being

$$\xi = \frac{\boldsymbol{p}_\pi \times \boldsymbol{p}_\ell \cdot \boldsymbol{p}_\gamma}{|\boldsymbol{p}_\pi \times \boldsymbol{p}_\ell \cdot \boldsymbol{p}_\gamma|} \tag{16}$$

in $K \to \pi\ell\nu\gamma$ decays, in which final-state interaction effects are estimated to be at the 10^{-4} level. A null result with modest accuracy for the above quantity was obtained on a thousand events by the ISTRA experiment [31] and a measurement is underway by NA48/2 with ten times more data and a better control of systematics.

Future and conclusions

After having contributed in such a significant way to the shaping of the Standard Model of particle physics, the kaon system still appears to be an endless source of opportunities for investigating the mysteries of flavour physics, and new horizons are being opened.

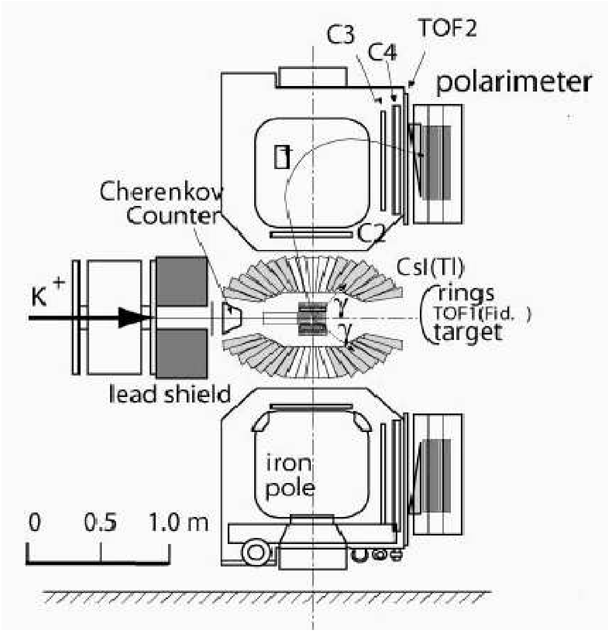


Figure 5: Schematic drawing of the principle of the KEK E246 experiment for measuring the transverse polarization of muons in $K^+ \to \pi^0 \mu^+ \nu$ decays.

It was realized that a class of flavour-changing neutral-current induced kaon decays holds a great potential as precision probe of the current picture of particle physics: these are decays of the kind $K \to \pi \ell \overline{\ell}$, in which strong-interaction effects are confined to one neutral-current vertex, which is accurately known from experiment, so that precise theoretical predictions can be made. Due to a well-known correlation in physics, unfortunately the expected branching ratios for these theoretically clean decays are in the 10^{-11} range, making their detection somewhat hard from an experimental point of view.

While $K^\pm \to \pi^\pm \ell^+ \ell^-$ are dominated by long-distance physics and thus useless for precision measurements, $K_L \to \pi^0 \ell^+ \ell^-$ are interesting, in that their short-distance part can be precisely computed; in this case, however, in order to isolate the interesting CP-violating short-distance contribution an indirect CP-violating and a CP-conserving part have to be measured (using ancillary decay modes) and subtracted. Experimental limits are available on these modes at the few 10^{-10} level, but further progress would require dedicated experiments.

Even more interesting are the decays with two neutrinos in the final state, in which no EM corrections are involved: $K^{\pm} \rightarrow \pi^{\pm}\nu\overline{\nu}$ and $K_L \rightarrow \pi^0\nu\overline{\nu}$. The former was detected (3 events) by the dedicated BNL experiment E787 [32], while only upper limits exist for the latter, which is CP-violating. For these two decay modes the theoretical predictions have the astounding non-parametric uncertainties of only a few percent, and thus they represent very powerful probes for testing the SM at very high precision.

The experimental challenges to be overcome in order to measure kinematically unconstrained decays with backgrounds 9-10 order of magnitudes larger than the signal are clearly formidable, but the high potential of such investigations deserves the efforts. Two new projects are underway to measure both the above "golden" decay modes in flight: NA62 at CERN [33] plans to collect ~ 80 K^+ decays in two years, and the follow-up to E391a at KEK [34], to be carried on at J-PARC, targets the K_L decays. Such experiments are really complementary to those at the energy frontier: when (if) LHC will eventually reveal the existence of new particles beyond the SM, their nature and flavour properties will only be acessible to experiments at the precision frontier, of which those mentioned above are the prime examples.

Acknowledgments

It is a pleasure to thank the organizers and participants to the Time and Matter conference – among the most challenging and mind-opening ones – for the very interesting atmosphere in the wonderful environment of Bled.

References

[1] J.H. Christenson *et al.*, Phys. Rev. Lett. **13** (1964) 138.
[2] M.S. Sozzi, *Discrete symmetries and CP violation* (Oxford University Press, 2008).
[3] L. Wolfenstein, Phys. Rev. Lett. **13** (1964) 562.
[4] A. Alavi-Harati *et al.* (KTeV collaboration), Phys. Rev. D **67** (2003) 012005.
[5] J.R. Batley *et al.* (NA48 collaboration), Phys. Lett. B **544** (2002) 97.
[6] G. Barr *et al.* (NA31 collaboration), Phys. Lett. B **317** (1993) 233.
[7] M.S. Sozzi and I. Mannelli, Riv. Nuovo Cim. **26(3)** (2003) 1, [arχiv:hep-ex/0312015].
[8] B. Aubert *et al.* (BABAR collaboration), Phys. Rev. Lett. **97** (2001) 091801.
[9] K. Abe *et al.* (BELLE collaboration), Phys. Rev. Lett. **97** (2001) 091802.

[10] M.L. Good, Phys. Rev. **121** (1961) 311.
[11] eds. L. Maiani *et al.*, The second DAΦNE physics handbook (INFN Frascati, 1995).
[12] E. Abouzaid *et al.* (KTeV collaboration), Phys. Rev. D **74** (2006) 032004.
[13] S. Goy Lopez (NA48/2 collaboration), Proceedings of 5th International Workshop on Chiral Dynamics, Theory and Experiment, Durham/Chapel Hill, USA, 18-22 Sep 2006.
[14] J.N. Matthews *et al.* (E731 collaboration), Phys. Rev. Lett. **75** (1995) 2803.
[15] A. Alavi-Harati *et al.* (KTeV collaboration), Phys. Rev. Lett. **84** (2000) 408.
[16] A. Lai *et al.* (NA48 collaboration), Eur. Phys. J. C **30** (2003) 33.
[17] A. Alavi-Harati *et al.* (KTeV collaboration), Phys. Rev. Lett. **86** (2001) 5425.
[18] A. Lai *et al.* (NA48 collaboration), Phys. Lett. B **615** (2005) 281.
[19] A. Alavi-Harati *et al.* (KTeV collaboration), Phys. Rev. Lett. **88** (2002) 181601.
[20] C. Lazzeroni (NA48 collaboration), Eur. Phys. J. C **33** (2004) s330.
[21] J.S. Bell and J. Steinberger, in Proceedings of Oxford International Conference on Elementary Particles, Rutherford Laboratory (1965), 195.
[22] F. Ambrosino *et al.* (KLOE collaboration), Phys. Lett. B **636** (2006) 173.
[23] A. Angelopoulos *et al.* (CPLEAR collaboration), Phys. Rep. **374** (2003) 165.
[24] A. Lai *et al.* (NA48 collaboration), Phys. Lett. B **610** (2005) 165.
[25] F. Ambrosino *et al.* (KLOE collaboration), Phys. Lett. B **619** (2005) 61.
[26] F. Ambrosino *et al.* (KLOE collaboration), JHEP **0612** (2006) 011.
[27] W.-M. Yao *et al.* (Particle Data Group), J. Phys. G **33** (2006) 1.
[28] J. R. Batley *et al.* (NA48/2 collaboration), Eur. Phys. J. C **52** (2007) 875.
[29] A. Angelopoulos *et al.* (CPLEAR collaboration), Phys. Lett. B **444** (1998) 43.
[30] M. Abe *et al.* (E246 experiment), Phys. Rev. D **73** (2006) 072005.
[31] V.N. Bolotov *et al.* (ISTRA+ collaboration), [arχiv:hep-ex/0510064]. (2005).
[32] S. Adler *et al.* (BNL E787 collaboration), Phys. Rev. Lett. **88** (2002) 041803.
[33] `http://na48.web.cern.ch/NA48/NA48-3`.
[34] `http://www-ps.kek.jp/e391`.

Time and Matter
2007 Bled, Slovenia

Experimental Tests of *CPT* Symmetry and Quantum Mechanics in the Neutral Kaon System

ANTONIO DI DOMENICO*
Dipartimento di Fisica, Università di Roma "La Sapienza", and I.N.F.N. Sezione di Roma, P.le A. Moro, 2, I-00185 Rome, Italy

Abstract: The neutral kaon system offers a unique possibility to perform fundamental tests of CPT invariance, as well as of the basic principles of quantum mechanics. The most recent limits on CPT violation are reviewed, including the ones related to possible decoherence mechanisms or Lorentz symmetry breaking, which might be induced by quantum gravity. Quantum coherence tests are also reviewed. The experimental results show no deviations from the expectations of quantum mechanics and CPT symmetry, while the accuracy in some cases reaches the interesting Planck's scale region. A possible demonstration of the quantum eraser principles using correlated kaon pairs at a ϕ-factory is briefly discussed. Finally, perspectives on this kind of experimental studies at an upgraded DAΦNE e^+e^- collider at Frascati are presented.

Introduction

The three discrete symmetries of quantum mechanics, C (charge conjugation), P (parity) and T (time reversal) are known to be violated in nature, both singly and in pairs. Only the combination of the three - CPT (in any order) - appears to be an exact symmetry of nature. An intuitive justification of this [1] can be based on the fact that our space-time is four dimensional, and that for an even dimensional space, from well known geometrical arguments, reflection of all axes is equivalent to a rotation. For instance, in the case of a plane, i.e. a two dimensional space, both coordinate axes change sign under total reflection, and exactly the same happens for a 180° rotation around the origin. It would therefore be tempting to assume that PT reflection is equivalent to a rotation in four dimensional

* antonio.didomenico@roma1.infn.it

space-time. In particular, for the rotation in question, all components of any 4-vector should change signs. However it can be easily verified that this does not happen, e.g. for the four-vector current j_μ. The reason is that our four dimensional space-time is *pseudo*euclidean, and the time coordinate is not exactly equivalent to a space coordinate. In order to restore the equivalence it can be shown [1] that it is necessary to add C conjugation, which e.g. changes the sign of the electromagnetic four-current, to the PT operation. So, it appears that in our pseudoeuclidean spacetime, it is indeed the CPT operation, and not simply PT, which is equivalent to the reflection of all four axes.

A rigorous proof of the CPT theorem can be found in Refs. [2, 3, 4, 5] (see also Refs. [6, 7, 8] for some recent developments). This theorem ensures that exact CPT invariance holds for any quantum field theory assuming (1) Lorentz invariance, (2) Locality and (3) Unitarity (i.e. conservation of probability). Testing the validity of CPT invariance therefore probes the most fundamental assumptions of our present understanding of particles and their interactions.

The neutral kaon doublet is one of the most intriguing systems in nature. During its time evolution a neutral kaon oscillates between its particle and antiparticle states with a beat frequency $\Delta m \approx 5.3 \cdot 10^9\ \mathrm{s}^{-1}$, where Δm is the small mass difference between the exponentially decaying states $\mathrm{K_L}$ and $\mathrm{K_S}$. The fortunate coincidence that Δm is about half the decay width of $\mathrm{K_S}$ makes possible to observe a variety of intricate interference phenomena in the production and decay of neutral kaons. In turn, such observations enable us to test quantum mechanics, the interplay of different conservation laws and the validity of various symmetry principles. In particular the extreme sensitivity of the neutral kaon system to a variety of CPT-violating effects makes it one of the best candidates for an accurate experimental test of this symmetry. As a figure of merit, the fractional mass difference $(m_{\mathrm{K}^0} - m_{\bar{\mathrm{K}}^0})/m_{\mathrm{K}^0}$ can be considered: it can be measured at the level of $\mathcal{O}(10^{-18})$ for neutral kaons, while, for comparison, a limit of $\mathcal{O}(10^{-14})$ can be reached on the corresponding quantity for the $\mathrm{B}^0 - \bar{\mathrm{B}}^0$ system, and only of $\mathcal{O}(10^{-8})$ for proton-antiproton [9].

The entanglement between the two kaons produced in $\phi \to \mathrm{K}^0\bar{\mathrm{K}}^0$ decays is a unique feature which could open up new horizons in the study of discrete symmetries, and of the basic principles of quantum mechanics. For instance possible CPT violations could manifest in conjunction with tiny decoherence effects, modifications of the initial correlation, or Lorentz symmetry violations, which, in turn, might be justified in a quantum theory of gravity. At a ϕ-factory the sensitivity to some observable effects can reach the level of the interesting Planck's scale region, i.e. $\mathcal{O}(m_K^2/M_{\text{Planck}}) \sim 2 \cdot 10^{-20}$ GeV, which is a very remarkable level of accuracy, presently unreachable in other similar systems (e.g. the $\mathrm{B}^0 - \bar{\mathrm{B}}^0$ system). Moreover recent theoretical studies demonstrated that entangled neutral kaons at a ϕ-factory are suitable to test the foundations of quantum mechanics, such as Bohr's complementarity principle, the quantum erasure and marking concepts, and the coherence of states over macroscopic distances, while for the more *classical* test with Bell's inequalities, new ideas have been put forward (see Ref. [10] for a detailed discussion on neutral kaon interferometry at a ϕ-factory).

The neutral kaon system

The time evolution of a neutral kaon that is initially a generic superposition of K^0 and $\bar{\mathrm{K}}^0$,

$$|\mathrm{K}(0)\rangle = a(0)|\mathrm{K}^0\rangle + b(0)|\bar{\mathrm{K}}^0\rangle \ , \tag{1}$$

can be described by the state vector

$$|\mathrm{K}(t)\rangle = a(t)|\mathrm{K}^0\rangle + b(t)|\bar{\mathrm{K}}^0\rangle + \sum_j c_j(t)|f_j\rangle \ , \tag{2}$$

where t is the time in the kaon rest frame, f_j's with $\{j = 1, 2, ...\}$ represent all possible decay final states, and $a(t)$, $b(t)$, and $c_j(t)$ are time dependent functions. In the Wigner-Weisskopf approximation [11], which is valid for times larger than the typical strong interaction formation time, the functions $a(t)$ and $b(t)$, describing the time evolution of the state in the $\{\mathrm{K}^0, \bar{\mathrm{K}}^0\}$ sub-space, obey the Schrödinger-like equation

$$i\frac{\partial}{\partial t}\begin{bmatrix} a(t) \\ b(t) \end{bmatrix} = \mathbf{H}\begin{bmatrix} a(t) \\ b(t) \end{bmatrix} , \tag{3}$$

where the effective Hamiltonian $\mathbf{H}$ is a 2×2 complex, not Hermitian, and time independent matrix. It can be decomposed in terms of its hermitian and anti-hermitian parts

$$\begin{aligned} \mathbf{H} &= \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = \\ &= \mathbf{M} - \frac{i}{2}\mathbf{\Gamma} = \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^* & M_{22} \end{bmatrix} - \frac{i}{2}\begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{12}^* & \Gamma_{22} \end{bmatrix} , \end{aligned} \tag{4}$$

where **M** and **Γ** are two hermitian matrices with positive eigenvalues, usually called *mass* and *decay* matrices, and indices 1 and 2 stand for K^0 and $\bar{\mathrm{K}}^0$, respectively.

The matrix **H** is characterized by eight independent real parameters; seven of them are observables, while an overall phase is arbitrary and unphysical, and can be fixed by convention.

The conservation of discrete symmetries constrains the matrix elements of **H**, and the following phase-invariant conditions hold[1]:

$$H_{11} = H_{22} \qquad \text{for } CPT \text{ conservation,} \tag{5}$$
$$|H_{12}| = |H_{21}| \qquad \text{for } T \text{ conservation,} \tag{6}$$
$$H_{11} = H_{22} \ \text{ and } \ |H_{12}| = |H_{21}| \qquad \text{for } CP \text{ conservation.} \tag{7}$$

The eigenvalues of **H** are

$$\begin{aligned} \lambda_S &= m_S - i\Gamma_S/2 \\ \lambda_L &= m_L - i\Gamma_L/2\,, \end{aligned} \tag{8}$$

where $m_{S,L}$ and $\Gamma_{S,L}$ are the masses and widths of the physical states, respectively. It is also useful to define the differences

$$\begin{aligned} \Delta m &= m_L - m_S > 0 \\ \Delta\Gamma &= \Gamma_S - \Gamma_L > 0 \end{aligned} \tag{9}$$

and the so called *superweak* phase

$$\tan\phi_{SW} = \frac{2\Delta m}{\Delta\Gamma}\,. \tag{10}$$

The physical states that diagonalize **H** are the short- and long-lived states; they evolve in time as pure exponentials

$$\begin{aligned} |\mathrm{K_S}(t)\rangle &= e^{-i\lambda_S t}|\mathrm{K_S}\rangle \\ |\mathrm{K_L}(t)\rangle &= e^{-i\lambda_L t}|\mathrm{K_L}\rangle\,, \end{aligned} \tag{11}$$

and are usually written as:

$$\begin{aligned} |\mathrm{K_S}\rangle &= \frac{1}{\sqrt{2(1+|\epsilon_S|^2)}}\{(1+\epsilon_S)|\mathrm{K}^0\rangle + (1-\epsilon_S)|\bar{\mathrm{K}}^0\rangle\} \\ |\mathrm{K_L}\rangle &= \frac{1}{\sqrt{2(1+|\epsilon_L|^2)}}\{(1+\epsilon_L)|\mathrm{K}^0\rangle - (1-\epsilon_L)|\bar{\mathrm{K}}^0\rangle\}\,, \end{aligned} \tag{12}$$

[1]For a general review on discrete symmetries in the neutral kaon system see Refs.[12, 13, 14, 15, 16, 17].

where $\epsilon_{S,L}$ are two small complex parameters describing the CP impurity in the physical states; one can equivalently define the parameters

$$\epsilon \equiv (\epsilon_S + \epsilon_L)/2 \ , \qquad \delta \equiv (\epsilon_S - \epsilon_L)/2 \, . \tag{13}$$

Ignoring negligible quadratic terms, they can be expressed in terms of the elements of $\mathbf{H}$ as:

$$\epsilon = \frac{H_{12} - H_{21}}{2(\lambda_S - \lambda_L)} = \frac{-i\Im M_{12} - \frac{1}{2}\Im\Gamma_{12}}{\Delta m + i\,(\Delta\Gamma)\,/2} \tag{14}$$

$$\delta = \frac{H_{11} - H_{22}}{2(\lambda_S - \lambda_L)} = \frac{\frac{1}{2}\left(M_{22} - M_{11} - \frac{i}{2}\,(\Gamma_{22} - \Gamma_{11})\right)}{\Delta m + i\,(\Delta\Gamma)\,/2} \, . \tag{15}$$

It is convenient to adopt a phase convention such that $\Im\Gamma_{12} = 0$, fixing the phase of ϵ to ϕ_{SW}, i.e. $\epsilon = |\epsilon| e^{i\phi_{SW}}$ (e.g. see Refs. [15, 18]). Then it is easy to show that

- $\delta \neq 0$ implies CPT violation;
- $\epsilon \neq 0$ implies T violation;
- $\epsilon \neq 0$ or $\delta \neq 0$ implies CP violation.

CPT violation in semileptonic decays

The semileptonic decay amplitudes can be parametrized as follows [12]:

$$\begin{aligned} \langle \pi^- l^+ \nu | T | \mathrm{K}^0 \rangle &= a + b \, , \qquad & \langle \pi^+ l^- \bar{\nu} | T | \bar{\mathrm{K}}^0 \rangle &= a^* - b^* \\ \langle \pi^+ l^- \bar{\nu} | T | \mathrm{K}^0 \rangle &= c + d \, , \qquad & \langle \pi^- l^+ \nu | T | \bar{\mathrm{K}}^0 \rangle &= c^* - d^* \end{aligned} \tag{16}$$

where a, b, c, d are complex quantities; CPT invariance implies $b = d = 0$, $\Delta S = \Delta Q$ rule implies $c = d = 0$, T invariance implies $\Im a = \Im b = \Im c = \Im d = 0$, while CP invariance implies $\Im a = \Re b = \Im c = \Re d = 0$. Then three measurable parameters can be defined:

$$y = -b/a \, , \qquad x_+ = c^*/a \, , \qquad x_- = -d^*/a \, ; \tag{17}$$

x_+ (x_-) describes the violation of the $\Delta S = \Delta Q$ rule in CPT conserving (violating) decay amplitudes, while y parametrizes CPT violation for $\Delta S = \Delta Q$ transitions. Then the semileptonic charge asymmetries for K_S and K_L states can be expressed as

$$\begin{aligned} A_{S,L} &= \frac{\Gamma(\mathrm{K_{S,L}} \to \pi^- l^+ \nu) - \Gamma(\mathrm{K_{S,L}} \to \pi^+ l^- \bar{\nu})}{\Gamma(\mathrm{K_{S,L}} \to \pi^- l^+ \nu) + \Gamma(\mathrm{K_{S,L}} \to \pi^+ l^- \bar{\nu})} \\ &= 2\Re\epsilon \pm 2\Re\delta - 2\Re y \pm 2\Re x_- \, . \end{aligned} \tag{18}$$

Any difference between A_S and A_L would signal a violation of the CPT symmetry:

$$A_S - A_L = 4(\Re\delta + \Re x_-) \ . \tag{19}$$

CPT violation in two pion decays

In the case of $K \to \pi^+\pi^-, \pi^0\pi^0$ decays, the decay amplitudes can be decomposed in terms of definite isospin $I = 0, 2$ of the final state:

$$\begin{aligned} \langle \pi\pi; I|T|\mathrm{K}^0\rangle &= (A_I + B_I)e^{i\delta_I} \\ \langle \pi\pi; I|T|\bar{\mathrm{K}}^0\rangle &= (A_I^* - B_I^*)e^{i\delta_I} \ . \end{aligned} \tag{20}$$

Here A_I (B_I) describe the CPT-conserving (CPT-violating) part of $\pi\pi$ decay amplitudes; δ_I is the $\pi\pi$ strong interaction phase shift for channel of total isospin I. The following observable decay amplitude ratios can be defined:

$$\begin{aligned} \eta_{+-} &\equiv |\eta_{+-}|e^{i\phi_{+-}} \equiv \frac{\langle \pi^+\pi^-|T|\mathrm{K_L}\rangle}{\langle \pi^+\pi^-|T|\mathrm{K_S}\rangle} = \epsilon_L + i\frac{\Im A_0}{\Re A_0} + \frac{\Re B_0}{\Re A_0} + \epsilon' \\ \eta_{00} &\equiv |\eta_{00}|e^{i\phi_{00}} \equiv \frac{\langle \pi^0\pi^0|T|\mathrm{K_L}\rangle}{\langle \pi^0\pi^0|T|\mathrm{K_S}\rangle} = \epsilon_L + i\frac{\Im A_0}{\Re A_0} + \frac{\Re B_0}{\Re A_0} - 2\epsilon' \end{aligned} \tag{21}$$

where

$$\epsilon' = \frac{1}{\sqrt{2}}e^{i(\delta_2-\delta_0)}\frac{\Re A_2}{\Re A_0}\left[i\left(\frac{\Im A_2}{\Re A_2} - \frac{\Im A_0}{\Re A_0}\right) + \left(\frac{\Re B_2}{\Re A_2} - \frac{\Re B_0}{\Re A_0}\right)\right] \ ; \tag{22}$$

$\epsilon' \neq 0$ signals direct CP violation (see Refs.[12, 19, 20, 21]). It can be shown that

$$\begin{aligned} \phi_{00} - \phi_{+-} &\simeq \frac{3}{\sqrt{2}|\eta_{+-}|}\frac{\Re A_2}{\Re A_0}\left(\frac{\Re B_2}{\Re A_2} - \frac{\Re B_0}{\Re A_0}\right) \approx -3\Im\left(\frac{\epsilon'}{\epsilon}\right) \ , \\ \phi_{+-} - \phi_{SW} &\simeq \frac{-1}{\sqrt{2}|\eta_{+-}|}\left(\frac{M_{11}-M_{22}}{2\Delta m} + \frac{\Re B_0}{\Re A_0}\right) \ . \end{aligned} \tag{23}$$

Therefore any phase difference between the η_{+-} and η_{00} parameters is a signal of CPT violation in the $\pi\pi$ decay, while a difference between ϕ_{+-} and ϕ_{SW} is a signal of CPT violation in the mixing and/or decay.

The following quantity, obtained combining semileptonic and two pion decays parameters,

$$\Re\left(\frac{2}{3}\eta_{+-} + \frac{1}{3}\eta_{00}\right) - \frac{A_L}{2} = \Re\left(y + x_- + \frac{\Re B_0}{\Re A_0}\right) \tag{24}$$

signals CPT violation if different from zero.

Experiments

In the CPLEAR experiment [18], a high flux of antiprotons is stopped in a gaseous hydrogen target. From the antiproton annihilation process, the following rare reactions with a branching fraction of about 0.4% are selected:

$$\begin{aligned}(\mathrm{p\bar{p}})_{\text{at rest}} &\rightarrow \mathrm{K}^0 + \mathrm{K}^- + \pi^+ \\ (\mathrm{p\bar{p}})_{\text{at rest}} &\rightarrow \bar{\mathrm{K}}^0 + \mathrm{K}^+ + \pi^- \ .\end{aligned} \tag{25}$$

The whole detector is embedded in a solenoidal magnet that provides a homogeneous longitudinal field of 0.44 T. The charged particles are measured by a series of cylindrical tracking detectors, followed by a particle identification detector (PID) for charged kaon identification, and electron/pion separation. The outermost detector is a 6 radiation lengths calorimeter used to detect photons from π^0 decays.
The CPLEAR detector was fully operational between 1992 and 1996, collecting a total of $1.1 \cdot 10^{13}$ antiproton interactions.

Two experiments, KTeV [22] at Fermilab and NA48 [23, 24] at CERN, have as a primary goal the measurement of direct CP violation in neutral kaon decays. Both experiments exploit the hadronic interactions of an intense high energy proton beam with a fixed target to produce neutral kaons.

In the KTeV experiment double kaon beams from a single BeO target are used to enable the simultaneous collection of $\mathrm{K_L}$ and $\mathrm{K_S}$ decays to minimize the systematics due to time variation of beam flux and detector inefficiencies. A precision magnetic spectrometer is used to minimize backgrounds in the $\pi^+\pi^-$ samples and to allow *in situ* calibration of the calorimeter with electrons. A high precision electromagnetic calorimeter, Cesium Iodide (CsI) array, is used for $\pi^0\pi^0$ reconstruction and better background suppression. Nearly hermetic photon vetoes (up to 100 mrad) are used for further background reduction for $\pi^0\pi^0$ mode. A $\mathrm{K_S}$ beam is produced by passing a $\mathrm{K_L}$ beam through a "regenerator" which is made of scintillator and is fully active to reduce the scattered background to the coherently regenerated $\mathrm{K_S}$.
The KTeV results presented here refers to data collected in the period 1996-1997.

DAΦNE, the Frascati ϕ-factory [25], is an e^+e^- collider working at a center of mass energy of $\sqrt{s} \sim 1020$ MeV, corresponding to the peak of the ϕ resonance. The ϕ-meson production cross section is $\sim 3\mu$b, and its decay into $\mathrm{K^0\bar{K}^0}$ has a branching fraction of 34%. The neutral kaon pair is produced in a coherent quantum state with the ϕ-meson quantum numbers $J^{PC} = 1^{--}$:

$$|i\rangle = \frac{1}{\sqrt{2}}\{|\mathrm{K}^0\rangle|\bar{\mathrm{K}}^0\rangle - |\bar{\mathrm{K}}^0\rangle|\mathrm{K}^0\rangle\} = \frac{N}{\sqrt{2}}\{|\mathrm{K_S}\rangle|\mathrm{K_L}\rangle - |\mathrm{K_L}\rangle|\mathrm{K_S}\rangle\}\,, \qquad (26)$$

where $N = \sqrt{(1+|\epsilon_S|^2)(1+|\epsilon_L|^2)}/(1-\epsilon_S\epsilon_L) \simeq 1$ is a normalization factor.

The detection of a kaon at large (small) times *tags* a $\mathrm{K_S}$ ($\mathrm{K_L}$) in the opposite direction. This is a unique feature at a ϕ-factory, not possible at fixed target experiments, that can be exploited to select pure $\mathrm{K_{S,L}}$ beams.

The KLOE detector consists mainly of a large volume drift chamber [26] surrounded by an electromagnetic calorimeter [27]. A superconducting coil provides a 0.52 T solenoidal magnetic field. At KLOE a $\mathrm{K_S}$ is tagged by identifying the interaction of the $\mathrm{K_L}$ in the calorimeter ($\mathrm{K_L}$-crash), while a $\mathrm{K_L}$ is tagged by detecting a $\mathrm{K_S} \to \pi^+\pi^-$ decay near the interaction point (IP). KLOE completed the data taking in March 2006 with a total integrated luminosity of $\sim 2.5\ \mathrm{fb}^{-1}$, corresponding to $\sim 7.5 \cdot 10^9$ ϕ-mesons produced.

"Standard" *CPT* symmetry tests

In this section the best experimental limits on the previously mentioned *CPT*-violating parameters (*standard* tests) are reviewed.

The CPLEAR collaboration measured the following decay rate asymmetry:

$$A_\delta(\tau) = \frac{\bar{R}_+(\tau) - \alpha R_-(\tau)}{\bar{R}_+(\tau) + \alpha R_-(\tau)} + \frac{\bar{R}_-(\tau) - \alpha R_+(\tau)}{\bar{R}_-(\tau) + \alpha R_+(\tau)} \qquad (27)$$

where

$$\begin{aligned}
R_+(\tau) &= R\left[\mathrm{K}^0{}_{t=0} \to (\pi^- e^+ \nu)_{t=\tau}\right], \\
R_-(\tau) &= R\left[\mathrm{K}^0{}_{t=0} \to (\pi^+ e^- \bar{\nu})_{t=\tau}\right], \\
\bar{R}_+(\tau) &= R\left[\bar{\mathrm{K}}^0_{t=0} \to (\pi^- e^+ \nu)_{t=\tau}\right], \\
\bar{R}_-(\tau) &= R\left[\bar{\mathrm{K}}^0_{t=0} \to (\pi^+ e^- \bar{\nu})_{t=\tau}\right],
\end{aligned} \qquad (28)$$

and $\alpha = (1 + 4\Re\epsilon_L)$. In the limit of large times one has $A_\delta(\tau \gg \tau_S) = 8\Re\delta$ (where $\tau_{S,L} = 1/\Gamma_{S,L}$), yielding the best measurement of $\Re\delta$ [28]:

$$\Re\delta = (0.30 \pm 0.33_{\mathrm{stat}} \pm 0.06_{\mathrm{syst}}) \cdot 10^{-3}\,. \qquad (29)$$

The KTeV collaboration exploited the coherent regeneration phenomenon, occurring when a kaon beam traverses a slab of material, which modifies a pure $\mathrm{K_L}$ beam into a coherent superposition of $\mathrm{K_L}$ and $\mathrm{K_S}$, i.e. $|\mathrm{K_L}\rangle \to |\mathrm{K_L}\rangle + \rho|\mathrm{K_S}\rangle$, where ρ is the *regeneration* (complex) parameter. The fit of the measured $\pi^+\pi^-$ (and $\pi^0\pi^0$) decay intensity downstream the regenerator with the function

$$\begin{aligned} R_{+-(00)}(t) \propto |\rho|^2 e^{-\Gamma_S t} + |\eta_{+-(00)}|^2 e^{-\Gamma_L t} + \\ + 2|\rho||\eta_{+-(00)}| e^{-(\Gamma_S+\Gamma_L)t/2} \cos\left[\Delta m t + \phi(\rho) - \phi_{+-(00)}\right], \end{aligned} \tag{30}$$

yields the best CPT tests using eqs. (23) [22],

$$\begin{aligned} \phi_{00} - \phi_{+-} &= (0.39 \pm 0.22_{\text{stat}} \pm 0.45_{\text{syst}})^\circ, \\ \phi_{+-} - \phi_{SW} &= (0.61 \pm 0.62_{\text{stat}} \pm 1.01_{\text{syst}})^\circ , \end{aligned} \tag{31}$$

consistent with no CPT violation. KTeV also measured the asymmetry A_L given in eq. (18) [29]

$$A_L = (3322 \pm 58_{\text{stat}} \pm 47_{\text{syst}}) \cdot 10^{-6} . \tag{32}$$

This result can be used in combination with two pion decay measurements in order to test CPT symmetry as in eq. (24) [29]:

$$\Re\left(y + x_- + \frac{\Re B_0}{\Re A_0}\right) = \Re\left(\frac{2}{3}\eta_{+-} + \frac{1}{3}\eta_{00}\right) - \frac{A_L}{2} = (-3 \pm 35) \cdot 10^{-6} . \tag{33}$$

The first measurement of the $\mathrm{K_S}$ semileptonic charge asymmetry has been performed by the KLOE collaboration analysing a part of the collected data (380 pb^{-1}) [30]:

$$A_S = (1.5 \pm 9.6_{\text{stat}} \pm 2.9_{\text{syst}}) \cdot 10^{-3}$$

The uncertainty on A_S can be reduced at the level of $\approx 3 \cdot 10^{-3}$ with the analysis of the full data sample of 2.5 fb^{-1}.

Using the values of A_L, $\Re\delta$, and $\Re\epsilon$ from other experiments the real part of the CPT violating parameters y and x_- (see eqs. (17)) can be evaluated [30]:

$$\begin{aligned} \Re x_- &= \frac{A_S - A_L}{4} - \Re\delta = (-0.8 \pm 2.5) \cdot 10^{-3} \\ \Re y &= \Re\epsilon - \frac{A_S + A_L}{4} = (0.4 \pm 2.5) \cdot 10^{-3} . \end{aligned} \tag{34}$$

The unitarity relation, originally derived by Bell and Steinberger [31]:

$$\left(\frac{\Gamma_S + \Gamma_L}{\Gamma_S - \Gamma_L} + i \tan\phi_{SW}\right)\left[\frac{\Re\epsilon}{1+|\epsilon|^2} - i\Im\delta\right] = \\ = \frac{1}{\Gamma_S - \Gamma_L}\sum_f A^*(K_S \to f)A(K_L \to f) \equiv \sum_f \alpha_f \quad (35)$$

can be used to bound $\Im\delta$, after having provided all the α_i parameters, Γ_S, Γ_L, and ϕ_{SW} as inputs. Using KLOE measurements, values from Particle Data Group (PDG) [9], and a combined fit of KLOE and CPLEAR data, the following result is obtained [32]:

$$\Re\epsilon = (159.6 \pm 1.3)\cdot 10^{-5}\,, \qquad \Im\delta = (0.4 \pm 2.1)\cdot 10^{-5}\,,$$

which is the best limit on $\Im\delta$, the main limiting factor of this result being the uncertainty on the phase ϕ_{+-} entering in the parameter $\alpha_{\pi^+\pi^-}$.
The limits on $\Im(\delta)$ and $\Re(\delta)$ can be used, through eq.(15), to constrain the mass and width difference between K^0 and $\bar{K}^0$. In the limit $\Gamma_{11} = \Gamma_{22}$, i.e. neglecting *CPT*-violating effects in the decay amplitudes, the best bound on the neutral kaon mass difference is obtained:

$$-5.3\cdot 10^{-19}\,\text{GeV} < M_{11} - M_{22} < 6.3\cdot 10^{-19}\,\text{GeV} \quad \text{at } 95\,\%\,\text{CL}\,.$$

Decoherence and *CPT* violation

The quantum interference between the two kaons initially in the *entangled* state in eq. (26) and decaying in the *CP* violating channel $\phi \to K_S K_L \to \pi^+\pi^-\pi^+\pi^-$, has been observed for the first time by the KLOE collaboration [33]. The measured Δt distribution, with Δt the absolute value of the time difference of the two $\pi^+\pi^-$ decays, can be fitted with the distribution:

$$I(\pi^+\pi^-,\pi^+\pi^-;\Delta t) \propto e^{-\Gamma_L\Delta t} + e^{-\Gamma_S\Delta t} \\ -2(1-\zeta_{SL})e^{-\frac{(\Gamma_S+\Gamma_L)}{2}\Delta t}\cos(\Delta m\Delta t)\,, \quad (36)$$

where the quantum mechanical expression in the $\{K_S, K_L\}$ basis has been modified with the introduction of a decoherence parameter ζ_{SL}, and a factor $(1-\zeta_{SL})$ multiplying the interference term. Analogously, a $\zeta_{0\bar{0}}$ parameter can be defined in the $\{K^0, \bar{K}^0\}$ basis [34]. After having included resolution and detection efficiency effects, having taken into account the background due to coherent and incoherent K_S-regeneration on the beam pipe wall, the small contamination of non-resonant $e^+e^- \to \pi^+\pi^-\pi^+\pi^-$ events, and keeping fixed in the fit Δm, Γ_S and Γ_L to the PDG values, the fit is performed on the Δt distribution with the following results:

$$\begin{aligned} \zeta_{SL} &= 0.018 \pm 0.040_{\text{stat}} \pm 0.007_{\text{syst}} \\ \zeta_{0\bar{0}} &= (1.0 \pm 2.1_{\text{stat}} \pm 0.4_{\text{syst}}) \cdot 10^{-6} \,, \end{aligned} \tag{37}$$

compatible with the prediction of quantum mechanics, i.e. $\zeta_{SL} = \zeta_{0\bar{0}} = 0$, and no decoherence effect. In particular the result on $\zeta_{0\bar{0}}$ has a high accuracy, $\mathcal{O}(10^{-6})$, due to the CP suppression present in the specific decay channel; it improves of five orders of magnitude the previous limit obtained by Bertlmann and co-workers [34] in a re-analysis of CPLEAR data [35]. This result can also be compared to a similar one recently obtained in the B meson system [36], where an accuracy of $\mathcal{O}(10^{-2})$ has been reached.

The decoherence mechanism can be made more specific in the case it is induced by quantum gravity effects. In fact one of the main open problem in quantum gravity is related to what is commonly known as the black hole *information-loss paradox*. In 1976 Hawking showed [37] that the formation and evaporation of black holes, as described in the semiclassical approximation, appear to transform pure states near the event horizon of black holes into mixed states. This corresponds to a loss of information about the initial state, in striking conflict with quantum mechanics and its unitarity description. At a microscopic level, in a quantum gravity picture, space-time might be subjected to inherent non-trivial quantum metric and topology fluctuations at the Planck scale ($\sim 10^{-33}$ cm), called generically *space-time foam*, with associated microscopic event horizons. As further suggested by Hawking himself [38], this space-time structure, might induce a pure state to evolve into a mixed one, i.e. decoherence of apparently isolated matter systems. This decoherence, in turn, necessarily implies, by means of a theorem [39], CPT violation, in the sense that the quantum mechanical operator generating CPT transformations cannot be consistently defined.

The information-loss paradox generated a lively debate during the last decades with no generally accepted solution. Even the recent proposed solutions in favor of no-loss and preservation of information do not completely solve the problem, some aspects of which still remaining a puzzle (see for instance Refs. [40, 41, 42]). It seems therefore extremely interesting to put experimental limits at the level of the Planck's scale region on possible decoherence effects.

The above mentioned decoherence mechanism lead Ellis and coworkers [43] to formulate a model in which a single kaon is described by a density matrix ρ that obeys a modified Liouville-von Neunmann equation:

$$\frac{d\rho}{dt} = -i\mathbf{H}\rho + i\rho\mathbf{H}^{\dagger} + L(\rho;\alpha,\beta,\gamma) \tag{38}$$

where the extra term $L(\rho;\alpha,\beta,\gamma)$ would induce decoherence in the system, and depends on three real parameters, α, β and γ, which violate CPT symmetry and quantum mechanics (they satisfy the inequalities α, $\gamma > 0$ and $\alpha\gamma > \beta^2$ - see Refs. [43, 44]). They have mass dimension and are guessed to be at most of $\mathcal{O}(m_K^2/M_{Planck}) \sim 2 \cdot 10^{-20}\,\text{GeV}$, where $M_{Planck} = 1\sqrt{G_N} = 1.22 \cdot 10^{19}$ GeV is the Planck mass.

The CPLEAR collaboration, studying the time behaviour of single neutral kaon decays to $\pi^+\pi^-$ and $\pi e \nu$ final states, obtained the following results [45]:

$$\begin{aligned} \alpha &= (-0.5 \pm 2.8) \cdot 10^{-17}\,\text{GeV} \\ \beta &= (2.5 \pm 2.3) \cdot 10^{-19}\,\text{GeV} \\ \gamma &= (1.1 \pm 2.5) \cdot 10^{-21}\,\text{GeV}\,. \end{aligned} \tag{39}$$

The KLOE collaboration, studying the same $I(\pi^+\pi^-, \pi^+\pi^-; \Delta t)$ distribution as in the ζ parameters analysis, obtained the following preliminary results [46]:

$$\begin{aligned} \alpha &= \left(-10^{+41}_{-31}\text{stat} \pm 9_{\text{syst}}\right) \cdot 10^{-17}\,\text{GeV} \\ \beta &= \left(3.7^{+6.9}_{-9.2}\text{stat} \pm 1.8_{\text{syst}}\right) \cdot 10^{-19}\,\text{GeV} \\ \gamma &= \left(-0.5^{+5.8}_{-5.1}\text{stat} \pm 1.2_{\text{syst}}\right) \cdot 10^{-21}\,\text{GeV} \end{aligned} \tag{40}$$

In the simplifying hypothesis of complete positivity [47], i.e. $\alpha = \gamma$ and $\beta = 0$, the KLOE result is [33]:

$$\gamma = \left(1.3^{+2.8}_{-2.4}\text{stat} \pm 0.4_{\text{syst}}\right) \cdot 10^{-21}\,\text{GeV}\,, \tag{41}$$

All results are compatible with no CPT violation, while the sensitivity approaches the interesting level of $\mathcal{O}(10^{-20}\,\text{GeV})$.

As discussed above, in a quantum gravity framework inducing decoherence, the CPT operator is *ill-defined*. This consideration lead Bernabeu, Mavromatos and Papavassiliou [48, 49] to investigate intriguing consequences in correlated neutral kaon states. In fact the resulting loss of particle-antiparticle identity could induce a breakdown of the correlation of state (26) imposed by Bose statistics. As a result the initial state (26) can be parametrized in general as:

$$|i\rangle = \frac{1}{\sqrt{2}}\left[|K^0\rangle|\bar{K}^0\rangle - |\bar{K}^0\rangle|K^0\rangle + \omega\left(|K^0\rangle|\bar{K}^0\rangle + |\bar{K}^0\rangle|K^0\rangle\right)\right] , \qquad (42)$$

where ω is a complex parameter describing a completely novel CPT violation phenomenon, not included in previous analyses. Its order of magnitude could be at most

$$|\omega| \sim \left[(m_K^2/M_{\text{Planck}})/\Delta\Gamma\right]^{1/2} \sim 10^{-3}$$

with $\Delta\Gamma = \Gamma_S - \Gamma_L$. A similar analysis performed by the KLOE collaboration on the same $I(\pi^+\pi^-, \pi^+\pi^-; \Delta t)$ distribution as before, including in the fit the modified initial state eq.(42), yields the first measurement of the complex parameter ω [33]:

$$\begin{aligned} \Re(\omega) &= \left(1.1^{+8.7}_{-5.3}\text{stat} \pm 0.9_{\text{syst}}\right) \cdot 10^{-4} \\ \Im(\omega) &= \left(3.4^{+4.8}_{-5.0}\text{stat} \pm 0.6_{\text{syst}}\right) \cdot 10^{-4} , \end{aligned} \qquad (43)$$

with an accuracy that already reaches the interesting Planck's scale region.

CPT violation and Lorentz symmetry breaking

CPT invariance holds for any realistic Lorentz-invariant quantum field theory. However a very general theoretical possibility for CPT violation is based on spontaneous breaking of Lorentz symmetry, as developed by Kostelecký [50, 51, 52], which appears to be compatible with the basic tenets of quantum field theory and retains the property of gauge invariance and renormalizability (Standard Model Extensions - SME). In SME for neutral kaons, CPT violation manifests to lowest order only in the parameter δ (e.g. B_I, y and x_- vanish at first order), and exhibits a dependence on the 4-momentum of the kaon:

$$\delta \approx i \sin\phi_{SW} e^{i\phi_{SW}} \gamma_K(\Delta a_0 - \vec{\beta}_K \cdot \Delta\vec{a})/\Delta m \qquad (44)$$

where γ_K and $\vec{\beta}_K$ are the kaon boost factor and velocity in the observer frame, and Δa_μ are four CPT- and Lorentz-violating coefficients for the two valence quarks in the kaon.

Following Ref. [51], the time dependence arising from the rotation of the Earth can be explicitly displayed in eq. (44) by choosing a three-dimensional basis $(\hat{X}, \hat{Y}, \hat{Z})$ in a non-rotating frame, with the $\hat{Z}$ axis along the Earth's rotation axis, and a basis $(\hat{x}, \hat{y}, \hat{z})$ for the rotating (laboratory) frame. The CPT violating parameter δ may then be expressed as:

$$\begin{aligned} \delta &= \frac{1}{2\pi}\int_0^{2\pi} \delta(\vec{p}, t) d\phi \\ &= \frac{i \sin\phi_{SW} e^{i\phi_{SW}}}{\Delta m} \gamma_K \{\Delta a_0 + \beta_K \Delta a_Z \cos\theta \cos\chi \\ &\quad + \beta_K (\Delta a_Y \sin\chi \cos\theta \sin\Omega t + \Delta a_X \sin\chi \cos\theta \cos\Omega t)\} , \end{aligned} \tag{45}$$

where Ω is the Earth's sidereal frequency, $\cos\chi = \hat{z} \cdot \hat{Z}$, θ and ϕ are the conventional polar and azimuthal angles defined in the laboratory frame about the $\hat{z}$ axis, and an integration on the azimuthal angle ϕ has been performed, assuming a symmetric decay distribution in this variable[2]. The sensitivity to the four Δa_μ parameters can be very different for fixed target and collider experiments, showing complementary features [51].

At KLOE the Δa_0 parameter can be measured through the difference $A_S - A_L$, by performing the measurement of each asymmetry with a symmetric integration over the polar angle θ, thus averaging to zero any possible contribution from the terms proportional to $\cos\theta$ in eq.(45),

$$A_S - A_L \simeq \left[\frac{4\Re\left(i \sin\phi_{SW} e^{i\phi_{SW}}\right)\gamma_K}{\Delta m} \right] \Delta a_0 \, . \tag{46}$$

In this way a first preliminary evaluation of the Δa_0 parameter can be obtained by KLOE [53],

$$\Delta a_0 = (0.4 \pm 1.8) \cdot 10^{-17}\,\text{GeV}. \tag{47}$$

With the analysis of the full KLOE data sample ($L = 2.5$ fb^{-1}) an accuracy $\sigma(\Delta a_0) \sim 7 \cdot 10^{-18}$ GeV could be reached.

At KLOE the Δa_Z parameter can be evaluated measuring the asymmetry:

$$A(\Delta t) = \tag{48}$$
$$\frac{I(\pi^+\pi^-(+), \pi^+\pi^-(-); \Delta t > 0) - I(\pi^+\pi^-(+), \pi^+\pi^-(-); \Delta t < 0)}{I(\pi^+\pi^-(+), \pi^+\pi^-(-); \Delta t > 0) + I(\pi^+\pi^-(+), \pi^+\pi^-(-); \Delta t < 0)},$$

where the two identical final states are distinguished by their emission in the forward ($\cos\theta > 0$) or backward ($\cos\theta < 0$) hemispheres (denoted by

[2]This simplifying assumption will be maintained throughout the following; however small non-symmetric ϕ angle effects could be easily included in the formulas without significantly modifying the main conclusions below.

the symbols $+$ and $-$, respectively), and Δt is the time difference between $(+)$ and $(-)$ $\pi^+\pi^-$ decays. A preliminary analysis based on the same data used for the measurement of the decoherence parameters [33], yields the first preliminary result on Δa_Z [53]:

$$\Delta a_Z = (-1 \pm 4) \cdot 10^{-17}\ \text{GeV}\ . \tag{49}$$

With the analysis of 2.5 fb^{-1} an accuracy $\sigma(\Delta a_Z) \sim 2 \cdot 10^{-17}$ GeV could be reached. An alternative and independent method for the evaluation of Δa_Z parameter at KLOE is based on the measurement of the A_L asymmetry (for details see Ref. [53]).

The Δa_X, Δa_Y and Δa_Z parameters can be all simultaneously measured by performing a proper sidereal time dependent analysis of the asymmetry (48). An accuracy $\sigma(\Delta a_{X,Y,Z}) = \mathcal{O}(1 \cdot 10^{-17}\ \text{GeV})$ could be reached with the analysis of the full KLOE data sample. However there exists a preliminary measurement performed by the KTeV collaboration [54] based on the search of sidereal time variation of the phase ϕ_{+-}, that constrains Δa_X and Δa_Y to less than $9.2 \cdot 10^{-22}$ GeV at 90% C.L.

Quantum eraser with neutral kaons at a ϕ-factory

The interference term in eq.(36) gives rise to a characteristic correlation between the two kaon decays. For instance, a complete destructive interference prevents the two kaons from decaying into the same final state at the same time, i.e. $I(\pi^+\pi^-, \pi^+\pi^-; \Delta t = 0) = 0$. This is a consequence of the antisymmetry of state (26). From an intuitive point of view, once produced, the two kaons can be viewed as two freely propagating independent particles. However even though the two decays can be regarded as separated space-like events (the kaons are produced with opposite momentum in the ϕ meson rest frame), it is like the kaon flying in the $+\vec{p}$ direction cannot "freely" decay into a certain final state f at a certain proper time t, but its behaviour depends on what the other kaon flying in the opposite $-\vec{p}$ direction does. This kind of correlation (*entanglement*) for neutral kaon pairs cannot be simply explained in terms of conservation laws, and is of the type first pointed out by Einstein, Podolsky and Rosen (EPR) in their famous paper [57].

This feature of the initial state (26) has long reaching consequences in terms of potentialities of the neutral kaon system in testing fundamental aspects of quantum mechanics. This can be easily understood by recognizing that the kaon state (26) is formally analogous to a singlet state of two spin 1/2 particles, the strangeness for a neutral kaon playing the same role of the spin *up* or *down* along a chosen direction. Recently it has been shown [58, 59] that kaon pairs produced at a ϕ-factory are suitable for the study of Bohr's complementarity principle with an interesting implementation of the quantum eraser [58, 59]. Under certain approximations, kaon strangeness or lifetime measurements project the state into two bases $\{K^0, \bar{K}^0\}$ or $\{K_S, K_L\}$, analogously to spin measurements along two orthogonal directions. The quantum erasure principle is established by measuring the strangeness of one kaon (object system) on one side, and the strangeness or lifetime of the other kaon (meter system) on the opposite side; the appearance of strangeness oscillations on the object depends on what is measured on the meter system, even if the measurement on the meter is performed at a later time than the measurement on the object (delayed choice). More details on the *kaonic* quantum eraser, and other interesting studies on possible Bell's inequality tests with neutral kaons at a ϕ-factory, can be found in Ref. [10] and references therein.

Future plans

A proposal [56] has been recently submitted for a physics program to be carried out with an upgraded KLOE detector, KLOE-2, at an upgraded DAΦNE machine, which is expected to deliver an integrated luminosity of the order of 50 fb^{-1}. The KLOE-2 program concerning neutral kaon interferometry is summarized in table 1, where the KLOE-2 statistical sensitivities to the main parameters that can be extracted from kaon decay time distributions (with different choices of final states) are listed in the hypothesis of an integrated luminosity $L = 50$ fb^{-1}, and compared to the best presently published measurements. At KLOE-2, an experimental demonstration of the *kaonic* quantum eraser, as mentioned above, is also foreseen.

Conclusions

The neutral kaon system constitutes an excellent laboratory for the study of the CPT symmetry and the basic principles of quantum mechanics. Several parameters related to possible CPT violations, including decoherence and Lorentz symmetry breaking effects, have been measured, in some cases with a precision reaching the interesting Planck's scale region. Simple quantum coherence tests have been also performed. All results are consistent with no violation of the CPT symmetry and/or quantum mechanics.

Table 1: KLOE-2 statistical sensitivities on several parameters.

parameter	best published meas.	KLOE-2 (50 fb^{-1})
A_S	$(1.5 \pm 11) \cdot 10^{-3}$	$\pm 1 \cdot 10^{-3}$
A_L	$(3322 \pm 58 \pm 47) \cdot 10^{-6}$	$\pm 25 \cdot 10^{-6}$
$\Re\frac{\epsilon'}{\epsilon}$	$(1.66 \pm 0.26) \cdot 10^{-3}$	$\pm 0.2 \cdot 10^{-3}$
$\Im\frac{\epsilon'}{\epsilon}$	$(1.2 \pm 2.3) \cdot 10^{-3}$	$\pm 3 \cdot 10^{-3}$
$(\Re\delta + \Re x_-)$	$\Re\delta = (0.29 \pm 0.27) \cdot 10^{-3}$	$\pm 0.2 \cdot 10^{-3}$
	$\Re x_- = (-0.8 \pm 2.5) \cdot 10^{-3}$	
$(\Im\delta + \Im x_+)$	$\Im\delta = (0.4 \pm 2.1) \cdot 10^{-5}$	$\pm 3 \cdot 10^{-3}$
	$\Im x_+ = (0.8 \pm 0.7) \cdot 10^{-2}$	
Δm	$5.288 \pm 0.043 \cdot 10^9 s^{-1}$	$\pm 0.03 \cdot 10^9 s^{-1}$
ζ_{SL}	$(1.8 \pm 4.1) \cdot 10^{-2}$	$\pm 0.2 \cdot 10^{-2}$
$\zeta_{0\bar{0}}$	$(1.0 \pm 2.1) \cdot 10^{-6}$	$\pm 0.1 \cdot 10^{-6}$
α	$(-0.5 \pm 2.8) \cdot 10^{-17}$ GeV	$\pm 2 \cdot 10^{-17}$ GeV
β	$(2.5 \pm 2.3) \cdot 10^{-19}$ GeV	$\pm 0.1 \cdot 10^{-19}$ GeV
γ	$(1.1 \pm 2.5) \cdot 10^{-21}$ GeV	$\pm 0.2 \cdot 10^{-21}$ GeV
		(compl. pos. hyp.)
		$\pm 0.1 \cdot 10^{-21}$ GeV
$\Re\omega$	$(1.1^{+8.7}_{-5.3} \pm 0.9) \cdot 10^{-4}$	$\pm 2 \cdot 10^{-5}$
$\Im\omega$	$(3.4^{+4.8}_{-5.0} \pm 0.6) \cdot 10^{-4}$	$\pm 2 \cdot 10^{-5}$
Δa_0	(prelim.: $(0.4 \pm 1.8) \cdot 10^{-17}$ GeV)	$\pm 1 \cdot 10^{-18}$ GeV
Δa_Z	(prelim.: $(-1 \pm 4) \cdot 10^{-17}$ GeV)	$\pm 3 \cdot 10^{-18}$ GeV
$\Delta a_X, \Delta a_Y$	(prelim.: $< 9.2 \cdot 10^{-22}$ GeV)	$\mathcal{O}(10^{-18})$ GeV

A ϕ-factory represents a unique opportunity to push forward these studies. It is also an ideal place to investigate the entanglement and correlation properties of the produced $K^0\bar{K}^0$ pairs. A proposal for continuing the KLOE physics program (KLOE-2) at an improved DAΦNE machine, able to deliver an integrated luminosity up to 50 fb^{-1}, has been recently presented. Improvements by about one order of magnitude in almost all present limits are expected, and an experimental demonstration of the *kaonic* quantum eraser is foreseen.

Acknowledgments

I would like to thank the organizing committee, and in particular Danilo Zavrtanik, Samo Stanič and Martin O'Loughlin for the invitation to this very interesting and successful conference, and the pleasant stay in Bled.

References

[1] I.B. Khriplovich and S.K. Lamoreaux, *CP Violation Without Strangeness. Electric Dipole Moments of Particles, Atoms and Molecules* (Springer, Berlin, 1997).
[2] G. Lueders, Ann. Phys. (NY) **2** (1957) 1, reprinted in Ann. Phys. (NY) **281** (2000) 1004.
[3] W. Pauli, *Exclusion principle, Lorentz group and reflexion of space-time and charge* in *Niels Bohr and the development of physics*, edited by W. Pauli (Pergamon, London, 1955), p.30.
[4] J.S. Bell, Proc. R. Soc. London A **231** (1955) 479.
[5] R. Jost, Helv. Phys. Acta **30** (1957) 409.
[6] O.W. Greenberg, Phys. Rev. Lett. **89** (2002) 231602.
[7] O.W. Greenberg, Found. Phys. **36** (2006) 1535, [arχiv:hep-ph/0309309].
[8] S. Hollands, Commun. Math. Phys. **244** (2004) 209.
[9] W.-M. Yao *et al.*, Particle Data Group, J. Phys. G **33** (2006) 1.
[10] *Handbook on neutral kaon interferometry at a ϕ-factory*, ed. A. Di Domenico, Frascati Physics Series **43**, INFN-LNF, Frascati, 2007.
[11] V. Weisskopf and E.P. Wigner, Z. Phys. **63** (1930) 54; see also appendix A of: P.K. Kabir, *The CP puzzle* (Academic Press, London, 1968) or appendix I of: O. Nachtmann, *Elementary Particle Physics: Concepts and Phenomena* (Springer-Verlag, Berlin, 1990).
[12] L. Maiani, in *The second DAΦNE handbook*, ed. L. Maiani, G. Pancheri and N. Paver, Vol. I, INFN-LNF, Frascati, 1995.
[13] G.C. Branco, L. Lavoura and J.P. Silva, *CP Violation* (Oxford University Press, Oxford, 1999).
[14] I.I. Bigi and A.I. Sanda, *CP Violation* (Cambridge University press, Cambridge, 2000).
[15] M. Fidecaro and H.J. Gerber, Rep. Progr. Phys. **69** (2006) 1713.
[16] L. Lavoura, Ann. Phys. **207** (1991) 428.
[17] J.P. Silva, Phys. Rev. D **62** (2000) 116008.
[18] A. Angelopoulos *et al.* (CPLEAR collaboration), Phys. Rep. **374** (2003) 165.
[19] C.D. Buchanan *et al.*, Phys. Rev. D **45** (1992) 4088.
[20] G. D'Ambrosio, G. Isidori and A. Pugliese, in *The second DAΦNE handbook*, ed. L. Maiani, G. Pancheri and N. Paver, Vol. I, INFN-LNF, Frascati, 1995.
[21] M. Hayakawa and A.I. Sanda, Phys. Rev. D **48** (1993) 1150.
[22] A. Alavi-Harati *et al.* (KTeV collaboration), Phys. Rev. D **67** (2003) 012005.
[23] V. Fanti *et al.* (NA48 Collaboration), Nucl. Instrum. Meth. A **574** (2007) 433.
[24] For details on the NA48 experiment and on measurements of direct *CP* violation see also M. Sozzi, these proceedings.

[25] S. Guiducci, *Status of DAΦNE*, Proc. of the 2001 Particle Accelerator Conference, Chicago, IL, 2001, p. 353.
[26] M. Adinolfi *et al.* (KLOE collaboration), Nucl. Instr. and Meth. A **488** (2002) 51.
[27] M. Adinolfi *et al.* (KLOE collaboration), Nucl. Instr. and Meth. A **482** (2002) 364.
[28] A. Angelopoulos *et al.* (CPLEAR collaboration), Eur. Phys. J. C **22** (2001) 55.
[29] A. Alavi-Harati *et al.* (KTeV collaboration), Phys. Rev. Lett. **88** (2002) 181601.
[30] F. Ambrosino *et al.* (KLOE collaboration), Phys. Lett. B **636** (2006) 173.
[31] J.S. Bell and J. Steinberger, Proc. Oxford Int. Conf. on Elementary Particles, 1965.
[32] F. Ambrosino *et al.* (KLOE collaboration), JHEP **12** 011 (2006).
[33] F. Ambrosino *et al.* (KLOE collaboration), Phys. Lett. B **642** (2006) 315.
[34] R.A. Bertlmann, W. Grimus and B.C. Hiesmayr, Phys. Rev. D **60** (1999) 114032.
[35] A. Apostolakis *et al.* (CPLEAR collaboration), Phys. Lett. B **422** (1998) 339.
[36] A. Go *et al.* (Belle collaboration), Phys. Rev. Lett. **99** (2007) 131802.
[37] S. Hawking, Phys. Rev. D **14** (1976) 2460.
[38] S. Hawking, Commun. Math. Phys. **87** (1982) 395.
[39] R. Wald, Phys. Rev. D **21** (1980) 2742.
[40] S. Hawking, Phys. Rev. D **72** (2005) 084013.
[41] J. Smolin and J. Oppenheim, Phys. Rev. Lett. **96** (2006) 081302.
[42] C. Kiefer, Annalen Phys. **15** (2005) 129.
[43] J. Ellis, J.S. Hagelin, D.V. Nanopoulos and M. Srednicki, Nucl. Phys. B **241** (1984) 381.
[44] J. Ellis, J.L. Lopez, N.E. Mavromatos and D.V. Nanopoulos, Phys. Rev. D **53** (1996) 3846.
[45] R. Adler *et al.* (CPLEAR collaboration), Phys. Lett. B **364** (1995) 239.
[46] A. Di Domenico, *Correlations in ϕ decays into $K^0\bar{K}^0$*, Proceedings of the International School of Physics “Enrico Fermi” CLXIII course “CP violation: from quarks to leptons”, 19-29 July 2005, Varenna, Italy.
[47] F. Benatti and R. Floreanini, Nucl. Phys. B **488** (1997) 335; Nucl. Phys. B **511** (1998) 550; Phys. Lett. B **468** (1999) 287.
[48] J. Bernabeu, N. Mavromatos and J. Papavassiliou, Phys. Rev. Lett. **92** (2004) 131601.
[49] J. Bernabeu, N. Mavromatos, J. Papavassiliou and A. Waldron-Lauda, Nucl. Phys. B **744** (2006) 180.
[50] V.A. Kostelecký, Phys. Rev. Lett. **80** (1998) 1818.
[51] V.A. Kostelecký, Phys. Rev. D **61** (1999) 016002.
[52] V.A. Kostelecký, Phys. Rev. D **64** (2001) 076001.
[53] A. Di Domenico (KLOE collaboration), *CPT and Lorentz Symmetry IV*, ed. V.A. Kostelecký (World Scientific, Singapore, 2008).
[54] H. Nguyen (KTeV collaboration), *CPT and Lorentz Symmetry II*, ed. V.A. Kostelecký (World Scientific, Singapore, 2002).
[55] B. Aubert *et al.* (BABAR collaboration), [arχiv:hep-ex/0607103]; [arχiv:0711.2713].
[56] R. Beck *et al.* (KLOE-2 collaboration), *Expression of interest for the continuation of the KLOE physics program at DAΦNE upgraded in luminosity and in energy*, March 31, 2006, available at

`http://www.lnf.infn.it/lnfadmin/direzione/roadmap/LoIKLOE.pdf`.
[57] A. Einstein, B. Podolsky and N. Rosen, Phys. Rev. **47** (1935) 777.
[58] A. Bramon, G. Garbarino and B.C. Hiesmayr, Phys. Rev. Lett. **92** (2004) 020405; Phys. Rev. A **69** (2004) 062111.
[59] see also B.C. Hiesmayr, A. Bramon, A. Di Domenico, G. Garbarino and A. Go, *Erasing the past and impacting the future with kaons at a ϕ-factory*, Kaon Int. Conf. 2007, PoS(KAON)044 and references therein.

Section V: Quantum Gravity

canonical quantum gravity
problem of time in quantum gravity
gravitational limitations on space and time measurements
fundamental loss of coherence in quantum theory
emergent space-time

TIME AND MATTER 2007

Time and Matter
2007 Bled, Slovenia

Conceptual Issues in Canonical Quantum Gravity and Cosmology

CLAUS KIEFER*
Institute for Theoretical Physics, University of Cologne,
Zülpicher Straße 77, 50937 Köln, Germany

Abstract: Existing approaches to quantum gravity exhibit plenty of startling conceptual issues. Here I restrict my attention to the canonical approach. Both classical and quantum canonical gravity are discussed. Most conceptual problems circle around the problem of time – the absence of any external time parameter. I then turn to quantum cosmology, where these and more problems can be discussed in a transparent way. I conclude with brief remarks on singularity avoidance, the arrow of time, and the interpretation of quantum theory in general.

Introduction

In his book *The Meaning of Relativity*, Albert Einstein emphasizes the following point [1]:

> Es widerstrebt dem wissenschaftlichen Verstande, ein Ding zu setzen, das zwar wirkt, aber auf das nicht gewirkt werden kann.[1]

This statement expresses our modern understanding of space and time. In contrast to Newtonian physics, no absolute fields exist in general relativity; space and time are fully dynamical. Spacetime acts on matter, but matter also acts on spacetime. The mutual relationship between both is described by the non-linear Einstein field equations. There is no fixed, absolute background any more; spacetime as described by the metric of general relativity is fully dynamical. This feature is called *background independence* and

* kiefer@thp.uni-koeln.de

[1] It is contrary to the scientific mode of understanding to postulate a thing that acts, but which cannot be acted upon.

is of central relevance for general relativity, but also for quantum gravity, since it is the dynamical fields that are subject to quantization. An absolute ('background') field can be characterized by the fact that there exist coordinates which assign universal values to all components of a field; such a field must not be quantized. A prominent example for a background field is the spacetime metric of special relativity, for which priviledged coordinates exist (the inertial coordinates) in which the metric assumes the standard form diag(-1,1,1,1).

In my contribution, I shall give a brief review of the central conceptual issues that arise in both classical and quantum gravity and that are all more or less directly connected with background independence and the related problem of time. Attention is restricted to the canonical approach, because there these issues are most transparent. A detailed exposition covering all aspects of quantum gravity can be found in my book [2]; an introduction into classical and quantum canonical gravity from a conceptual point of view can also be found in our essay [3].

A theory of quantum gravity seems to be needed because the singularity theorems predict that general relativity cannot be fundamentally complete. In particular, the origin of our Universe and the final fate of black holes do not seem to be comprehensible without quantum gravity. Unfortunately, no final theory exists to date, so discussing conceptual issues in quantum gravity means to discuss them in existing *approaches* to such a theory. However, one can put forward various arguments in support of the generality of these issues in most approaches. This should become clear from the following discussion.

What are the main approaches to quantum gravity? There exist presently two main classes:

- *Quantum general relativity*: this includes all approaches that arise from an application of formal quantization rules to general relativity. They can be subdivided further into:
  - *Covariant approaches* (such as perturbation theory, path integrals, and others), which are characterized by the fact that spacetime covariance plays a crucial role in some parts of the formalism, and
  - *Canonical approaches* (such as geometrodynamics, connection dynamics, loop dynamics, and others), in which a Hamiltonian formalism is being used.
- *String theory*: this is intended to be a unified quantum theory of all interactions, in which the quantized gravitational field can be distin-

guished as a separate field only in appropriate limits, for example, for energies lower than the fundamental string energy scale.

There also exist other even more fundamental approaches (such as the quantization of topology) which, however, have not been developed as far as the main approaches. In the following, I shall restrict myself to the canonical formalism. I start with classical canonical gravity, proceed then to quantum canonical gravity, and conclude with quantum cosmology where the conceptual issues of the full approach (and further issues) are exhibited clearly and explicitly.

Canonical classical gravity

The canonical formalism starts with the '3+1 decomposition' of general relativity [2]. Spacetime is assumed to be globally hyperbolic, that is, to be of the form $\mathbb{R} \times \Sigma$, where Σ denotes a three-dimensional manifold; spacetime is thus foliated into a set of spacelike hypersurfaces Σ_t. The dynamical variable is the three-dimensional metric, h_{ab}, which can be obtained as the metric that is induced by the spacetime metric $g_{\mu\nu}$ on each Σ_t. Instead of considering the three-metrics on each Σ_t, one can adopt the equivalent viewpoint and consider a t-dependent three-metric on the given manifold Σ. For each choice of Σ in the topological sense one obtains a different version of canonical gravity.

This lends itself to a more fundamental viewpoint: instead of starting with a four-dimensional spacetime, M, to be foliated, we assume that only Σ is given. Then, only *after* solving the field equations can we construct spacetime and interpret the time dependence of the metric h of Σ as being brought about by 'wafting' Σ through M via a one-parameter family of hypersurfaces Σ_t. The field equations can actually be divided into

- six dynamical evolution equations for h_{ab} and its canonical momentum π^{ab}, and

- four constraints, which are restrictions on the initial data, that is, restrictions on the allowed choices for h_{ab} and π^{ab} on an 'initial' hypersurface.

After having solved these equations, spacetime can be interpreted as a 'trajectory of spaces'. The origin of the constraints is the diffeomorphism invariance of general relativity. They have the explicit form

$$
\begin{aligned}
H[h,\pi] &= 2\kappa\, G_{ab\,cd}\pi^{ab}\pi^{cd} \;-\; (2\kappa)^{-1}\sqrt{h}({}^{(3)}\!R-2\Lambda) \;+\; \sqrt{h}\rho \approx 0\,, &(1)\\
D^a[h,\pi] &= -2\nabla_b\pi^{ab} \;+\; \sqrt{h}j^a \approx 0\,, &(2)
\end{aligned}
$$

where $\kappa = 8\pi G/c^4$, Λ is the cosmological constant, ${}^{(3)}\!R$ is the three-dimensional Ricci scalar, and

$$
G_{ab\,cd} = \frac{1}{2\sqrt{h}}(h_{ac}h_{bd} + h_{ad}h_{bc} - h_{ab}h_{cd}) \qquad (3)
$$

is called the DeWitt metric; it plays the role of a metric on the space of all three-metrics. While (1) is called Hamiltonian constraint, the three constraints (2) are called momentum or diffeomorphism constraints.

There are two important theorems which connect the constraints with the evolution. The first one states:

> Constraints are preserved in time $\Longleftrightarrow$ energy–momentum tensor of matter has vanishing covariant divergence.

This can be compared with the corresponding situation in electrodynamics: the Gauss constraint $\nabla\mathbf{E} = 4\pi\rho$ is preserved in time if and only if electric charge is conserved in time. The second theorem states:

> Einstein's equations are the unique propagation law consistent with the constraints.

Again, this can be compared with the situation in electrodynamics: Maxwell's equations are the unique propagation law consistent with the Gauss constraint. Constraints and evolution equations are thus inextricably mixed.

A central conceptual issue in quantum gravity is the 'problem of time'. Part of this problem is already present in the classical theory. Namely, if we restrict ourselves for simplicity to *compact* three-spaces Σ, the total Hamiltonian is a combination of pure constraints; all of the evolution will therefore be generated by constraints and is thus, in a sense, pure gauge. How can this be reconciled with the usual interpretation of the Hamiltonian as a generator of time translations? The point is that the evolutions along different spacelike hypersurfaces are equivalent and lead to the same spacetime satisfying Einstein's equations. This freedom is expressed by the fact that the Hamiltonian generates both gauge transformations and time translations (hypersurface deformations). In other words, *no* external time parameter exists. All physical time parameters are to be constructed from within the system, that is, as a functional of the canonical variables; a priori there is no preferred choice of such an intrinsic time parameter. The absence of an extrinsic time and the non-preference of an intrinsic time is known as the *problem of time* in classical canonical gravity. As we shall see below, this leads in quantum gravity to stationary fundamental equations for a wave function which only depends on variables defined on the three-dimensional space Σ.

Above we have defined the DeWitt metric – the metric on the space of all three-metrics, see (3). Its signature is responsible for the structure of the kinetic term in the Hamiltonian constraint (1). As it turns out, the DeWitt-metric possesses an indefinite structure [2, 3]. It can be viewed at each spacepoint as a symmetric 6×6-matrix. This matrix can be diagonalized, and it is found thereby that the diagonal contains one minus and five plus signs; the DeWitt-metric is thus indefinite, and the kinetic term in (1) is indefinite, too. Due to the pointwise Lorentzian signature of $G^{ab\,cd}$, it is of a *hyper-Lorentzian structure* with infinitely many negative, null, and positive directions. This property will be of central relevance in the quantum theory.

Canonical quantum gravity

The classical constraints (1) and (2) can be implemented in the quantum theory in various ways; after all, the quantum theory can only be heuristically constructed and never be derived from the underlying classical theory. One possibility would be to try to solve the constraints classically and quantize only the remaining physical variables. However, this 'reduced quantization' leads to many problems; in particular, it is hard to perform in practice except for the simplest models [4, 2]. The alternative method is to implement the constraints à la Dirac as restrictions on physically allowed wave functionals. Replacing the canonical momenta by $-\mathrm{i}/\hbar$ times functional derivatives with respect to the three-metric, the Hamiltonian constraint becomes the *Wheeler–DeWitt equation*:[2]

$$\hat{H}\Psi \equiv \left(-2\kappa\hbar^2 G_{abcd} \frac{\delta^2}{\delta h_{ab}\delta h_{cd}} - (2\kappa)^{-1}\sqrt{h}\left({}^{(3)}R - 2\Lambda \right) \right) \Psi = 0. \qquad (4)$$

[2]For simplicity, we neglect here the non-gravitational fields, which occur in addition to the three-metric.

The kinetic term can only be interpreted as being formal, because the factor-ordering problem and the connected problem of regularizing the functional derivatives have not been addressed.

The quantization of the momentum constraints (2) leads to the equations

$$\hat{D}^a\Psi \equiv -2\nabla_b \frac{\hbar}{\mathrm{i}} \frac{\delta\Psi}{\delta h_{ab}} = 0, \tag{5}$$

which are called *quantum diffeomorphism (momentum) constraints*. They express the invariance of the wave functional under infinitesimal coordinate transformations on the three-dimensional space.

The 'problem of time', which was already discussed above in the context of the classical theory, is being enforced in the quantum theory. Namly, spacetime as such has completely disappeared! All that remains in the formalism is a wave functional which depends on the metric (and matter fields) on a three-dimensional manifold Σ. In retrospect, this is not surprising. A classical spacetime as a succession of three-dimensional hypersurfaces is fully analogous to a particle trajectory in mechanics (a succession of positions). In the same way that the particle trajectory vanishes in quantum mechanics, the spacetime vanishes in quantum gravity. This feature is independent of this particular quantization of Einstein's equations; it holds for any theory which at the classical level does not contain an external time parameter.

The absence of an external time and of spacetime does not necessarily mean that no sensible notions of time can be defined. Regarding the kinetic term in (4), one recognizes in view of (3) that is possesses an indefinite structure [2, 3]. The Wheeler–DeWitt equation thus has the form of a wave equation (more precisely, a local hyperbolic equation). Part of the three-metric thus comes with a positive sign in the kinetic term and can therefore be called *intrinsic time*, in full analogy to the time variable occurring in a standard wave equation. It turns out that it is just the scale part of the three-metric (the three-volume in simple models) which plays the role of intrinsic time.

A conceptual problem that is related to the problem of time, is the Hilbert-space problem: which inner product, if any, does one have to choose between wave functionals? From the point of view of the standard Schrödinger picture, one would like to employ the Schrödinger inner product (square integrable wave functionals). On the other hand, in view of the fact that (4) resembles more a wave equation, one would prefer a Klein–Gordon type of inner product, which is, however, indefinite. Every choice has its merits and its disadvantages, so it seems difficult to make a definite choice [2, 4].

What is an observable in quantum gravity? This question, too, is related with the problem of time. One would assume that all observables have to commute with both the Hamiltonian and momentum constraints. But this would mean that all observables would be constants of motion, because the total Hamiltonian is a sum of these constraints. Is therefore all change observed in Nature a pure illusion? The answer is no because we view the world from inside. As we shall see below, observers in the semiclassical approximation will have an approximate time variable at their disposal, which can be approximately identified with the standard time of non-relativistic physics. The timeless view of constants of motion would correspond to a hypothetical perspective from outside the world where everything (all branches of the wave functional) would be present at once.

So far we have restricted our discussion to quantum geometrodynamics. Of course, other canonical variables can be chosen. The most prominent choice at the moment is to choose holonomies and fluxes, leading to loop quantum gravity [5, 6]. While the details differ from quantum geometrodynamics, the timeless nature of the constraints remains, with all of the above consequences. A mathematically sound Hilbert-space structure can be constructed at least on the kinematical level, that is, before all the constraints are implemented; this inner product is of the Schrödinger type. For more details on loop quantum gravity, I refer to the literature [5, 2].

Quantum cosmology

Most of the conceptual issues in quantum gravity can be discussed in a transparent way in quantum cosmology. Quantum cosmology is the application of quantum theory to the Universe as a whole. That the whole cosmos must be fundamentally described in quantum terms follows from very general arguments, which are independent of gravity. All systems, except the most microscopic ones, are quantum entangled with their natural environment; this leads to their classical appearance – through a process called *decoherence* [7]. Since every environment has again an environment, this leads to the conclusion that the whole Universe must be described in quantum terms. It is only because the gravitational interaction dominates on large scales that we need a theory of quantum gravity to cope with quantum cosmology [2, 8, 9].

It is important to note that quantum effects in cosmology are not *a priori* restricted to the Planck scale, which reads

$$
\begin{aligned}
l_{\mathrm{P}} &= \sqrt{\frac{\hbar G}{c^3}} \approx 1.62 \times 10^{-33}\ \mathrm{cm}, && (6)\\
t_{\mathrm{P}} &= \frac{l_{\mathrm{P}}}{c} = \sqrt{\frac{\hbar G}{c^5}} \approx 5.40 \times 10^{-44}\ \mathrm{s}, && (7)\\
m_{\mathrm{P}} &= \frac{\hbar}{l_{\mathrm{P}} c} = \sqrt{\frac{\hbar c}{G}} \approx 2.17 \times 10^{-5}\ \mathrm{g} \approx 1.22 \times 10^{19}\ \mathrm{GeV}. && (8)
\end{aligned}
$$

The reason is that the *superposition principle*, which allows to form non-trivial quantum states, holds at any scale, not only the Planck scale.[3]

The simplest model of quantum cosmology is obtained if one quantizes directly a Friedmann–Lemaître universe; it is characterized by a scale factor, a, and we choose in addition a homogeneous massive scalar field, ϕ. The classical spacetime metric is of the form

$$\mathrm{d}s^2 = -N^2(t)\mathrm{d}t^2 + a^2(t)\mathrm{d}\Omega_3^2, \tag{9}$$

where $N(t)$ is the lapse function which encodes the invariance of the classical theory under reparametrizations of the time coordinate $t \to f(t)$. In the quantum theory, t disappears and only a and ϕ remain as the variables on which the wave function $\psi(a,\phi)$ depends.

The Wheeler–DeWitt equation then reads (with a convenient choice of units $2G/3\pi = 1$ and for the case of a closed universe)

$$\frac{1}{2}\left(\frac{\hbar^2}{a^2}\frac{\partial}{\partial a}\left(a\frac{\partial}{\partial a}\right) - \frac{\hbar^2}{a^3}\frac{\partial^2}{\partial \phi^2} - a + \frac{\Lambda a^3}{3} + m^2 a^3 \phi^2\right)\psi(a,\phi) = 0. \tag{10}$$

The factor ordering has been chosen in order to achieve covariance in the two-dimensional configuration space comprised of a and ϕ. The wave-nature of this equation is evident, and it is seen that the intrinsic time variable is given by the scale factor itself. The quantum diffeomorphism constraint is automatically satisfied by the ansatz (9). Solutions of the Wheeler–DeWitt equation in the context of quantum cosmology are often called ‘wave function of the universe’.

The new concept of time has far-reaching consequences: the classical and the quantum model exhibit two drastically different concepts of determinism, see Figure 1.

[3]This is, of course, an assumption, but an assumption which is applied in almost all discussions of quantum cosmology.

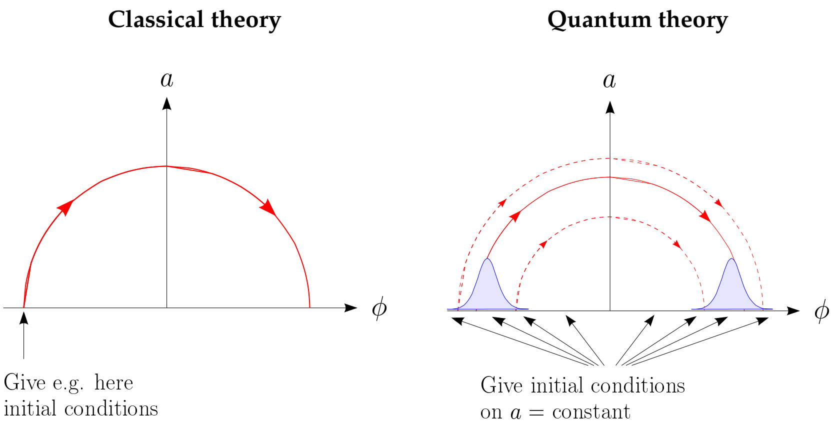


Figure 1. (a) Recollapsing part is deterministic successor of expanding part. (b) 'Recollapsing' wave packet must be present 'initially'.

Consider the case of a classically recollapsing universe. In the classical case (left) we have a trajectory in configuration space: although it can be parametrized in many ways, the important point is that it *can* be parametrized by some time parameter. Therefore, upon solving the classical equations of motion, the recollapsing part of the trajectory is the deterministic successor of the expanding part: the model universe expands, reaches a maximum point, and recollapses.

Not so for the quantum model. There is no classical trajectory and no classical time parameter and one must take the wave equation (10) as it stands. The wave function only distinguishes small a from large a, not earlier t from later t. There is thus no intrinsic difference between big bang and big crunch. If one wants to construct a wave packet following the classical trajectory as a narrow tube, one has to impose the presence of two packets as an initial condition at small a; if one chose only one packet, one would obtain a wave function which is spread out over configuration space and which does not resemble anything close to a narrow wave packet.

Wave packets are of crucial importance when studying the validity of the semiclassical approximation. In quantum mechanics, narrow wave packets remain narrow if the WKB approximation holds. In quantum cosmology, this issue has to be studied from the new viewpoint of the Wheeler–DeWitt equation (10). If the classical model describes a recollapsing universe, one has to impose in the quantum theory onto the wave function the restriction that it go to zero for $a \to \infty$; with respect to intrinsic time, this corresponds to a 'final condition'. Calculations show that it is then *not* possible to have narrow wave packets all along the classical trajectory; the packet disperses, see [2, 9] and the references therein. Again, this is a consequence of the novel concept of time.

But how do classical properties arise if wave packets necessarily disperse? The answer to this question is again decoherence [7]. In order to study this process, additional degrees of freedom must be introduced. They can then act as an 'environment' which becomes quantum entangled with a and ϕ, causing their classical appearance.

A straightforward way to introduce a large number of additional degrees of freedom is to consider small inhomogeneities superimposed on the homogeneous three-sphere of the closed Friedmann universe [10]. These inhomogeneities are described by small multipoles denoted by the set of variables $\{x_n\}$; they describe small gravitational waves and density perturbations. One then finds the more general Wheeler–DeWitt equation

$$\left(H_0 + \sum_n H_n(a,\phi,x_n)\right)\Psi(a,\phi,\{x_n\}) = 0, \tag{11}$$

where

$$\Psi(a,\phi,\{x_n\}) = \psi_0(a,\phi)\prod_n \psi_n(a,\phi,x_n),$$

and $H_0\psi_0 = 0$ is the original 'unperturbed' Wheeler–DeWitt equation (10). If ψ_0 is of WKB form, $\psi_0 \approx C\exp(\mathrm{i}S_0/\hbar)$ (with a slowly varying prefactor C), one gets [10, 11]

$$\mathrm{i}\hbar\frac{\partial\psi_n}{\partial t} \approx H_n\psi_n \tag{12}$$

with

$$\frac{\partial}{\partial t} \equiv \nabla S_0 \cdot \nabla\ . \tag{13}$$

The multipoles therefore obey separate Schrödinger equations with respect to some approximate time parameter t. This parameter is called '*WKB time*' – it controls the dynamics in this approximation and corresponds to the Friedmann time parameter of the classical model. This is the limit where the standard formalism of quantum theory with its Hilbert-space structure applies. A Hilbert space is needed to implement the probability interpretation of quantum theory, in particular, the conservation of probability with respect to external time t. Whether a Hilbert-space structure is needed in timeless quantum gravity, too, is thus an open issue.

This 'emergence of time' from timeless quantum gravity is one of the satisfactory features of quantum geometrodynamics. A corresponding recovery is not yet known in loop quantum gravity. An analogous feature can be discussed in Euclidean quantum gravity where the fundamental concept is a Euclidean path integral. A suggestion to find the quantum state of the Universe is encoded, for example, in the no-boundary condition of Hartle and Hawking [12, 2]. The time parameter t appears there in the limit where the saddle point approximation holds (corresponding to the WKB approximation) and where the saddle point gives a complex time in the Euclidean formulation – corresponding to the ordinary real time t as in (13).

In order to study the decoherence for a and ϕ, one has to solve the full Wheeler–DeWitt equation (11) and trace out all the multipoles from the resulting full quantum state. This gives a density matrix whose diagonal terms are suppressed in the generic case, which means that interferences between universes of different sizes can be neglected and the universe can be treated classically for most of its evolution [2, 9]. Moreover, one can also understand why and how interferences between the $\exp(\mathrm{i}S_0/\hbar)$- and $\exp(-\mathrm{i}S_0/\hbar)$-branches of a wave function become suppressed by decoherence. A calculation within a particular model leads, for example, to the following suppression term for the interference between these two branches [13]:

$$\exp\left(-\frac{\pi m H_0^2 a^3}{128\hbar}\right) \sim \exp\left(-10^{43}\right),$$

where the number arises for today's universe if some natural values are inserted for the Hubble parameter H_0 and the mass m of a fundamental Higgs field. One recognizes that today the universe behaves indeed very classically!

Once a classical background is established as an approximate concept, one can then address the quantum-to-classical transition for the relevant part of the multipoles itself on this background, which are the density fluctuations serving as the seeds for galaxy formation [14].

Using the above introduced concepts, one can also discuss the issues of singularity avoidance and arrow of time. Both issues can have to do with quantum effects far away from the Planck scale. As for the former example, classical models exist which exhibit a singularity at large scale factor, that is, far away from the Planck scale. For example, by choosing a scalar-field potential of the form

$$V(\phi) = V_0 \left(\sinh(|\phi|) - \frac{1}{\sinh(|\phi|)} \right) ,$$

one can obtain a 'big-brake singularity' – the universe suddenly stops its expansion in the future, while keeping both the scale factor and its time derivative finite, but leading to an infinite value for the deceleration. Discussing the corresponding quantum model, it was shown upon solving the Wheeler–DeWitt equation that *all normalizable solutions* vanish at the classical singularity, thus entailing complete singularity avoidance [15]. Singularity avoidance is also a central feature of loop quantum cosmology, which is discussed in another contribution to this volume [6].

As for the arrow of time, its origin can in principle be understood from quantum cosmology. The reason is that the Wheeler–DeWitt equation (10) is asymmetric with respect to intrinsic time a. Choosing a simple initial wave function which factorizes between the a and ϕ-part and the higher multipoles, the full solution is a wave function whose quantum entanglement between these two parts increases with a. Integrating out the multipoles leads to a density matrix whose impurity increases with a. This, in turn, leads to an increasing entanglement entropy which could be at the heart of the Second Law of thermodynamics [9]. An interesting consequence would be a formal reversal of the arrow of time at the region of the classical turning point [9, 16] – another quantum effect far from the Planck scale.

Last but not least, quantum cosmology has an important bearing on our understanding of quantum theory itself. Both quantum general relativity and string theory preserve the linear structure for the quantum states, that is, stick to the strict validity of the superposition principle. Since the Universe as a whole by definition contains all degrees of freedom, it must also describe all observers in quantum terms. The only interpretation known so far is the 'Everett interpretation', with decoherence as an essential part [7]. I thus want to conclude with the following quote from one of the pioneering papers on canonical quantum gravity [17]:

> Everett's view of the world is a very natural one to adopt in the quantum theory of gravity, where one is accustomed to speak without embarassment of the 'wave function of the universe.' It is possible that Everett's view is not only natural but essential.

Acknowledgements

I would like to thank the organizers of "Time and Matter 2007" for inviting me to this inspiring conference amidst beautiful surroundings.

References

[1] A. Einstein, *The Meaning of Relativity* (Princeton University Press, Princeton, 1922).
[2] C. Kiefer, *Quantum Gravity*, Second edition (Oxford University Press, Oxford, 2007).
[3] D. Giulini and C. Kiefer, *Approaches to fundamental physics – an assessment of current theoretical ideas*, ed. E. Seiler and I.-O. Stamatescu (Springer, Berlin, 2007), 131–150.
[4] K.V. Kuchař, Proceedings of the 4th Canadian Conference on General relativity and Relativistic Astrophysics, ed. G. Kunstatter, D. Vincent and J. Williams (World Scientific, Singapore), 211–314.
[5] C. Rovelli, *Quantum Gravity* (Cambridge University Press, Cambridge, 2004).
[6] M. Bojowald, contribution to this volume.
[7] 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*, Second edition (Springer, Berlin, 2003); see also `www.decoherence.de`.
[8] C. Kiefer and B. Sandhöfer, *Beyond the Big Bang – Prospects for an Eternal Universe*, ed. R. Vaas (Springer, Berlin, 2008, to appear).
[9] H.D. Zeh, *The physical basis of the direction of time*, Fifth edition (Springer, Berlin, 2007); see also `www.time-direction.de`.
[10] J.J. Halliwell and S.W. Hawking, Phys. Rev. D **31** (1985) 1777–1791.
[11] C. Kiefer, Class. Quantum Grav. **4** (1987) 1369–1382.
[12] J.B. Hartle and S.W. Hawking, Phys. Rev. D **28** (1983) 2960–2975.
[13] C. Kiefer, Phys. Rev. D **46** (1992) 1658–1670.
[14] C. Kiefer, I. Lohmar, D. Polarski and A.A. Starobinsky, Class. Quantum Grav. **24** (2007) 1699–1718.
[15] A.Y. Kamenshchik, C. Kiefer and B. Sandhöfer, Phys. Rev. D **76** (2007) 064032.
[16] C. Kiefer and H.D. Zeh, Phys. Rev. D **51** (1995) 4145–4153.
[17] B.S. DeWitt, Phys. Rev. **160** (1967) 1113–1148.

Time and Matter
2007 Bled, Slovenia

Records Theory

EDWARD ANDERSON*
Peterhouse and DAMTP, Cambridge, UK

Abstract:

In quantum gravity, one seeks to combine quantum mechanics and general relativity. In attempting to do so, one comes across the 'problem of time' impasse: the notion of time is conceptually different in each of these theories. In this seminar, I consider the timeless records approach toward resolving this. Records are localized, information-containing subconfigurations of a single instant. Records theory is the study of these and of how science (or history) is to be abstracted from correlations between them. I explain how to motivate this approach, provide a ground-level structure for it and discuss what kind of further tools are needed. For a more comprehensive account with many more references, see [1].

Introduction

Records play a role in Quantum Cosmology and Quantum Gravity. The literature on this is a heterogeneous, consisting of 1) reinterpretations [2, 3] of how α-particle tracks form in a bubble chamber [4] that may be analogous to Quantum Cosmology [5, 6, 7, 8]. Therein, Barbour's approach also involves reformulating classical physics in timeless terms [9, 10, 11, 12] and places emphasis on the configuration of the universe as a whole and on timelessness casting mystery [3] upon why 'ordinary physics' works. 2) The *Conditional Probabilities Interpretation* for Quantum Cosmology [13, 14, 15] places its emphasis on subconfigurations (SC's) of the universe within a single instant. Here, ordinary physics of SC's ends up familiarly explained through other SC's providing approximate time standards for them, and what is habitually observed is the dynamics of subsystems rather than of the whole universe [18, 10, 3, 19]. 3) While Histories Theory (see e.g. [16]) is not primarily timeless, a Records Theory sits

* ea212@cam.ac.uk

within it [16, 17], and benefits from inheriting part of the structural framework developed for Histories Theory.

Records are "*somewhere in the universe that information is stored when histories decohere*" (p 3353 of [16]) . (0)

This seminar's Records Theory is a synthesis of elements drawn from 1 to 3. In outline, I consider *records* to be information-containing SC's of a single instant. *Records Theory* is then the study of these and how dynamics (or history or science) is to be abstracted from correlations between same-instant records. It is to make this abstraction meaningful that I insist on records being SC's rather than whole instants, thus getting round the abovementioned 'mystery' by a similar argument to the Conditional Probability Interpretation's.

For adopting a Records Theory approach to profitable, I argue that records should have the following properties. *Useability*, in that A) their whereabouts [c.f. (0)] should be spatially-localized SC's of the universe, for whatever notion of space that one's theory has and restricted to the observationally accessible part thereof. B) They should also belong to a part of the SC space for which observational imprecisions in identifying SC's do not distort the extraction of information too much *Usefulness*: their information content [c.f. (0)] should be high enough and of the right sort of quality to enable reliable measures of correlation to be computed. Not all systems have instants solely of this nature, so Records Theory may not always be profitable. 1) and 2) additionally require *semblance of dynamics* to emerge from timeless records.

Sec 2 summarizes motivation for Records Theory. Sec 3 mentions some illustrative toy models. Sec 4 proposes a ground-level structure for Records Theory which parallels some of that of Histories Theory. I then comment on the useability, usefulness and correlation aspects of records in Sec 4–6, and more speculative aspects in Sec 7.

Some motivations for Records Theory

Records Theory should be motivated as follows [1]. 1) The Problem of Time (POT) in Quantum Gravity is an incompatibility between the roles played by 'time' in GR and in QM [20]. One conceptually clear way of dealing with this problem is to recast both GR and QM in a timeless mold. [While the

Wheeler–DeWitt equation (WDE)'s timelessness might specifically prompt some physicists toward Timeless Records Theory, this equation has numerous technical problems and may not be trustworthy. Despite e.g. [18, 10, 3], nor should one turn to Timeless Records Theory due to earlier detailed documentation of problems with the other POT approaches, but rather judge it due to its own merits and shortcomings (Sec 7).] 2) One can in principle treat all of change, processes, dynamics, history and the scientific enterprise in these timeless terms. The classification and subsequent partial elimination of question types in this seminar is to demonstrate that (and its impracticality in some cases). 3) Records Theory is of potential use in removing some unclarities (see e.g. [6]) from the foundations of Quantum Cosmology (which in turn is what Inflationary Theory is to rest on). Records Theory makes contact with this area in e.g. the following ways. A) Such as CMB inhomogeneities or the pattern and spectra of galaxies may be considered to be useful records. B) Within a histories perspective, the decoherence process makes records, but information is in general lost in the making. E.g. mixed states necessarily produce imperfect records [17]. Furthermore, finding out where in the universe the information resides (i.e. where the records are) should be capable of resolving in which cases gravity decoheres matter or vice versa. Decoherence is habitually linked with the emergence of (semi)classicality, so there may well be some bridge between Records Theory and the Semiclassical Approach. C) See also Sec 7. 4) Records Theory is (alongside Histories Theory) a universal scheme in that all types of theory or system admit a such.

It has also been suggested that records are more operationally meaningful than the histories. For, study of records is how one does science (and history) in practise? Unfortunately, this last suggestion fails *as motivation*, because of the difference between the notion of records as in (some of the) SC's that the system provides and as in things which are localized, accessible and of significant information content. As effective reconstruction of history requires the SC's in question to have these in general unestablished properties, it is a *question to address* rather than a preliminary motivation to have that records are more operationally meaningful than histories through possessing good enough qualities to permit a meaningful such reconstruction. Thus what one should do is 1) pin down where the "somewhere" in (0) is (the central motivation in some of Halliwell's papers, e.g. [17]). 2) Determine whether the record thereat is useful – Gell-Mann–Hartle assert that what they call records "*may not represent records in the usual sense of being constructed from quasiclassical variables accessible to us*" (p 3353 of [16]). Also, it may be that the α-particle track in the bubble chamber is atypical in its neatness and localization. For, bubble chambers are carefully selected environments for revealing tracks – much human trial and error has gone into finding a piece of apparatus that does just that. α-tracks being useful records could then hinge on this careful pre-selection, records in general then being expected to be (far) poorer, as suggested e.g. by Joos–Zeh's paradigm [21] of a dust particle decohering due to the microwave background photons. In this situation, records are exceedingly diffuse as the information is spread around by the CMB photons.

Toy models for Records Theory

Ordinary (conservative) mechanics already has a simple analogue of the Hamiltonian constraint: a homogeneous quadratic energy constraint which gives a time-independent Schrödinger equation (TISE) analogue of the WDE at the quantum level. There are furthemore other mechanics that share more features with GR: relational particle mechanics (RPM's) are such [9, 19, 23, 24]. These have additional linear constraints [such as a zero angular momentum constraint, the physics then being encoded solely in the relative separations and angles rather than in any absolute angles, in analogy with how the linear momentum constraint of GR is interpretable in terms of the physics being in the shape of 3-space and not in its coordinatization.] Scale-invariant RPM's additionally have a linear zero dilational constraint that is analogous to the maximal slicing condition in GR. Full reductions are available [23] for 2d RPM's (of which the scale-invariant one is better behaved) allowing us to do quite a lot more with these particular models. The kinetic term then contains the positive-definite Fubini–Study metric.

One could include a harmonic oscillator detector within one's mechanics model, or couple it to an up–down detector. These can hold information about one Fourier mode in the signal, thus showing that even very simple systems can make imperfect records [17]. Finally, the inhomogeneous perturbations about homogeneous cosmologies [25] are a more advanced toy model (with which RPM's nevertheless share various features).

SC space structure and useable records

First level of classical structure

A *configuration* $Q_\Delta(p)$ is a set of particle positions and/or field values, where Δ is a multi-index which covers both particle and field species labels and whatever 'tensorial' indices each of these may carry, and p is a fixed label.

Hierarchical, nonunique splittings into subsystems can then be construed: $Q_\Gamma(p)$ is a subsystem of $Q_\Delta(p)$ if Γ is a subset of the indexing set Δ. The *finest* such subdivision is into individual degrees of freedom.

Two question-types that may be considered at this level are: **Be**1$'$), does $Q_\Delta(p)$ have acceptable properties? (That covers both mathematical consistency and physical reasonableness). **Be**2$'$) If properties of $Q_\Delta(p)$ are known, does this permit deduction of any observable properties of some $Q_{\Delta'}(p)$ for Δ' disjoint from Δ? In other words, are there observable correlations between SC's of a single instant?

Many notions and constructions that theoretical physicists use (see e.g.

[20]) additionally require consideration of sets of instants. A *configuration space of instants* is $\mathsf{Q}_\Delta = \{Q_\Delta(p) :\ p$ a label running over a (generally stratified) manifold $\}$. This is a *heap* of instants. One defines SC spaces similarly. The counterpart of decomposition into subsystems is now a break-down into subspaces. An ordinary (absolute) particle mechanics configuration space is the set of possible positions of N particles, Q(N, d). Relational configuration space $\mathcal{R}$(N, d) is the set of possible relative separations and relative angles between particles. Preshape space P(N, d) is the set of possible scale-free relative particle positions. Shape space S(N, d) is the set of possible scale-free relational configurations. As another example, for geometrodynamics, a rather redundant configuration space is Riem(Σ): the space of positive-definite 3-metrics $h_{\mu\nu}(x_\gamma)$ on the 3-space of fixed topology Σ. Less redundant ones are superspace(Σ) = Riem(Σ)/Diff(Σ) and (something like) conformal superspace(Σ) = Riem(Σ)/Diff(Σ) $\times$ Conf(Σ), for Diff(Σ) the diffeomorphisms of Σ and Conf(Σ) the conformal transformations of Σ.

While each Q_Δ corresponds to a given model with a fixed list of contents, one may not know which model a given (e.g. observed) SC belongs to, or the theory may admit operations that alter the list of contents of the universe. Then one has a *grand heap* of SC spaces of instants. For example, use 1) $\bigcup_{\mathrm{N}\in\,\mathbb{N}_0}$ Q(N, d) for a mechanics theory that allows for variable particle number. 2) $\bigcup_{\text{various }\Sigma}$ superspace(Σ) for a formulation of GR that allows for spatial topology change.

A second type of hierarchical splitting are *grainings*: the various ways that Q can be partitioned. These define a partial order $\prec$ on the subsets of Q. $A \prec B$ is termed 'A is finer grained than B', while $C \succ D$ is termed 'C is coarser-grained than D'. The coarsest grained set is Q itself, while the finest grained sets are each individual $q(p)$ (the constituent points of Q).

Localization in space continues to be formulable in the relational context [1]. *Localization on configuration spaces* [1, 27] can sometimes be attained by augmenting the configuration space to be equipped with a norm. E.g. on Q(N, d), these are the obvious unweighted and (inverse) mass-weighted $\mathbb{R}^{\mathrm{Nd}}$ norms, which still play a role in more reduced configuration spaces through these 'inheriting' structures such as the $\mathbb{R}^{nd}$ norm for R(N, d) or the chordal norm for P(N, d). If the configuration space has a natural metric more complicated than the Euclidean one, one might be able to extend the above notion to the norm corresponding to that. E.g., one could use the Fubini–Study norm on S(N, 2), or the inverse DeWitt line element on Riem(Σ) (but its indefiniteness causes some problems).

Another way is to intrinsically compute on each configuration a finite number of quantities, i: Q $\longrightarrow \mathbb{R}^n$, and then use the $\mathbb{R}^n$ norm $D^{\mathrm{i}}_{\mathrm{Eucl}}(Q_\Delta, Q'_\Delta)$ $= ||\mathrm{i}(Q_\Delta) - \mathrm{i}(Q'_\Delta)||^2$ (though this is limited for some purposes by i having a nontrivial kernel). E.g. one can compare SC's in Q(N, d), R(N, d) or $\mathcal{R}$(N, d) by letting i be the total moment of inertia for each SC (a mass-weighted norm). In geometrodynamical theories, one could additionally compute geometrical quantities to serve as i, or embed N points in a uniformly random way in each geometry and then use the pairwise metric distances between the points to furbish a vectorial i. Or, one could use total volume, anisotropy parameter or a vector made out of these, or use curvature invariants such as maximal or average curvatures of a given 3-space (e.g. objects related to the Weyl tensor which are also perported measures of gravitational information, see Sec 5). Or, for nonhomogeneous GR, one could compute eigenvalues of an operator D associated with that geometry (e.g. Laplacian or Yano–Bochner operators) and construct a spectral measure i from these. Another measure of inhomogeneity that could be used as an i would be an energy density contrast type quantity F[$\varepsilon/\langle\varepsilon\rangle$] (for ε the energy density distribution and $\langle\ \rangle$ denoting average over some volume) such as $\varepsilon/\langle\varepsilon\rangle$ or $\left\langle \frac{\varepsilon}{\langle\varepsilon\rangle}\log\left(\frac{\varepsilon}{\langle\varepsilon\rangle}\right)\right\rangle$, which particular functional form [22] also has information content connotations (see Sec 6). One can readily supply a notion of 'within ϵ of' for each above structure (contingent to what distance axioms it obeys), thus obtaining examples of grainings. RPM's with their local particle clusters, and inhomogeneous perturbations about minisuperspace with their localized bumps, are two such settings.

Four further question types can then be addressed. Two generalize their primed counterparts to model the imperfection of observation. **Be**1), does $q_\Delta(P)$ have acceptable properties? This is now for a graining set P rather than for an individual instant p. **Be**2), if properties of $q_\Delta(P)$ are known, does this permit deduction of any properties of $q_{\Delta'}(P)$ for Δ' disjoint from

Δ? The other two involve the Q space of the theory or theories that the observations are perported to belong to. **BeS**1) is: what is P($q_\Delta(P)$) within the collection of SC spaces? **BeS**2) is: what is P($q_{\Delta'}(P)$ has properties $\mathcal{P}'|q_\Delta(P)$ has properties $\mathcal{P}$)?[1] Examples of such questions are: what is P(space is almost flat)? What is P(space is almost isotropic)? What is P(space is almost homogeneous)?

Configuration comparers and decorated instants

The above single-configuration notion of closeness may not suffice for some

purposes (whether in principle or through lack of mathematical structure leaving one bereft of theorems through which to make progress). Other notions of closeness on the collection may depend on a fuller notion of comparison *between* instants, i.e. their joint consideration rather than a subsequent comparison of real numbers extracted from each individually. That may serve as a means of judging which instants are similar, or of which instants can evolve into each other along dynamical trajectories. Some criteria to determine which notion to use [27] are adherence to the axioms of distance, gauge or 3-Diffeomorphism invariance as suitable, and, for some applications, whether it can be applied to grand heaps.

One way of providing comparers is to upgrade the previous subsection's normed spaces and geometries to inner product spaces, metric spaces and topological spaces [23, 1]. In the case of inner products or metrics, $\mathcal{M}^{\Gamma\Delta}Q_\Gamma Q'_\Delta$ then supplies a primitive comparer of unprimed and primed objects Q_Γ, Q'_Δ.

Also consider replacing Q_Δ by the tangent bundle $\mathsf{T}(\mathsf{Q}_\Delta)$ (configuration-velocity space [10]), or the unit tangent bundle $\mathsf{T}_u(\mathsf{Q}_\Delta)$ (configuration-direction space), or the cotangent bundle $\mathsf{T}^*(\mathsf{Q}_\Delta)$ (configuration-momentum space, which, if augmented by a symplectic structure, is phase space). Such notions continue to exist for restricted configuration spaces in cases with constraints. This last feature involves quotienting operations, which can considerably complicate structure in practise. Envisage all these as *'heaps of decorated instants'*, H, which more general notion I use to supercede Q.

A common situation is to compare not configurations Q_Γ and Q'_Δ but rather the corresponding velocities $\dot{Q}_\Gamma$ and $\dot{Q}'_\Delta$, with the comparer employing the kinetic metric An example of such a comparer is the Lagrangian

[1] P denotes probability and | denotes 'given that', i.e. conditional probability.

$\mathcal{L}$: T(G-bundle over Q) $\longrightarrow \mathbb{R}$ $\mathcal{L}[Q_\Delta, g_\Lambda, \dot{Q}_\Delta, \dot{g}_\Lambda] = 2\sqrt{\mathsf{T}\{\mathsf{U}+\mathsf{E}\}}$, where, in this seminar's examples, U is minus the potential term $\mathsf{V}(Q_\Delta)$ and T is the kinetic term $\mathsf{T}[Q_\Delta, g_\Lambda, \dot{Q}_\Delta, \dot{g}_\Lambda] = M^{\Gamma\Delta}(Q_\Theta)\{\overrightarrow{G}\ \dot{Q}_\Gamma\}\ \overrightarrow{G}\ \dot{Q}_\Delta/2$ for $\overrightarrow{G}$ the action of the group G of redundant motions whose generators are parametrized by auxiliary variables g_Λ. [Here, the dot denotes the derivative with respect to label-time, an overall time that is meaningless because the actions considered are invariant under label change (= reparametrization)].

This also exemplifies that one often corrects the Q_Γ or $\dot{Q}_\Gamma$ with respect to a group G of transformations under which they are held to be physically unchanged. That involves the group action of G on the Q_Γ or $\dot{Q}_\Gamma$. E.g. for particle velocities $\dot{q}_{i\alpha}$, the infinitesimal action of the rotations (generated by b_α) is $\dot{q}_{i\alpha} \longrightarrow \overrightarrow{R}\ \dot{q}_{i\alpha} = \dot{q}_{i\alpha} + q_{i\alpha} \times \dot{b}_\alpha$ E.g. for 3-metric velocities $\dot{h}_{\mu\nu}$, the infinitesimal action of the 3-diffeomorphisms (generated by B_α) is $\dot{h}_{\mu\nu} \longrightarrow \overrightarrow{\mathrm{Diff}}\ \dot{h}_{\mu\nu} = \dot{h}_{\mu\nu} - \pounds_{\dot{B}} h_{\mu\nu}$. One furthermore often then minimizes with respect to the group generator (arbitrary frame 'shuffling auxiliary'). This ensures the physical requirement of G-invariance (i.e. gauge invariance, including 3-diffeomorphism invariance in geometrodynamics).

Then one has e.g. 1) The *Kendall-type comparer* [1] $\min_{g \,\in\, \mathrm{G}} \mathcal{M}^{\Gamma\Delta}\mathcal{Q}_\Gamma\ \overrightarrow{G}\ \mathcal{Q}'_\Delta$ for $\overrightarrow{G}$ the *finite* group action. 2) Construct $\mathcal{M}^{\Gamma\Delta}\{\overrightarrow{G}\ \dot{Q}_\Gamma\}\ \overrightarrow{G}\ \dot{Q}_\Delta$ for $\overrightarrow{G}$ the *infinitesimal* group action. Then weight by $\mathsf{U}+\mathsf{E}$, square-root, then integrate with respect to spatial extent if required and with respect to label time so as to produce the corresponding action. Variation of this ensures G-independence. Actions of this form include [1] the Jacobi action for mechanics, Barbour–Bertotti type actions for RPM, and the Baierlein–Sharp–Wheeler type actions for geometrodynamics. The variational procedure then entails minimization with respect to g_Λ. One could also weight by $1/\{\mathsf{U}+\mathsf{E}\}$ and square-root. This gives Leibniz–Mach–Barbour timefunctions (c.f. [10, 26]). Another variant is the DeWitt measure of distance: let one $\dot{h}_{\alpha\beta}$ and 1 of the 2 metrics in each factor of the DeWitt supermetric be with respect to primed coordinates, integrate with respect to both primed and unprimed space, and *then* square-root. One then obtains a semi-Riemannian metric functional (in the sense of 'Finslerian metric function'). In inhomogenous geometrodynamics, one can likewise decompose combined measures of local size and shape into separate comparers (c.f. [27] and techniques in [28]). In each case, the individual rather than combined comparers are better-behaved as distances.

Comparers along the lines of 1) and 2) are universal, insofar as they apply both to RPM's and to GR. However, the GR version has an indefinite inner product which does not confer good distance properties in contrast to the positive definite one in mechanics. Thus one might need different tools in each case, or use only the *shape part* of the GR inner product, which is itself positive definite [27]. Instead of using highly redundant variables alongside gauge auxiliaries and a shuffling procedure, one could work with reduced gauge-invariant configurations Q_Ω, for Ω a smaller indexing set than Δ, and a Lagrangian $\widetilde{\mathcal{L}} : \widetilde{\mathsf{T}}(\mathsf{Q}_\Omega) \longrightarrow \mathbb{R}$ constructed from these, $\widetilde{\mathcal{L}}[Q_\Omega, \dot{Q}_\Omega] = 2\sqrt{\widetilde{\mathsf{T}}\{\widetilde{\mathsf{U}} + \widetilde{\mathsf{E}}\}}$ for $\widetilde{\mathsf{T}}[Q_\Omega, \dot{Q}_\Omega]$ a suitable, 'more twisted' kinetic term. While, one seldom has this luxury of explicit gauge-invariant variables being available, it is available [23] for the 2d RPM of pure shape. The reduced configuration space metric is the Fubini–Study metric, from which this example's 'more twisted' kinetic term is formed. The associated notion of distance is then useable between 2d shapes. Alternatively, one could work with (more widely available) secondary quantities that are guaranteed to have the suitable invariances, e.g. further spectral measures.

If there's a sense of more than one instant, there is one becoming question type per question type above, by the construction Prob(if $Q_\Delta(p)$ has properties $\mathcal{P}$ then it becomes $Q'_\Delta(p')$ with properties $\mathcal{P}'$). I denote each such question type as above but with '**Become**' rather than 'Be'.

If there were a notion of time

Then yet further question types would emerge. Each of the non-statespace questions can now involve each instant being *prescribed to be at a time*. I denote this by appending a **T**. The new Be questions concern 'being at a particular time', while the new Become questions are of the form 'X at time 1 becomes Y at time 2'. For questions concerning heaps there is a further ambiguity: 'at any time' now makes sense as well as 'at a particular time'. Thus for each BeS question there are two **BeST** questions (denoted **a**, **b**), and for each BecomeS question there are four **BecomeST** questions (denoted **a**, **b**, **c**, **d**). Thus 32 question-types have been uncovered.

Further analysis of question-types and of time

First note **Suppression 1**: the 8 primed questions are clearly just subcases of their more realistic unprimed counterparts. Next note that the previous subsection crucially does not say what time is. Ordinary classical physics is easily excused: there is a real number valued external time, so that each

H is augmented to an extended heap space $\mathsf{H} \times \mathbb{R}$. One key lesson from GR, however, is that there is no such external time. Stationary spacetimes (including SR's Minkowski spacetime) do possess a timelike Killing vector, permitting a close analogue of external time to be used, but the generic GR solution permits no such construction. The generic solution of GR has a vast family of coordinate timefunctions, none of which has a privileged status unlike that associated with a stationary spacetime's timelike Killing vector. Questions along the lines of those above which involve time need thus specify *which* time. Using 'just any' time comes with the multiple choice and functional evolution [20] subaspects of the POT – this ambiguity tends to lead to inequivalent physics at the quantum level.

Another way of partly adhering to the above key lesson, which can be modelled at the level of nonrelativistic but temporally-relational mechanical models, is that 'being, at a time t_0' is *by itself* meaningless if one's theory is time label reparametrization invariant.

Alternatives that render particular times, whether uniquely or in families up to frame embedding variables, meaningful are specific internal, emergent or apparent time approaches. Therein, time is but a property that can be read off the (decorated sub)configuration. E.g. the notion of time in [13] can be thought of in this way. Thereby one has **Subsumption 2**: all question types involving a T are turned into the corresponding question types without one. This property might concern a clock within the environment/background, within the subsystem under study, or partly within both. Indeed, one could have a universe-time to which all parts of the configuration contribute rather than a clock *sub*system.

Subsumption 3: Each BecomeST b, c pair becomes a single question type if there is time reversal invariance. **Subsumption 4**: If the time used is globally defined on H, BeSTb questions and BecomeSTd questions are redundant. This can in any case be attained by considering restricted H defined so that this is so. (Whether that excludes interesting physics is then pertinent). At this stage, one is left with 8 question types.

Subsumption 5 has been suggested by Page (e.g [14]) and also to some extent Barbour [3]. It consists in supplanting all becoming questions by more operationally accurate being questions as follows. It is not the past instant that is involved, but rather this appearing as a memory/subrecord in the present instant, alongside the subsystem itself. Thus this is in fact a correlation within the one instant. In this scheme, one does not have a sequence of events but rather one present event that contains memories or other evidence of 'other events'.[2]

If subsumption 5 is adopted, the remaining question types are Be2 about how likely a correlation between two subsystems within the one grained subinstant is, theory-observation question type BeS2 about how likely an instant is within a statespace, and two 'consistency' question types Be1 and BeS1 about properties of a subinstant. If subsumption 5 is not adopted (or not adoptable in practise), there are additionally four corresponding types of becoming questions. Reasons why subsumption 5 might not be adopted, or might not be a complete catch-all of what one would like to be explained include I) impracticality: studying a subsystem S now involves studying a larger subsystem containing multiple imprints of S. Models involving memories would be particularly difficult to handle (see footnote 2). II) If one wants a scheme that can explain the Arrow of Time, then Page's scheme looks to be unsatisfactory. While single instants such as that in footnote 2 could be used to simulate the scientific process as regards 'becoming questions', N.B. that these single instants correspond to the *latest* stage of the investigation (in the 'becoming' interpretation), while 'earlier instants' will not have this complete information. III) Additionally, important aspects of the scientific enterprise look to be incomplete in this approach – in interpreting present correlations, one is in difficulty if one cannot affirm that one did in fact prime the measuring apparatus would appear to retain its importance. I.e. as well as the 'last instant' playing an important role in the interpretation, initial conditions implicit in the 'first instant' also look to play a role [1].

QM and beyond?

At the classical level, one could either take certainty to be a subcase of probability, or note that even classically it is probabilities that are relevant in practise – e.g. due to limits on precision of observations. 2) A notion of P(trajectory goes through a subregion Δ for each space H) is then required. This is particularly common in the literature in the case in which H is phase space. Then if one canonically-quantizes, the Hamiltonian provides a TISE such as the WDE in the case of GR.

[2] As an illustrative sketch, one can imagine a configuration in which the record actually under study is the naïve record plus the observer next to it, whose memory includes a SC which encodes himself peering at the record 'at an earlier time' and a SC in which he has this first memory and a prediction 'derived from it'.

Because they refer to configurations, such as the almost flat, almost isotropic and almost homogeneous questions have obvious counterparts in configuration–representation QM; in concerning pieces of the configuration space these questions lie outside the usual domain of QM. The Naïve Schrödinger Interpretation and the Conditional Probabilities Interpretation are two interpretations outside or beyond conventional QM formalism suggested to answer such questions. The former serves to address the BeS1 version of this paragraph's questions, such as what is P(universe is almost flat) or what is P(Inflation) [1]. The latter addresses Be2 or BeS2 questions such as P(One part of the sky is smooth | another is) [all within a given instantaneous configuration].

Are records typically useful?

Records Theory requires A) for SC's to be capable of holding enough information to address whatever issues are under investigation. Thus Information Theory is pertinent. Information being (more or less) negentropy, a starting classical notion is the Boltzmann-like $I_{\mathrm{Boltzmann}} = -\log W$ (using $k_{\mathrm{B}} = 1$ units) for W the number of microstates. One could furthermore use such as *Shannon information*, $I_{\mathrm{Shannon}}(p_x) = \sum_x p_x \log p_x$ for p_x a discrete probability distribution for the records, or $I_{\mathrm{Shannon}}[\sigma] = \int \mathrm{d}\Omega \sigma \log \sigma$ for σ a continuous probability distribution. If one considers records at the quantum level, then one could instead use such as *von Neumann information*,

$I_{\mathrm{von\ Neumann}}[\rho] = \mathrm{Tr}(\rho \log \rho)$ for ρ the QM density matrix. These notions have suitable properties and remain applicable [1] in passing to QFT and GR contexts. One contention in interpreting (0) at the general level required for developing a POT strategy is that information is minus entropy and classical (never mind quantum) gravitational entropy is a concept that is not well understood or quantified for general spacetimes [1]. Quantum gravity may well have an information notion $I[\rho_{\mathrm{QGrav}}] = \mathrm{Tr}\rho_{\mathrm{QGrav}} \log \rho_{\mathrm{QGrav}}$, but either the quantum-gravitational density matrix is an unknown object since the underlying microstates are unknown, or, alternatively, one would need to provide an extra procedure for obtaining this, such as how to solve and interpret the WDE, which would be fraught with numerous further technical and conceptual problems.

Rather than a notion of gravitational information that is completely general, a notion of entropy suitable for approximate classical and quantum cosmologies may suffice for the present study. Quite a lot of candidate objects of this kind have been proposed. However, it is unclear how some of these would arise from the above fundamental picture, while for others it is not clear that the candidate does in fact possess properties that make it a bona fide entropy [1]. Cosmologically relevant information notions proposed to date include some that are manifestly related to the above conventional notions of information, and also [22] use $I_{\text{HBM}}[\varepsilon] = \int \text{d}\Omega\varepsilon\log(\varepsilon/\langle\varepsilon\rangle)$ and $I'_{\text{HBM}}[\varepsilon] = \langle\varepsilon\log(\varepsilon/\langle\varepsilon\rangle)\rangle$, the first of which is a relative information type quantity (see Sec 6).

B) However, whether there is a pattern in a record or collection of records (and whether that pattern is significant rather than random) involves more than just how much information is contained within. Two placings of the same pieces on a chessboard could be, respectively, from a grandmasters' game and frivolous. What one requires is a general quantification of there being a pattern. This should be linked at least in part to information content, in that the realization of at least some complicated patterns requires a minimum amount of information. Records Theory is, intuitively, about drawing conclusions from similar patterns in different records.

Consider also the situation in which information in a curve or in a wave pulse that is detectable by/storeable in a detector in terms of approximands or modes. As regards localized useable information content per unit volume, considering the Joos–Zeh dust–CMB and α-track–bubble chamber side by side suggests that most records in nature/one's model will be poor or diffuse. For the Joos–Zeh [21] example the 'somewhere' is all over the place: "*in the vastness of cosmological space*". Detectors, such as the extension of Halliwell's 1-piece detector model (Sec 3.3, [17]) to a cluster, could happen to be tuned to pick up the harmonics that are principal contributors in the signal. In this way one can obtain a good approximation to a curve from relatively little information. E.g. compare the square wave with the almost-square wave that is comprised of the first 10 harmonics of the square wave. That is clearly specific information as opposed to information storage capacity in general. Likewise, a bubble chamber is attuned to seeing tracks, a detector will often only detect certain (expected) frequencies. Through such specialization, a record that 'stands out' can be formed. One should thus investigate is quantitatively which of the α-track and 'dust grain' paradigms is more common.

C) Information can be lost from a record 'after its formative event' – the word "stored" in (0) can also be problematic. Photos yellow with age and can be defaced or doctored.

Correlations between records

One concept of possible use is *mutual information*: this is a notion $M(A, B) = I(A) + I(B) - I(AB)$ for AB the joint distribution of A and B for each of classical Shannon or QM von Neumann information. This is a quantity of the *relative information type* [1], $I_{\mathrm{relative}}[p, q] = \sum_x p_x \mathrm{log}(p_x/q_x)$ (discrete case), $I_{\mathrm{relative}}[\sigma, \tau] = \int \mathrm{d}\Omega\sigma\mathrm{log}(\sigma/\tau)$ (continuous case), (the object in Sec 4.1 is a special case of the continuous case of this in which the role of the second distribution is played by the average of the first). The QM counterpart of relative information is $I_{\mathrm{relative}}[\rho_1, \rho_2] = \mathrm{Tr}(\rho_1\{\mathrm{log}\rho_1 - \mathrm{log}\rho_2\})$; mutual information also has QM analogues. It is not clear that these notions cover all patterns. Two records could be part of a discernible common pattern even if their constituent information is entirely different, e.g. the pattern to spot on two chessboards could be interprotection, manifest between rooks on one and between knights on the other.

Another is the family of notions of correlator/n-point function in the cosmological or QFT senses (or both at once).

Further features of Records Theory

Barbour furthermore asks [3] whether there are any *selection principles* for such records (which he calls 'time capsules'; the bubble chamber with the α-particle track within is a such). If these features are to be incorporated, one would additionally need a (relative) measure of semblance of dynamics. How does a record achieve this encodement? Are SC's that encode this generic? Let us suppose that this is actually a special rather than generic feature for a SC to have. This would be the case if the dust grain–CMB photon paradigm is more typical than the α-particle–bubble chamber one. Then one would have the problem of explaining why the universe around us nevertheless contains a noticeable portion of noticeably history-encoding records, i.e. a selection principle would be needed.

Barbour suggests a selection principle based on the following layers [3]. 1) There are some distinctive places in the configuration space. 2) The wavefunction of the universe peaks around these places, making them probable. 3) These parts of the configuration space contain records that bear a semblance of dynamics ('time capsules'). [Following my arguments in Sec 1, this should be rephrased in terms of SC's.] Some doubts are cast on this scheme in [1]. In particular, A) Barbour supplies no concrete mathematical model evidence for there being any correlation between SC's being time capsules and their being near a distinctive feature of SC space such as a change of stratum or a point of great uniformity. B) Semiclassicality might either explain or supplant Barbour's selection principle, while there are additionally two further a priori unrelated selection principles in the literature, which could be viewed either as competitors or as features that Barbour's scheme should be checked to be able to account for: I) *branching processes* and II) *consistency conditions in the Histories Theory framework*.

One reason that Barbour favours the above scheme is so as to be open to the possibility of explaining the Arrow of Time, unlike I), II), [8] (which builds in a time asymmetry in the choice of admitted solutions), and Page's scheme (which is subject to the difficulties pointed out in Sec 4.4). These various interesting issues should be further investigable using RPM's.

As regards Records Theory as a POT resolution, limitations exposed in this seminar are as follows. Records are "somewhere in the universe that information is stored when histories decohere". But a suitable notion of localization in space and in configuration space may be hard to come by and/or to use for quantum gravity in general – 'where' particular records are can be problematic to quantify, and the records can be problematic to access and use too, since the relevant information may be 'all over the place'. Also, 'information' is problematic both as it may be of too poor a quality to reconstruct the history and because a suitably general notion of information is missing from our current understanding of classical gravity, never mind quantum gravity with its unknown microstates (mechanical toy models are useful in not having this last obstruction). Finally, the further Records Theory notions of significant correlation patterns and how one is to deduce dynamics/history from them looks to be a difficult and unexplored area even in simpler contexts than gravitation.

Acknowledgments

I thank Peterhouse for funding and the organizers of TAM 2007 for inviting me.

References

[1] See E. Anderson, [arχiv:0709.1892] and references therein.
[2] J.S. Bell, in *Quantum Gravity 2. A Second Oxford Symposium* ed. C.J. Isham, R. Penrose and D.W. Sciama (Carendon, Oxford 1981).
[3] J.B. Barbour, Class. Quant. Grav. **11** (1994) 2875; *The End of Time*, (Oxford University Press, Oxford 1999).

[4] N.F. Mott, Proc. Roy. Soc. Lon. A **126** (1929) 79.
[5] J.J. Halliwell, Phys. Rev. D **64** (2001) 044008.
[6] J.J. Halliwell, *The Future of Theoretical Physics and Cosmology (Stephen Hawking 60th Birthday Festschrift volume)*, ed. G.W. Gibbons, E.P.S. Shellard and S.J. Rankin (Cambridge University Press, Cambridge 2003).
[7] M. Castagnino and R. Laura, Int. J. Theor. Phys. **39** (2000) 1737.
[8] M. Castagnino, Phys. Rev. D **57** (1998) 750.
[9] J.B. Barbour and B. Bertotti, Proc. Roy. Soc. Lond. A **382** (1982) 295.
[10] J.B. Barbour, Class. Quant. Grav. **11** (1994) 2853.
[11] J. Barbour, B. Foster and N.Ó. Murchadha, Class. Quant. Grav. **19** (2002) 3217.
[12] E. Anderson, Stud. Hist. Philos. Mod. Phys. **38** (2007) 15.
[13] D.N. Page and W.K. Wootters, Phys. Rev. D **27** (1983) 2885.
[14] D.N. Page, in *Physical Origins of Time Asymmetry* ed. J.J. Halliwell, J. Perez-Mercader and W.H. Zurek (Cambridge University Press, Cambridge 1994).
[15] D.N. Page, Int. J. Mod. Phys. D **5** (1996) 583, [arχiv:quant-ph/9507024]; *Consciousness: New Philosophical Essays*, ed. Q. Smith and A. Jokic (Oxford University Press, Oxford 2002), [arχiv:quant-ph/0108039].
[16] M. Gell-Mann and J.B. Hartle, Phys. Rev. D **47** (1993) 3345.
[17] J.J. Halliwell, Phys. Rev. D **60** (1999) 105031.
[18] J.B. Barbour and L. Smolin, unpublished, dating from 1989.
[19] E. Anderson, Class. Quant. Grav. **23** (2006) 2469; Class. Quant. Grav. **23** (2006) 2491.
[20] K.V. Kuchař, *Proceedings of the 4th Canadian Conference on General Relativity and Relativistic Astrophysics*, ed. G. Kunstatter, D. Vincent and J. Williams (World Scientific, Singapore 1992); C.J. Isham, *Integrable Systems, Quantum Groups and Quantum Field Theories* ed. L.A. Ibort and M.A. Rodríguez (Kluwer, Dordrecht 1993).
[21] E. Joos and H.D. Zeh, Z. Phys. B **59** (1985) 223.
[22] A. Hosoya, T. Buchert and M. Morita, Phys. Rev. Lett. **92** (2004) 141302.
[23] E. Anderson, Class. Quant. Grav. **24** (2007) 5317; [arχiv:gr-qc/0706.3934].
[24] E. Anderson, 2 forthcoming papers.
[25] J.J. Halliwell and S.W. Hawking, Phys. Rev. D **31** (1985) 1777.
[26] E. Anderson, Class. Quant. Grav. **24** (2007) 2935; Class. Quant. Grav. **24** (2007) 2971.
[27] E. Anderson, forthcoming technical papers on Records Theory.
[28] E. Anderson *et al.*, Class. Quant. Grav. **20** (2003) 157; Class. Quant. Grav. **22** (2005) 1795.

Time and Matter
2007 Bled, Slovenia

Towards a New Paradigm: Relativity in Configuration Space

MATEJ PAVŠIČ*
Jožef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia

Abstract: We consider the possibility that the basic space of physics is not spacetime, but configuration space. We illustrate this on the example with a system of gravitationally interacting point particles. It turns out that such system can be described by the minimal length action in a multidimensional configuration space $\mathcal{C}$ with a block diagonal metric. Allowing for more general metrics and curvatures of $\mathcal{C}$, we step beyond the ordinary general relativity in spacetime. The latter theory is then an approxiamtion to the general relativity in $\mathcal{C}$. Other sorts of configuration spaces can also be considered, for instance those associated with extended objects, such as strings and branes. This enables a deeper understanding of the geometric principle behind string theory, and an insight on the occurrence of Yang-Mills and gravitational fields at the 'fundamental level'.

Introduction

After many decades of intensive research there is still no general consensus on the major persisting puzzles such as the unification of fundamental interactions, quantum gravity, the problem of time, the cosmological constant problem, the nature of dark matter and energy, etc. From history we know that such situation calls for a 'paradigm shift'. We also know that often a formalism is more powerful than initially envisaged. For example, in the Hamilton-Jacobi function there is a hint of quantum mechanics, which could have been guessed much earlier before its experimental discovery. The line element in Minkowski spacetime suggested its generalization to curved spacetime and thus the theory of gravity. Clifford algebra led to the Dirac theory of electron. In all those cases the formalism itself pointed to its own generalization! This introduced important new physics.

* matej.pavsic@ijs.si

Having in mind such lessons from history it seems reasonable to do something analogous with the currently available formalisms, and to step beyond the existing paradigm. We will first examine the formalism that describes a system of point particles in the presence of gravity. We will then consider a generalization of the theory of relativity in which spacetime M_4 is replaced by the configuration space $\mathcal{C}$ associated with a given physical system. The system will be considered as a point that traces a geodetic line in configuration space. Such theory predicts in general a different dynamical behavior of a many particle system than does the ordinary theory. But in particular, for a suitable metric on $\mathcal{C}$, we obtain the ordinary many particle action in the presence of gravitational field. In general, the configuration space can have non vanishing curvature. From the point of view of 4-dimensional spacetime, which is a subspace of $\mathcal{C}$, there exist extra forces that act on a particle, besides the ordinary gravity. Observations suggest that the ordinary theory of gravity cannot be straightforwardly applied to large scale systems, such as galaxies, clusters of galaxies, and the universe. Instead, one has to introduce the concept of dark matter [1] and dark energy [1], or alternatively, to consider suitable modifications of the theory of gravity (MOND) [2]. We propose to explore the possibility that general relativity, not in spacetime M_4, but in multidimensional configuration space $\mathcal{C}$ might solve such astrophysical puzzles. The theory can also be applied to other kinds of configuration spaces, e.g., those associated with extended objects such as strings [3] and branes [4]. This enables a deeper understanding of the geometric principle behind string theory, and the insight on the occurrence of the Yang-Mills and gravitational fields.

Generalizing relativity

Configuration space replaces spacetime

Let us consider a system of point particles in the presence of a gravitational field $g_{\mu\nu}$. The action is the sum of the individual point particle actions

$$I[\dot{X}_i^\mu] = \sum_i \int \mathrm{d}\tau \sqrt{\dot{X}_i^\mu \dot{X}_i^\nu \, m_i \, g_{\mu\nu}(X_i^\mu)} \tag{1}$$

We will now rewrite this into an equivalent form.

Let us recall that a point particle action

$$I[X^\mu] = \int \mathrm{d}\tau \, m \sqrt{\dot{X}^\mu \dot{X}^\nu \, g_{\mu\nu}} \tag{2}$$

has its equivalent in the Schild action

$$I[X^\mu] = \int \mathrm{d}\tau \, \frac{m}{k} \, \dot{X}^\mu \dot{X}^\nu \, g_{\mu\nu} \tag{3}$$

which is a gauge fixed action with

$$\dot{X}^\mu \dot{X}^\nu g_{\mu\nu} = k^2 = \text{const.} \tag{4}$$

where k is a constant.[1]

The Schild action for a system of point particles is

$$I[\dot{X}_i^\mu] = \int \mathrm{d}\tau \sum_i \dot{X}_i^\mu \dot{X}_i^\nu \, \frac{m_i}{k_i} \, g_{\mu\nu}(X_i^\mu) \tag{5}$$

This can be considered as a quadratic form in a multidimensional space $\mathcal{C}$ whose dimension is 4 times the number N of particles in the system. To see this more clearly, it is convenient to introduce a more compact notation:

$$\dot{X}_i^\mu \equiv \dot{X}^{(i\mu)} \equiv \dot{X}^M, \quad M = (i\mu) \tag{6}$$

$$\frac{m_i}{k_i} \, g_{\mu\nu} \equiv \frac{M}{K} \, g_{(i\mu)(j\nu)} \equiv \frac{M}{K} \, g_{MN} \tag{7}$$

Then the action (5) becomes

$$I[X^M] = \int \mathrm{d}\tau \, \dot{X}^M \dot{X}^N \, \frac{M}{K} \, g_{MN}(X^M) \tag{8}$$

which is the Schild action in $\mathcal{C}$. The $4N$-dimensional space $\mathcal{C}$ is *the configuration space* associated with a system. From the context it should be clear when M is a double index $M \equiv (i\mu)$, and when it is a constant, analogous to single particle mass m.

From the equations of motion derived from the action (8) it follows

$$\dot{X}^M \dot{X}^M \, g_{MN} = K^2 \tag{9}$$

[1]Variation of the action (2) gives $\mathrm{d}(\dot{X}^\nu \, g_{\mu\nu})/\mathrm{d}\tau - g_{\alpha\beta,\mu} \, \dot{X}^\alpha \dot{X}^\beta/2 = 0$. This can be rewritten in the forms $(1/\sqrt{\dot{X}^2}) \, \mathrm{d}(\dot{X}^\nu g_{\mu\nu}/\sqrt{\dot{X}^2})/\mathrm{d}\tau - g_{\alpha\beta,\mu} \dot{X}^\alpha \dot{X}^\beta/2\dot{X}^2 - (1/\sqrt{\dot{X}^2}) \, \mathrm{d}(1/\sqrt{\dot{X}^2})/\mathrm{d}\tau \; \dot{X}^\nu g_{\mu\nu} = 0$. If we multiply this by $\dot{X}^\mu$ (and sum over μ), then the first two terms give identically zero, so that we find $\sqrt{\dot{X}^2} \, \mathrm{d}(1/\sqrt{\dot{X}^2})/\mathrm{d}\tau = 0$, or $\dot{X}^2 \equiv g_{\mu\nu} \dot{X}^\mu \dot{X}^\nu = C^2$, with C being a constant. On the other hand, the momentum belonging to the Schild action is $p_\mu = m\dot{X}_\mu/k$. Thus $m^2 = p_\mu p^\mu = m^2 \dot{X}_\mu \dot{X}^\mu/k^2 = m^2 C^2/k^2$ which implies $C^2 = k^2$.

where K is a constant. Explicitly this reads

$$\dot{X}_1^2 + \dot{X}_2^2 + \ldots + \dot{X}_N^2 = K^2 \tag{10}$$

Rewriting the latter equation as

$$\frac{\dot{X}_1^2}{K^2} = 1 - \frac{\dot{X}_2^2}{K^2} - \frac{\dot{X}_3^2}{K^2} - \ldots - \frac{\dot{X}_N^2}{K^2} \tag{11}$$

multiplying it by M^2, and using the expression

$$p_M = \frac{M\dot{X}_M}{K} \equiv \frac{M\dot{X}_{i\mu}}{K} \tag{12}$$

we find

$$\frac{M^2\dot{X}_1^2}{K^2} = M^2 - p_2^2 - p_3^2 - \ldots - p_N^2 = p_1^2 \equiv m_1^2 \tag{13}$$

$$\frac{M^2}{K^2} = \frac{m_1^2}{k_1^2} \quad \text{or} \quad \frac{M}{K} = \frac{m_1}{k_1} \tag{14}$$

where

$$k_i^2 = \dot{X}_i^2 = g_{(i\mu)(i\nu)}\dot{X}_i^\mu \dot{X}_i^\nu \quad \text{(no sum over) } i, i = 1, 2, \ldots, N \tag{15}$$

Since the above derivation can be repeated for any $i = 1, 2, \ldots, N$, we have

$$\frac{M}{K} = \frac{m_i}{k_i} \tag{16}$$

Eqs. (9),(16) and (16) thus imply

$$M^2 = \sum_i p_i^2 \tag{17}$$

$$p_{i\mu} = \frac{M\dot{X}_{i\mu}}{\sqrt{\dot{X}^2}} = \frac{m_i\dot{X}_{i\mu}}{\sqrt{\dot{X}_i^2}} \tag{18}$$

The Schild action in $\mathcal{C}$ is equivalent to the reparametrization invariant action $\mathcal{C}$:

$$I[X^M] = M \int \mathrm{d}\tau \sqrt{\dot{X}^M \dot{X}^N g_{MN}(X^M)} \tag{19}$$

which is proportional to the length of a worldline in $\mathcal{C}$. The constant M has the role of *mass* in $\mathcal{C}$.

Having arrived at the action (19), we will now assume that the metric g_{MN} need not be of the block diagonal form (7). We will assume that the configuration space $\mathcal{C}$ is a manifold equipped with metric G_{MN}, connection and curvature (that in general does not vanish).

In particular, for the block diagonal metric

$$G_{MN} \equiv G_{(i\mu)(j\nu)} = g_{(i\mu)(j\nu)} = \begin{bmatrix} g_{\mu\nu}(x_1) & 0 & 0 & \cdots \\ 0 & g_{\mu\nu}(x_2) & 0 & \cdots \\ 0 & 0 & g_{\mu\nu}(x_3) & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \tag{20}$$

we obtain the ordinary relativistic theory for a many particle system in a given gravitational field.

By allowing for a more general metric, that cannot be transformed into the form (20) by a choice of coordinates in $\mathcal{C}$, we go beyond the ordinary theory.

Configuration space $\mathcal{C}$ is the space of possible "instantaneous" configurations in M_4. Its points are described by coordinates $x^M \equiv x_i^\mu$. A given configuration traces a worldline $x^M = X^M(\tau)$ in $\mathcal{C}$ (see Fig. 1).

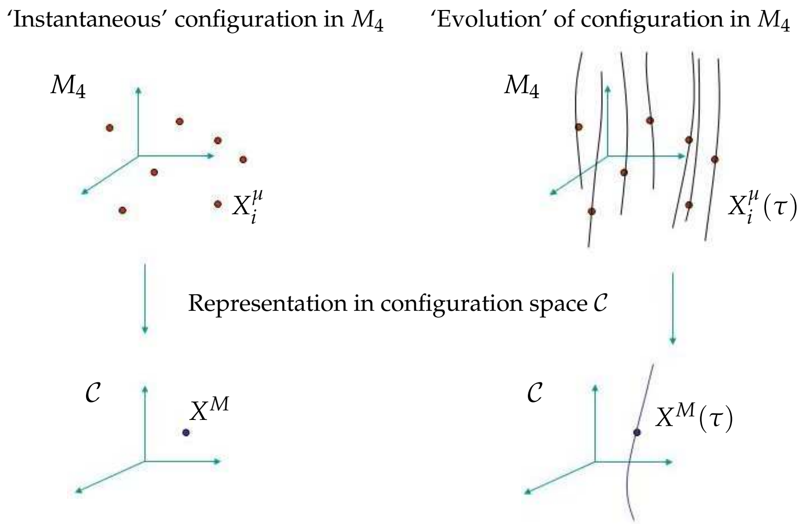


Figure 1: An 'instantaneous' configuration can be represented as a set of points in spacetime M_4, or as a point in configuration space $\mathcal{C}$. Analogously, a 'moving' configuration can be represented as a set of worldines in M_4, or a single worldline in $\mathcal{C}$.

A dynamically possible worldline in $\mathcal{C}$ is a geodesic in $\mathcal{C}$, and it satisfies the variation principle based on the action (19).

Instead of considering a fixed metric on $\mathcal{C}$, let us assume that the metric G_{MN} is dynamical, so that the total action contains a kinetic term for G_{MN}:

$$I[X^M, G_{MN}] = I_m + I_g \tag{21}$$

where

$$I_m = \int \mathrm{d}\tau \, M \sqrt{G_{MN}\dot{X}^M \dot{X}^N} = \tag{22}$$

$$= \int \mathrm{d}\tau \, M \sqrt{G_{MN}\dot{X}^M \dot{X}^N} \, \delta^D \left(x - X(\tau)\right) \mathrm{d}^D x \tag{23}$$

and

$$I_g = \frac{1}{16\pi G_D} \int \mathrm{d}^D x \sqrt{|G|} \, \mathcal{R} \tag{24}$$

Here $\mathcal{R}$ is the curvature scalar in $\mathcal{C}$. So we have *general relativity in configuration space* $\mathcal{C}$. We have arrived at a theory which is analogous to Kaluza-Klein theory. Configuration space is a higher dimensional space, whereas spacetime M_4 is a 4-dimensional subspace of $\mathcal{C}$, associated with a chosen particle.

The concept of configuration space can be used either in macrophysics or in microphysics. Configuration space associated with a system of point particles is finite dimensional. Later we will discuss infinite dimensional configuration spaces associated with strings and branes.

Equations of motion for a configuration of point particles

The equations of motion derived from the action (21) are the Einstein equations in configuration space $\mathcal{C}$. Let us now split the coordinates of $\mathcal{C}$ into 4-coordinates $X^\mu \equiv X^{1\mu}$, $\mu = 0, 1, 2, 3$ associated with position of a chosen particle, labeled by 1, and the remaining coordinates $X^{\bar{M}}$,

$$X^M = (X^\mu, X^{\bar{M}}) \tag{25}$$

The quadratic form occurring in the action (8) can then be split—according to the well known procedure of Kaluza-Klein theories—into a 4-dimensional part plus the part due to the extra dimensions of configuration space $\mathcal{C}$:

$$\dot{X}^M \dot{X}^N G_{MN} = \dot{X}^\mu \dot{X}^\nu g_{\mu\nu} + \text{extra terms} \tag{26}$$

More precisely, if for the metric on $\mathcal{C}$ we take the ansatz

$$G_{MN} = \begin{bmatrix} g_{\mu\nu} + A_\mu^{\bar{M}} A_\nu^{\bar{N}} \phi_{\bar{M}\bar{N}} & A_\mu^{\bar{N}} \phi_{\bar{M}\bar{N}} \\ A_\nu^{\bar{N}} \phi_{\bar{M}\bar{N}} & \phi_{\bar{M}\bar{N}} \end{bmatrix} \tag{27}$$

then we obtain

$$\dot{X}^M \dot{X}^N G_{MN} = \dot{X}^\mu \dot{X}^\nu g_{\mu\nu} + \dot{X}_{\bar{M}} \dot{X}_{\bar{N}} \phi^{\bar{M}\bar{N}} \tag{28}$$

where

$$\dot{X}_{\bar{M}} = G_{\bar{M}N} \dot{X}^N = A_{\bar{M}\mu} \dot{X}^\mu + \phi_{\bar{M}\bar{N}} \dot{X}^{\bar{N}} \tag{29}$$

Inserting expression (28) into the action (23), we have

$$I[X^\mu, X^{\bar{M}}] = M \int \mathrm{d}\tau \Big[\dot{X}^\mu \dot{X}^\nu g_{\mu\nu} + \tag{30}$$
$$+ \phi^{\bar{M}\bar{N}} (A_{\bar{M}\mu} \dot{X}^\mu + \phi_{\bar{M}\bar{J}} \dot{X}^{\bar{J}})(A_{\bar{N}\nu} \dot{X}^\nu + \phi_{\bar{N}\bar{K}} \dot{X}^{\bar{K}}) \Big]^{1/2}$$

where we have omitted subscript m.
Variation of the latter action with respect to X^μ gives

$$\frac{1}{\sqrt{\dot{X}^2}} \frac{\mathrm{d}}{\mathrm{d}\tau} \left(\frac{\dot{X}^\mu}{\sqrt{\dot{X}^2}} \right) + \frac{1}{\dot{X}^2} \Gamma^\mu_{\rho\sigma} \dot{X}^\rho \dot{X}^\sigma + \text{extra terms} = 0, \tag{31}$$

where $\dot{X}^2 \equiv g_{\rho\sigma} \dot{X}^\rho \dot{X}^\sigma$. This is just the 4-dimensional geodesic equation plus extra terms due to the extra coordinates of $\mathcal{C}$.
For the explicit derivation it is convenient to use, instead of (23), an equivalent action, namely the phase space action

$$I[X^M, P_M, \Lambda] = \int \mathrm{d}\tau \, (P_M \dot{X}^M - H) \tag{32}$$

where

$$H = \frac{\Lambda}{2M} \left(P_M P_N \, G^{MN} - M^2 \right) \tag{33}$$

is the "Hamiltonian" which—due to reparametrization invariance—is identically zero. Variation of the action (32) with respect to P_M and Λ, respectively, gives

$$P_M = \frac{M \dot{X}_M}{\Lambda}, \qquad \Lambda = \dot{X}_M \dot{X}_N \, G_{MN} \tag{34}$$

Splitting variables X^M according to (25), and analogously for P_M, we obtain

$$I[X^\mu,\, X^{\bar M},\, p_\mu,\, P_{\bar M},\, \Lambda] = \int \mathrm{d}\tau \, \left[p_\mu \dot X^\mu + P_{\bar M} \dot X^{\bar M} - H \right] \tag{35}$$

The Hamiltonian becomes

$$H = \frac{\Lambda}{2M} \left[g^{\mu\nu} \left(p_\mu - A^{\bar J}_\mu P_{\bar J} \right) \left(p_\nu - A^{\bar K}_\nu P_{\bar K} \right) + \phi^{\bar M \bar N} P_{\bar M} P_{\bar N} - M^2 \right] \tag{36}$$

where $p_\nu = P_\nu$ is 4-dimensional momentum.

Let us now assume that the "internal" subspace of $\mathcal{C}$ admits isometries given by the Killing vector fields $k^{\bar J}_\alpha$. Index α runs over the independent Killing vectors, whereas $\bar J$, like $\bar M$, $\bar N$, runs over the "internal" coordinates. Then, as it is customary in Kaluza-Klein theories, we write

$$A^{\bar J}_\mu = k^{\bar J}_\alpha A^\alpha_\mu \tag{37}$$

The metric $\phi^{\bar M \bar N}$ of the internal space can be rewritten in terms of a metric $\varphi^{\alpha\beta}$ in the space of isometries:

$$\phi^{\bar M \bar N} = \varphi^{\alpha\beta} k^{\bar M}_\alpha k^{\bar N}_\beta + \phi^{\bar M \bar N}_{\text{extra}} \tag{38}$$

Here $\phi^{\bar M \bar N}_{\text{extra}}$ are additional terms due to the directions that are orthogonal to isometries. For particular internal spaces $\bar{\mathcal{C}}$ those additional terms may vanish.

Introducing projections of momentum onto Killing vectors

$$p_\alpha \equiv k^{\bar J}_\alpha P_{\bar J} \tag{39}$$

and chosing a coordinate system in $\mathcal{C}$ in which

$$k^M_\alpha = \left(k^\mu_\alpha,\, k^{\bar M}_\alpha \right), \qquad k^\mu_\alpha = 0, \qquad k^{\bar M}_\alpha \neq 0. \tag{40}$$

The Hamiltonian (36) reads

$$H = \frac{\Lambda}{2M} \left[g^{\mu\nu} \left(p_\mu - A^\alpha_\mu p_\alpha \right) \left(p_\nu - A^\beta_\nu p_\beta \right) + \varphi^{\alpha\beta} p_\alpha p_\beta - M^2 \right] \tag{41}$$

For simplicity we will omit the extra terms $\phi^{\bar M \bar N}_{\text{extra}}$.

Now we can use the Hamilton equations of motion:

$$\dot p_\alpha = \{ p_\alpha,\, H \} \tag{42}$$

$$\dot{p}_\mu = \{p_\mu, H\} \tag{43}$$

Calculating the Poisson brackets

$$\{p_\alpha, p_\beta\} = \frac{\partial p_\alpha}{\partial X^J}\frac{\partial p_\beta}{\partial X_J} - \frac{\partial p_\beta}{\partial X^J}\frac{\partial p_\alpha}{\partial X_J} = \left(k^M_{\alpha,J}k^J_\beta - k^M_{\beta,J}k^J_\alpha\right)p_M = -C^\gamma_{\alpha\beta}p_\gamma \tag{44}$$

introducing the kinetic momentum

$$p_\mu - A^{\bar{J}}_\mu P_{\bar{J}} \equiv \pi_\mu, \qquad g^{\mu\nu}\pi_\nu = \frac{M}{\Lambda}\dot{X}^\mu \tag{45}$$

and the gauge field strength

$$F^\alpha_{\mu\nu} = \partial_\mu A^\alpha_\nu - \partial_\nu A^\alpha_\mu + C^\alpha_{\alpha'\beta'}A^{\alpha'}_\mu A^{\beta'}_\nu \tag{46}$$

we obtain

$$\dot{p}_\alpha = C^\gamma_{\alpha\beta}p_\gamma A^\beta_\mu \dot{X}^\mu - \frac{\Lambda}{2M}\varphi^{\alpha'\beta'}_{,\bar{J}}p_{\alpha'}p_{\beta'}k^{\bar{J}}_\alpha \tag{47}$$

$$\dot{\pi}_\mu - \frac{\Lambda}{2M}g_{\rho\sigma,\mu}\pi^\rho\pi^\sigma + F^\alpha_{\mu\nu}p_\alpha\dot{X}^\nu + \frac{\Lambda}{2M}\left(\varphi^{\alpha\beta}_{,\mu} - \varphi^{\alpha\beta}_{,\bar{J}}k^{\bar{J}}_{\alpha'}A^{\alpha'}_\mu\right)p_\alpha p_\beta = 0. \tag{48}$$

This is the well known *Wong equation* [5], with additional terms due to the presence of scalar fields $\varphi^{\alpha\beta}$.

Relation between the higher dimensional and 4-dimensional mass

If we rewrite the quadratic form (28) as

$$\frac{\dot{X}^\mu\dot{X}^\nu g_{\mu\nu}}{\dot{X}^M\dot{X}^N G_{MN}} = 1 - \frac{\dot{X}_{\bar{M}}\dot{X}_{\bar{N}}\phi^{\bar{M}\bar{N}}}{\dot{X}^M\dot{X}^N G_{MN}} \tag{49}$$

and multiply by M^2 we find

$$M^2\frac{\dot{X}^\mu\dot{X}^\nu g_{\mu\nu}}{\dot{X}^M\dot{X}^N G_{MN}} = M^2 - \phi^{\bar{M}\bar{N}}p_{\bar{M}}p_{\bar{N}} = g^{\mu\nu}p_\mu p_\nu = m^2 \tag{50}$$

where

$$P_M = \frac{M\dot{X}_M}{\sqrt{\dot{X}^J\dot{X}^K G_{JK}}} \tag{51}$$

From eq. (50) we obtain the ratio of the mass m in M_4 to the mass M in $\mathcal{C}$ expressed in terms of the corresponding velocity quadratic form,

$$\frac{m}{M} = \sqrt{\frac{\dot{X}^\mu X^\nu g_{\mu\nu}}{\dot{X}^M\dot{X}^N G_{MN}}} \tag{52}$$

For the 4-dimensional momentum we have

$$P_\mu = \frac{M\dot{X}_\mu}{\sqrt{\dot{X}^J\dot{X}^K G_{JK}}} = \frac{m\dot{X}_\mu}{\sqrt{\dot{X}^\mu\dot{X}^\nu g_{\mu\nu}}} = p_\mu \tag{53}$$

This is the same as eq. (18), but now derived for a more general metric of $\mathcal{C}$. Using eqs.(50),(53), together with (34), the equation of motion (48), after raising free indices, we obtain

$$\frac{1}{\lambda}\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{\dot{X}^\mu}{\lambda}\right) + {}^{(4)}\Gamma^\mu_{\rho\sigma}\frac{\dot{X}^\rho\dot{X}^\sigma}{\lambda^2} + \frac{p_\alpha}{m}F^\alpha_{\mu\nu}\frac{\dot{X}^\nu}{\lambda} \tag{54}$$

$$+\frac{1}{2m^2}\left(\varphi^{\alpha\beta}_{,\mu} - \varphi^{\alpha\beta}_{,\bar{J}}k^{\bar{J}}_{\alpha'}A^{\alpha'}_\mu\right)p_\alpha p_\beta + \frac{1}{\lambda m}\frac{\mathrm{d}m}{\mathrm{d}\tau} = 0, \tag{55}$$

where $\lambda = \sqrt{\dot{X}^\mu\dot{X}^\nu g_{\mu\nu}}$.

From eq. (55), in which p_α play the role of gauge charges, we see that m has the role of *inertial mass* in 4-dimensions. Four dimensional mass m is given by higher dimensional mass M and the contribution due to the extra components of momentum $P_{\bar{M}}$:

$$m^2 = g^{\mu\nu}p_\mu p_\nu = M^2 - \phi^{\bar{M}\bar{N}}p_{\bar{M}}p_{\bar{N}} = M^2 - \varphi^{\alpha\beta}p_\alpha p_\beta \tag{56}$$

These extra components $P_{\bar{M}}$ are in fact momenta of all other particles within the configuration. In general m is not constant, but in configuration spaces with suitable isometries it may be constant.

A configuration under consideration can be the universe. Then, according to this theory, the motion of a subsystem, approximated as a point particle, obeys the law of motion (55). Besides the usual 4-dimensional gravity, there are extra forces. They come from the generalized metric, i.e. the metric of configuration space. Since the inertial mass of a given particle depends on momenta of other particles and their states of motion (their momenta), the Mach principle is automatically incorporated in this theory. Such approach opens a Pandora's box of possibilities to revise our current views on the universe. Persisting problems, such as the horizon problem, dark matter, dark energy, the Pioneer effect, etc., can be examined afresh within this theoretical framework based on the concept of configuration space.

Locality, as we know it in the usual 4-dimensional relativity, no longer holds in this new theory, at least not in general. But in particular, when the metric on $\mathcal{C}$ assumes the block diagonal form (20), we recover the ordinary relativity (special and general), together with locality. However, it is reasonable to expect that metric (20) may not be a solution of the Einstein equations in $\mathcal{C}$. Then the ordinary relativity, i.e., the relativity in M_4, could be recovered as an approximation only. Even before going into the intricate work of solving the equations of general relativity in $\mathcal{C}$, we already have a crucial prediction, namely that locality in spacetime holds only approximately. When considering the universe within this theory, we have to bear in mind that the concept of spacetime has to be replaced by the concept of configuration space $\mathcal{C}$. Locality in M_4 has thus to be replaced by locality in $\mathcal{C}$. More technically this means that, instead of differential equations in M_4 (e.g., the Einstein equations), we have differential equations in $\mathcal{C}$: a given configuration (a point in $\mathcal{C}$) can only influence a nearby configuration (a nearby point in $\mathcal{C}$). Only in certain special cases this translates into the usual notion of locality in M_4 (a subspace of $\mathcal{C}$). The so called 'horizon problem' does not arise in this theory.

Strings, branes

Theories of strings and higher dimensional objects – branes – are very promising in explaining the origin and interrelationship of the fundamental interactions, including gravity [3, 4] .

But there is a cloud. A question arises as to what is the geometric principle behind string and brane theories, and how to formulate them in a background independent way [6].

$$I[g_{\mu\nu}] = \int \mathrm{d}^4x \, \sqrt{-g}\, R$$

?

Figure 2: To point particle there corresponds the Einstein-Hilbert action in spacetime. What is a corresponding space and action for a closed string?

Since such a fundamental issue has been left unsettled in the course of the development of string theory, it is not difficult to imagine that the latter theory is not yet finished. Recent serious criticism of string theory refers to an incomplete theory [7]. In the following we will consider the possibility that string/brane theories should take into account the concept of configuration space.

Configuration space for infinite dimensional objects – branes

A brane can be considered as a point in an infinite dimensional space $\mathcal{M}$ with coordinates

$$X^\mu(\xi^a) \equiv X^{\mu(\xi)} \equiv X^M \tag{57}$$

where $X^\mu(\xi^a)$, $\mu = 0, 1, 2, \ldots, N-1$; $a = 0, 1, 3, \ldots, n-1$; $n < N$, the embedding fuctions of the branes [8, 9] This includes classes of tangentially deformed branes, which we can interpret as being physically different objects, not just as being related by reparametrizations of the brane's world manifold [8] (see Fig. 3).

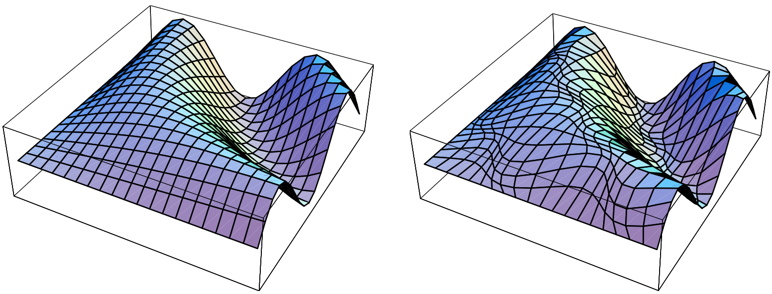


Figure 3: Examples of tangentially deformed membranes. Mathematically the surface on the left is the same as the surface on the right. Physically the two surfaces are different.

All such objects are represented by different points of $\mathcal{M}$-space. The latter space is the configuration space associated with a brane. This is the space of all (infinitely many) possible configurations of a brane.

Instead of one brane we can take a 1-parameter family of branes $X^\mu(\tau, \xi^a) \equiv X^{\mu(\xi)}(\tau) \equiv X^M(\tau)$, i.e. a curve (trajectory) in $\mathcal{M}$. In principle every trajectory is kinematically possible. A particular dynamical theory then selects which amongst those kinematically possible branes and trajectories are dynamically possible. We assume that dynamically possible trajectories are *geodesics* in $\mathcal{M}$ determined by the minimal length action [8, 9]

$$I[X^M] = \int \mathrm{d}\tau \sqrt{\rho_{MN} \dot{X}^M \dot{X}^N}. \tag{58}$$

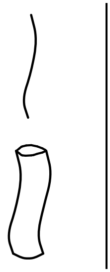
Here ρ_{MN} is the metric of $\mathcal{M}$.
In particular, if the metric is

$$\rho_{MN} \equiv \rho_{\mu(\xi')\nu(\xi'')} = \kappa \frac{\sqrt{|f(\xi')|}}{\sqrt{\dot{X}^2(\xi')}}\, \delta(\xi' - \xi'')\, \eta_{\mu\nu} \tag{59}$$

where $f_{ab} \equiv \partial_a X^\mu \partial_b X^\nu \eta_{\mu\nu}$ is the induced metric on the brane, $f \equiv \det f_{ab}$, $\dot{X}^2 \equiv \dot{X}^\mu \dot{X}^\nu g_{\mu\nu}$, ($\eta_{\mu\nu}$ being the Minkowski metric of the embedding spacetime), then the equations of motion derived from (58) are precisely those of a Dirac-Nambu-Goto brane [8, 9].

In this theory we assume that the metric (59) is just one particular choice amongst many other possible metrics of $\mathcal{M}$. But dynamically possible metrics are not arbitrary. We assume that they must be solutions of the Einstein equations in $\mathcal{M}$ [8, 9].

We take the brane space $\mathcal{M}$ as an arena for physics. The arena itself is a part of the dynamical system, it is not prescribed in advance. The theory is thus background independent. It is based on the geometric principle which has its roots in the brane space $\mathcal{M}$.



$$I[g_{\mu\nu}] = \int \mathrm{d}^4x\, \sqrt{-g}\, R$$

$$I[\rho_{\mu(\phi)\nu(\phi')}] = \int \mathcal{D}X\, \sqrt{|\rho|}\, \mathcal{R}$$

Figure 4: Brane theory is formulated in $\mathcal{M}$-space. The action is given in terms of the $\mathcal{M}$-space curvature scalar $\mathcal{R}$. We use the abbreviation $\phi \equiv \phi^A = (\tau, \xi^a)$.

In summary, the infinite dimensional brane space $\mathcal{M}$ has in principle any metric that is a solution to the Einstein's equations in $\mathcal{M}$. For the particular diagonal metric (59) we obtain the ordinary branes, including strings. But it remains to be checked whether such a particular metric is also a solution of this generalized dynamical system. If not, then this would mean that the ordinary string and brane theory is not exactly embedded into the theory based on dynamical $\mathcal{M}$-space. The proposed theory goes beyond that of the usual strings and branes. It resolves the problem of background independence and the geometric principle behind the string theory (Fig.

4). The geometric principle of string theory is based on the concept of brane space $\mathcal{M}$, i.e. the configuration space for branes. The occurrence of gauge and gravitational fields in string theories is also elucidated. Such fields are due to string configurations. They occur in the expansion of a string state functional in terms of the Fock space basis. This can now be understood as well within the classical string theory based on the action (58) with $\mathcal{M}$-space metric ρ_{MN}, which is dynamical and which satisfies the Einstein equations in $\mathcal{M}$. Multidimensionality of ρ_{MN} allows for extra gauge interactions, besides gravity. In the following we will point out how in the infinite dimensional space $\mathcal{M}$ one can factor out a finite dimensional subspace.

Finite dimensional description of extended objects

The Earth has a huge (practically infinite) number of degrees of freedom. And yet, when describing the motion of the Earth around the Sun, we neglect them all, except for the coordinates of *the centre of mass.*

Instead of infinitely many degrees of freedom associated with an extended object, we may consider *a finite number of degrees of freedom*.

Strings and branes have infinitely many degrees of freedom. But at first approximation we can consider just *the centre of mass* (Fig. 5a).

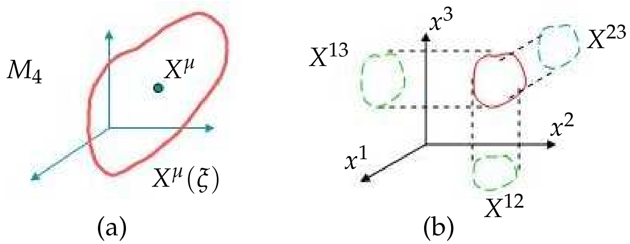


Figure 5: With a closed string one can associate the centre of mass coordinates (a), and the area coordinates (b)).

Next approximation is in considering the holographic coordinates $X^{\mu\nu}$ of the *oriented area* enclosed by the string (Fig. 5b).

We may go further and search for eventual thickness of the object. If the string has finite thickness, i.e., if actually it is not a string, but a 2-brane, then there exist the corresponding volume degrees of freedom $X^{\mu\nu\rho}$ (Fig. 6).

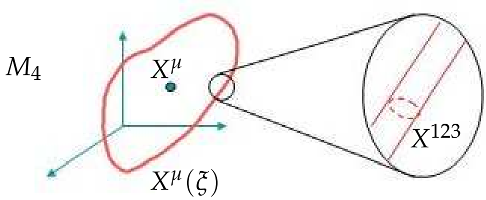


Figure 6: Looking with a sufficient resolution one can detect eventual presence of volume degrees of freedom.

In general, for an extended object in M_4, we have 16 coordinates

$$X^M \equiv X^{\mu_1 \cdots \mu_r}, \qquad r = 0, 1, 2, 3, 4 \tag{60}$$

They are projections of r-dimensional volumes (areas) onto the coordinate planes.

Oriented r-volumes can be elegantly described by Clifford algebra [10]. Instead of the usual relativity, formulated in spacetime in which the interval is

$$\mathrm{d}s^2 = \eta_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu \tag{61}$$

one can consider the theory in which the interval is extended to the space of r-volumes, called pandimensional continuum [11] or Clifford space [12, 8, 13],

$$\mathrm{d}S^2 = G_{MN} \mathrm{d}x^M \mathrm{d}x^N \tag{62}$$

Coordinates of Clifford space can be used to model extended objects [12, 14]. They are a generalization of the concept of center of mass. Instead of describing an extended object in "full detail, we can describe it in terms of the center of mass, area and volume coordinates. In particular, the extended object can be a fundamental string/brane.

Dynamics Let the action for an extended object described in terms of the coordinates of Clifford space be

$$I = \int \mathrm{d}\tau \sqrt{G_{MN} \dot{X}^M \dot{X}^N} \tag{63}$$

If $G_{MN} = \eta_{MN}$ is the Minkowski metric, then the equations of motion are

$$\ddot{X}^M \equiv \frac{\mathrm{d}^2 X^M}{\mathrm{d}\tau^2} = 0 \tag{64}$$

They hold for tensionless branes. For the branes with tension one has to replace η_{MN} with a generic metric G_{MN} with non vanishing curvature. Eq. (64) then generalizes to the corresponding geodesic equation

$$\frac{1}{\sqrt{\dot{X}^2}}\left(\frac{\dot{X}^M}{\sqrt{\dot{X}^2}}\right) + \Gamma^M_{JK}\frac{\dot{X}^J\dot{X}^K}{\dot{X}^2} = 0 \tag{65}$$

Such higher dimensional configuration space, associated with branes, enables unification of fundamental interactions à la Kaluza-Klein [15]. For alternative, although related approaches see [16].

In quantizing such classical theory, based on the action (63) which describes 'point particle' in $\mathcal{C}$, one expects to be able to proceed as usual, and arrive at the quantum field theory in $\mathcal{C}$. So there would be no necessity to increase the dimensionality of (4-dimensional) spacetime, from which we start in building the 16-dimensional Clifford space $\mathcal{C}$.

Consider now the possibility of choosing to describe the object of Figs. 5,6, not with a set of coordinates $X^M = \{X,\ X^{\mu_1},\ X^{\mu_1\mu_2},\ X^{\mu_1\mu_2,\mu_3},\ X^{\mu_1\mu_2,\mu_3,\mu_4}\}$, but rather employ a more detailed description. Let ξ be a parameter along the "centroid" loop, i.e., a closed string of Fig. 5a, and let $X^\mu(\xi)$ be its embedding functions in M_4. But if we look more closely (Fig. 6), we find that at every value of parameter ξ the string has a thickness: At every value of ξ, instead of a point, there may be a loop, described by $X^{\mu_1\mu_2}$. If we look even closer, we may find even more structure, encoded in coordinates $X^{\mu_1\mu_2,\mu_3}$ and $X^{\mu_1\mu_2,\mu_3,\mu_4}$. Altogether, taking into account as well a time like parameter τ, our object is described by 16 functions $X^M(\tau,\xi)$. These are functions describing the embedding of a string's worldsheet into a 16-dimensional Clifford space $\mathcal{C}$. Instead of the string worldsheet $X^\mu(\tau,\xi)$, $\mu = 0,1,2,3$, embedded in M_4, we have a worldsheet in a higher dimensional space which is not spacetime, but Clifford space (i.e. a configuration sapce) $\mathcal{C}$. It was shown that such a string living in $\mathcal{C}$ – which happens to have signature $(8,8)$ – can be be consistently quantized [17] by employing the Jackiw definition of vacuum state [18].

Hence, it is possible to have a consistent string theory without employing extra dimensions of spacetime, provided that one does not consider infinitely thin strings, but allows for their thickness, encoded in the coordinates of Clifford space.

Summary

We have considered a theory in which spacetime is replaced by a larger space, namely the configuration space $\mathcal{C}$ associated with the system under consideration. The ordinary special and general relativity are recovered for particular classes of metrics on $\mathcal{C}$. Since configuration space has extra dimensions, its metric provides description of additional interactions, beside the 4-dimensional gravity, just as in Kaluza-Klein theories. They can occur in *macro physics* and in *micro physics*. Eventual modification of the dynamics at the levels from galaxies to the Universe is thus based on the same underlying principle as the dynamics of elementary particles (leaving quantum features aside). In this theory there is no need for extra dimensions of *spacetime*. The latter space is a subspace of the *configuration space* $\mathcal{C}$, and all dimensions of $\mathcal{C}$ are physical. Therefore, there is no need for compactification of the extra dimensions of $\mathcal{C}$.

We have presented some theoretical justifications for the insight that the basic space of physics could be associated with configurations of physical systems. The search for possible realistic solutions to the theory and their comparison with observations is a project worth pursuing in the future.

References

[1] See e.g. John Ellis, *Royal Society of London Transactions Series A*, **1812** (2003) 2607; P.J.E. Peebles and Bharat Ratra, Rev. Mod. Phys. **75** (2003) 559.

[2] M. Milgrom, *Does Dark Matter Really Exist?*, Scientific American, (Aug. 2002) 42-50, 52; J.D. Bekenstein, *Modified Gravity vs Dark Matter: Relativistc theory for MOND*, JHEP Conference Proceedings (2005), [arχiv:astro-ph/0412652].

[3] M.B. Green, J.H. Schwarz and E. Witten, *Superstring theory* (Cambridge University Press, Cambridge, 1987); M. Kaku, *Introduction to Superstrings* (Springer-Verlag, New York, 1988); M.J. Duff, Scientific American (February 1998) 69; U. Danielsson, Rep. Progr. Phys. **64** (2001) 51.

[4] M.J. Duff, *Benchmarks on the brane*, [arχiv:hep-th/0407175]; M.J. Duff, Nucl. Phys. B **335** (1990) 610; M.J. Duff and J.X. Lu, Nucl. Phys. B **347** (1990) 394.

[5] S.K. Wong, Nuovo Cim. A **65** (1983) 79; K.S. Viswanathan and B. Wong, Phys. Rev. D **32** (1985) 3108.

[6] E. Witten, Phys. Rev. D **46** (1992) 5467; *Quantum Background Independence in String Theory*, [arχiv:hep-th/9306122]; *Reflections on the Fate of Spacetime*, Physics Today (April 1996).

[7] Lee Smolin, *The Trouble With Physics: The Rise of String Theory, The Fall of a Science and What Comes Next* (Houghton Mifflin, Boston 2006).

[8] M. Pavšič, *The Landscape of Theoretical Physics: A Global View; From Point Particle to the Brane World and Beyond, in Search of Unifying Principle* (Kluwer Academic, Dordrecht 2001).

[9] M. Pavšič, *General Principles of Brane Kinematics and Dynamics*, Proceedings of the EURESCO Conference 'What comes beyond the Standard Model', 12-17 July 2003, Portorož, Slovenia, [arχiv:hep-th/0311060].

[10] D. Hestenes, *Space-Time Algebra* (Gordon and Breach, New York, 1966); D. Hestenes and G. Sobcyk, *Clifford Algebra to Geometric Calculus* (D. Reidel, Dordrecht, 1984).

[11] W. Pezzaglia, *Physical Applications of a Generalized Geometric Calculus*, in *Dirac Operators in Analysis*, Pitman Research Notes in Mathematics, Number **394**, eds. J. Ryan and D. Struppa, (Longmann 1997) 191–202, [arχiv:gr-qc/9710027].
[12] C. Castro, Chaos Solitons Fractals **11** (2000) 1721, [arχiv:hep-th/9912113].
[13] M. Pavšič, Found. Phys. **31** (2001) 1185, [arχiv:hep-th/0011216].
[14] M. Pavšič, Found. Phys. **33** (2003) 1277, [arχiv:gr-qc/0211085].
[15] M. Pavšič, Phys. Lett. B **614** (2005) 85, [arχiv:hep-th/0412255]; Int. J. Mod. Phys. A **21** (2006) 5905, [arχiv:gr-qc/0507053]; Found. Phys. **37** (2007) 1197, [arχiv:hep-th/0605126].
[16] C. Castro, J. Math. Phys. **47** (2006) 112301; Found. Phys. **35** (2005) 971.
[17] M. Pavšič, Found. Phys. **35** (2005) 1617, [arχiv:hep-th/0501222].
[18] D. Cangemi, R. Jackiw and B. Zwiebach, Annals of Physics **245** (1996) 408; E. Benedict, R. Jackiw and H.-J. Lee, Phys. Rev. D **54** (1996) 6213.
[19] M. Pavšič, Phys. Lett. A **254** (1999) 119, [arχiv:hep-th/9812123].

Time and Matter
2007 Bled, Slovenia

Fundamental Decoherence from Quantum Gravity

RODOLFO GAMBINI[1], RAFAEL PORTO[2] AND JORGE PULLIN[3*]
[1] *Instituto de Física, Facultad de Ciencias, Universidad de la República, Iguá 4225, CP 11400 Montevideo, Uruguay*
[2] *Department of Physics, University of California, Santa Barbara, CA 93106, USA*
[3] *Department of Physics and Astronomy, Louisiana State University, Baton Rouge, LA 70803-4001, USA*

Abstract: We discuss the fundamental loss of unitarity that appears in quantum mechanics when one uses a physical apparatus to measure time. This induces a decoherence effect that is independent of any interaction with the environment and appears in addition to any usual environmental decoherence. We discuss the theoretical and potential experimental implications of this process of decoherence.

Introduction

In the usual formulation, quantum mechanics involves an idealization. The idealization is the use of a perfect classical clock to measure times. Such a device clearly does not exist in nature, since all measuring devices are subject to some level of quantum fluctuations. Therefore the equations of quantum mechanics, when cast in terms of the variable that is really measured by a clock in the laboratory, will differ from the traditional Schrödinger description. Although this is an idea that arises naturally in ordinary quantum mechanics, it is of paramount importance when one is discussing quantum gravity. This is due to the fact that general relativity is a generally covariant theory where one needs to describe the evolution in a relational way. One ends up describing how certain objects change when other objects, taken as clocks, change. At the quantum level this relational description will compare the outcomes of measurements of quantum objects. Quantum gravity is expected to be of importance in regimes (e.g.

* pullin@lsu.edu

near the big bang or a black hole singularity) in which the assumption of the presence of a classical clock is clearly unrealistic. The question therefore arises: is the difference between the idealized version of quantum mechanics and the real one just of interest in situations when quantum gravity is predominant, or does it have implications in other settings? We will argue that indeed it does have wider implications. Some of them are relevant to conceptual questions (e.g. the problem of measurement in quantum mechanics or the black hole information paradox) and there might even be experimental implications.

A detailed discussion of these ideas can be found in previous papers [1, 2, 3], and in particular in the pedagogical review [4]. Here we present an abbreviated discussion.

The plan of this paper is as follows: in the next section we will discuss the form of the evolution equation of quantum mechanics when the time variable, used to describe it, is measured by a real clock. In section III we will consider a fundamental bound on how accurate can a real clock be and the implications it has for quantum mechanics in terms of real clocks and its consequences. Section IV discusses the implications of the formalism.

Quantum mechanics with real clocks

Given a physical situation we start by choosing a "clock". By this we mean a physical quantity (more precisely a set of quantities, like when one chooses a clock and a calendar to monitor periods of more than a day) that we will use to keep track of the passage of *time*. An example of such a variable could be the angular position of the hand of an analog watch. Let us denote it by T. We then identify some physical variables that we wish to study as a function of time. We shall call them generically O ("observables"). We then proceed to quantize the system by promoting all the observables and the clock variable to self-adjoint quantum operators acting on a Hilbert space. The latter is defined once a well defined inner product is chosen in the set of all physically allowed states. Usually it consists of squared integrable functions $\psi(q)$ with q the configuration variables.

Notice that we are not in any way modifying quantum mechanics. We assume that the system has an evolution in terms of an external parameter t, which is a classical variable, given by a Hamiltonian and with operators evolving with Heisenberg's equations (it is easier to present things in the Heisenberg picture, though it is not mandatory to use it for our construction). Then the standard rules of quantum mechanics and its probabilistic nature apply.

We will call the eigenvalues of the "clock" operator T and the eigenvalues of the "observables" O. We will assume that the clock and the measured system do not interact (if one considered an interaction it would produce additional effects to the one discussed). So the density matrix of the total system is a direct product of that of the system under study and the clock $\rho = \rho_{\rm sys} \otimes \rho_{\rm cl}$, and the system evolves through a unitary evolution operator that is of the tensor product form $U = U_{\rm sys} \otimes U_{\rm cl}$. The quantum states are described by a density matrices at a time t. Since the latter is unobservable, we would like to shift to a description where we have density matrices as functions of the observable time T. We define the probability that the resulting measurement of the clock variable T correspond to the value t,

$$\mathcal{P}_t(T) \equiv \frac{\mathrm{Tr}\, P_T(0) U_{\rm cl}(t) \rho_{\rm cl} U^\dagger_{\rm cl}(t)}{\int_{-\infty}^{\infty} \mathrm{d}t\ \mathrm{Tr}\, P_T(t)\rho_{\rm cl}}, \tag{1}$$

where $P_T(0)$ is the projector on the eigenspace with eigenvalue T evaluated at $t = 0$. We note that $\int_{-\infty}^{\infty} \mathrm{d}t\, \mathcal{P}_t(T) = 1$. We now define the evolution of the density matrix,

$$\rho(T) \equiv \int_{-\infty}^{\infty} \mathrm{d}t\ U_{\rm sys}(t) \rho_{\rm sys} U^\dagger_{\rm sys}(t) \mathcal{P}_t(T) \tag{2}$$

where we dropped the "sys" subscript in the left hand side since it is obvious we are ultimately interested in the density matrix of the system under study, not that of the clock.

We have therefore ended with an "effective" density matrix in the Schrödinger picture given by $\rho(T)$. It is possible to reconstruct entirely in a relational picture the probabilities using this effective density matrix, for details we refer the reader to the lengthier discussion in [4]. By its very definition, it is immediate to see that in the resulting evolution unitarity is lost, since one ends up with a density matrix that is a superposition of density matrices associated with different t's and that each evolve unitarily according to ordinary quantum mechanics.

Now that we have identified what will play the role of a density matrix in terms of a "real clock" evolution, we would like to see what happens if we assume the "real clock" is behaving semi-classically. To do this we assume that $\mathcal{P}_t(T) = f(T - T_{\rm max}(t))$, where f is a function that decays very rapidly for values of T far from the maximum of the probability distribution $T_{\rm max}$. We refer the reader to [4] for a derivation, but the resulting evolution equation for the probabilities is,

$$\frac{\partial \rho(T)}{\partial T} = \mathrm{i}[\rho(T), H] + \sigma(T)[H, [H, \rho(T)]]. \tag{3}$$

and the extra term is dominated by the rate of change $\sigma(T)$ of the width of the distribution $f(t - T_{\max})(t)$.

An equation of a form more general than this has been considered in the context of decoherence due to environmental effects, it is called the Lindblad equation. Our particular form of the equation is such that conserved quantities are automatically preserved by the modified evolution. Other mechanisms of decoherence coming from a different set of effects of quantum gravity have been criticized in the past because they fail to conserve energy [5]. It should be noted that Milburn arrived at a similar equation as ours from different assumptions [6]. Egusquiza, Garay and Raya derived a similar expression from considering imperfections in the clock due to thermal fluctuations [7]. It is to be noted that such effects will occur in addition to the ones we discuss here. Corrections to the Schrödinger equation from quantum gravity have also been considered in the context of WKB analyses [8].

In a real experiment, there will be decoherence in the system under study due to interactions with the environment, that will be superposed on the effect we discuss. Such interactions might be reduced by cleverly setting up the experiment. The decoherence we are discussing here however, is completely determined by the quality of the clock used. It is clear that if one does experiments in quantum mechanics with poor clocks, pure states will evolve into mixed states very rapidly. The effect we are discussing can therefore be magnified arbitrarily simply by choosing a lousy clock. This effect has actually been observed experimentally in the Rabi oscillations describing the exchange of excitations between atoms and field [9].

Fundamental limits to realistic clocks

We have established that when we study quantum mechanics with a physical clock (a clock that includes quantum fluctuations), unitarity is lost, conserved quantities are still preserved, and pure states evolve into mixed states. The effects are more pronounced the worse the clock is. Which raises the question: is there a fundamental limitation to how good a clock can be? This question was first addressed by Salecker and Wigner [10]. Their reasoning went as follows: suppose we want to build the best clock we can. We start by insulating it from any interactions with the environment. An elementary clock can be built by considering a photon bouncing between two mirrors. The clock "ticks" every time the photon strikes one of the mirrors. Such a clock, even completely isolated from any environmental effects, develops errors. The reason for them is that by the time the photon travels between the mirrors, the wavefunctions of the mirrors spread. Therefore the time of arrival of the photon develops an uncertainty. Salecker and Wigner calculated the uncertainty to be $\delta t \sim \sqrt{t/M}$ where M is the mass of the mirrors and t is the time to be measured (we are using units where $\hbar = c = 1$ and therefore mass is measured in 1/second). The longer the time measured the larger the error. The larger the mass of the clock, the smaller the error.

So this tells us that one can build an arbitrarily accurate clock just by increasing its mass. However, Ng and Van Dam [11] pointed out that there is a limit to this. Basically, if one piles up enough mass in a concentrated region of space one ends up with a black hole. Some readers may ponder why do we need to consider a concentrated region of space. The reason is that if we allow the clock to be more massive by making it bigger, it also deteriorates its performance (see the discussion in [12] in response to [13]).

A black hole can be thought of as a clock since it is an oscillator. In fact it is the "fastest" oscillator one can have, and therefore the best clock for a given size. It has normal modes of vibration that have frequencies that are of the order of the light travel time across the Schwarzschild radius of the black hole. (It is amusing to note that for a solar sized black hole the frequency is in the kilohertz range, roughly similar to that of an ordinary bell). The more mass in the black hole, the lower the frequency, and therefore the worse its performance as a clock. This therefore creates a tension with the argument of Salecker and Wigner, which required more mass to increase the accuracy. This indicates that there actually is a "sweet spot" in terms of the mass that minimizes the error. Given a time to be measured, light traveling at that speed determines a distance, and therefore a maximum mass one could fit into a volume determined by that distance before one forms a black hole. That is the optimal mass. Taking this into account one finds that the best accuracy one can get in a clock is given by $\delta T \sim T_{\text{Planck}}^{2/3} T^{1/3}$ where $t_{\text{Planck}} = 10^{-44}$ s is Planck's time and T is the time interval to be measured. This is an interesting result. On the one hand it is small enough for ordinary times that it will not interfere with most known physics. On the other hand is barely big enough that one might contemplate experimentally testing it, perhaps in future years.

With this absolute limit on the accuracy of a clock we can quickly work out an expression for the $\sigma(T)$ that we discussed in the previous section [14, 3]. It turns out to be $\sigma(T) = T_{\text{Planck}} \sqrt[3]{T_{\text{Planck}}/(T_{\max} - T)}$. With this estimate of the absolute best accuracy of a clock, we can work out again the evolution of the density matrix for a physical system in the energy eigenbasis. One gets

$$\rho(T)_{nm} = \rho_{nm}(0)\, \mathrm{e}^{-\mathrm{i}\omega_{nm}T} \exp(-\omega_{nm}^2 T_{\mathrm{Planck}}^{4/3} T^{2/3}). \tag{4}$$

So we conclude that *any* physical system that we study in the lab will suffer loss of quantum coherence at least at the rate given by the formula above. This is a fundamental inescapable limit. A pure state inevitably will become a mixed state due to the impossibility of having a perfect classical clock in nature.

Possible experimental implications

Given the conclusions of the previous section, one can ask what are the prospects for detecting the fundamental decoherence we propose. At first one would expect them to be dim. It is, like all quantum gravitational effects, an "order Planck" effect. But it should be noted that the factor accompanying the Planck time can be rather large. For instance, if one would like to observe the effect in the lab one would require that the decoherence manifest itself in times of the order of magnitude of hours, perhaps days at best. That requires energy differences of the order of 10^{10} eV in the Bohr frequencies of the system. Such energy differences can only be achieved in "Schrödinger cat" type experiments, but are not outrageously beyond our present capabilities. Among the best candidates today are Bose–Einstein condensates, which can have 10^6 atoms in coherent states. However, it is clear that the technology is still not there to actually detect these effects, although it could be possible in forthcoming years.

A point that could be raised is that atomic clocks currently have an accuracy that is less than a decade of orders of magnitude worse than the absolute limit we derived in the previous section. Couldn't improvements in atomic clock technology actually get better than our supposed absolute limit? This seems unlikely. When one studies in detail the most recent proposals to improve atomic clocks, they require the use of entangled states [15] that have to remain coherent. Our effect would actually prevent the improvement of atomic clocks beyond the absolute limit!

Another point to be emphasized is that our approach has been quite naive in the sense that we have kept the discussion entirely in terms of non-relativistic quantum mechanics with a unique time across space. It is clear that in addition to the decoherence effect we discuss here, there will also be decoherence spatially due to the fact that one cannot have clocks perfectly synchronized across space and also that there will be fundamental uncertainties in the determination of spatial positions. This is discussed in some detail in our paper [16].

Finally, if one has doubts about the effect's existence, one must recall that one can make it arbitrarily large just by picking a lousy clock. This is of course, not terribly interesting and is not what is normally done in physics labs. But it should be noted that experiments of Rabi oscillations in rubidium atoms measure certain correlations which can be interpreted as having the atom work as a lousy clock. The oscillations show experimentally the exponential decay we discuss. See Bonifacio et al. [17] for a discussion.

Conceptual implications

The fact that pure states evolve naturally into mixed states has conceptual implications in at least three interesting areas of physics. We will discuss them separately.

The black hole information paradox

The black hole information paradox appeared when Hawking [18] noted that when quantum effects are taken into account, black holes emit radiation like a black body with a temperature $T_{\rm BH} = \hbar/(8\pi kGM)$ where M is the black hole mass, k is Boltzmann's constant and G is Newton's constant. As the black hole radiates, it loses mass, and therefore its temperature increases. This process continues until the black hole eventually evaporates completely and the only thing left is outgoing purely thermal radiation. Now, suppose one had started with a pure quantum state of enough mass that it collapses into a black hole. After the evaporation process, one is left with a mixed state (the outgoing purely thermal radiation). In ordinary quantum mechanics this presents a problem, since pure states cannot evolve into mixed states. (For further discussion and references on the paradox see [19]).

On the other hand, we have argued that due to the lack of perfectly classical clocks, quantum mechanics really implies that pure states do evolve into mixed states. The question is: could the effect be fast enough to render the black hole information paradox effectively unobservable? On one hand we have argued that our effect is small. But it is also true that black holes usually take a very long time to evaporate. Of course, a full calculation of the evaporation of a black hole would require a detailed modeling including quantum effects of gravity that no one is in a position of carrying out yet. We have done a very naive estimate [14, 3] of how our effect would take place in the case of an evaporating black hole. To this aim we have assumed the black hole is a system with energy levels (this is a common assumption in many quantum gravity scenarios), and that most of the Hawking radiation is coming from a transition between two dominant energy levels separated by a characteristic frequency dependent on the temperature. A detailed calculation based on this naive model [3] for the evolution of the density matrix shows that,

$$|\rho_{12}(T_{\max})| \sim |\rho_{12}(0)| \left(\frac{M_P}{M_{\mathrm{BH}}}\right)^{2/3}. \tag{5}$$

For astrophysical sized black holes, where M_{BH} is of the order of the mass of the Sun, this indicates that the off diagonal elements are suppressed by the time of evaporation by 10^{-28}, rendering the information puzzle effectively unobservable. What happens for smaller black holes? The effect is smaller. So can one claim that there still is an observable information puzzle for smaller black holes? This is debatable. After all, we do expect decoherence from other environmental effects to be considerably larger than the one we are considering here. If one makes the holes too small, then none of these calculations apply, and in fact the traditional Hawking evaporation is not an adequate description, since one has to take into account full quantum gravity effects. So we can say that at this naive level of ball-park figure calculation the paradox can be rendered unobservable for large black holes and we cannot say for sure for smaller ones using this simplified analysis. It should be noted that the paradox is not "solved", since it still exists at the level of the Schrödinger time t. A better calculation than the one we did could probably be attempted, since both in string theory and loop quantum gravity there is some understanding of the energy levels of a black hole, even though the evaporation process is not well understood. Using such levels one could get a better estimate of how much coherence is lost.

The measurement problem in quantum mechanics

A potential conceptual application of the fundamental decoherence that we discussed that has not been exploited up to now is in connection with the measurement problem in quantum mechanics. The latter is related to the fact that in ordinary quantum mechanics the measurement apparatus is assumed to be always in an eigenstate after a measurement has been performed. The usual explanation [20] for this is that there exists interaction with the environment. This selects a preferred basis, i.e., a particular set of quasi-classical states that commute, at least approximately, with the Hamiltonian governing the system-environment interaction. Since the form of the interaction Hamiltonians usually depends on familiar "classical" quantities, the preferred states will typically also correspond to the small set of "classical" properties. Decoherence then quickly damps superpositions between the localized preferred states when only the system is considered. This is taken as an explanation of the appearance to a local observer of a "classical" world of determinate, "objective" (robust) properties.

The main problem with such a point of view is how is one to interpret the local suppression of interference in spite of the fact that the total state describing the system-environment combination retains full coherence. One may raise the question whether retention of the full coherence could ever lead to empirical conflicts with the ascription of definite values to macroscopic systems. The usual point of view is that it would be very difficult to reconstruct the off diagonal elements of the density matrix in practical circumstances. However, at least as a matter of principle, one could indeed reconstruct such terms (the evolution of the whole system remains unitary [21]) by "waiting long enough".

Our mechanism of fundamental decoherence could contribute to the understanding of this issue, since it implies that coherence is irreversibly lost and therefore one cannot reconstruct the off diagonal elements. Some people claim that we have just changed the environment by the clock as responsible for the loss of coherence and therefore the original criticism applies. But in the case of the clock, the minimum "size" of it in terms of its degrees of freedom if one wishes to view it as "a particular form of an environment" is determined by the length of the experiment and guarantees that in that length one will not be able to reconstruct the off diagonal elements. There is not the luxury of "waiting long enough" in this setting. For further discussion see our paper [22].

Quantum computing

In quantum computing, when one performs operations one is evolving quantum states. If one wishes the computers to perform faster, one needs to expend extra energy to evolve the quantum states. Based on this premise, Lloyd [23] presented a fundamental limitation to how fast quantum computers can be. Using the Margolus–Levitin [24] theorem he notes that in order to perform a computation in a time δT one needs to expend at least an energy $E \geq \pi\hbar/(2\Delta T)$. As a consequence, a system with an average energy E can perform a maximum of $n = 2E/(\pi\hbar)$ operations per second. For an "ultimate laptop" (a computer of a volume of one liter and one kilogram of weight) the limit turns out to be 10^{51} operations per second.

Such results assume the evolution is unitary. When it is not, as we have argued in this paper, erroneous computations are carried out. Since the rate of decoherence we discussed increases with increased energy differences, the rate of erroneous computations increases the faster one wishes to make the computer.

Can't one error correct? After all, one expects quantum computers to have errors due to decoherence from environmental factors. One can indeed error-correct. But there are limitations to how fast this can be done. At its most basic level error correction is achieved by duplicating calculations and comparing results. This requires spatial communication, which is limited by the speed of light. Our point is that one cannot simply error correct one's way out of the fundamental decoherence effects.

We have to distinguish a bit between serial and parallel computing. In serial computing one achieves speed by increasing the energy in each qubit. This enhances our decoherence effect and significantly affects the performance. In a parallel machine one increases the speed by operating simultaneously on many qubits with lower energies per qubit therefore lowering the importance of the effect we introduced. For a machine with L qubits and a number of simultaneous operations d_P one gets,

$$n \leq \left(\frac{1}{t_{\mathrm{P}}}\right)^{4/7} \left(\frac{cL}{R}\right)^{3/7} d_p^{4/7} \sim 10^{47} \mathrm{op/s}, \tag{6}$$

where the last estimate was obtained by taking the values of parameters for the "ultimate laptop" (for more details see [25]).

This is actually four orders of magnitude stronger than the bound that Lloyd found. If one had chosen a serial machine, the bound would have been tighter, 10^{42} operations per second.

We therefore see that although the effect we introduced is far from being achievable in quantum computers built in the next few years, it can limit the ultimate computing power of quantum computer. This is quite remarkable, given that it is a limit obtained involving gravity. Few people could have foreseen that gravity would play any role in quantum computation.

Discussion

We have argued that the use of realistic clocks in quantum mechanics implies that pure states evolve into mixed states. Another way of putting this is that we are allowing quantum fluctuations in our clock. Similar ideas have been considered by Bonifacio, with a different formulation [26]. In quantum gravity and quantum cosmology it is natural to consider the clock to be part of the system under study. This is what motivated our interest in these issues, but it is clear that the core of the phenomenon can be described without references to quantum gravity, and that is what we have attempted to do in this presentation.

Even in the absence of a direct possibility of detecting these effects, they can have important conceptual implications, as we have illustrated with the black hole information puzzle, quantum computing and the problem of measurement in quantum mechanics.

Acknowledgements

This work was supported in part by grants NSF-PHY0244335, DOE-ER-40682-143 and DEAC02-6CH03000, and by funds of the Horace C. Hearne Jr. Institute for Theoretical Physics, PEDECIBA (Uruguay) and CCT-LSU.

References

[1] R. Gambini, R. Porto and J. Pullin, Class. Quant. Grav. **21** (2004) L51, [arχiv:gr-qc/0305098].
[2] R. Gambini, R. Porto and J. Pullin, New J. Phys. **6** (2004) 45, [arχiv:gr-qc/0402118].
[3] R. Gambini, R. Porto and J. Pullin, Braz. J. Phys. **35** (2005) 266, [arχiv:gr-qc/0501027].
[4] R. Gambini, R. Porto and J. Pullin, Gen. Rel. Grav. **39** (2007) 1143, [arχiv:gr-qc/0603090].
[5] See for instance J. Ellis, J. Hagelin, D.V. Nanopoulos, and M. Srednicki, Nucl. Phys. B **241** (1984) 381; T. Banks, M.E. Peskin, and L. Susskind, Nucl. Phys. B **244** (1984) 125.
[6] G.J. Milburn, Phys. Rev. A **44** (1991) 5401.
[7] I. Egusquiza, L. Garay and J. Raya, Phys. Rev. A **59** (1999) 3236, [arχiv:quant-ph/9811009].
[8] C. Kiefer and T. Singh, Phys. Rev. D **44** (1991) 1061.
[9] D.M. Meekhof, C. Monroe, B.E. King, W.M. Itano and D.J. Wineland, Phys. Rev. Lett **76** (1996) 1796; M. Brune, F. Schmidt-Kaler, A. Maali, J. Dreyer, E.

Hagley, J.M. Raimond and S. Haroche, Phys. Rev. Lett. **76** (1996) 1800; R. Bonifacio, S. Olivares, P. Tombesi and D. Vitali, Phys. Rev. A **61** (2000) 053802.
[10] E. Wigner, Rev. Mod. Phys. 29 (1957) 255.
[11] Y.J. Ng and H. van Dam, Annals N.Y. Acad. Sci. **755** (1995) 579, [arχiv:hep-th/9406110]; Mod. Phys. Lett. A **9** (1994) 335.
[12] Y.J. Ng and H. van Dam, Class. Quant. Grav. **20** (2003) 393, [arχiv:gr-qc/0209021].
[13] J.C. Baez and S.J. Olson, Class. Quant. Grav. **19** (2002) L121, [arχiv:gr-qc/0201030].
[14] R. Gambini, R.A. Porto and J. Pullin, Phys. Rev. Lett. **93** (2004) 240401, [arχiv:hep-th/0406260].
[15] See for instance A. Andre, A. Sorensen and M. Lukin, Phys. Rev. Lett **92** (2004) 230801, [arχiv:quant-ph/0401130].
[16] R. Gambini, R.A. Porto and J. Pullin, Int. J. Mod. Phys. D **15** (2006) 2181, [arχiv:gr-qc/0611148].
[17] R. Bonifacio *et al.*, Phys. Rev. A **61** (2000) 053802.
[18] S. Hawking, Commun. Math. Phys. **43** (1975) 199.
[19] See for instance S. Giddings and L. Thorlacius, in *Particle and nuclear astrophysics and cosmology in the next millennium*, ed. E. Kolb (World Scientific, Singapore, 1996), [arχiv:astro-ph/9412046]; for more recent references see S.B. Giddings and M. Lippert, [arχiv:hep-th/0402073] and D. Gottesman and J. Preskill, JHEP **0403** (2004) 026, [arχiv:hep-th/0311269].
[20] M. Schlosshauer, Rev. Mod. Phys. **76** (2004) 1267, [arχiv:quant-ph/0312059].
[21] R. Omnès, *The interpretation of quantum mechanics*, Princeton Series in Physics (Princeton, NJ, 1994).
[22] R. Gambini and J. Pullin, Found. Phys. **37** (2007) 1074.
[23] S. Lloyd, Nature **406** (2000) 1047.
[24] N. Margolus and L. Levitin, Physica D **120** (1998) 188.
[25] R. Gambini, R. Porto and J. Pullin, in *Gravity, astrophysics and strings at the Black Sea*, eds. P. Fiziev and M. Todorov (St. Kliment Ohridski Press, 2006), [arχiv:quant-ph/0507262].
[26] R. Bonifacio, Nuovo Cim. D **114** (1999) 473.

TIME AND MATTER 2007

Time and Matter
2007 Bled, Slovenia

Emergent Rainbow Spacetimes: Two Pedagogical Examples

MATT VISSER*
School of Mathematics, Statistics and Computer Science
Victoria University of Wellington
PO Box 600, Wellington, New Zealand

Abstract: There is a possibility that spacetime itself is ultimately an emergent phenomenon, a near-universal ”low-energy long-distance approximation”, similar to the way in which fluid mechanics is the near-universal low-energy long-distance approximation to quantum molecular dynamics. If so, then direct attempts to quantize spacetime are misguided — at least as far as fundamental physics is concerned.

In particular, this implies that we may have totally mis-identified the fundamental degrees of freedom that need to be quantized, and even the fundamental nature of the spacetime arena in which the physics takes place. Based on this and other considerations, there has recently been a surge of interest in the notion of energy-dependent and momentum-dependent “rainbow” geometries. Motivations for such a concept vary widely, from attempts at applying the renormalization group to cosmology in the large, through to attempts at interpreting the DSR models in terms of energy-dependent transformations on phase space. All of these models suffer from the fact that there is considerable disagreement and confusion as to what exactly an energy-dependent “rainbow” geometry might actually entail.

In the present article I will not discuss these exotic ideas in any detail, instead I will present two specific and concrete examples of situations where an energy-dependent “rainbow” geometry makes perfectly good mathematical and physical sense. These simple examples will then serve as templates suggesting ways of proceeding in situations where the underlying physics may be more complex. The specific models I will deal with are (1) acoustic spacetimes in the presence of nontrivial dispersion, and (2) a mathematical reinterpretation of Newton’s second law for a non-relativistic conservative force, which is well-known to be equivalent to the differential geometry of an energy-dependent conformally flat three-manifold.

These two models make it clear that there is nothing wrong with the concept of an energy-dependent “rainbow” geometry *per se*. Whatever

* matt.visser@mcs.vuw.ac.nz

problems may arise in the implementation of any specific quantum-gravity-inspired proposal for an energy-dependent spacetime are related to deeper questions regarding the compatibility of that specific proposal with experimental reality.

Introduction

If the physical spacetime described by Einstein's theory of gravity is "emergent", and one should be aware that this is a very big "if", then the fundamental short-distance degrees of freedom can be *radically* different from the long-distance "emergent" degrees of freedom [1, 2]. A prime example of this type of behaviour is fluid mechanics where the short distance physics (quantum molecular dynamics) is radically different from the long-distance degrees of freedom (density and velocity fields) which appear in the Euler and continuity equations. If a similar scenario holds for gravity, then, just as one cannot hope to get quantum molecular dynamics from quantizing fluid mechanics and the degrees of freedom appearing in the Euler's equation, one could not hope to get quantum gravity from quantizing Einstein's theory of gravity and the degrees of freedom appearing in the Einstein equation (the metric, or tetrad). In fact the metric (or tetrad) would lose their status as fundamental variables, being (like density and velocity fields) only defined in some "mean field" sense once one averages over appropriate microscopic degrees of freedom — whatever they might turn out to be [1, 2].

We already have at least one very concrete and specific example of such a behaviour in the "acoustic spacetimes" that emerge upon linearizing the equations of (non-relativistic, irrotational, barotropic, inviscid) fluid dynamics [3, 4, 5, 6], and there are a large number of more general situations in which "analogue spacetimes" can be constructed [7, 8, 9, 10, 11, 12, 13]. Once one attempts to generalize the acoustic spacetimes to nontrivial dispersion relations, where the phase and group velocities can differ and have nontrivial energy and momentum dependence [10, 11, 12, 13], then one is very naturally lead to one specific class of incarnations of the notion of "rainbow geometry" [1]. Furthermore under a plausible set of working hypotheses this class of "rainbow geometries" is remarkably similar to the class naturally arising from "quantum gravity phenomenology" [1].

Further afield energy-dependent rainbow geometries are currently of interest in cosmology, where a number of authors have tried to develop the notion of a scale-dependent metric within the context of a renormalization group flow on the space of all metrics. Very roughly speaking the idea is that if one averages the spacetime geometry over a cosmologically large length scale L, then the averaged metric $g_{ab}(x, L)$ should obey the Einstein equations for an effective scale-dependent Newton constant $G_N(L)$, at least to lowest order in a curvature expansion based upon the renormalization group [14, 15, 16, 17]. While "running" coupling constants are well-understood in particle physics, the general relativity community has traditionally been much more conservative when it comes to considering a "running" Newton constant or a "running" cosmological constant [18, 19, 20, 21, 22].

Even further afield, in some of the models loosely based on the ideas of "doubly special relativity" (also called "distorted special relativity", and in either case abbreviated as DSR) there is a notion often called "gravity's rainbow" wherein at short distances (corresponding to individual elementary particles) there is a feature that seems to imply that different spacetime geometries are seen by particles of different energy-momentum [23, 24, 25, 26, 27, 28, 29, 30, 31, 32].

In many of these models there are a number of ambiguities (and depending on one's attitude, potentially significant controversies) that serve to confuse the situation. In this article I hope to clarify the situation somewhat by explicitly constructing a pair of elementary and pedagogically useful examples of energy-dependent "rainbow" geometries in physical situations where all of the relevant physics is both well understood and straightforward. By doing so I hope ultimately to clarify the situation for the more complicated situations arising in cosmology, DSR, and quantum gravity phenomenology.

The two pedagogical models I will specifically deal with in this article are:

- Acoustic spacetimes with nontrivial dispersion relations [1].
- Newtonian mechanics reinterpreted as geodesic motion on an energy-dependent conformally flat 3-manifold.

These two models will teach us slightly different things about what it means to be a rainbow geometry, and hopefully will eventually lead us to a useful abstract definition of rainbow geometry.

Acoustic rainbow geometries

The acoustic rainbow geometries are based on extensions of the following rigorous theorem [1, 4, 5, 6]:

Theorem: Consider a non-relativistic irrotational, inviscid, barotropic perfect fluid, governed by the Euler equation, continuity equation, and an equation of state. The dynamics of the linearized perturbations (sound waves, phonons) is governed by a D'Alembertian equation

$$\Delta_g \Phi = \frac{1}{\sqrt{-g}} \, \partial_a \left(\sqrt{-g} \, g^{ab} \, \partial_b \Phi \right) = 0 \tag{1}$$

where g^{ab} is an (inverse) "acoustic metric" that depends algebraically on the background flow one is linearizing around:

$$g^{ab}(t, \vec{x}) \equiv \frac{1}{\rho_0 \, c_0} \left[\begin{array}{ccc} -1 & \vdots & v_0^j \\ \cdots\cdots & \cdot & \cdots\cdots\cdots\cdots \\ v_0^i & \vdots & (c_0^2 \, \delta^{ij} - v_0^i \, v_0^j) \end{array} \right]. \tag{2}$$

Here c_0 is the (hydrodynamic) speed of sound given by $c_0^2 = \partial p / \partial \rho$, while ρ_0 is the background density, and v_0 is the background velocity of the fluid. $\square$

It is important to realise that this is a rigorous theorem of abstract mathematical physics that leads to an *a priori* unexpected occurrence of Lorentzian-signature spacetime in a fluid mechanical setting [1, 4, 5, 6]. The (covariant) acoustic metric is

$$g_{ab}(t, \vec{x}) \equiv \frac{\rho_0}{c_0} \left[\begin{array}{ccc} -(c_0^2 - v_0^2) & \vdots & -v_0^j \\ \cdots\cdots\cdots\cdots & \cdot & \cdots\cdots \\ -v_0^i & \vdots & \delta_{ij} \end{array} \right], \tag{3}$$

and the "line element" can be written as

$$\mathrm{d}s^2 \equiv g_{ab} \, \mathrm{d}x^a \, \mathrm{d}x^b = \frac{\rho_0}{c_0} \left[-c_0^2 \, \mathrm{d}t^2 + (\mathrm{d}x^i - v_0^i \, \mathrm{d}t) \, \delta_{ij} \, (\mathrm{d}x^j - v_0^j \, \mathrm{d}t) \right]. \tag{4}$$

The relevance to rainbow geometries comes once one replaces the hydrodynamic speed of sound c_0 by any generalized wavenumber-dependent notion of propagation speed. Specifically, replace the hydrodynamic speed of sound by any one of:

$$c_0 \to c(k^2) \to \begin{cases} c_{\text{phase}}(k^2); \\ c_{\text{group}}(k^2); \\ c_{\text{geometric}}(k^2) = \sqrt{c_{\text{phase}}(k^2) \, c_{\text{group}}(k^2)}; \\ c_{\text{signal}}. \end{cases} \tag{5}$$

Here we define the signal speed by [33]

$$c_{\text{signal},1} = \lim_{k\to\infty} c_{\text{phase}}(k^2), \tag{6}$$

if we wish to focus on the propagation of discontinuities, and by

$$c_{\text{signal},2} = \max_{k}\; c_{\text{group}}(k^2), \tag{7}$$

if we wish to focus on information transfer via wave packets.

The point now is that these generalized wavenumber-dependent versions of acoustic geometry are all well-defined but *distinct* and convey different information about the physics as one moves beyond the hydrodynamic region [1]:

- The rainbow metric based on phase velocity contains information about the dispersion relation.
- The rainbow metric based on group velocity contains information about the propagation of wavepackets.
- The rainbow metric based on the geometric mean of group and phase velocities is in some sense the "best" local Lorentz approximation to the dispersion relation — see further discussion below.
- The non-rainbow metric based on signal velocity contains information about the overall causal structure.

Indeed if the signal velocity is finite then despite the complicated rainbow metric the overall causal structure is similar to that of general relativity, but with signal cones replacing light cones. If on the other hand the signal speed is infinite then the overall causal structure is similar to that of Newtonian physics, with a preferred global time. A secondary point to extract from this discussion is that, in contrast to general relativity, rainbow spacetimes are typically "multi-metric" with several different metrics encoding different parts of the physics.

All in all, the present discussion is sufficient to guarantee the well-defined mathematical and physical existence of at least one wide class of rainbow geometries — it is not necessarily true that all rainbow geometries can be put into this "acoustic" form, indeed the arguments in [1] identify at least one slightly wider class of rainbow geometries inspired by quantum gravity phenomenology — we could simply think of replacing $\delta_{ij} \leftrightarrow h_{ij}$ in the acoustic rainbow geometries, where h_{ij} is some Riemannian 3-metric.

The "geometric mean" rainbow geometry: The rainbow geometry based on

$$c_{\text{geometric}}(k^2) = \sqrt{c_{\text{phase}}(k^2)\; c_{\text{group}}(k^2)} \tag{8}$$

is perhaps more unusual than the others. To see why this might be a useful object to consider, write the dispersion relation in the form

$$(\omega - \vec{v}\cdot\vec{k})^2 = c_{\text{phase}}(k^2)\; k^2 = F(k^2), \tag{9}$$

and now expand $F(k^2)$ as a function of k^2 around some convenient reference point k_*^2. Then

$$F(k^2) = F(k_*^2) + F'(k_*^2)\; [k^2 - k_*^2] + \mathcal{O}([k^2 - k_*]^2), \tag{10}$$

which we can rewrite as

$$F(k^2) = \left\{ F(k_*^2) - F'(k_*^2)\; k_*^2 \right\} + F'(k_*^2)\; k^2 + \mathcal{O}([k^2 - k_*]^2). \tag{11}$$

But

$$F'(k_*^2) = \left.\frac{\partial(\omega^2)}{\partial(k^2)}\right|_{k_*^2} = \left.\frac{\omega}{k}\frac{\partial\omega}{\partial k}\right|_{k_*^2} = c_{\text{phase}}(k_*^2)\; c_{\text{group}}(k_*^2). \tag{12}$$

Furthermore, we can define a "mass term"

$$\omega_0(k_*^2) = F(k_*^2) - F'(k_*^2)\; k_*^2 = \left\{ c^2_{\text{phase}}(k_*^2) - c^2_{\text{geometric}}(k_*^2) \right\} k_*^2, \tag{13}$$

and so write

$$(\omega - \vec{v}\cdot\vec{k})^2 = \omega_0(k_*^2) + c^2_{\text{geometric}}(k_*^2)\; k^2 + \mathcal{O}([k^2 - k_*]^2). \tag{14}$$

If we simply truncate the expansion, writing

$$(\omega - \vec{v}\cdot\vec{k})^2 = \omega_0(k_*^2) + c^2_{\text{geometric}}(k_*^2)\; k^2, \tag{15}$$

then this is the best-fit "Lorentz invariant" dispersion relation that is tangent to the full dispersion relation at k_*^2. It is intriguing that it is this geometric mean velocity that seems to be governing the effective Hawking temperature in recent numerical calculations by Unruh [34], which are a natural extension of his earlier work in [35]. (Though in those calculations $\omega_0(k_*^2)$ does not seem to occur in any natural manner.)

The differential geometry of Newton's second law

I will now change gear quite drastically, and in counterpoint present a somewhat unusual route from Newton's second law to Maupertuis' variational principle, ending up with an energy-dependent conformally flat three-geometry. More standard presentations of various parts of this analysis can be found in the textbooks [36, 37], and research articles [38, 39].

Consider Newton's second law

$$\vec{F} = m\,\vec{a}, \tag{16}$$

so that for a body subject to a "conservative" force field

$$m\frac{\mathrm{d}^2\vec{x}}{\mathrm{d}t^2} = -\frac{\partial V(x)}{\partial \vec{x}}. \tag{17}$$

Now suppose that you have good surveying equipment but very bad clocks. So you can tell *where* the particle is, and its *path* through space, but you have poor information on *when* it is at a particular point. Can you reformulate Newton's second law in such a way as to nevertheless be able to get good information about the *path* the body follows?

Since (for simplicity) we are in Euclidean geometry we can adopt Cartesian coordinates, and so we can write the distance travelled in physical space as

$$\mathrm{d}s = \sqrt{\mathrm{d}\vec{x}\cdot\mathrm{d}\vec{x}}. \tag{18}$$

Can we now find a differential equation for the unit tangent vector $\mathrm{d}\vec{x}/\mathrm{d}s$ (instead of the velocity $\mathrm{d}\vec{x}/\mathrm{d}t$)? By using the chain rule we find

$$\frac{\mathrm{d}^2\vec{x}}{\mathrm{d}t^2} = \frac{\mathrm{d}s}{\mathrm{d}t}\frac{\mathrm{d}}{\mathrm{d}s}\left[\frac{\mathrm{d}s}{\mathrm{d}t}\frac{\mathrm{d}\vec{x}}{\mathrm{d}s}\right], \tag{19}$$

which implies

$$\frac{\mathrm{d}^2\vec{x}}{\mathrm{d}t^2} = \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2\left[\frac{\mathrm{d}^2\vec{x}}{\mathrm{d}s^2}\right] + \frac{1}{2}\left[\frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2\right]\frac{\mathrm{d}\vec{x}}{\mathrm{d}s}. \tag{20}$$

Putting this into Newton's second law

$$m\left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2\left[\frac{\mathrm{d}^2\vec{x}}{\mathrm{d}s^2}\right] = -\frac{\partial V(x)}{\partial \vec{x}} - \frac{1}{2}m\left[\frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2\right]\frac{\mathrm{d}\vec{x}}{\mathrm{d}s}. \tag{21}$$

Now let's simplify this a little. From the conservation of energy we have

$$\frac{1}{2}m\left(\frac{\mathrm{d}\vec{x}}{\mathrm{d}t}\right)^2 + V(x) = E. \tag{22}$$

But this means

$$\left(\frac{\mathrm{d}\vec{x}}{\mathrm{d}t}\right)^2 = \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2 = \frac{2[E - V(x)]}{m}, \tag{23}$$

so that Newton's second law becomes (note that m drops out)

$$2[E - V(x)] \left[\frac{\mathrm{d}^2\vec{x}}{\mathrm{d}s^2}\right] = -\frac{\partial V(x)}{\partial \vec{x}} - \frac{1}{2}\left[\frac{\mathrm{d}}{\mathrm{d}s}2[E - V(x)]\right]\frac{\mathrm{d}\vec{x}}{\mathrm{d}s}. \tag{24}$$

This can be rewritten in terms of a projection operator as

$$\frac{\mathrm{d}^2\vec{x}}{\mathrm{d}s^2} = \frac{1}{2}\left[I - \frac{\mathrm{d}\vec{x}}{\mathrm{d}s}\otimes\frac{\mathrm{d}\vec{x}}{\mathrm{d}s}\right]\frac{\partial \ln[E - V(x)]}{\partial \vec{x}}. \tag{25}$$

This completes the job of removing "time" from the equation of motion. One now has an equation strictly in terms of position x and physical distance along the path s.

But now let's go one step further and re-write this in terms of differential geometry — you should not be too surprised to see a three-dimensional conformally flat geometry drop out. To see this, consider a conformally flat three-geometry with metric

$$g_{ab} = \Omega^2(x)\,\delta_{ab}, \tag{26}$$

and note that the geodesic equations are (in arbitrary *non-affine* parameterization)

$$\frac{\mathrm{d}^2x^a}{\mathrm{d}\lambda^2} + \Gamma^a{}_{bc}\frac{\mathrm{d}x^b}{\mathrm{d}\lambda}\frac{\mathrm{d}x^a}{\mathrm{d}\lambda} = f(\lambda)\,\frac{\mathrm{d}x^c}{\mathrm{d}\lambda}, \tag{27}$$

where (with indices being raised and lowered using the flat metric δ_{ab}) we have

$$\Gamma^a{}_{bc} = \Omega^{-1}\left\{\delta^a{}_b\Omega_{,c} + \delta^a{}_c\Omega_{,b} - \delta_{bc}\Omega^{,a}\right\}. \tag{28}$$

Now if we chose our parameter λ to be arc-length s, *as measured by the flat metric* δ_{ab}, then

$$\delta_{ab}\frac{\mathrm{d}x^a}{\mathrm{d}s}\frac{\mathrm{d}x^c}{\mathrm{d}s} = 1, \tag{29}$$

so that differentiating

$$\delta_{ab}\frac{\mathrm{d}^2x^a}{\mathrm{d}s^2}\frac{\mathrm{d}x^b}{\mathrm{d}s} = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}s}\left[\delta_{ab}\frac{\mathrm{d}x^a}{\mathrm{d}s}\frac{\mathrm{d}x^b}{\mathrm{d}s}\right] = 0. \tag{30}$$

But this permits us to evaluate

$$f(s) = \Gamma^a{}_{bc} \frac{\mathrm{d}x_a}{\mathrm{d}s} \frac{\mathrm{d}x^b}{\mathrm{d}s} \frac{\mathrm{d}x^c}{\mathrm{d}s} = [\ln \Omega]_{,a} \frac{\mathrm{d}x^a}{\mathrm{d}s}, \tag{31}$$

and consequently the geodesic equation *in this particular parameterization* is

$$\frac{\mathrm{d}^2 x^a}{\mathrm{d}s^2} = \left[\delta^{ab} - \frac{\mathrm{d}x^a}{\mathrm{d}s} \frac{\mathrm{d}x^b}{\mathrm{d}s} \right] \partial_b \ln \Omega. \tag{32}$$

This is exactly the from of the equations previously derived for the path of a particle subjected to Newton's second law, provided we identify the conformal factor as

$$\Omega = \sqrt{E - V(x)}. \tag{33}$$

That is: the paths of particles subject to Newton's second law follow geodesics of the conformally flat three-geometry defined by

$$g_{ab} = [E - V(x)]\ \delta_{ab} \tag{34}$$

If we denote "distance" as measured by this conformal metric as ℓ then we have

$$\mathrm{d}\ell^2 = [E - V(x)]\ \mathrm{d}s^2. \tag{35}$$

While we have completely eliminated time from the equation for the paths there is very definitely a price to be paid — you now need to consider a separate geometry for each value of the energy. Minimizing the "distance" ℓ is what is commonly called Maupertuis' constant-energy variational principle. (Though note that Landau and Lifshitz attribute this version of the variational principle to Jacobi [36].) Also note that the classically forbidden regions where $E < V(x)$ have opposite signature (−−−) to the allowed regions (+++) and correspond to an imaginary $\mathrm{d}\ell$.

***N*-particle systems:** Furthermore, for N coupled particles of mass m_i with ($i \in [1..n]$) Newton's second law becomes the system of equations

$$m_i \frac{\mathrm{d}^2 \vec{x}_i}{\mathrm{d}t^2} = -\frac{\partial V(\vec{x}_1, \ldots, \vec{x}_N)}{\partial \vec{x}_i}. \tag{36}$$

It is easy to show that the paths swept out by this system of ODEs are geodesics in a conformally flat $3N$ dimensional configuration space with metric

$$\mathrm{d}\ell^2 = \{E - V(\vec{x}_1, \ldots, \vec{x}_N)\} \left\{ \sum_{i=1}^{N} m_i\ \mathrm{d}s_i^2 \right\}, \tag{37}$$

where $\mathrm{d}s_i$ is ordinary physical distance for the i'th particle. Note that we now need to explicitly keep track of individual particles masses.

Generalized configuration manifolds: In an even more general context where we have a mechanical system with N degrees of freedom, that has a kinetic energy quadratic in velocities, we may write

$$\mathcal{L} = \frac{1}{2} m_{ij}(q^k)\, \dot{q}^i \dot{q}^j - V(q^k). \tag{38}$$

Here $m_{ij}(q)$ is the configuration-dependent "mass matrix", and the q^k are generalized coordinates (living in the configuration manifold $\mathcal{M}$) that do not need to have the physical interpretation of being particle positions. A similar analysis to the above now yields an energy-dependent geometry

$$\mathrm{d}\ell^2 = g_{ij}(q^k)\, \mathrm{d}q^i\, \mathrm{d}q^j = \left\{ E - V(q^k) \right\}\; m_{ij}(q^k)\, \mathrm{d}q^i\, \mathrm{d}q^j, \tag{39}$$

that lives on the classically accessible submanifold,

$$\mathcal{M}(E) = \left\{ q^k : E > V(q^k) \right\}, \tag{40}$$

of the original configuration manifold $\mathcal{M} = \mathcal{M}(\infty)$.

Geodesic deviation: Now consider two initially parallel curves, in the 1-particle case, both corresponding to individual particles of energy E, separated by $\Delta x^a(s)$. Then by considering the difference of two geodesic equations we see that their separation grows according to the rule

$$\frac{\mathrm{d}^2 \Delta x^a}{\mathrm{d}s^2} = \frac{1}{2} \left[\delta^{ab} - \frac{\mathrm{d}x^a}{\mathrm{d}s} \frac{\mathrm{d}x^b}{\mathrm{d}s} \right] \;\; \partial_b \partial_c \ln[E - V(x)] \;\; \Delta x^d(s). \tag{41}$$

But this is now easily turned into a statement about the focussing effect of the Riemann tensor,

$$\frac{\mathrm{d}^2 \Delta x^a}{\mathrm{d}s^2} = R^a{}_{bcd}\, \frac{\mathrm{d}x^b}{\mathrm{d}s}\, \frac{\mathrm{d}x^d}{\mathrm{d}s}\, \Delta x^c(s), \tag{42}$$

with the Riemann tensor for the conformally flat 3-geometry being

$$R_{abcd} = -2 \left\{ g_{a[d}\, R_{c]b} + g_{b[c}\, R_{d]a} \right\} - R\, g_{a[c} g_{d]b}. \tag{43}$$

The Ricci tensor is given by

$$R_{ab} = \frac{1}{2} \frac{\partial_a \partial_b V}{E - V} + \frac{3}{4} \frac{\partial_a V\, \partial_b V}{(E - V)^2} + \delta_{ab} \left\{ \frac{1}{2} \frac{\nabla^2 V}{E - V} + \frac{1}{4} \frac{\partial_c V\, \partial^c V}{(E - V)^2} \right\}, \tag{44}$$

while the Ricci scalar is

$$R = 2\frac{\nabla^2 V}{E - V} + \frac{3}{2}\frac{\partial_c V\, \partial^c V}{(E - V)^2}. \tag{45}$$

The formalism has straightforward generalizations to multi-particle systems and mechanical systems with general quadratic kinetic energies.

Optical–mechanical analogy: There is of course a deep relationship between the differential geometric formulation above, Fermat's principle of minimum time, and the optical-mechanical analogy often used in situations where index-gradient methods are useful. If the refractive index is a function of position, then it is well known that the path of a light ray can be determined by minimizing the optical distance

$$\ell = \int n(x)\, \mathrm{d}s. \tag{46}$$

This is certainly equivalent to dealing with a conformally flat 3-geometry but the major difference here is that there is no simple way to turn Fermat's principle *directly* into a energy-dependent differential geometry, at least not without going all the way back to Newton's equations to cook up a specific mechanical system that effectively mimics the refractive index $n(x)$. This is often the main task in developing a specific instance of the optical-mechanical analogy [38, 40, 41, 42, 43].

Discussion

The acoustic rainbow geometries and the differential geometric version of Newton's second law have the great virtue that all relevant physics is clearly completely under control. Thus they provide an existence proof for the notion of an energy-dependent geometry, without the additional theoretical complications inherent in the running cosmology, DSR, or quantum gravity phenomenology frameworks. When it comes to the notion of a scale-dependent running metric in cosmology, or an energy-dependent metric in DSR, or in quantum gravity phenomenology, these present toy models make it clear that it is not the kinematics of single-particle motion in an energy-dependent spacetime that is in any way questionable. The technical difficulties lie at a deeper level. Specifically:

- Multi-particle kinematics: In the Newton law scenario multi-particle kinematics was best dealt with by going to a $3N$ dimensional configuration space. This would be extremely unnatural in the context of general relativity, both mathematically and physically. Observationally the various Eötvös-inspired experiments now verify the universality of free fall to at least one part in 10^{13}. The standard way of building this observational fact into general relativity [and any plausible extension of general relativity] is via the Einstein equivalence principle wherein one explicitly enforces "one spacetime for all individual particles", not one (nonlocal?) spacetime based on the configuration space. While as we have seen in the current article, there is nothing particularly radical in the proposal of using an energy-dependent geometry *per se*, there are real and fundamental experimental issues that must be addressed when it comes to positing an energy-dependent extension for general relativity.

- Geometrodynamics: There is a whole additional level of complexity that comes in to play when one wishes to ascribe a spacetime dynamics to the energy-dependent geometry. In DSR-inspired "gravity's rainbow" scenarios [23] this would seem to require an extension of the Einstein equations (and indeed an extension of the entire notion of spacetime curvature) onto the entire tangent bundle, typically with the individual tangent spaces becoming curved manifolds in their own right. In "running cosmology" scenarios there is the delicate question of exactly what renormalization point to pick when solving the FRW equations for the universe as a whole. Should we (self-consistently?) pick the Hubble scale $L = c/H_0$? Or the scale defined by the space curvature? Or something else?

It is these harder problems that must be confronted when trying to make sense of more complicated specific models built on the notion of an energy-dependent spacetime.

Independent of the questions raised by these more complicated models, the energy-dependent acoustic geometries and energy-dependent conformal geometry implicit in Newton's second law are of interest in their own rights, both as a simple illustration of the mathematical formalism of differential geometry, and as a different and unusual way of looking at wave propagation and classical mechanics.

Acknowledgments

This Research was supported by the Marsden Fund administered by the Royal Society of New Zealand. I also particularly wish to thank SISSA/ISAS (Trieste) for ongoing hospitality. Various versions of the ideas in this

article were also presented at the following conferences: “From quantum to emergent gravity: theory and phenomenology”, June 2007, SISSA/ISAS, Trieste, Italy; at ”Enrageing ideas”, September 2007, Utrecht University, the Netherlands; and at “Experimental search for quantum gravity”, November 2007, Perimeter Institute, Canada.

References

[1] M. Visser and S. Weinfurtner, *Analogue spacetimes: Toy models for 'quantum gravity'*, [arχiv:gr-qc/0712.0427].

[2] C. Barceló, M. Visser and S. Liberati, *Einstein gravity as an emergent phenomenon?*, Int. J. Mod. Phys. D **10** (2001) 799, [arχiv:gr-qc/0106002].

[3] W.G. Unruh, *The analogue between Rimfall and black holes*, Lect. Notes Phys. **718** (2007) 1.

[4] W.G. Unruh, *Experimental black hole evaporation*, Phys. Rev. Lett. **46** (1981) 1351.

[5] M. Visser, *Acoustic black holes: Horizons, ergospheres and Hawking radiation*, Class. Quant. Grav. **15** (1998) 1767, [arχiv:gr-qc/9712010]; *Acoustic propagation in fluids: An unexpected example of Lorentzian geometry*, [arχiv:gr-qc/9311028].

[6] C. Barceló, S. Liberati and M. Visser, *Analogue gravity*, Living Rev. Rel. **8** (2005) 12, [arχiv:gr-qc/0505065].

[7] R. Schützhold and W.G. Unruh, *Gravity wave analogs of black holes*, Phys. Rev. D **66** (2002) 044019, [arχiv:gr-qc/0205099].

[8] C. Barceló, S. Liberati and M. Visser, *Analog gravity from field theory normal modes?*, Class. Quant. Grav. **18** (2001) 3595, [arχiv:gr-qc/0104001].

[9] C. Barceló, S. Liberati and M. Visser, *Refringence, field theory and normal modes*, Class. Quant. Grav. **19** (2002) 2961, [arχiv:gr-qc/0111059]; M. Visser, C. Barceló and S. Liberati, *Bi-refringence versus bi-metricity*, [arχiv:gr-qc/0204017].

[10] M. Visser, C. Barceló and S. Liberati, *Acoustics in Bose-Einstein condensates as an example of broken Lorentz symmetry*, [arχiv:hep-th/0109033].

[11] M. Visser and S. Weinfurtner, *Massive phonon modes from a BEC-based analog model*, [arχiv:cond-mat/0409639]; *Massive Klein-Gordon equation from a BEC-based analogue spacetime*, Phys. Rev. D **72** (2005) 044020, [arχiv:gr-qc/0506029]; S. Liberati, M. Visser and S. Weinfurtner, *Analogue quantum gravity phenomenology from a two-component Bose-Einstein condensate*, Class. Quant. Grav. **23** (2006) 3129, [arχiv:gr-qc/0510125]; S. Weinfurtner, S. Liberati and M. Visser, *Analogue model for quantum gravity phenomenology*, J. Phys. A **39** (2006) 6807, [arχiv:gr-qc/0511105]; *Modelling Planck-scale Lorentz violation via analogue models*, J. Phys. Conf. Ser. **33** (2006) 373, [arχiv:gr-qc/0512127].

[12] S. Liberati, M. Visser and S. Weinfurtner, *Naturalness in emergent spacetime*, Phys. Rev. Lett. **96** (2006) 151301, [arχiv:gr-qc/0512139].

[13] S. Weinfurtner, S. Liberati and M. Visser, *Analogue spacetime based on 2-component Bose-Einstein condensates*, Lect. Notes Phys. **718** (2007) 115, [arχiv:gr-qc/0605121].

[14] M. Reuter, *Nonperturbative Evolution Equation for Quantum Gravity*, Phys. Rev. D **57** (1998) 971, [arχiv:hep-th/9605030]; *Newton's constant isn't constant*, [arχiv:hep-th/0012069]; M. Reuter and J.M. Schwindt, *Scale dependent metric and minimal length in QEG*, J. Phys. A **40** (2007) 6595, [arχiv:hep-th/0610064]; *Scale-dependent metric and causal structures in quantum Einstein gravity*, JHEP **0701** (2007) 049, [arχiv:hep-th/0611294]; M. Reuter and H. Weyer, *On the possibility of quantum gravity effects at astrophysical scales*, Int. J. Mod. Phys. D **15** (2006) 2011, [arχiv:hep-th/0702051]; A. Bonanno and M. Reuter, *Entropy signature of the running cosmological constant*, JCAP **0708** (2007) 024, [arχiv:hep-th/0706.0174]; M. Reuter and F. Saueressig, *Functional Renormalization Group Equations, Asymptotic Safety and Quantum Einstein Gravity*, [arχiv:hep-th/0708.1317].

[15] D. Dou and R. Percacci, *The running gravitational couplings*, Class. Quant. Grav. **15** (1998) 3449, [arχiv:hep-th/9707239]; R. Percacci and D. Perini, *Constraints on matter from asymptotic safety*, Phys. Rev. D **67** (2003) 081503, [arχiv:hep-th/0207033]; *Should we expect a fixed point for Newton's constant?*, Class. Quant. Grav. **21** (2004) 5035, [arχiv:hep-th/0401071]; R. Percacci, *The Renormalization Group, Systems of Units and the Hierarchy Problem*, J. Phys. A **40** (2007) 4895, [arχiv:hep-th/0409199]; F. Girelli, S. Liberati, R. Percacci and C. Rahmede, *Modified dispersion relations from the renormalization group of gravity*, Class. Quant. Grav. **24** (2007) 3995, [arχiv:gr-qc/0607030]; R. Percacci, *Asymptotic Safety*, [arχiv:hep-th/0709.3851].

[16] V. Periwal, *Cosmological and astrophysical tests of quantum gravity*, [arχiv:astro-ph/9906253].

[17] F. Cardone, M. Francaviglia and R. Mignani, *Five-Dimensional Relativity With Energy As Extra Dimension*, Gen. Rel. Grav. **31** (1999) 1049; F. Cardone and R. Mignani, *Broken Lorentz Invariance And Metric Description Of Interactions In A Deformed Minkowski Space*, Found. Phys. **29** (1999) 1735; F. Cardone, M. Francaviglia and R. Mignani, *Energy As A Fifth Dimension*, Found. Phys. Lett. **12** (1999) 347.

[18] C. Espana-Bonet, P. Ruiz-Lapuente, I.L. Shapiro and J. Sola, *Testing the running of the cosmological constant with type Ia supernovae at high z*, JCAP **0402** (2004) 006, [arχiv:hep-ph/0311171].

[19] I.L. Shapiro and J. Sola, *A Friedmann-Lemaitre-Robertson-Walker cosmological model with running Lambda*, [arχiv:astro-ph/0401015].

[20] B. Guberina, R. Horvat and H. Štefančić, *Renormalization-group running of the cosmological constant and the fate of the universe*, Phys. Rev. D **67** (2003) 083001, [arχiv:hep-ph/0211184].

[21] E. Elizalde, S.D. Odintsov and I.L. Shapiro, *Asymptotic regimes in quantum gravity at large distances and running Newtonian and cosmological constants*, Class. Quant. Grav. **11** (1994) 1607, [arχiv:hep-th/9404064].

[22] D.A. Kosower, *The Running Of The Cosmological Constant*, Mod. Phys. Lett. A **4** (1989) 2323.

[23] J. Magueijo and L. Smolin, *Gravity's Rainbow*, Class. Quant. Grav. **21** (2004) 1725, [arχiv:gr-qc/0305055].

[24] G. Amelino-Camelia, *Planck-scale structure of spacetime and some implications for astrophysics and cosmology*, [arχiv:astro-ph/0312014].

[25] G. Amelino-Camelia, *The three perspectives on the quantum-gravity problem and their implications for the fate of Lorentz symmetry*, [arχiv:gr-qc/0309054].

[26] G. Amelino-Camelia, *Fundamental physics in space: A quantum-gravity perspective*, Gen. Rel. Grav. **36** (2004) 539, [arχiv:astro-ph/0309174].

[27] G. Amelino-Camelia, L. Smolin and A. Starodubtsev, *Quantum symmetry, the cosmological constant and Planck scale phenomenology*, Class. Quant. Grav. **21** (2004) 3095, [arχiv:hep-th/0306134].

[28] J. Magueijo, *New varying speed of light theories*, Rept. Prog. Phys. **66** (2003) 2025, [arχiv:astro-ph/0305457].

[29] D. Kimberly, J. Magueijo and J. Medeiros, *Non-Linear Relativity in Position Space*, Phys. Rev. D **70** (2004) 084007, [arχiv:gr-qc/0303067].

[30] S. Hossenfelder, *Deformed Special Relativity in Position Space*, Phys. Lett. B **649** (2007) 310, [arχiv:gr-qc/0612167].

[31] S. Hossenfelder, *Multi-Particle States in Deformed Special Relativity*, Phys. Rev. D **75** (2007) 105005, [arχiv:hep-th/0702016].

[32] F. Girelli, S. Liberati and L. Sindoni, *Phenomenology of quantum gravity and Finsler geometry*, Phys. Rev. D **75** (2007) 064015, [arχiv:gr-qc/0611024].

[33] L. Brillouin, *Wave Propagation and Group Velocity* (Academic Press, New York, 1960).

[34] W.G. Unruh, unpublished (Conference presentations at "From Quantum to Emergent Gravity: Theory and Phenomenology", June 2007, SISSA/ISAS, Trieste, Italy; and at "Effective Models of Quantum Gravity", November 2007, Perimeter Institute, Canada.)

[35] W.G. Unruh, *Sonic Analog Of Black Holes And The Effects Of High Frequencies On Black Hole Evaporation*, Phys. Rev. D **51** (1995) 2827; *Dumb Holes And The Effects Of High Frequencies On Black Hole Evaporation*, [arχiv:gr-qc/9409008].

[36] L.D. Landau and E.M. Lifshitz, *Mechanics* [third edition] (Butterworth-Heinenann, Oxford, England, 2000); See especially §44, p. 140–143.

[37] V.I. Arnold, *Mathematical methods of classical mechanics* (Springer–Verlag, New York, 1978); See especially p. 247, 253.

[38] M. Biesiada and S.E. Rugh, *Maupertius principle, Wheeler's superspace and an invariant criterion for local instability in general relativity*, [arχiv:gr-qc/9408030].

[39] M. Szydlowski and A. Krawiec, *Average rate of separation of trajectories near the singularity in mixmaster models*, Phys. Rev. D **47** (1993) 5323; *Description of chaos in simple relativistic systems*, Phys. Rev. D **53** (1996) 6893; M. Szydlowski and J. Szczesny, *Invariant chaos in mixmaster cosmology*, Phys. Rev. D **50** (1994) 819.

[40] A. Tsiganov, *The Maupertuis principle and integrable systems*, [arχiv:nlin.SI/0009044].

[41] K.K. Nandi, Y.Z. Zhang, P.M. Alsing, J.C. Evans and A. Bhadra, *Analog of the Fizeau Effect in an Effective Optical Medium*, Phys. Rev. D **67** (2003) 025002, [arχiv:gr-qc/0208035].

[42] J.C. Evans, P.M. Alsing, S. Giorgetti and K.K. Nandi, *Matter waves in a gravitational field: An index of refraction for massive particles in general relativity*, Am. J. Phys. **69** (2001) 1103, [arχiv:gr-qc/0107063].

[43] M. Marklund, D. Anderson, F. Cattani, M. Lisak, L. Lundgren, *Fermat's principle and variational analysis of an optical model for light propagation exhibiting a critical radius*, Am. J. Phys. **70** (2002) 680, [arχiv:physics/0102019].

Section VI: Big Bang Evolution and Structure Formation of the Universe

origin of time and its error
cosmological constant
new matter

Time and Matter
2007 Bled, Slovenia

A Discrete Space and Time Before the Big Bang

MARTIN BOJOWALD*
Institute for Gravitation and the Cosmos, The Pennsylvania State University, 104 Davey Lab, University Park, PA 16802, USA

Abstract: A discrete structure of space and time is often expected to result from a quantum theory of gravity. Loop quantum gravity realizes this clearly at least for space while time, in such a canonical quantization, is handled more indirectly. Cosmological scenarios show how the discreteness of time, unnoticeable at current scales, becomes an important feature in the early universe. It plays a crucial role in resolving the classical big bang singularity and in opening the door to a universe before the big bang. While this happens independently of what matter is prevalent at the big bang, parity violating effects do have a bearing on the relation between pre- and post-big bang branches. Even in the absence of parity violation, quantum effects in dynamical coherent states indicate similar asymmetries. All this may be important to discern the origin of the universe.

The gravitational field is the only known fundamental force not yet quantized completely, despite more than six decades of research. Difficulties in the construction arise mainly due to two key properties: Although gravity is weak in usual regimes of particle physics, it becomes the dominant player on cosmic scales. Strong quantum gravity effects should appear in large gravitational fields such as the very early universe and black holes. Then, however, the classical field grows without bound, implying space-time singularities.

Secondly, there is the equivalence principle: gravity is a manifestation of space-time geometry rather than a field propagating on a given space-time. The full space-time metric $g_{\mu\nu}$ is thus the physical object to be quantized non-perturbatively, rather than using perturbations $h_{\mu\nu}$ on a background space-time $\eta_{\mu\nu}$ such as Minkowski space.

As a third difficulty, not in the construction but in interpretations of the theory, one could add the fact that space-time is now described by a quan-

* bojowald@gravity.psu.edu

tum state with all the well-known potential pitfalls and counter-intuitive effects. One may avoid this in initial steps of developing the theory and in analyzing semiclassical aspects. But, as we will indicate in the end of this contribution, state properties are already becoming crucial in applications and indicate new properties of a universe.

One possibility to quantize the full space-time is canonical quantization, which goes back to Wheeler and DeWitt [1]. Here, one combines general relativity and quantum physics but, compared to quantum mechanics, one substitutes (q, p) by (q_{ab}, p^{ab}). This makes use of gravitational canonical variables where q_{ab} is the spatial metric appearing in the line element

$$\mathrm{d}s^2 = -N^2 \mathrm{d}t^2 + \sum_{a,b=1}^{3} q_{ab}(\mathrm{d}x^a + N^a \mathrm{d}t)(\mathrm{d}x^b + N^b \mathrm{d}t)\,. \tag{1}$$

for a dynamical, curved space-time, and p^{ab} its momentum related geometrically to extrinsic curvature $K_{ab} = \frac{1}{2N}(\dot{q}_{ab} - D_a N_b - D_b N_a)$. Compared to the Minkowski line element $\mathrm{d}s^2 = -\mathrm{d}t^2 + (\mathrm{d}x^1)^2 + (\mathrm{d}x^2)^2 + (\mathrm{d}x^3)^2$, all components in (1) in general depend on space coordinates x^a and time t such that measurements of space and time intervals depend on the position in space-time. This captures general relativistic effects, describing the gravitational field in a geometrical way. All the coefficients N, N^a and q_{ab} in (1) are space-time dependent, but it turns out that only q_{ab} is dynamical and has a non-vanishing momentum p^{ab}, while N and N^a determine the behavior of observers in space-time.

For the wave function, we correspondingly substitute $\psi(q)$ by a functional $\psi[q_{ab}]$. In contrast to the Schrödinger equation $i\hbar\frac{\partial}{\partial t}\psi(q) = \hat{H}\psi(q)$ for $\psi(q)$, $\psi[q_{ab}]$ is subject to constraints $\hat{H}\psi[q_{ab}] = 0$ and $\hat{D}\psi[q_{ab}] = 0$. The equation $\hat{H}\psi[q_{ab}] = 0$, where $\hat{H}$ corresponds to the sum of gravitational and matter Hamiltonians, is called the Wheeler–DeWitt equation. In addition, there is a second equation $\hat{D}\psi[q_{ab}] = 0$ which ensures that physical solutions are invariant under spatial coordinate changes. This equation is comparatively easy to implement and will only play a minor role in what follows. The Hamiltonian constraint equation $\hat{H}\psi[q_{ab}] = 0$, on the other hand, is not only the main difficulty of quantum gravity but also most important, as it determines the dynamics of a quantum space-time.

Space-time structure

To make the Wheeler–DeWitt equation $\hat{H}\psi[q_{ab}] = 0$ precise and explicit, one has to step over several mathematical hurdles. Tensor fields such as

q_{ab} are subject to transformation laws under changing coordinates. The classical invariance of the theory under such transformations must be preserved by the quantization. However, in generally covariant systems such as general relativity, a non-linear change of coordinates

$$q_{ab} \mapsto \sum_{a',b'=1}^{3} \frac{\partial x'^{a'}}{\partial x^a} \frac{\partial x'^{b'}}{\partial x^b} q_{a'b'} \tag{2}$$

would lead to coordinate dependent factors not represented on the Hilbert space: There is no operator for x^a in quantum gravity, but only those for the dynamical field q_{ab} and its momentum exist, just as the vector potential A_a and its momentum, the electric field E^a, become the operators of quantum electrodynamics but not x^a or t. Starting with an operator $\hat{q}_{ab}$, for instance in the Wheeler–DeWitt equation, and then doing a coordinate transformation would imply the occurrence of coordinate dependent functions which lack physical meaning. Such difficulties do not arise for quantizations on a background Minkowski space, where only linear Poincaré transformations are allowed as coordinate changes. Factors in coordinate transformations (2) of tensor fields are then merely space-time independent constants.

Such conceptual difficulties in setting up a well-defined theory are useful guides in the search for suitable building blocks of quantum gravity, especially given the absence of observational input to date. The distinguishing features of general relativity as a classical theory should be reflected in its quantization. One solution of the coordinate problem is to reformulate the classical theory in an equivalent way using only index-free objects. This is provided by loop quantum gravity [2, 3, 4], based on holonomies and fluxes as elementary objects analogous to magnetic and electric fluxes, rather than space-time tensors such as q_{ab}. The precise implementation sketched below has an immediate consequence: The configuration space given by holonomies is compact, and geometrical fluxes become derivative operators on a compact space with discrete spectra. Since the fluxes of loop quantum gravity by construction describe the quantum geometry of space, as they quantize what is classically determined by q_{ab}, space itself is discrete.

The definition of scalar objects used in loop quantum gravity is based on new variables, obtained by a complicated-looking canonical transformation

$$(q_{ab}, p^{ab}) \mapsto (-\sum_{j,k,b=1}^{3} \epsilon^{ijk} e_j^b (\partial_{[a} e_{b]}^k + \tfrac{1}{2} e_k^c e_a^l \partial_{[c} e_{b]}^l) + \gamma \sum_{b=1}^{3} e_i^b K_{ab}, |\det(e_b^j)| e_i^a)\,.$$

In this notation, one first takes a "matrix square-root" of $q_{ab} = \sum_{i=1}^{3} e_a^i e_b^i$, introducing the co-triad e_a^i. Its inverse, the triad, is denoted as e_i^a with reversed indices. One way to visualize the triad is as a set of three spatial vector fields $\vec{e}_1 = (e_1^a)_{a=1,2,3}$, $\vec{e}_2 = (e_2^a)_{a=1,2,3}$, $\vec{e}_3 = (e_3^a)_{a=1,2,3}$ which are declared to be orthogonal to each other and have unit norm. This provides the same information as specifying metric components q_{ab} for a line element (1).

More compactly, we have $(q_{ab}, p^{cd}) \mapsto (A_a^i, E_j^b)$ with the Ashtekar connection $A_a^i := \Gamma_a^i + \gamma K_a^i$ and the densitized triad E_j^b [5, 6]. These are variables of a gauge theory with gauge group SO(3) for spatial rotations which can change the triad vectors without affecting the metric. In A_a^i, we have a sum of contributions Γ_a^i, which determines spatial curvature, and K_a^i which determines the bending of space in space-time. The combination is weighted by the so-called Barbero–Immirzi parameter γ [6, 7], which does not have much of a classical effect but does play a role in the quantization. With such variables as in non-Abelian gauge theories one can use a "lattice" formulation, which assigns basic variables to curves and transversal surfaces but does not require curves to lie on any regular graph [8]. For any curve C and surface S in space, we define holonomies and fluxes by

$$h_C(A) = \mathcal{P}\exp\left(\int_C \sum_{a,i=1}^{3} A_a^i \tau_i \mathrm{d}s^a\right), \qquad F_S(E) = \int_S \sum_{a,i=1}^{3} E_i^a \tau_i \mathrm{d}^2\sigma_a$$

with Pauli matrices τ_i. (The symbol $\mathcal{P}$ indicates that the non-commuting su(2) elements in the integrand are ordered along the path as the integration is performed.)

Holonomies provide the field A_a^i, which can be reconstructed if holonomies for all curves are known. It can thus be used as a non-tensorial configuration variable, such that wave functions are $\psi[h_C]$, depending on the holonomies along all curves in space. Since holonomies take values in SU(2), they indeed form a compact manifold even for this infinite dimensional quantum field theory where all possible curves are allowed. The quantization thus has properties parallel to quantum mechanics on a compact space such as a circle. In this example, wave functions $\psi_n(\phi) = \exp(in\phi)$ with integer n are periodic (or periodic up to a phase, which introduces a 1-parameter quantization ambiguity in the choice of the representation). Momentum eigenvalues are discrete: $\hat{p}\psi_n(\phi) = -i\hbar\mathrm{d}\psi_n(\phi)/\mathrm{d}\phi = \hbar n\psi_n(\phi)$.

Similarly, fluxes as the canonically conjugate momenta of holonomies become derivative operators on SU(2) and, analogously to angular momentum operators in quantum mechanics, have discrete spectra. By construction, we started with the spatial metric q_{ab}, replaced it by the densitized triad E^a_i and integrated it to obtain fluxes. Thus, flux operators with discrete spectra imply that spatial geometry is discrete. For instance, volumes of point sets can only increase in discrete steps when they are enlarged [9, 10].

When $\psi[h_C] = \text{const}$, i.e. there is no dependence on holonomies at all, then all fluxes as derivative operators are zero. This can be viewed as the fundamental ground state, in which not even geometry is excited: there is no space at all for any measurement of distances in this state. Fluxes and thus geometry are switched on for surfaces transversal to at least one C for which ψ depends non-trivially on h_C. Clearly, we need a non-trivial dependence on many h_C for a macroscopic geometry to result, where fluxes change almost smoothly when their surfaces are enlarged. Thus, any macroscopic space-time, including Minkowski space, appears as a highly excited state in this framework. In fact, also dynamical growth such as universe expansion appears in discrete steps [11] by exciting more and more atoms of space. In particular our present classical state of the universe has to grow out of a small, very quantum state at the big bang.

The scale of the discreteness is important for applications and potential implications of quantum gravity, for instance in cosmology. A common dimensional argument leads to the Planck length $\ell_{\mathrm{P}} = \sqrt{G\hbar} \approx 10^{-35}\mathrm{m}$, with Newton's constant G, as the only parameter of the dimension length which can be constructed from fundamental constants. The argument is thus similar to dimensional arguments in quantum mechanics which can be used to estimate the Bohr radius $a_0 \propto \hbar^2/m_e e^2$ without much calculation. However, just as calculations are necessary to verify the precise numerical prefactor of the Bohr radius, and to see which spatial extensions occur for excited states, the role of the Planck length in quantum gravity is to be determined from calculations. Loop quantum gravity shows that the precise discreteness scale of an elementary excitation of geometry is $\sqrt{\gamma}\ell_{\mathrm{P}}$ with a value $\gamma \approx 0.238$ obtained from black hole entropy [12, 13, 14, 15]. The numerical value of γ is obtained by requiring consistency between quantum gravity calculations and semiclassical calculations of Hawking radiation. This is indeed close to the Planck scale, although slightly smaller than the Planck length itself.

There is thus a derived scale for the ground state of geometrical quanta, much like the Bohr radius of a hydrogen atom. In this context, however, it is important to note that states of a quantum space are dynamical and should be expected to change in time at a fundamental scale, just as the classical geometry changes during cosmic expansion. Unlike atomic systems which in many situations settle quickly into their ground state whose scale provides a good estimate for the atomic dimensions, there is no clear notion of a ground state for the quantum state of a dynamical universe. Thus, the possibility of excited states contributing to the understanding of the universe in an essential way needs to be taken into account [16, 17]. In this case, discreteness levels of geometry can be higher than the lowest level, for instance as $\sqrt{\gamma}\ell_{\mathrm{P}}n$ with integer n. Which quantities are nearly equidistantly spaced by integer multiples of an elementary quantum is another property which has to be determined from calculations in the theory. In loop quantum gravity it is areas which are nearly equidistantly spaced for a single excitation created by holonomy operators [9, 18].

The dynamical aspect of the underlying discreteness combined with the geometrical nature of gravity has an additional implication [19]. As usually, a state is more semiclassical for higher excitations, and thus larger n. However, higher n also raise the spatial scale of excitations of geometry, implying a coarser discreteness. We have strong quantum behavior at small n and noticeable discreteness at large n, both of which gives deviations from the classical theory. Only a finite window is left in which one has agreement between classical continuum geometry and quantum geometry. To both sides, quantum gravity effects would be noticeable which has certainly not been observed. This window of observationally viable parameters is currently large, but its finite size provides crucial leverage to be exploited by observations despite the smallness of ℓ_{P}. Especially in cosmology with its long evolution times, magnification effects of quantum corrections can result [20].

Quantum cosmology

Space-time discreteness must have dynamical implications for the evolution of the universe, which are most easily seen for isotropic spaces. Here, the fields of basic variables reduce to a single component $|E| = a^2/4$ of the densitized triad, which is expressed in a more familiar way through the scale factor a, and a single connection component $A = \dot{a}/2$. This setting corresponds to a Friedmann–Robertson–Walker line element $\mathrm{d}s^2 = -\mathrm{d}t^2 + a(t)^2((\mathrm{d}x^1)^2 + (\mathrm{d}x^2)^2 + (\mathrm{d}x^3)^2)$ where, compared to 1, only spatial scales change in time but do so isotropically and homogeneously. Also these components are canonically conjugate to each other, i.e. E can be considered as the momentum of A. Notice that we explicitly wrote an absolute value $|E|$ in the relation to the scale factor because E itself can have any sign. Unlike the scale factor, which only determines the size of space, E as a triad component also has information about the orientation of space via the handedness of the triad.

A loop quantization of isotropic cosmological models [21, 22, 23] leads to an orthonormal basis of states $\langle A|\mu\rangle = e^{i\mu A/2}$ in the A-representation, where the quantum number $\mu \in \mathbb{R}$ can take any real value. The Hilbert space is thus non-separable, having a continuum of basis states. Basic operators act as

$$\begin{aligned} \widehat{e^{i\mu' A/2}}|\mu\rangle &= |\mu+\mu'\rangle \\ \hat{E}|\mu\rangle &= \tfrac{4\pi}{3}\gamma\ell_{\mathrm{P}}^2\mu|\mu\rangle \end{aligned}$$

the latter being obtained from $\hat{E} = -\frac{1}{3}i\hbar\gamma G\frac{\partial}{\partial A}$ where factors of γ and Newton's constant G appear due to analogous factors in the classical Poisson bracket $\{A, E\} = \frac{8\pi}{3}\gamma G$. Also here, even in this highly reduced setting, one can recognize the discreteness of quantum geometry: Only shift operators of the quantum number μ are represented, while a derivative of $\exp\widehat{(i\mu' A/2)}$ by μ' does not exist (which one can easily verify for matrix elements of this operator). This mimics the construction of the full theory where only holonomies, i.e. exponentials of the integrated A_a^i, are represented but not A_a^i itself. Moreover, $\hat{E}$ has a discrete spectrum because all its eigenstates are normalizable (and thus correspond to bound states rather than scattering states even though the eigenvalues form a continuum).

These properties of isotropic operators follow from the full holonomy and flux operators, whose representation is highly restricted. In an isotropic model there would be alternative ways for a quantum representation such as that due to Wheeler and DeWitt, but it is inequivalent to the loop representation. Thus, physical properties can easily differ especially in strong quantum regimes. The loop representation is distinguished through its close relation to the full holonomy-flux algebra [24, 16], which is lacking in the Wheeler–DeWitt case.

Dynamical properties are determined by solving the equation $\hat{H}|\psi\rangle = 0$, where $\hat{H}$ quantizes the Hamiltonian of the gravitational field and matter. The wave function $\psi(a,\phi)$ of a Wheeler–DeWitt quantization is now replaced by a function $\psi_\mu(\phi)$ in loop quantum cosmology where the quantum number μ replaces the continuous scale factor a. Imposing $\hat{H}\psi_\mu(\phi) = 0$, the state is subject to a difference equation [22]

$$\begin{aligned} &3(V_{\mu+5}-V_{\mu+3})\psi_{\mu+4}(\phi) - 6(V_{\mu+1}-V_{\mu-1})\psi_\mu(\phi) \\ &3(V_{\mu-3}-V_{\mu-5})\psi_{\mu-4}(\phi) = -32\pi^2 G\gamma^3\ell_{\mathrm{P}}^2\hat{H}_{\mathrm{matter}}(\mu)\psi_\mu(\phi) \end{aligned} \tag{3}$$

with volume eigenvalues $V_\mu = (4\pi\gamma\ell_{\mathrm{P}}^2|\mu|/3)^{3/2}$. Also the matter Hamiltonian $\hat{H}_{\mathrm{matter}}(\mu)$ occurs and is well-defined in a loop quantization [25]. (Several other versions of this equation have been defined, reflecting the incomplete status of the full construction of loop quantum gravity and its reduction to isotropy. None of the differences between the alternatives matter for the discussions here.)

At this point it becomes important that μ, the eigenvalue of the triad component E, takes both signs: this fact allows evolution of the wave function across the classical big bang singularity at $\mu = 0$, continuing to a new branch preceding the big bang [26]. A Wheeler–DeWitt wave function $\psi(a, \phi)$ subject to a differential equation, by contrast, is not extended beyond the classical singularity at $a = 0$. All ingredients for non-singular evolution are provided automatically by the quantization without further input. For instance, we are forced to use the triad in index-free scalar objects which are not known in any other suitable form. The role in singularity removal was realized only much later than the basic construction. The sign $\mathrm{sgn}(\mu)$ originates from the handedness of the triad, and thus determines the spatial orientation. Physically, then, the orientation flips while the universe traverses the classical big bang singularity, and the universe "turns its inside out."

It is interesting to highlight the role of spatial discreteness in this context. (See [27] for extended discussions.) It directly led to the absence of a singularity by a difference equation, uniquely connecting the underlying wave function at both sides of vanishing volume $\mu = 0$. Classical evolution as well as the Wheeler–DeWitt wave function, none of which incorporates spatial discreteness, would stop there. The discreteness is also responsible for making the inverse scale factor in a matter Hamiltonian, e.g. $H_\phi = \frac{1}{2}a^{-3}p_\phi^2 + a^3V(\phi)$ for a scalar ϕ with momentum p_ϕ, finite [28]. In fact, $\widehat{a^{-3}}$ becomes a bounded operator when expressed through holonomies and fluxes, as illustrated in Fig. 1. The discreteness thus directly implies regular evolution to a universe before the big bang:

$$\text{discrete space} \longrightarrow \left\{ \begin{array}{l} \bullet\text{discrete steps of} \\ \;\text{wave function evolution} \\ \bullet\text{no diverging } \widehat{a^{-3}} \\ \;\text{in matter Hamiltonian} \end{array} \right\} \longrightarrow \begin{array}{l} \text{time before} \\ \text{the big bang} \end{array}$$

The absence of a singularity is independent of the form of matter because the matter term does not affect the recurrence scheme of the difference equation (3). (This hold true even for non-minimal coupling, where a change in the recurrence could initially be expected [30].) But as we saw,

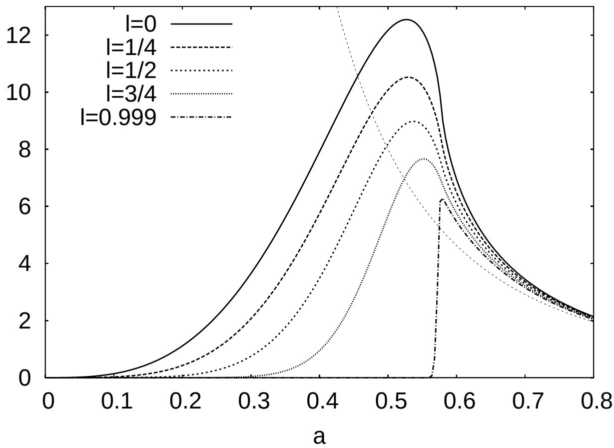


Figure 1: Eigenvalues of an operator quantizing a^{-3} in isotropic loop quantum cosmology. While the values depend on quantization ambiguities such as those parameterized by l, they remain finite and approach zero where the classical expression would diverge. See also [29] for a discussion of the generality of this behavior.

orientation reversal is key in the transition through the singularity. Parity violation on the matter side, i.e. $\hat{H}_{\rm matter}(\mu) \neq \hat{H}_{\rm matter}(-\mu)$, makes the equation non-invariant under $\mu \mapsto -\mu$ and states of the universe before and after the big bang must then be different. This behavior is of importance to understand the origin of the universe before the big bang.

Effective picture

It is quite difficult to analyze the fundamental difference equation more generally, which would also require an understanding of the measurement process in quantum cosmology and the meaning of the wave function. But for many purposes, effective equations are available and very powerful. In this case, one solves the dynamics for expectation values, say $\langle \hat{V} \rangle = \langle |\hat{E}|^{3/2} \rangle$ of the volume, directly, rather than using wave functions with their interpretational problems [31, 32]. Properties of the state are determined simultaneously with expectation values but only show up implicitly. The procedure is analogous to perturbation techniques in quantum field theory, which allow one to introduce important effects of an interacting vacuum state without actually computing it. A full interacting vacuum would be difficult if not impossible to find explicitly, but perturbation expansions such as the Feynman series do take its properties into account. Similarly, more general schemes of effective equations allow one to include properties of dynamical coherent states in perturbative calculations.

This procedure has the usual starting point as it is well-known from the Ehrenfest theorem in quantum mechanics:

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\hat{V}\rangle = \frac{\langle[\hat{V},\hat{H}]\rangle}{i\hbar} = \cdots$$

allows one to express the change in time of, say, the volume expectation value in terms of other degrees of freedom such as $\langle\widehat{\exp(iA)}\rangle$ but also higher moments of the state, in particular its fluctuations, e.g. $(\Delta V)^2 = \langle\hat{V}^2\rangle - \langle\hat{V}\rangle^2$.

In general, this results in infinitely many coupled equations, which is difficult to solve unless there are approximations (or, in lucky cases, exact procedures) to decouple most equations. This is feasible if the system of interest is closely related to a solvable model where commutators such as $[\hat{V},\hat{H}]$ are linear in basic operators. Then, the Ehrenfest equations involve only expectation values without coupling terms to fluctuations or higher moments. Also fluctuations themselves are then subject to only a finite set of coupled equations. A well-known example where this occurs is the harmonic oscillator, while any anharmonicity couples all state moments to the expectation values.

In cosmology, a solvable model is available in the form of an isotropic space-time with a free, massless scalar [33]. This model plays the role of the harmonic oscillator for cosmology. It does not arise in an obvious solvable form, for which one rather has to formulate it in special, non-canonical variables such as $V = |E|^{3/2}$ and $J := |E|^{3/2}\exp(iA/\sqrt{|E|})$. When quantized, with the given ordering for the factors in J, one obtains an operator algebra of $\mathrm{sl}(2,\mathbb{R})$: $[\hat{V},\hat{J}] = \hbar\hat{J}$, $[\hat{V},\hat{J}^\dagger] = -\hbar\hat{J}^\dagger$ and $[\hat{J},\hat{J}^\dagger] = -2\hbar\hat{V} - \hbar^2$. Moreover, the Hamiltonian underlying (3) becomes linear in a suitable factor ordering defined by $\hat{H} = -\frac{1}{2}i(\hat{J}-\hat{J}^\dagger)$, making the whole system solvable with decoupled equations of motion for expectation values and fluctuations. (The fact that complex exponentials of A are used in the definition of J implies that the Hamiltonian is indeed a difference operator and not differential, implementing the discreteness of space.)

By now, explicit solutions of this model have been used to find precise properties of dynamical coherent states [34] and to shed light on the transition through the big bang singularity by a quantum state [35, 36]. Fig. 2 provides an illustration of the effects which can happen even in this rather

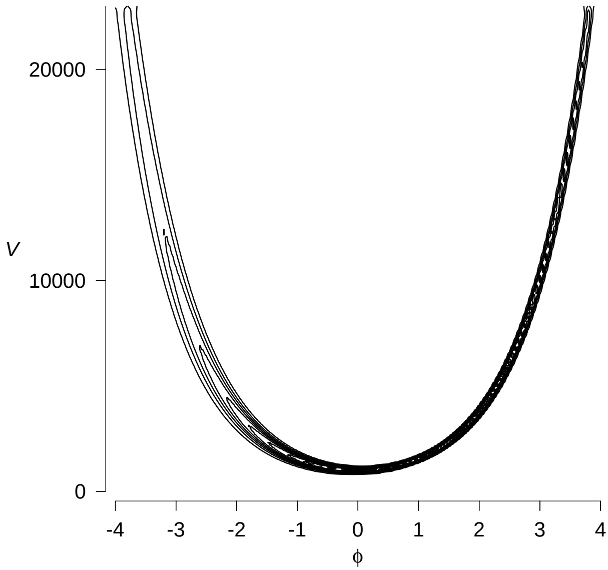


Figure 2: Contour plot of a wave function following a trajectory $\langle \hat{V} \rangle(\phi)$ of volume (vertical) as a function of the free scalar matter ϕ (horizontal) which can be used as a clock variable. The volume expectation value never reaches zero (bottom edge) but instead bounces and avoids the classical singularity. The spread of the state demonstrates fluctuations, which generically differ before and after the bounce. Moreover, the size of fluctuations before the bounce very sensitively depends on initial values, and would thus be nearly impossible to determine observationally [35, 36].

simple solvable system. The expectation value $\langle \hat{V} \rangle$ bounces, i.e. has a positive lower bound, in every state which is semiclassical at least at one time. This clearly illustrates how the singularity is avoided in this model. Moreover, the expectation value is symmetric around the bounce point, with the pre-bounce collapse simply being the time reverse of the post-bounce expansion. Also quantum fluctuations of $\hat{V}$ in a generic coherent state (i.e., near saturation of the uncertainty relation) are indicated by the spread in Fig. 2, for which quite a different behavior is realized: While states with symmetric fluctuations around the bounce do exist, they are very special. Generic coherent states are asymmetric, implying different quantum properties before and after the bounce. One can explicitly see that the ratio of fluctuations (or similarly higher moments) before and after the bounce is very sensitive to small changes in the chosen state such as the precise squeezing [36]. Observations at one side of the bounce could not be precise enough to determine the precise form of the wave function of the universe, making it impossible to determine completely what the state of the universe was before the big bang [35].

Most importantly for future developments, a solvable model of this type can, just as the harmonic oscillator in particle physics, be used as zeroth order of systematic perturbation expansions. This has been started for self-interacting or massive matter in [37]. Completion of this program will result in effective equations for cosmological structure formation which can be used to determine the potential of cosmological observations of quantum gravity effects at least in an indirect manner.

Conclusions

Loop quantum gravity establishes a discrete fundamental picture of space based on a well-defined canonical quantization of gravity. Its resulting space-time dynamics is more complicated but can be analyzed in cosmological models. Then, one obtains a dynamical equation for quantum states which is free of big bang singularities. Thus, there is a direct route from a discrete space to non-singular evolution and time before the big bang.

The absence of a singularity is realized independently of the precise form of matter, but details do play a role in the transition. In particular orientation reversal takes place at the classical singularity. Thus, parity violating matter affects how pre- and post-big bang physics connect. In addition, quantum effects such as the spreading of states can imply further differences between pre- and post-big bang states.

Parity violating matter is yet to be put into equations, where especially the torsion coupling in a first order theory of gravity is to be taken into account. (See [25, 38, 39] for canonical descriptions.) Effective equations allow a more straightforward analysis and show the possibility of parity asymmetric solutions even from gravity itself: a state spreads during evolution, making fluctuations in general non-symmetric [33]. Surprisingly, the asymmetry depends on initial values of the state in a highly sensitive way, presenting a form of cosmic forgetfulness which shrouds any precise statements one could make about the quantum state before the big bang [35].

Acknowledgements

This work was supported in part by NSF grant PHY06-53127.

References

[1] B.S. DeWitt, *Quantum Theory of Gravity. I. The Canonical Theory*, Phys. Rev. **160** (1967) 1113–1148.
[2] C. Rovelli, *Quantum Gravity*, Cambridge University Press (Cambridge, UK) 2004.
[3] A. Ashtekar and J. Lewandowski, *Background independent quantum gravity: A status report*, Class. Quantum Grav. **21** (2004) R53–R152, [arχiv:gr-qc/0404018].
[4] T. Thiemann, *Introduction to Modern Canonical Quantum General Relativity*, [arχiv:gr-qc/0110034].
[5] A. Ashtekar, *New Hamiltonian Formulation of General Relativity*, Phys. Rev. D **36** (1987) 1587–1602.
[6] J.F. Barbero G., *Real Ashtekar Variables for Lorentzian Signature Space-Times*, Phys. Rev. D **51** (1995) 5507–5510, [arχiv:gr-qc/9410014].
[7] G. Immirzi, *Real and Complex Connections for Canonical Gravity*, Class. Quantum Grav. **14** (1997) L177–L181.
[8] C. Rovelli and L. Smolin, *Loop Space Representation of Quantum General Relativity*, Nucl. Phys. B **331** (1990) 80–152.
[9] C. Rovelli and L. Smolin, *Discreteness of Area and Volume in Quantum Gravity*, Nucl. Phys. B **442** (1995) 593–619, [arχiv:gr-qc/9411005]; Erratum: Nucl. Phys. B **456** (1995) 753.
[10] A. Ashtekar and J. Lewandowski, *Quantum Theory of Geometry II: Volume Operators*, Adv. Theor. Math. Phys. **1** (1997) 388–429, [arχiv:gr-qc/9711031].
[11] M. Bojowald, *Loop Quantum Cosmology IV: Discrete Time Evolution*, Class. Quantum Grav. **18** (2001) 1071–1088, [arχiv:gr-qc/0008053].
[12] A. Ashtekar, J.C. Baez, A. Corichi and K. Krasnov, *Quantum Geometry and Black Hole Entropy*, Phys. Rev. Lett. **80** (1998) 904–907, [arχiv:gr-qc/9710007].
[13] A. Ashtekar, J.C. Baez and K. Krasnov, *Quantum Geometry of Isolated Horizons and Black Hole Entropy*, Adv. Theor. Math. Phys. **4** (2000) 1–94, [arχiv:gr-qc/0005126].
[14] M. Domagala and J. Lewandowski, *Black hole entropy from Quantum Geometry*, Class. Quantum Grav. **21** (2004) 5233–5243, [arχiv:gr-qc/0407051].
[15] K.A. Meissner, *Black hole entropy in Loop Quantum Gravity*, Class. Quantum Grav. **21** (2004) 5245–5251, [arχiv:gr-qc/0407052].
[16] M. Bojowald, *Loop quantum cosmology and inhomogeneities*, Gen. Rel. Grav. **38** (2006) 1771–1795, [arχiv:gr-qc/0609034].
[17] M. Bojowald, *The dark side of a patchwork universe*, Gen. Rel. Grav. (2007) to appear, [arχiv:0705.4398].
[18] A. Ashtekar and J. Lewandowski, *Quantum Theory of Geometry I: Area Operators*, Class. Quantum Grav. **14** (1997) A55–A82, [arχiv:gr-qc/9602046].
[19] M. Bojowald, *Quantum gravity and cosmological observations*, AIP Conf. Proc. **917** (2007) 130–137, [arχiv:gr-qc/0701142].

[20] M. Bojowald, H. Hernández, M. Kagan, P. Singh and A. Skirzewski, *Formation and evolution of structure in loop cosmology*, Phys. Rev. Lett. **98** (2007) 031301, [arχiv:astro-ph/0611685].
[21] M. Bojowald, *Loop Quantum Cosmology*, Living Rev. Relativity **8** (2005) 11, [arχiv:gr-qc/0601085]; `http://relativity.livingreviews.org/Articles/lrr-2005-11/`
[22] M. Bojowald, *Isotropic Loop Quantum Cosmology*, Class. Quantum Grav. **19** (2002) 2717–2741, [arχiv:gr-qc/0202077].
[23] A. Ashtekar, M. Bojowald and J. Lewandowski, *Mathematical structure of loop quantum cosmology*, Adv. Theor. Math. Phys. **7** (2003) 233–268, [arχiv:gr-qc/0304074].
[24] M. Bojowald and H.A. Kastrup, *Symmetry Reduction for Quantized Diffeomorphism Invariant Theories of Connections*, Class. Quantum Grav. **17** (2000) 3009–3043, [arχiv:hep-th/9907042].
[25] T. Thiemann, *QSD V: Quantum Gravity as the Natural Regulator of Matter Quantum Field Theories*, Class. Quantum Grav. **15** (1998) 1281–1314, [arχiv:gr-qc/9705019].
[26] M. Bojowald, *Absence of a Singularity in Loop Quantum Cosmology*, Phys. Rev. Lett. **86** (2001) 5227–5230, [arχiv:gr-qc/0102069].
[27] M. Bojowald, *Singularities and Quantum Gravity*, AIP Conf. Proc. **910** (2007) 294–333, [arχiv:gr-qc/0702144].
[28] M. Bojowald, *Inverse Scale Factor in Isotropic Quantum Geometry*, Phys. Rev. D **64** (2001) 084018, [arχiv:gr-qc/0105067].
[29] M. Bojowald, *Degenerate Configurations, Singularities and the Non-Abelian Nature of Loop Quantum Gravity*, Class. Quantum Grav. **23** (2006) 987–1008, [arχiv:gr-qc/0508118].
[30] M. Bojowald and M. Kagan, *Singularities in Isotropic Non-Minimal Scalar Field Models*, Class. Quantum Grav. **23** (2006) 4983–4990, [arχiv:gr-qc/0604105].
[31] M. Bojowald and A. Skirzewski, *Effective Equations of Motion for Quantum Systems*, Rev. Math. Phys. **18** (2006) 713–745, [arχiv:math-ph/0511043].
[32] M. Bojowald and A. Skirzewski, *Quantum Gravity and Higher Curvature Actions*, Int. J. Geom. Meth. Mod. Phys. **4** (2007) 25–52, [arχiv:hep-th/0606232].
[33] M. Bojowald, *Large scale effective theory for cosmological bounces*, Phys. Rev. D **75** (2007) 081301(R), [arχiv:gr-qc/0608100].
[34] M. Bojowald, *Dynamical coherent states and physical solutions of quantum cosmological bounces*, Phys. Rev. D **75** (2007) 123512, [arχiv:gr-qc/0703144].
[35] M. Bojowald, *What happened before the big bang?*, Nature Physics **3** (2007) 523–525.
[36] M. Bojowald, *Harmonic cosmology: How much can we know about a universe before the big bang?*, [arχiv:0710.4919].
[37] M. Bojowald, H. Hernández and A. Skirzewski, *Effective equations for isotropic quantum cosmology including matter*, Phys. Rev. D **76** (2007) 063511, [arχiv:0706.1057].
[38] S. Mercuri, *Fermions in Ashtekar-Barbero Connections Formalism for Arbitrary Values of the Immirzi Parameter*, Phys. Rev. D **73** (2006) 084016, [arχiv:gr-qc/0601013].
[39] M. Bojowald and R. Das, *Canonical gravity with fermions*, [arχiv:0710.5722].

Time and Matter
2007 Bled, Slovenia

Expectation and Challenges from Future CMB Science

CARLO BACCIGALUPI*
SISSA/ISAS, Via Beirut 4, I-34014 Trieste, Italy

Abstract: We review the cosmological concordance model and the most important observations which are consistent with its predictions. We focus on the status of Cosmic Microwave Background (CMB) anisotropy observations, describing how the traces of processes occurred in the early universe are stored, and pointing out those w hich still evade our knowledge, as well as the most important expectations and obstac les concerning forthcoming and future experiments.

Introduction

The term "precision cosmology" has finally come to a common use among cosmologists and theoretical physicists in general, meaning that observations have being progressing so much in the recent years to allow the measure of the main cosmological observables with percent precision. Those observations concern the abundance of primordial elements in the early universe, as well as the early and late stage of the evolution of cosmological structures, as seen in the image brought to us by the Cosmic Microwave Background (CMB) and in the present distrubution of large scale structures, respectively. The technology, data acquisition, reduction and interpretation in modern cosmology are without any doubt one of the main achievements of this century for physics as a whole. On the other hand, still it has to be remembered that the extraction of physical information from cosmological measurements is extremely complex even in the sort of easy context provided by the linearization of general relativity, i.e. small deviations around the homogeneous and isotropic picture, known as Friedmann Robertson Wolker (FRW) spacetime. Thus, the interpretation of a given experiment in cosmology relies on a variety of hypotheses and assumptions, and often does depend on complementary data, so that

* `bacci@sissa.it`

practically any claim in cosmology is possible only by combining different datasets. On the other hand, the variety and independence of the main cosmological observables make us confident that the picture we have reached is solid, to the point of being called "cosmological concordance model", although from the point of view of its theoretical understanding it poses unanswered and outstanding questions for physics as a whole. In the next section we review the basic properties of the cosmological concordance model; in section 3 we point out the role that the CMB is having in its estabilishment, pointing out the future expectations and challenges posed by the next generation experiments trying to progress in the unveiling of this observable. In section 4 we make some concluding remarks.

Cosmological concordance model

From the point of view of a pure description, i.e. giving up our ambition to understand it theoretically, the present cosmological picture is relatively simple. The expansion is described by assuming an highly symmetric spacetime on large scales, isotropic and homogeneous as described by the FRW metric, say larger than the biggest cosmological structures, galaxy clusters, extending over tens of Mpc. Other than that, the cosmological concordance model is specified by three main characters: perturbations, components and geometry.

The description of perturbations is supported and simplified by the evidence of small perturbations with respect to the mean temperature of the CMB photons; as it is well known, fluctuations are of the order of one part over 10^5 with respect to the mean 2.7 K, over the whole sky and down to the smallest angular scale probed so far. This evidence supports the assumption that in their mathematical description only small deviations from homogeneity and isotropy are allowed [1]. Those perturbations are supposed to originate in a phase of quasi-exponential expansion in the very early universe, the inflation, driven by a non-zero vacuum energy similar to a cosmological constant provided by the potential energy of one or more fundamental fields: quantum vacuum states on microscopic scales are stretched by the expansion itself, approaching the scale corresponding to the Hubble expansion rate, where they appear as perturbations, via a mechanism which is simular to the Hawking process for black hole evaporation, and related to the non-invariance of the vacuum for quantum field in curved spacetimes [2]. Other than that, linear cosmological perturbations are specified by a distribution, which the quantum nature within the inflationary picture predicts to be Gaussian, i.e. specified only by the two points correlation function, which is conveniently described in terms of their power spectrum in the Fourier domain. Since inflation is essentially specified by a single scale which sets the amount of vacuum energy responsible for the expansion itself, it predicts a sort of white spectrum, also known as scale invariant or Harrison Zel'dovich, i.e. with the same power on all scales; more precisely, the perturbation variance in the energy density on a given scale when the latter equals the Hubble expansion rate is a constant [3].

At least three different cosmological components are required. Relativistic matter is made by photons and neutrinos, while non-relativistic species, forming galaxies and their clusters is fundamentally divided into two subclasses. The matter belonging to the standard model of particle physics, dominated in mass by baryons and constituting the luminous part of cosmological structures, and the Cold Dark Matter (CDM), interacting at most weakly with the rest, and characterized by a kinetic energy which is negligible with respect to the mass, forming the dark haloes around galaxies and galaxy clusters. Recently, a third, and actually dominant cosmological component has been discovered and added to the picture, as some kind of vacuum or dark energy similar to the one driving inflation and imprinting an acceleration to the cosmic expansion during the last couple of billion years. The simplest description, also consistent with the present data, is the cosmological constant; this is not the first time that such a term was considered and studied: it was actually appearing and disappearing from the Einstein equations several times in this century, initially introduced by Einstein himself, wrongly asking for stationarity of the universe, and later by quantum mechanical arguments, predicting it to correspond to the value obtained combining fundamental constants, which is about 123 orders of magnitude larger than the cosmological density today; caused by our incapability of explaining this discrepancy, this issue is known since tens of years as the cosmological constant problem, probably related to our ignorance in describing the vacuum in physics. Although the cosmological constant is a viable candidate for the dark energy, it cannot be for inflation, as the latter has to finish and decay in order to allow structures to form and grow, and therefore cannot be constant; this brought cosmologists to guess that the dark energy responsible for the acceleration today may be similarly dynamical, and most of the cosmological probes today are oriented to constrain such a dynamics. In terms of relative weight, baryons and leptons constitute about 4 % of the total cosmological density, dark matter about 20 %, while the dark energy is at the level of 76 %.

Let us come to geometry now. That is also most simple, in the sense that inflation predicts it to be zero within errors, and indeed so far that is the case in the cosmological observations. Simply, the curvature term in the

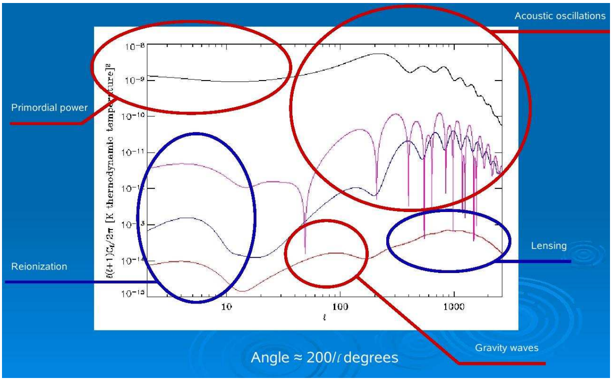


Figure 1: A representation of the angular power spectrum of CMB anisotropies as predicted by the cosmological concordance model, as a function of the angular scale at which it is measured. Also, the main processes activating the different regions of the spectrum are highlighted.

Friedmann equation describing the expansion gets more and more irrelevant as the inflationary expansion proceeds, as it happens in any physical system being stretched in size. As there is no reason to consider a particularly short amount of inflation, the generic prediction is that the expansion lasts enough to make the curvature irrelevant.

Observationally, several independent observables support the cosmological concordance model. The Abundance of light elements, which is the earliest observable we have at the moment, probing the Nucleosynthesis epoch through the abundance of light elements measured today in nearby structures [4]. The CMB anisotropies, subject of the next section, culminating with the observations from the Wilkinson Microwave Anisotropy (WMAP) probe [5]. The Large Scale Structure, reconstructing the pattern of cosmological perturbations at present via large redshift surveys made of three dimensional maps of millions of galaxies, extending in a volume which measures several hundreds of Mpc [6, 7, 8], as well as from weak distortion that background light undergoes due to the lensing generated by forming cosmological structures [9, 10]. The Type Ia supernovae, originated in a star binary system in which the actual explosion threshold of the supernova is reached gradually due to the accretion from the companion star, and thus known to possess an almost constant luminosity, regardless of the enviroment or epoch in which they explode, probing the cosmic geometry from their location to us [11].

Cosmic microwave background

Photons decouple from the rest of the system at a temperature of about 3000 K, corresponding to an age of the universe of about 300 000 years. They bring to us the records of cosmological perturbations at that epoch, in the form of anisotropies in their black body distribution. The latter possess temperature fluctuations, as well as linear polarization, since the Thomson scattering onto free electrons at the decoupling epoch is able to transform an anisotropic incident radiation into an outgoing lineary polarized wave, characterized by its total intensity T and Stokes parameters Q and U, function of the direction in the sky. Anisotropies are sensitive to all kinds of cosmological perturbations, which according to general relativity may be scalar, like density perturbations, vector, like vorticity, and tensor, equivalent to gravitational waves [1]. The Stokes parameters may be combined together into a gradient or E component, affected by all kinds like the intensity T, and a curl one, also labeled as B, sensitive to vectors and tensors only [12, 13]. Since the growth of vectors is suppressed by the cosmological expansion, detecting a B component would be a strong indication of the existence of gravitational waves of cosmological origin. The cosmological information from the CMB are usually extracted in the harmonic domain, i.e. considering the coefficients of harmonic expansion on the sphere, $a_{\ell m}^{T,E,B}$, and compressing them in order to reduce the effects of noise and systematics, in the quantities

$$C_\ell^{XY} = \frac{1}{2\ell+1}\sum_m a_{\ell m}^X (a_{\ell m}^Y)^*, \qquad X, Y = T, E, B. \tag{1}$$

The multipole ℓ is conveniently expressed in terms of the angular scale θ at which it probes the anisotropy power, i.e. $\theta \simeq 200/\ell$ degrees. In figure 1 the prediction of the C_ℓ distribution from the cosmological concordance model is shown, together with the main cosmological processes activating the different parts of the spectrum. From top to bottom, the curves represent C_ℓ^{TT}, C_ℓ^{TE}, C_ℓ^{EE} and C_ℓ^{BB}; the correlation between T and E is so strong because the E polarization modes are activated by the same density fluctuations at decoupling, which dominate the T modes. On large angular scales, the TT spectrum is dominated essentially by the unperturbed perturbations generated in the early universe, generating the long plateau on $\ell \leq 100$. On sub-degree angular scales, corresponding to scales lower than the sound speed for photons at decoupling, a series of oscillations is observed, the so called acoustic peaks; those are activated by photons tightly coupled to charged particles via Thomson scattering falling into the potential wells provided by the dark matter, and bouncing back due to their own pressure; therefore, they are most sensitive to the amount of baryonic matter. The angular scales corresponding to the transition between the plateau and the acoustic oscillations, marked by the peak at $\ell \simeq 200$ is one of the most important standard ruler in cosmology; its physical scale corresponds to the CMB sound horizon at decoupling, when the effects of curvature or dark energy are negligible; the angle it subtends depends therefore on the distance from last scattering, which is sensitive precisely to curvature or dark energy effects on the cosmic expansion at low redshifts. The angular scales at which the first peak is observed, together with the type Ia supernovae and large scale structure data, is the basis of the evidence for the acceleration in the cosmic expansion. Acoustic oscillations also occur in the component of the polarization which is correlated with T, i.e. the EE and TE spectra, because of the strong correlation between the two processes. Since polarization is caused by CMB photons hitting a last scatterer electron with anisotropic intensity, the polarization exists only on scales corresponding to the mean free path at last scattering, i.e. less than one degree in the sky. On the other hand, large bumps are visible on the EE and BB spectra on multipoles corresponding to tens of degrees in the sky; the reason is the re-scattering process of CMB photons onto electrons coming from the rionization process occurring at redshift of about 10. The induced polarization peaks on the scales subtended by the photon free path at the corresponding epoch, which is sensibly larger, explaining why the reionization bump occurs at such small multipoles. A leakage of EE modes in BB due to gravitational lensing is well apparent in the broad peak in the latter spectra, resembling the EE with smaller amplitude; indeed, even a pure E mode pattern of CMB anisotropy, passing through forming cosmological structures and getting deflected via gravitational lensing, get distorted acquiring a B component, as a result of the incoherence of the underlying lensing structures, see [14] and references therein. Finally, a tensor component with 10 % amplitude with respect to the scalar one has been added and manifests as peak at a multipole of about 100 in the BB spectrum. This is the imprint of the polarization anisotropies caused by the corresponding gravitational waves at last scattering; the latter are massless, diffusing as radiation on sub-degree angular scales, as their oscillations are not supported by interaction with any other component.

At the present, the main cosmological consequences are derived from the knowledge of the TT and TE spectrum obtained up to about $\ell \simeq 500$ from satellite measurements, and up to $\ell \simeq 2000$ in TT plus data on EE from sub-orbital experiments, see [5], [15] and references therein. A large fraction of the spectrum, including the EE and BB modes, remains hidden to our knowledge. In addition, almost nothing is known about the remaining

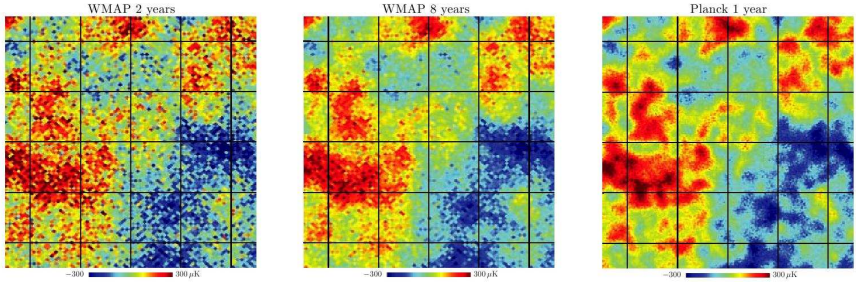


Figure 2: A representation of the improvement expected from Planck with respect to WMAP over limited regions of the sky, in terms of instrumental sensitivity. The image is taken from the Planck blue book available at www.rssd.esa.int/Planck.

statistical moments of the distribution of CMB anisotropies. Forthcoming probes promise to reduce our lack of knowledge on CMB anisotropies. The Planck satellite[1] will be launched in late 2008, and will produce all sky maps at 9 frequency channels, extending between 30 and 857 GHz, with unprecedented angular sensitivity and resolution, reaching a signal to noise ratio of a few on angular scales of a few arcminutes. The Planck expectations concerning the measure of the CMB angular power spectrum are shown in figur 3. Planck promises to be an imager, i.e. not limiting its insights into the two point correlation function, but going to the higher order statistics, detecting the distortion from gravitational lensing [9], as well as a possible primordial non-Gaussianity. A glimpse of the progress which is expected from Planck concerning the observation of the pattern of CMB anisotropies is shown in figure 2. In addition, a number of sub-orbital probes[2] are ongoing or planned to look for polarization CMB anisotropies on sub-degree angular scales and on low foreground regions of the sky, targeting the E modes as well as having the potential sensitivity to detect the B modes from cosmological gravitational waves if they are at least a few percent of the scalar amplitude; figure 4 reveals the expectations of one of those, the E and B experiment (EBEx[3]), a balloon borne probe scheduled for the North-American flight within 2008 [16]: the figure also reveals one of the possible ultimate limitation to CMB observations, represented by the diffuse emission from our own Galaxy, i.e. the synchrotron radiation emitted by electrons spiralizing around the Galactic magnetic field, as

[1] `www.rssd.esa.int/Planck`

[2] `lambda.gsfc.nasa.gov`

[3] `groups.physics.umn.edu/cosmology/ebex`

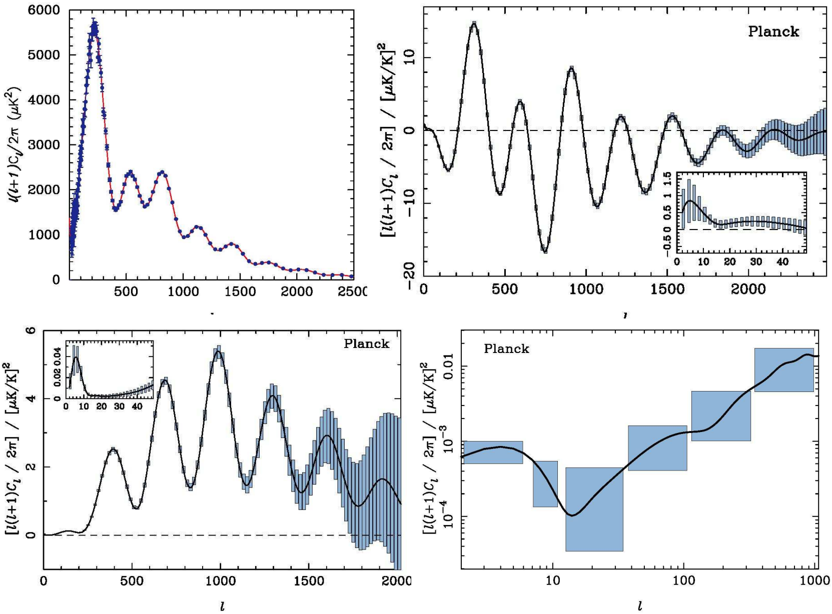


Figure 3: The forecasted Planck sensitivity, represented in 1σ bins in blue, for the TT (top left), TE (top right), EE (bottom left) and BB (bottom right) power spectrum observables. The images are taken from the Planck blue book available at `www.rssd.esa.int/Planck`.

well as the thermal dust emission from magnetized dust grains. The latter emission even in the frequency band of its minimum emission, might be comparable to the cosmological signal across the whole sky [17]. Robust data analysis techniques have to be designed in order to achieve a convincing foreground cleaning before any claim for the detection of B modes from primordial gravitational waves can be made [18]. If these observations are successful, in the control of instrumental systematics as well as the removal of the foreground emission, a way toward a post-Planck CMB satellite may open.

Concluding remarks

We briefly outlined the cosmological concordance model and the main experimental evidences on which it is based. The model is at odd with known physics within the standard model of particle physics, and requires sub-

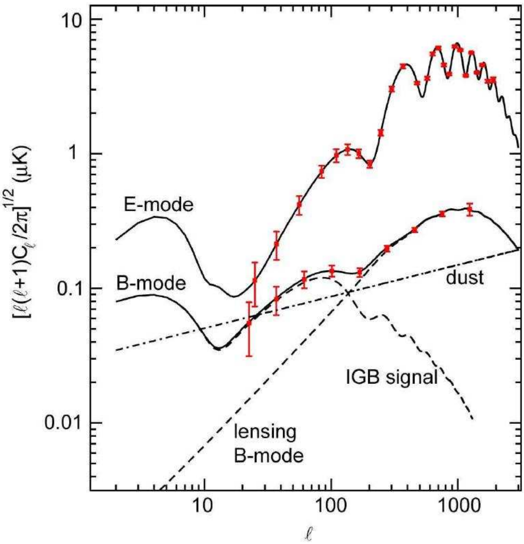


Figure 4: The forecasted performance of the EBEx experiment for the measurement of CMB polarization. The different components of the BB spectrum, namely lensing and the Inflationary Gravitational Background (IGB), are also shown, as well as the forecasted foreground emission from the Galactic dust in the sky area probed by the experiment [16].

stantial extensions. Known particles make only 4 % of the total density; the remaining part is in some form of pressureless dark matter, about 20 % of the total density, forming dark haloes around galaxies and clusters of them, and about 76 % of a dark energy component which is causing the expansion to accelerate, like a cosmological constant in its simplest form, although no hint about how to relate its energy scales with the fundamental ones exists. Primordial perturbations are Gaussian and with almost equal power on all scales, seeded by some form of accelerated expansion occurring in the early universe as well, which we describe semi-classically by means of a scalar field, the inflaton, driving and inflationary expansion with its potential energy, eventually decaying in matter and radiation. The cosmological concordance model is made credible and trusted by the community due to the marked independence of the observations leading to its estabilishment. Those are the measured abundances of primordial elements, microwave background radiaton, large scale distribution of galaxies, light coming from distant supernovae traveling cosmological distances. Most of these observables are still far from revealing their full potentiality. We focused on the case of the cosmological fossil radiation, where plenty of experiments with markedly different strategies are trying to catch the tiniest imprints of the cosmological perturbations which are coded in those. This aspect, together with the compelling evidence that the cosmological concordance model, if correct, is pointing to a entirely new world for physics as a whole, puts us roughly in the middle of the road toward the understanding of the messages which are stored in the cosmological observables. Those are likely to unveil at least a fraction of their hidden signatures in the forthcoming decades, the most spectacular ones coming probably in just a few years from the next generation of satellite and sub-orbital missions for the study of the cosmic microwave background anisotropies.

References

[1] H. Kodama and M. Sasaki, *Cosmological perturbation theory*, Progr. Theor. Phys. Suppl. **78** (1984) 1.

[2] N.D. Birrel and P.C.W. Davies, *Quantum Fields in Curved Space* (Cambridge University Press, Cambridge, 1982).

[3] A.R. Liddle and D.H. Lyth, *Cosmological Inflation and Large-Scale Structure*, (Cambridge University Press, Cambridge, 2000).

[4] B.D. Fields and K.A. Olive, *Big bang nucleosynthesis*, Nucl. Phys. A **777** (2006) 208.

[5] D.N. Spergel *et.al.*, *Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology*, Astrophys. J. Suppl. **170** (2007) 377.

[6] W. Percival *et al.*, *The Shape of the Sloan Digital Sky Survey Data Release 5 Galaxy Power Spectrum*, Astrophys. J. **657** (2007) 645.

[7] S. Cole *et al.*, *The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications*, MNRAS **362** (2005) 505.

[8] D. Eisenstein *et al.*, *Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies*, Astrophys. J. **633** (2005) 560.

[9] M. Bartelmann and P. Schneider, *Weak gravitational lensing*, Phys. Rept. **340** (2001) 291.

[10] A. Refregier, *Weak Gravitational Lensing by Large-Scale Structure*, Ann. Rev. Astron. Astrophys. **41** (2003) 645.

[11] A.G. Riess *et al.*, *New Hubble Space Telescope Discoveries of Type Ia Supernovae at z >= 1: Narrowing Constraints on the Early Behavior of Dark Energy*, Astrophys. J. **659** (2007) 98.

[12] M. Zaldarriaga and U. Seljak, *All-sky analysis of polarization in the microwave background*, Phys. Rev. D **55** (1997) 1830.

[13] M. Kamionkowski, A. Kosowsky and A. Stebbins, *Statistics of cosmic microwave background polarization*, Phys. Rev. D **55** (1997) 7368.

[14] A. Lewis and A. Challinor, *Weak gravitational lensing of the CMB*, Phys. Rept. **429** (2006) 1.

[15] C.L. Reichardt *et al.*, *High resolution CMB power spectrum from the complete ACBAR data set*, submitted to Astrophys. J. (2008), `babbage.sissa.it/abs/0801.1491`.
[16] P. Oxley *et al.*, *The EBEx experiment*, Proc. SPIE Int. Soc. Opt. Eng. **5543** (2004) 320.
[17] C. Baccigalupi, *Cosmic microwave background polarisation: foreground contrast and component separation*, New Astron. Rev. **47** (2003), 1127.
[18] F. Stivoli, C. Baccigalupi, D. Maino and R. Stompor, *Separating polarized cosmological and galactic emissions for cosmic microwave background B-mode polarization experiments*, MNRAS **372** (2006) 615.

Time and Matter
2007 Bled, Slovenia

Holography and Chronology Protection

GIUSEPPE MILANESI[1] AND MARTIN O'LOUGHLIN[2]*
[1] *Institut für Theoretische Physik, ETH, Zurich, CH-8093, Switzerland.*
[2] *University of Nova Gorica, Vipavska 13, SI-5000 N. Gorica, Slovenia*

Abstract: We review the motivations for an holographic duality between type IIB string theory on the space-time $AdS_5 \times S^5$ and the $\mathcal{N} = 4$ conformal Yang-Mills theory on the boundary of the AdS_5 factor. We illustrate a precise example of this correspondence constructed by Lin, Lunin and Maldacena. We then show that a slight generalisation of the LLM solutions can lead to singular geometries and possible closed time-like curves. This enables us to relate chronology protection in the string theory to properties of operators in the dual Supersymmetric gauge theory, leading to the possibility of a holographic proof of chronology protection.

Introduction

In any theory that contains general relativity, there is the possibility that the solution space contains various space-times that present interpretational difficulties. These difficulties are related to singularities and geometries that contain closed time-like curves. Now these in turn can be divided into two subclasses depending on whether they are naked or clothed. A singularity clothed by a global horizon is commonly known as a black hole. It is true that these also produce some problems when one looks at quantum physics but at the classical level they are not so troublesome. Similarly one can have a closed time-like curve in a space-time that cannot reach out into the asymptotic regions and thus the causality violation is confined to a part of the space-time again that is behind a global horizon.

The apparently more problematic structures are naked (time-like) singularities and closed time-like curves that can pass through any point in space-time. It turns out that these two can arise simultaneously in the case that

* martin.oloughlin@p-ng.si

one has a rotating naked singularity and the classical example of this is the negative mass Kerr black hole.

We will discuss in these proceedings some insights into these singular structures of gravitational theories from the point of view of holography, a conjectured relationship between gravitational theories and non-gravitational theories in a lower space-time dimension (thus holographic realisations of the higher dimensional gravitational theory).

To study these questions in more detail we will consider a class of solutions to type IIB string theory that preserve $1/2$ of the possible supersymmetries and their known holographic dual conformal field theory (CFT) descriptions. The basic trick of [1] is to note that assuming a certain amount of symmetry in the ansatz for metric and five-form field strength, and demanding that one half of the total supersymmetry is preserved, the remaining equations of motions reduce to an elliptic equation for a scalar function z on $\mathbb{R}^2 \times \mathbb{R}^+$. The value of $\rho = 1/2 - z$ on the 2-plane boundary of $\mathbb{R}^2 \times \mathbb{R}^+$ can be interpreted as a semiclassical fermion density, thus providing direct contact to the CFT dual Yang-Mills theory on $\mathbb{R} \times S^3$ [2, 3]. Indeed if this density takes on only the values 0 and 1 then the solutions are guaranteed to be singularity free space-times.

In this proceedings we consider the most general allowed (on the supergravity side) boundary conditions for this elliptic equation. This means that we study the full set of moduli of this sector of supergravity that consists of metrics asymptotic to $\mathrm{AdS}_5 \times S^5$, with an $\mathrm{SO}(4) \times \mathrm{SO}(4)$ isometry group and preserving half of the supersymmetry of type IIB string theory. The supergravity solutions in general will be singular. The space-time singularities appearing are always naked and fall into two distinct classes: null and timelike. The null ones can be considered as seeded by a fermion density between 0 and 1 and have been studied for example in [4, 5, 6, 7, 8] together with the possible *local* quantum effects responsible for their resolution - the singularity is due to an average over configurations of N fermions in a gas with average density less than one. An individual configuration with the same asymptotics can actually be seen to have as source a collection of N giant gravitons [9, 10] separated one from the other. In the supergravity theory, the resolution of the singularity thus appears as a sort of space-time foam [11] while in the dual CFT one sees that such a configuration corresponds to the Coulomb branch of the theory. As such these null singularities are effectively not singular in the full quantum theory.

The AdS/CFT correspondence has maybe something more interesting to tell us about the fate of the timelike singularities which arise when the density ρ is greater than 1 and as such no longer has an interpretation as a fermion density. The solutions with this kind of singularity are "pathological": they have closed timelike curves passing through *any point* of the spacetime and they include unbounded from below negative mass excitations of $AdS_5 \times S^5$.

It has been conjectured, [12, 13, 14], that geometries with these features should be considered as truly unphysical via *global* considerations in the setting of a full quantum theory of gravity. The AdS/CFT correspondence applied to the space-times of [1] suggests one particular mechanism for the global removal of solutions containing timelike singularities. The deformations of the geometry which produce these singularities apparently correspond to negative dimension operators in the dual field theory. The unitarity of the representations of the superalgebra of $N = 4$ supersymmetric Yang-Mills theory [15] indicate in particular that unitary operators must have a positive conformal dimension. Our observations indicate that there should actually exist a general proof of the chronology protection conjecture [16] in this sector of supergravity. A first indication of this mechanism linking unitarity to chronology protection can be found in [17] and in the current context in [18, 19].

Extending these works, in this paper we discuss the theorem proven in [22] that closed timelike curves (CTCs) are unavoidable in any solution with a timelike singularity and that they are excluded in the case of regular and null singular solutions, these being the spacetimes that can be represented in terms of dual fermions, a result anticipated but not proven in [18]. This provides a clear division between these two classes of singular spacetimes which is also reflected in the two different mechanisms responsible for the resolution of their respective spacetime singularities.

We begin with a review of the concept of holography as motivated by 't Hooft and Susskind in discussions of black hole entropy and black hole evaporation. We then show how this led Maldacena to conjecture a precise realisation of this holography concept relating string theory on a spacetime that includes an anti-de Sitter component to a conformal field theory on the boundary of that space-time.

We then look more closely at a particular class of solutions found by Lin, Lunin and Maldacena [1] and we show the most general allowed boundary conditions for a supergravity solution satisfying the symmetry requirements. We clarify the role of the boundary conditions in determining the mass of the solution in the asymptotic $AdS_5 \times S^5$ and we show the relation between the boundary conditions and the appearance of spacetime singularities.

We review the theorem of [22] regarding the appearance and the properties of CTCs and their genericity for the classes of solutions that contain timelike singularities. Moreover we an overview of the proof of the theorem which relates the appearance of CTCs to the boundary conditions responsible for these time-like singularities.

In the final section we return to a discussion of the meaning of these results, and in particular the possibility of proving the chronology protection conjecture for this class of geometries, by showing that the AdS/CFT correspondence relates chronology protection to unitarity in the CFT.

Holography and quantum black holes

The possibility that quantum gravity may have an holographic description was raised by 't Hooft in his construction of models for quantum gravity based on the properties of black holes and the demand that space-time physics, even with the inclusion of quantum gravity, is unitary. The hypothesis was then formulated in a more precise manner by Susskind, as we will now discuss. An obvious problem arises in this context when one considers the possibility of black hole evaporation which is a natural consequence of the coupling of quantum fields to the black hole background.

We will begin therefore with a discussion of black-hole thermodynamics which will lead us to a bound on the information (as characterised by entropy) that can be contained in a given volume in space-time.

Classical black holes and their "almost" thermodynamics.

For a spacetime with a Black Hole there is a relationship between the change in Mass, the surface gravity κ and the horizon area A, which is in direct anology with the First Law of Thermodynamics upon identification of κ with the temperature and the horizon area A with entropy,

$$\mathrm{d}M = \frac{\kappa}{8\pi}\,\mathrm{d}A + \Omega\,\mathrm{d}J \tag{1}$$

where Ω is the angular velocity and J the angular momentum. This is a completely general relation that can be derived by calculating the change in the mass and angular momentum of a black hole with given asymptotic charges M and J and the dependence of these variations on the horizon area. Comparing to the First Law of Thermodynamics (with a chemical potential μ)

$$\mathrm{d}E = T\,\mathrm{d}S + \mu\,\mathrm{d}N \tag{2}$$

leads to the identification for black holes,

$$E = M, \quad \mu = \Omega, \quad N = J \tag{3}$$

and the observation that the entropy is proportional to the area of the horizon. However, without some additional information, this analogy is incomplete because there is no classical sense in which the black hole can have a non-zero temperature and there are no "classical" microscopic degrees of freedom that could lead to a statistical entropy. This is where one needs to pass to quantum fields in the black hole background and in particular to the calculation of Hawking radiation.

Quantum fields in black hole - Hawking radiation

Hawking first calculated black-body radiation from a black hole by considering quantum fields in a background space-time that corresponds to a star collapsing to form a black hole [23]. The correct treatment of the boundary conditions on quantum fields in this background geometry, in particular with respect to the definition of vacuum in the past and future reveals that a past vacuum with collapsing star evolves to a black-hole with a flux of radiation corresponding to that of a black body at the temperature

$$T = \frac{\kappa}{2\pi}, \tag{4}$$

where again κ is the surface gravity of the black hole and is actually proportional to the inverse of the mass for a Schwarzschild black hole. This allows us to put another piece in the puzzle of the First Law of Black Hole Thermodynamics giving the ratio between horizon area and entropy,

$$S = \frac{A}{4} \tag{5}$$

This completes the analogy between the first law of thermodynamics and black hole mechanics. To truly identify the horizon area with entropy however we should also demand that it satisfies the second law of thermodynamics, that entropy always increases or remains constant in any physical process. Furthermore, this picture would still not be completely satisfactory without a means of identifying the microscopic degrees of freedom that will give us a true statistical interpretation of the black hole entropy. The completion of this picture is the subject of the next two subsections.

Entropy and information

One can prove geometrically that the total horizon area of a collection of black holes increases in any classical evolution and since we have just found an identification between the area of a black hole and its thermodynamic entropy we thus have a Second Law of Black Hole mechanics [25]. However one immmediately realises that this cannot be the full story as *quantum* processes like Hawking radiation *violate* this principle. A black hole evaporates due to the stream of black body radiation flowing from it. This will cause the black hole to lose energy and subsequently mass. The temperature is inversely proportional to the mass and the evaporation thus accelerates as the black hole gets smaller. The total evaporation time can be calculated and is proportional to the cube root of the initial black hole mass.

Clearly this is problematic for the interpretation of the entropy as it violates the Second Law of Thermodynamics which states that in every physical process entropy remains constant or increases.

If we consider however the *total* entropy of black hole plus environment then one can actually prove the Generalised Second Law of Black Hole Mechanics [26]. More precisely it can be shown that including quantum processes of particle creation in a gravitational field the total entropy of fields plus black hole

$$S_{\mathrm{T}} = S_{\mathrm{bh}} + S_{\mathrm{fields}} = \sum \frac{A}{4} + S_{\mathrm{fields}} \tag{6}$$

does indeed satisfy the second law,

$$\Delta S_{\mathrm{T}} \geq 0 \tag{7}$$

and thus the analogy between the laws of black hole mechanics and those of thermodynamics is complete.

However, the result is surprising from a statistical mechanics perspective. Entropy is usually an extensive quantity, related to information contained in a physical volume. The laws of black hole mechanics seem to indicate that this is no longer the case and that the entropy that is contained inside a surface is bounded by the area of that surface not by the volume inside that surface. There is a non-locality in the distribution of information in a theory that includes gravitational degrees of freedom.

Information and evolution

What are the microscopic degrees of freedom which could lead to a statistical picture of black hole entropy? Related to this question there is another problem that arises from our above considerations and one that was immediately noted by Hawking after completing the calculation of black hole evaporation. This is the so called Information Loss Problem of black hole physics and the reasoning that leads to this problem is based on the fact that Hawking radiation is thermal which leads to a quantum evolution involving a density matrix and subsequent loss of coherence. The space-time containing an evaporating black hole in a semi-classical treatment of quantum gravity is Cauchy incomplete. The conclusion that pure states must evolve to a density matrix is unavoidable unless there is some type of non-locality in the description of degrees of freedom in the gravitational theory. This puzzle is commonly known as the Black Hole Information Loss Paradox. That non-locality may be needed to fully understand quantum gravity was already noted at the end of the previous section and a more precise formulation of this non-locality can be found in the Holography Conjecture.

Holography conjecture, ('t Hooft [30], Susskind [29]).

- As a consequence of the applicability of the laws of classical thermodynamics to black-hole space-times, and in particular the generalised second law, for a volume in space-time the true degrees of freedom
  - are bounded by the area surrounding this volume with this bound saturated when a black hole forms inside the volume with horizon corresponding to the boundary
  - It was further conjectured [29], by analogy with the physics of D-branes that in string theory there should be a precise realisation of these degrees of freedom as corresponding to black hole microstates

Thus there should be a holographic realisation of degrees of freedom in theories that include gravitation.

AdS/CFT, a concrete realisation of Holography

In this section we will give an overview of the realisation of the Holography conjecture in string theory via the AdS/CFT correspondence. For more details one should read the excellent review article [27] and the original references cited there.

The first step towards a more concrete realisation of a holographic theory was made in the study of properties of Dirichlet-branes in string theory. Dirichlet branes are submanifolds of space-time on which string ends may attach themselves. In a standard closed string theory this means that we can also find open string states but only when the ends of the open string are attached to a Dirichlet brane. The simplest example of such a brane is the D3-brane, the world-volume of which is $3+1$–dimensional Minkowski space.

The first important step towards the construction of a holographic description of a gravitational theory was made when some very interesting simplifications were observed in the calculation of string scattering from a D3-brane. In particular such a scattering can be represented holographically in a Super Yang-Mills theory on the D3-brane world-volume this in turn being the limit of a Dirac-Born-Infeld model. It was found that the grey body factor for Hawking radiation from a black D-brane (a higher-dimensional generalisation of a black hole), can be obtained from the near horizon geometry of the D3-brane or via a *dual* calculation in the world volume Yang-Mills theory. The dual calculation involves the decay of a massive particle into a pair of gauge bosons.

The greybody factor is the factor σ_{abs} in the formula for radiation from a black hole

$$\mathrm{d}\Gamma_{\text{emit}} = \frac{\sigma_{\text{abs}}}{\exp(\omega/T) \pm 1} \frac{\mathrm{d}^n k}{2\pi^n} \tag{8}$$

and represents the deviation of the radiation from a pure black body spectrum, due to the presence of non-trivial fields in the vicinity of a black hole.

The observation that the greybody factor can be calculated in the gauge theory on the D3-brane and equivalently by a calculation in the geometry produced by the D3-brane backreaction in spacetime, and the fact that furthermore only the near horizon geometry is important in this calculation led to a duality conjecture between string theory in the near-horizon spacetime and the quantum field theory of the four-dimensional gauge theory.

Near-Horizon limit

The metric for the extremal D3-brane is obtained by solving the equations for type IIB supergravity. The metric has Lorentz symmetry in a $3+1$–dimensional submanifold and in the transverse space to the brane one finds a six-dimensional flat space which we will write in polar coordinates.

$$ds^2 = f^{-1/2}(-dt^2 + d\mathbf{x}^2) + f^{1/2}(dr^2 + r^2 d\Omega_5^2) \tag{9}$$

where

$$F_5 = (1 + *)dt \wedge dx^1 \wedge dx^2 \wedge dx^3 \wedge df^{-1} \tag{10}$$

$$f = 1 + \frac{R^4}{r^4}, \qquad R^4 = 4\pi g_s \alpha'^2 N \tag{11}$$

and the functional dependence of $f(r)$ can easily be found by plugging the above ansatz into the equations of motion. In the near horizon limit (which corresponds in these coordinates to $r \ll R$) this geometry becomes that of $AdS_5 \times S^5$. This is easily seen by noting that in this limit

$$f = \frac{R^4}{r^4}, \qquad ds^2 = \frac{r^2}{R^2}(-dt^2 + d\mathbf{x}^2) + \frac{R^2}{r^2}(dr^2 + r^2 d\Omega_5^2). \tag{12}$$

which is precisely the metric of an $AdS_5 \times S^5$ space-time with radius R.

Explicit mapping of AdS/CFT

A simple example of the holographic mapping is the coupling constant of the Yang-Mills theory which is related to the string coupling constant $g_s = \exp\phi$ by $4\pi g_s = g_{YM}^2$. Therefore the marginal operator in the gauge theory that corresponds to changing the coupling constant is related to the boundary conditions on the dilaton field. The precise relation was conjectured to be

$$\langle \exp \int \phi_0(\mathbf{x})\mathcal{O}(\mathbf{x})\rangle_{CFT} = \mathcal{Z}[\phi(\mathbf{x}, z)|_{z=0} = \phi_0(\mathbf{x})] \tag{13}$$

and this conjecture and generalisations to other fields has been supported by a large number of calculations the details of which can be found in [27].

One half supersymmetric holography

In the beautiful paper of Lin, Lunin and Maldacena [1] a full holographic mapping for 1/2 supersymmetric fields dual to non-singular geometries was constructed. It was found that these correspond, via a simple construction, to the class of marginal operators that are annihilated by half of the supersymmetry generators. In slightly more detail the correspondence of LLM is constructed in the following way.

Asymptotically $AdS_5 \times S^5$ geometries with 1/2 supersymmetry and respecting an $SO(4) \times SO(4)$ isometry can be found explicitly due to the power of the preserved supersymmetries. The most general ansatz for such a metric is

$$\begin{aligned} \mathrm{d}s^2 = & -\frac{1}{h^2}(\mathrm{d}t + \mathbf{V}\cdot \mathrm{d}\mathbf{x})^2 + h^2(\mathrm{d}y^2 + \mathrm{d}\mathbf{x}^2) + \\ & + y\,\mathrm{e}^G\,\mathrm{d}\Omega_3^2 + y\,\mathrm{e}^{-G}\,\mathrm{d}\tilde{\Omega}_3^2 \end{aligned} \tag{14}$$

and a similar equation for the Ramond-Ramond five-form field. The conditions imposed by the preservation of one half of the possible supersymmetries effectively linearise the non-linear Einstein equations rendering them completely solvable in terms of Green Functions. The complete set of geometries that satisfy these conditions are given by one function $\Phi(\mathbf{x}, y)$ that satisfies the equation

$$(\partial_1^2 + \partial_2^2)\Phi + \frac{1}{y^3}\,\partial_y(y^3\partial_y\Phi) = -4\pi^2\rho(\mathbf{x})\,\delta^{(4)}(y) \tag{15}$$

for which the solution has the form

$$\Phi = \pi\int_{y=0}\frac{\rho(\mathbf{x}')\,\mathrm{d}^2x'}{[(\mathbf{x}-\mathbf{x}')^2 + y^2]^2} \tag{16}$$

where

$$\rho(\mathbf{x}) = y^2\,\Phi(\mathbf{x}, y=0) \tag{17}$$

The functions, h, $\mathbf{V}$ and G and the five-form can all be found from the solution for Φ.

The mass which is equal to angular momentum is given by

$$M = \frac{1}{8\pi^2\ell_P^8}\left[\frac{1}{2\pi}\int\rho(\mathbf{x})\mathbf{x}^2\,\mathrm{d}^2x - \left(\frac{1}{2\pi}\int\rho(\mathbf{x})\,\mathrm{d}^2x\right)^2\right] \tag{18}$$

Those geometries for which $\rho = 1$ or 0 at $y = 0$ are in one to one corespondence, via AdS/CFT, to a class of operators in the boundary Superconformal Yang-Mills theory, which in particular are invariant under 1/2 of the original supersymmetry transformations and all of these geometries have $M > 0$. The mass of the space-time is equal to the conformal dimension of the corresponding operator, $M = \Delta = J$. Actually, one can slightly relax the boundary condition and require simply that ρ is always non-negative and bounded by 1. All such geometries have a dual description in terms of operators and the various Coulomb branches of the gauge theory. Indeed one can construct a precise mapping between the geometries and operators by noting that ρ can be interpreted as a fermion density in an harmonic oscillator potential, and the gauge theory on a space-time of topology $\mathbb{R} \times S^3$ (the topology of the boundary of AdS_5) can be mapped to the same system of free fermions.

Clearly outside the class of smooth and well-behaved geometries that are discussed here there is a potentially large class of other singular geometries. We would like to know precisely what such geometries could correspond to in the dual gauge theory if we follow the usual dictionary of the AdS/CFT correspondence.

Closed time-Like curves and causality/unitarity

Closed Time-Like curves are a feature of some interesting solutions to the Einstein equations. The two most well-known and typical examples are the Kerr metric and the Gödel space-time. In general such geometries contain angular momentum, such being necessary to cause the light-cones to tip over and effectively lie in the plane of what was a space-like region. In particular we require, for the existence of closed time-like curves, that there is some non-trivial time-like loop along which we can travel in such a way as to return to our beginning. For a discussion of closed time-like curves in Gödel and Kerr space-times see the beautiful treatise on General Relativity by Hawking and Ellis [24] and for a general discussion of closed time-like curves see the book by Visser [28].

Closed time-Like curves and singular geometries

This section is based on the paper Milanesi and O'Loughlin [22] where it is proven that if one relaxes the boundary conditions then there are geometries in the class of solutions based on the ansatz (14) of LLM that contain closed time-like curves.

Sufficiently close to any region of the $y = 0$ boundary in which the local density $\rho(\mathbf{x}) > 1$, the light-cone tips over and lies in a coordinate hypersurface that is asymptotically space-like as shown in figure 1. In the case that the region of $\rho(\mathbf{x}) > 1$ is compact one can always find a closed time-like curve that encircles this region as shown in figures 1 and 2. In the free fermion picture (where ρ is the fermion density), we see that the presence of a closed-time like curve is associated with a violation of the Pauli Exclusion Principle.

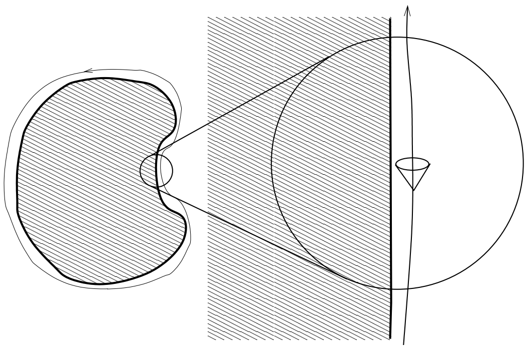


Figure 1: Distributions corresponding to boundary conditions at $y = 0$. The shaded region has $\rho > 1$, elsewhere $\rho = 0$.

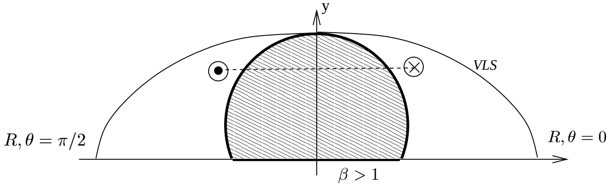


Figure 2: Cross section of rotationally symmetric configuration with $\rho > 1$ inside a disk. $y^2\Phi = 1$ on the boundary of the shaded region at $y > 0$. The space-time does not extend past the time-like singularity on the boundary of this shaded region.

The metric for these geometries has the same form as for LLM (14) and a closed time-like curve will arise when a compact direction in the $(\mathbf{x}, y)$ hypersurface becomes time-like. In the case of rotationally symmetric configurations in $\mathbf{x}$ and with $\mathbf{V} = (\mathbf{0}, V_\phi)$, the relevant parts of the metric are

$$ds^2 = -h^{-2}dt^2 - 2h^{-2}V_\phi\, dt\, d\phi + (h^2 - h^{-2}V_\phi^2)\, d\phi^2 + \cdots \qquad (19)$$

and when the discriminant $h^2 - h^{-2}V_\phi^2$ becomes negative a closed time-like curve will form. The boundary of the region of CTCs is called the velocity of light surface (VLS), and is indicated for this particular example of rotational symmetry in figure 2 ($\rho(\mathbf{x}, 0) = \beta > 1$).

In the case of constant density $\rho(\mathbf{x}, 0) = \beta$ inside a disk of radius R_0 and zero outside this disk, the mass (as measured at infinity) is

$$M = \frac{1}{8\pi^2 \ell_P^8} \left[\frac{1}{2\pi} \int \rho\, \mathbf{x}^2 \, \mathrm{d}^2 x - \left(\frac{1}{2\pi} \int \rho \, \mathrm{d}^2 x \right)^2 \right] = \frac{(\beta R_0^2)^2}{32\pi^2 \ell_P^8} \left(\frac{1}{\beta} - 1 \right) \tag{20}$$

and we see immediately that $\beta > 1$ corresponds to negative mass compared to the vacuum (AdS_5 in this case). We will refer to $\rho(\mathbf{x})$ as the local energy density and so:

Theorem[22]: Any solution of LLM type that has somewhere negative local energy density, $\rho(x_1, x_2) > 1$, contains a Closed Time-like curve and a Naked Singularity.

Unitarity and chronology protection

Now that we know what the closed time-like curves correspond to in the boundary data of the string theory we would like to see what they would correspond to in the dual gauge theory. A theorem of Dobrev and Petkova [15] states that operators in $N = 4$ SYM theories with negative conformal dimension are non-unitary.

In the previous section we have found that the conformal dimension of an operator in the $N = 4$ SYM theory is mapped to the mass of the corresponding geometry via the AdS/CFT correspondence. We arrive directly at the result then that a deformation of the geometry to a locally negative mass corresponds to a non-unitary operator in the boundary conformal field theory.

So we see that there is good evidence that constraining the operators of the CFT to be unitary corresponds in the holographic dual string theory to chronology protection as it should guarantee the absence of closed time-like curves.

The important role played by unitarity is not surprising as it is this same constraint that leads to a possible resolution of the black-hole evaporation puzzle. In fact, in that case unitarity of the holographic boundary theory implies that black hole evaporation is unitary and thus holography resolves the Information Loss Paradox of Black Hole physics at least for a family of black holes in AdS space.

Summary

The conjecture of holography for quantum gravity theories has lead to the construction of a correspondence between gauge theory and string theory. In the simplest case this maps string theory in a space-time that is five-dimensional times a finite volume internal space to a gauge theory in four-dimensions. The simplest model is supersymmetric which helps us calculate many features analytically. Thus holography has provided us with a powerful tool for the study of both gauge theory and string theory.

We have found in particular that unitarity of the dual gauge theory implies chronology protection and actually cosmic censorship in the sector of the theory that we have studied.

Our study leaves many questions unanswered in particular on the generality of these results as the explicit solutions that we studied can be found thanks to the large number of preserved supersymmetries. It is true however that naked singularities together with non zero angular momentum often gives rise to the conditions required to produce closed time-like curves and also naked singularities are usually related to the violation of some positive energy condition. Furthermore, it appears to be quite general that conformal dimensions in the gauge theory are related to mass (energy density) in the holographic dual string theory picture so maybe there is a way by which holography together with well understood properties of field theories may lead to a proof of chronology protection and possibly also cosmic censorship.

References

[1] H. Lin, O. Lunin and J. Maldacena, *Bubbling AdS space and 1/2 BPS geometries*, [arχiv:hep-th/0409174].

[2] D. Berenstein, *A toy model for the AdS/CFT correspondence*, JHEP **07** (2004) 018, [arχiv:hep-th/0403110].

[3] S. Corley, A. Jevicki and S. Ramgoolam, *Exact correlators of giant gravitons from dual N = 4 SYM theory*, Adv. Theor. Math. Phys. **5** (2002) 809–839, [arχiv:hep-th/0111222].

[4] R.C. Myers and O. Tafjord, *Superstars and giant gravitons*, JHEP **11** (2001) 009, [arχiv:hep-th/0109127].

[5] A. Buchel, *Coarse-graining 1/2 BPS geometries of type IIB supergravity*, [arχiv:hep-th0409271].

[6] P. Horava and P.G. Shepard, *Topology changing transitions in bubbling geometries*, JHEP **02** (2005) 063, [arχiv:hep-th/0502127].

[7] S.S. Gubser and J.J. Heckman, *Thermodynamics of R-charged black holes in AdS(5) from effective strings*, JHEP **11** (2004) 052, [arχiv:hep-th/0411001].

[8] D. Bak, S. Siwach and H.-U. Yee, *1/2 BPS geometries of M2 giant gravitons*, [arχiv:hep-th/0504098].

[9] J. McGreevy, L. Susskind and N. Toumbas, *Invasion of the giant gravitons from anti-de Sitter space*, JHEP **06** (2000) 008, [arχiv:hep-th0003075].

[10] A. Hashimoto, S. Hirano and N. Itzhaki, *Large branes in AdS and their field theory dual*, JHEP **08** (2000) 051, [arχiv:hep-th/0008016].

[11] V. Balasubramanian, V. Jejjala and J. Simon, *The library of Babel*, [arχiv:hep-th/0505123].

[12] S.S. Gubser, *Curvature singularities: The good, the bad and the naked*, Adv. Theor. Math. Phys. **4** (2002) 679–745, [arχiv:hep-th/0002160].

[13] R.C. Myers, *Stress tensors and Casimir energies in the AdS/CFT correspondence*, Phys. Rev. **D60** (1999) 046002, [arχiv:hep-th/9903203].

[14] G.T. Horowitz and R.C. Myers, *The value of singularities*, Gen. Rel. Grav. **27** (1995) 915–919, [arχiv:gr-qc/9503062].

[15] V.K. Dobrev and V.B. Petkova, *On the group theoretical approach to extended conformal supersymmetry: classification of multiplets*, Lett. Math. Phys. **9** (1985) 287; *All positive energy Unitary Irreducible Representations of extended conformal supersymmetry*, Phys. Lett. **B162** (1985) 127.

[16] S.W. Hawking, *The Chronology protection conjecture*, Phys. Rev. D **46** (1992) 603.

[17] J.C. Breckenridge, R.C. Myers, A.W. Peet and C. Vafa, *D-branes and spinning black holes*, Phys. Lett. **B391** (1997) 93–98, [arχiv:hep-th/9602065].

[18] M. M. Caldarelli, D. Klemm and P.J. Silva, *Chronology protection in anti-de Sitter*, [arχiv:hep-th/0411203].

[19] M. Boni and P.J. Silva, *Revisiting the D1/D5 system or bubbling in AdS(3)*, [arχiv:hep-th/0506085].

[20] C.A.R. Herdeiro, *Special properties of five dimensional BPS rotating black holes*, Nucl. Phys. **B582** (2000) 363–392, [arχiv:hep-th/0003063].

[21] L. Jarv and C.V. Johnson, *Rotating black holes, closed time-like curves, thermodynamics and the enhancon mechanism*, Phys. Rev. **D67** (2003) 066003, [arχiv:hep-th/0211097].

[22] G. Milanesi and M. O'Loughlin, *Singularities and closed time-like curves in type IIB 1/2 BPS geometries*, JHEP **08** (2005) 509, [arχiv:hep-th/0507056].

[23] S. Hawking, *Particle Creation by Black Holes*, Commun. Math. Phys. **43** (1975) 199-220; Erratum *ibid.* **46** (1976) 206.

[24] S.W. Hawking and G.F.R. Ellis, *The large scale structure of space-time*, Cambridge Monographs on Mathematical Physics, (1989).

[25] J.M. Bardeen, B. Carter and S.W. Hawking, *The Four laws of black hole mechanics*, Comm. Math. Phys. **31** (1973) 161.

[26] W.G. Unruh and R.M. Wald, *Acceleration Radiation and Generalized Second Law of Thermodynamics*, Phys. Rev. D **25** (1982) 942.

[27] O. Aharony, S.S. Gubser, J.M. Maldacena, H. Ooguri and Y. Oz, *Large N field theories, string theory and gravity*, Phys. Rept. **323** (2000) 183, [arχiv:hep-th/9905111].

[28] M. Visser, *Lorentzian Wormholes: From Einstein to Hawking*, (American Institute of Physics, New York, 1995).

[29] L. Susskind, *The World as a Hologram*, J. Math. Phys. **36** (1995) 6377, [arχiv:hep-th/9409089].

[30] G. 't Hooft, *Dimensional reduction in quantum gravity*, in Salamfestschrift, World Scientific series in 20th Century Physics, v. 4, [arχiv:gr-qc/9310026].