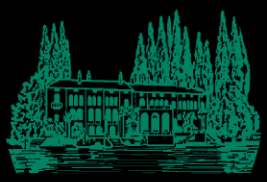


PROCEEDINGS
INTERNATIONAL SCHOOL OF PHYSICS «ENRICO FERMI»

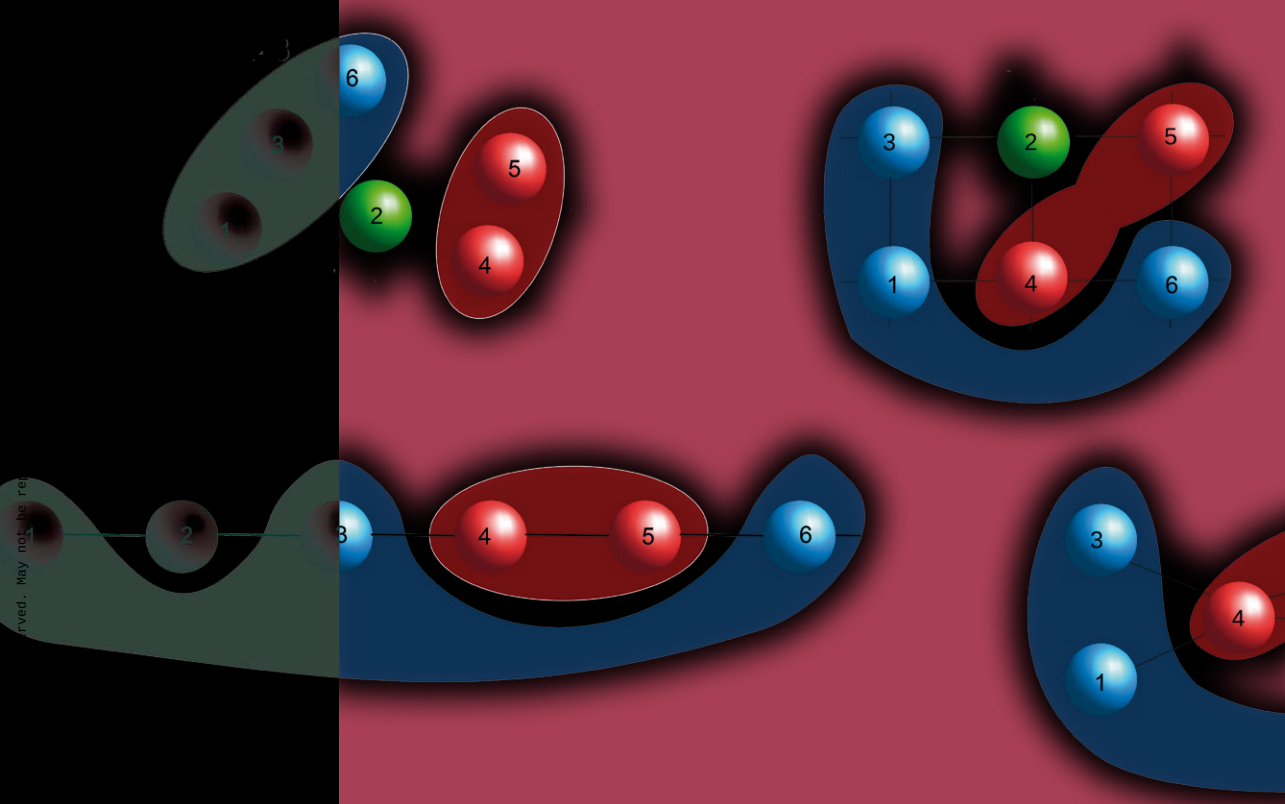
COURSE 197

Foundations of quantum theory

edited by E. M. Rasel, W. P. Schleich and S. Wölk

8 - 13 July 2016

Villa Monastero
Varenna, Lake Como



SOCIETÀ ITALIANA DI FISICA

SOCIETÀ ITALIANA DI FISICA

RENDICONTI

DELLA

SCUOLA INTERNAZIONALE DI FISICA

"ENRICO FERMI"

CXCVII Corso

a cura di E. M. Rasel e W. P. Schleich

Direttori del Corso

e di

S. Wölk

VARENNA SUL LAGO DI COMO

VILLA MONASTERO

8 – 13 Luglio 2016

Fondamenti della teoria dei quanti

2019

SOCIETÀ ITALIANA DI FISICA
BOLOGNA-ITALY

ITALIAN PHYSICAL SOCIETY

PROCEEDINGS

OF THE

INTERNATIONAL SCHOOL OF PHYSICS

"ENRICO FERMI"

COURSE 197

edited by E. M. RASEL and W. P. SCHLEICH

Directors of the Course

and

S. WÖLK

VARENNA ON LAKE COMO

VILLA MONASTERO

8 – 13 July 2016

Foundations of Quantum Theory

2019

IOS Press

AMSTERDAM, OXFORD, TOKIO, WASHINGTON DC

ISSN 0074-784X (print)
ISSN 1879-8195 (online)

ISBN 978-1-61499-936-2 (print) (IOS Press)
ISBN 978-1-61499-937-9 (online) (IOS Press)
ISBN 978-88-7438-115-9 (SIF)

LCCN 2018964161

jointly published and distributed by:

IOS PRESS
Nieuwe Hemweg 6B
1013 BG Amsterdam
The Netherlands
fax: +31 20 687 0019
info@iospress.nl

SOCIETÀ ITALIANA DI FISICA
Via Saragozza 12
40123 Bologna
Italy
fax: +39 051 581340
order@sif.it

Distributor in the USA and Canada
IOS Press, Inc.
4502 Rachael Manor Drive
Fairfax, VA 22032
USA
fax: +1 703 323 3668
sales@iospress.com

Supported by

Camera di Commercio di Lecco

Istituto Nazionale di Fisica Nucleare (INFN)

Istituto Nazionale di Geofisica e Vulcanologia (INGV)

UNIVERLECCO

Museo Storico della Fisica e Centro Studi e Ricerche “Enrico Fermi”

Wilhelm und Else Heraeus Foundation

Produced by the SIF Editorial Staff

Production Editor: Marcella Missiroli

Cover: see S. Wölk, *Revealing quantum properties with simple measurements*, p. 235

Graphic elaboration by Simona Oleandri

Printed in Italy by Labanti e Nanni Industrie Grafiche S.r.l. - Crespellano (BO)

CONTENTS

The electronic version of volumes from 124 to 198 is available online at the IOS Press web site

http://ebooks.iospress.nl/bookseries/proceedings-of-the-international-school-of-physics-enrico-fermi

Figures with colour source files will appear in colour in the online version.

SIF Members are granted free access to these volumes at
https://members.sif.it

For a complete list of published courses see
https://en.sif.it/books/series/rendiconti_fermi

Preface

The *International School of Physics "Enrico Fermi"* on the *Foundations of Quantum Theory* was organized by the Italian Physical Society in Villa Monastero, Varenna, Italy, during July 8-13, 2016 in collaboration with the *Wilhelm und Else Heraeus-Stiftung*. In the great tradition of the Fermi Schools the main goal was to provide an overview of the recent theoretical and experimental developments of an active field of research, in this case the foundations of quantum mechanics. The timing is especially appropriate considering the fact that the last "Enrico Fermi" Summer School on this topic took place in 1977.

Quantum mechanics is characterized by a dichotomy of unparalleled agreement between theory and experiment on the one hand, and an enormous variety of interpretations of the underlying mathematical formalism on the other. David Mermin([1]) proposed a very pragmatic approach to this situation: "Shut up and calculate"!

However, triggered by the rapid advance of experimental techniques in quantum optics, and the development of the field of quantum technology which takes advantage of the correlations of entangled quantum systems, the question of the interpretation of quantum mechanics has recently again received a lot of attention. Moreover, the old conundrum of the physical reality of the wave function has now been tested by experiments using single photons.

These two examples are only to serve as an illustration of this active field. Indeed, the topics discussed at our school included but were not limited to the history and interpretations of quantum theory, the principle of complementarity and wave-particle duality, quantum theory from first principles, the reality of the wave function, the concept of the photon, measurement in quantum theory, the interface of quantum theory and general relativity, and quantum optical tests of quantum theory.

The present volume summarizes the lectures presented at our School which was attended by more than 80 participants including students, lecturers and seminar speakers from all over the world. All young scientists presented their research in two poster

([1]) See for example: N. D. Mermin, *What's wrong with this pillow?*, *Physics Today*, **42** (4) (1989) 9. It is interesting that this quote is often attributed to Richard P. Feynman but we recommend N. D. Mermin, *Could Feynman have said this?*, *Physics Today*, **57** (5) (2004) 10.

sessions which were each introduced by a "poster-flash" session. We are most grateful to *Europhysics Letters* for supporting a prize for the best posters. Dennis Rätzel (First Prize), Yael Avni (Second Prize), and Da-Wei Wang (Third Prize) together with the winners of the Fourth Prize Sven Abend, Lorenzo Catani and Piotr Roztocki were invited to summarize their contributions in this volume.

The *Proceedings of the International School of Physics "Enrico Fermi"* on the *Foundations of Quantum Theory* start from a historical perspective by two contributions by Nancy Thorndike Greenspan describing the life and the science of Max Born based on her book entitled *The End of the Certain World*.

In the spring of 1925 Werner Heisenberg had escaped from Göttingen to the island of Helgoland to cure his hay fever. It was there that he discovered quantum mechanics. Considering the radical changes in the principles it is not surprising that his seminal article in *Zeitschrift für Physik* is hard to read. Manfred Kleber in his contribution summarizes in a pregnant way the underlying ideas of Heisenberg's transition from Born's *Atommechanik* to *Matrizenmechanik*.

Although quantum mechanics originally started from matrix mechanics, today we almost always employ the formulation pioneered by Erwin Schrödinger based on a time-dependent wave function. The history and the derivation of the Schrödinger equation constitute the topic of the lectures by Wolfgang P. Schleich. His main theme is the linearization of the nonlinear wave equation of statistical mechanics.

Gerd Leuchs in his contribution complements the previous more mathematical approach towards the Schrödinger wave equation by a more intuitive one. He employs a combination of the analogy between matter and water waves, and dispersion relations.

In 1935 Schrödinger identified entanglement as *the* trademark of quantum mechanics. Indeed, it is at the very heart of Bell's inequalities and many alien features of quantum theory can be traced back to it. Edward Fry in his lectures summarizes in an impressive way the early history of this field emphasizing the important role of Grete Hermann. Already in 1935 she discovered the flaw in von Neumann's proof that it is impossible to complete quantum mechanics. Da-Wei Wang extends these two-particles considerations to three particles and proposes a new way to generate mesoscopic Greenberger-Horne-Zeilinger states. In the same spirit Piotr Roztocki employs in his contribution a frequency comb to create scalable quantum states.

The lectures by Marlan O. Scully illuminate the problem of time in the process of a quantum measurement. When and how does the observer change or reduce the state vector? Three frequently employed scenarios illustrate how "before" and "after" arguments can be misleading: The Einstein-Podolsky-Rosen situation, Wigner's friend and the quantum eraser.

John Archibald Wheeler (1911-2008) in his seminal article *It from bit* has vividly argued that quantum theory is information theory. In this way he can be considered the father of quantum information. He often stated that it should be possible to derive quantum mechanics from information theory. Christopher A. Fuchs has followed this path and has proposed the new interpretation of quantum mechanics, Quantum Bayesianism (QBism). In his lectures he first summarizes the most prominent interpretations of

quantum mechanics and then provides an introduction into QBism. In the same spirit Giacomo Mauro D'Ariano derives from elementary principles of information theory free quantum field theories.

Across from Varenna at the West end of Lake Como is the place where Niels Bohr in 1927 introduced the principle of complementarity stimulated by Heisenberg's uncertainty relation. Sabine Wölk in her lecture notes employs simple measurements on quantum systems to compare complementarity and entanglement both of which have their roots in non-commuting operators.

The field of experimental quantum optics has opened new avenues towards tests of the foundations of quantum mechanics. Here the process of spontaneous parametric down conversion (SPDC) plays a central role and has created an avalanche of applications. Ralf Menzel in his lectures emphasizes the role of the mode function of the electromagnetic field as the carrier of the photon, and reviews his experiments on stimulated coherence and complementarity.

Quantum imaging is another product of SPDC. In his lecture Robert W. Boyd summarizes this active field by giving three examples: ghost imaging, imaging based on interaction-free measurements and imaging based on Mandel's induced coherence.

The reality of the wave function is an often debated question. So far it has been part of more philosophical discussions. However, SPDC has moved this realm from *Gedanken experiments* to real ones. The lectures by Andrew White addressed these issues. Unfortunately, due to time constraints he was not able to provide us with a paper. Likewise Aephraim Steinberg could not contribute. Fortunately, the notes by Lorenzo Catani *et al.* address some aspects of these questions. In particular, they discuss contextuality as a resource in quantum computation.

We recall that quantum mechanics originated from the analysis of blackbody radiation. In contrast to conventional wisdom Max Planck did not quantize the light in the resonator but the mechanical oscillators in the walls. Quantized electrodynamics had to wait till the *Drei-Männer-Arbeit* of Born, Heisenberg and Pascal Jordan. They rederived the result of Albert Einstein concerning the fluctuations of the radiation field in the thermal state from a field theoretical approach. The Casimir effect, that is the attraction of two uncharged conducting plates, is another consequence of these fluctuations. Yael Avni in her contribution summarizes her recent work with Ulf Leonhardt on Casimir forces in spherically symmetric dielectric media.

The theories of special and general relativity together with quantum mechanics are rightfully considered the major revolutions in physics of the 20th century. Despite the fact that by now they are almost 100 years old, general relativity and quantum mechanics have not been unified yet. The lectures of Daniel M. Greenberger provide insight into the reasons for the resistance and identify the strange roles of proper time and mass. He also discusses consequences originating from considering them as dynamical variables.

The field of atom optics is perfectly suited to probe this interface of quantum mechanics and general relativity. On the one hand we use the wave nature of atoms for interferometry, on the other hand due to their mass the atoms feel gravity. Ernst M. Rasel in his lectures provides an introduction into atom interferometry, discusses a quan-

tum test of the equivalence principle and gives an outlook to experiments in space. The contribution of Sven Abend *et al.* expands on this theme and discusses a new avenue based on an atom-chip gravimeter.

Since light represents energy it must also gravitate. Already in 1931 Richard Tolman together with Paul Ehrenfest and Boris Podolsky showed that a pencil of light leads to a curvature of spacetime. The contribution by Dennis Rätzel *et al.* summarizes gravitational properties of light.

Not included in this volume are two other highlights of our school. Nancy Thorndike Greenspan had discovered a movie taken by the Nobel Prize winner Irving Langmuir at the Solvay Meeting of 1927. It was impressive to see the famous quantum physicists of the time "in action" rather than sitting around a table. Moreover, our school ended with the playing of the Mozart piano concerto A-major KV 488 recorded around 1965 by the Bavarian Radio Symphony Orchestra under the conductor Rudolf Albert. The soloist was Werner Heisenberg. We owed this pleasure to Manfred Kleber who had found this treasure. Many thanks, Manfred, for sharing it with us!

All activities were inspired by the breathtaking beauty of Lake Como, the Villa Monastero and its gardens, and by the rich heritage by the *Enrico Fermi International School of Physics*. The success of the school measured by the exceptionally large number of interactions between the participants and the extremely lively discussions, during and immediately after the talks, in the park and on several excursions, is also due to the excellent organizational and administrative support provided by the staff of the Italian Physical Society. We are also most grateful to the *Wilhelm und Else Heraeus-Stiftung* for its generous monetary support.

E. M. Rasel, W. P. Schleich and S. Wölk

This page intentionally left blank

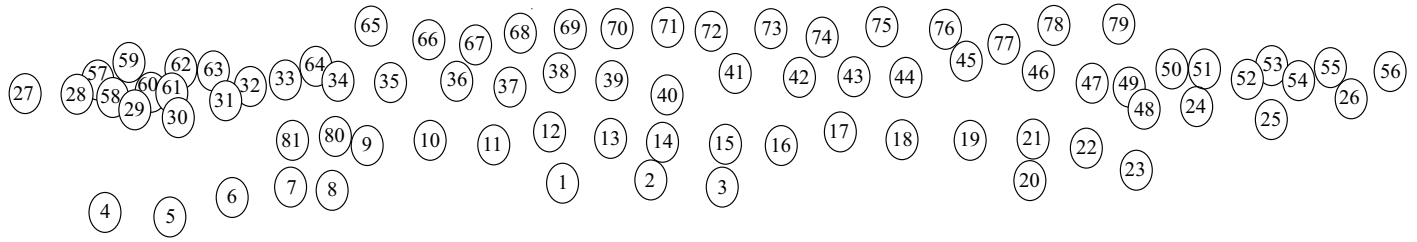
Italian Physical Society

INTERNATIONAL SCHOOL OF PHYSICS «E. FERMI»

COURSE 197

8 - 13 July 2016

VILLA MONASTERO − VARENNA, LAKE COMO



1) Maria Bianchi
2) Ramona Brigatti
3) Barbara Alzani
4) Maria Chiara Braidotti
5) John DeBrota
6) Natalia Salome Moller
7) Da-Wei Wang
8) Stefania Sciara
9) Amirhossein Sadeghi
10) Ernst Rasel
11) Sabine Wölk
12) Wolfgang Schleich
13) Marlan Scully
14) Gerd Leuchs
15) Wolfgang Ertmer
16) Manfred Kleber
17) Robert Boyd
18) Ralf Menzel
19) Christopher Fuchs
20) Marta Greselin
21) Edward Fry
22) Nancy Greenspan
23) Andrea Petrucci
24) Sahar Sahebdivan
25) Yael Avni
26) Yi Fan Wang
27) Stefano Bacchi
28) Matteo Schiavon
29) Sanah Altenburg
30) Sorayya Shadi
31) Ramona Walter
32) Lorenzo Catani
33) Michael Taheri
34) Hendrik Siebeneich
35) Patrick Huber
36) Michele Sacerdoti
37) Thomas Galley
38) Radu Ionicioiu
39) Jonathan Ben-Benjamin
40) Gabriel Fagundes
41) Aephraim Steinberg
42) Matthias Zimmermann
43) Alexander Friedrich
44) Pavel Jefremov
45) Matthias Dahlmanns
46) Manuel Daiber
47) Adrien Feix
48) Paolina Corona-Ugalde
49) Flaminia Giacomini
50) Stefan Jorda
51) Marduk Bolaños Puchet
52) Piotr Roztocki
53) Pawel Blasiak
54) Luo Qi
55) Vyome Singh
56) Dennis Rätzel
57) Reinhold Walser
58) Martin Malachov
59) Michal Jex
60) Iva Bezdekova
61) Nuria Munoz Gargante
62) Philipp Griebling
63) Athanasios C. Tzemos
64) Jannik Hoffmann
65) Tristan M. Kraft
66) Matthias Germann
67) Hai-Chau Nguyen
68) Murat H. Uygunol
69) Oleksandr Burlayenko
70) Hendrik J. Roßkamp
71) Robin Corgier
72) Sven Abend
73) Marco Erba
74) Ömer Bayraktar
75) Kevin Günthner
76) Lucas Happ
77) Esteban Castro Ruiz
78) Julian Fleniken
79) Nandan Jha
80) Anton Krieger
81) Fabiano Lever

This page intentionally left blank

Proceedings of the International School of Physics "Enrico Fermi"
Course 197 "Foundations of Quantum Theory", edited by E. M. Rasel, W. P. Schleich and S. Wölk
(IOS, Amsterdam; SIF, Bologna) 2019
DOI 10.3254/978-1-61499-937-9-1

Science in tumultuous times

Nancy Thorndike Greenspan(*)

Washington DC, USA

Summary. — Germany and German science underwent seismic upheavals from the start of World War I to the end of World War II. The political history exemplifies nationalistic aggression. The scientific history showcases breath-taking insights into understanding nature. Both histories revolutionized the world's future. How did the individual scientist maintain his focus and discipline to make these break-throughs in the midst of chaos? The lives of Max Born and his friends and colleagues give a glimpse into what they endured and what they accomplished.

Introduction

In the first half of the 20th century, physicists, particularly Germans, faced a chaotic environment—war, hyper-inflation, anti-Semitism, economic depression, and constant moral dilemmas. In one way or another, these conditions held for all German physicists. Well-funded research and social stability were not the order of any day. Yet through it all, they pierced some of nature's most mysterious phenomena, general relativity and quantum physics being the two most obvious. Understanding something of this tur-

(*) Nancy Thorndike Greenspan is a writer based in Washington DC. Her most recent book is *The End of the Certain World: The Life and Science of Max Born*.
E-mail: nancy@nancygreenspan.com

bulence lends added respect for what they accomplished. Some historians argue that the unpredictability surrounding them contributed to their understanding of this new perception of science [1].

As a German, a scientist, and a Jew, the life of Max Born—and those of his friends—reflects the major historical events of early 20th century.

1. – The years of the First World War (1914-1918)

Berlin, Spring 1918. Two men, one meticulous in the uniform of the Prussian Army, the other slightly disheveled in the recognizable black suit of an academic. The two, obviously friends and close in age — one is 36 and the other 39 — sit at an outdoor café drinking ersatz black coffee. Both are lean from the severe deprivations of the war. The years in Berlin had been physically tough. By war's end, the inhabitants of Berlin consumed about 1000 calories per day and 3/10 of a pound of meat per week. Two years earlier, they had largely existed on as many pounds of turnips as they could buy.

Throughout the war years, the army officer, Max Born, and the rumpled scholar, Albert Einstein, spent many an hour discussing science and war at such a café. Whatever problems occurred in everyday life, they had a panacea — theorizing about and expounding on the foundations of nature, looking for new insights into the mysteries of the universe. Born watched Einstein pull together the last pieces of his general theory of relativity, later describing it as "the greatest feat of human thinking about nature" [2].

At the start of their friendship in 1915, the science they agreed on; the war they did not. Einstein, a pacifist, was convinced that Germany should lose the war. He argued his views forcefully. Born struggled between love of country and a distain for Prussian militarism learned from his father. Eventually Born rationalized his patriotic feelings and ended his muddle. Einstein persuaded him that the country's future was as a republic under an elected socialist democratic government.

Born and his friends in Berlin worked for the war effort under some of Germany's top scientists. Einstein did not serve and had the freedom to pursue his science at will. But weapons and warfare did not stymie the research interests of junior scientists such as Max Born and his close friend James Franck.

Born headed a weapons research branch in the Artillerie-Prüfungs (*Testing*)-Kommission (the APK). His assigned project was "sound ranging": determining the position of the enemy's artillery. With data on the arrival time of the discharge sound from a gun at various points one could calculate where the sound waves intersected, *i.e.*, the enemy's gun. In addition to intricate calculations on timing, adjustments for interference factors such as wind, background noise, and the impact of the shells' bow waves on the propagation of sound had to be made.

The urgent needs of the army required the research to be conducted quickly, creating the need to recruit other scientists. Born's office swarmed with old friends from his university days in Breslau and his lecturer days in Göttingen — Erwin Madelung, Alfred

Landé, Rudi Minkowski (nephew of Hermann), and Fritz Reiche. (See Appendix A.) To his dismay there was no position for Otto Stern at the APK, but he helped to place him out of harm's way at a research division at the University of Berlin.

For Born, the primary benefit of the APK was sheltering physicists from the menace of the front and saving them for Germany's future. He stayed awake at night worrying about how to do more.

At the beginning of the sound-ranging project, Born tried to keep up with research by reading physics articles at night or attending various colloquia in the city. It frustrated him. But as the work for the army waned, opportunities opened up. Born and Landé emptied desk drawers of military graphs and calculations and re-filled them with calculations of the lattice energy of ionic crystal.

Niels Bohr had sent Born his article "On the Quantum Theory of the Line Spectrum, (Pt1)". Born with Landé—and some input from Minkowski— immediately began applying the information on the stability of the atom to Born's earlier research on the structure of solids. Investigating the simple ionic crystal, they calculated the forces between the lattice points that would determine the structure and stability of it. They found that using the Bohr-Sommerfeld ring-model to calculate one of the variables determining compressibility gave incorrect results. This particular research raised doubts in Born's mind about Bohr's ring model and further aroused his curiosity about quanta.

Working across town at the Kaiser Wilhelm Institute, James Franck took a different path. He joined the team of the renowned chemist Fritz Haber who was the Germans' lead scientist on gas warfare. Experiments by Franck and his colleague Gustav Hertz had already confirmed Einstein's interpretation of the photoelectric effect. (They won the Nobel Prize in 1925 for discovering laws governing the collision of an electron with an atom.) Throughout the war years they continued their research with Franck even managing to submit a paper with Hertz while hospitalized with a severe case of pleurisy. At the Institute, Franck with Otto Hahn, worked on the defensive side of the research. They were human guinea pigs, testing masks in a room filled with poison gas.

Early on Haber had offered Born the same opportunity. Born rejected it. In his reply to Haber, he condemned the use of gas warfare and created a rift between them.

At war's end, in the midst of the turmoil from Germany's surrender and a threatened Bolshevik revolution, with the trams shut down because of street fighting, Born walked the few miles across the fields to Dahlem from his apartment. His purpose was to talk to Franck about verifying his calculations on lattice energies. He wanted to compare the theoretical ones to measured data on their chemical companion, heats of formation.

Shortly after Born arrived at Franck's office, Haber happened by. An awkwardness hung between them for a minute but quickly dissolved into a friendly rapport. Haber took an immediate interest in Born's ideas. Working together over the next few weeks, they dissected the process by which metals form an ionic compound. Their work resulted in the Born-Haber cycle, a means of calculating lattice energies. The value for lattice energies that Born derived from theory was so accurate that it is now often used to calculate electron affinities.

During these same weeks at war's end, Einstein and Born set out to rescue officials

from the University of Berlin who had been arrested by students. They succeeded after somewhat bizarre events involving an audience with Germany's new Chancellor Friedrich Ebert and conflict with "red" soldiers who barred them from entering the Reichstag. At the end, they felt that they had witnessed what they had hoped for: the birth of a free, democratic, socialist Germany.

The shift in power did not bring harmony or stability to Germany. In fact, after the surrender in November 1918, the German people endured an eight-month blockade kept in place by the British to ensure that Germany met the Allied demands presented at Versailles. Children starved, babies went without milk, suicides floated down the River Spree, everyone was cold for lack of heating fuel, thousands died. Destitution similar to the war years continued and bitterness churned inside the people.

2. – Post-War years (1919-1921)

In 1919, the German government granted Born a professorship as head of the Department of Theoretical Physics at the University of Frankfurt. Born found a disorganized department consisting of two assistants, a mechanic, and a lecturer —small, but an exemplary group. The lecturer was Otto Stern who was shortly joined by Walther Gerlach. Stern was trying to verify that the distribution and the mean of the velocities of atoms followed the Maxwell-Boltzmann distribution. At this point he was creating the molecular beam method. But creeping inflation threatened the continuation of his research. The department research budget had completely run out.

At just that moment Albert Einstein came to the rescue—indirectly. In November 1919, British physicist Arthur Eddington announced that he had confirmed an element of Einstein's general theory — the deflection of light through the gravitational field of the Sun. Overnight, Einstein and his theory became international celebrities. Born held three lectures on relativity for the general public in Frankfurt and filled a large auditorium at the university. Charging admission, he raised almost 7000 marks. The funds financed the research that later won Stern a Nobel Prize. Einstein, impressed with Born's industry, teased him:

> "And you, Max, are giving lectures on relativity to save the institute from penury, and writing papers as if you were a single young man living in splendid isolation in his own specially heated apartment, with none of the worries of a *paterfamilias*. How do you do it?" [3].

Germany continued to suffer through waves of political instability with French troops occupying the Rhineland, including Frankfurt, and politicians falling to assassination. Europe's eastern countries offered no better and heightened the volatility. The plight of Paul Epstein, who had already done seminal research on quantum theory, namely on the Stark effect, is a microcosm of insecurity and an embodiment of its effect on young scientists.

Epstein was a man with many countries and no place to go. He was Russian by birth, had earned a Ph.D. at the University of Munich under Arnold Sommerfeld and after the

war carried a Polish passport because Poland's redrawn border with Russia now included his birthplace. His dilemma was that Russia was too dangerous; Polish universities did not want a German-educated professor; and German universities did not want a person with a Polish passport, having just ceded its northeast area to Poland.)

Born did not know Epstein well, but he committed himself to finding him a position. But where to look? Born himself was a tainted emissary. The countries to the west—France, England, and Belgium—were out because they still had no official contact with Germany. Physicists in the US harbored similar animosity. After lengthy searching, he was able to find Epstein a position at California Institute of Technology through an American friend—a rarity for Germans to have in those days.

One factor inhibiting German research was the refusal of scientists in Allied countries to reconcile, which even included circulating journal reprints to them. The impasse arose because of the Manifesto of Ninety-Three signed by German scientists and intellectuals, such as Max Planck, at the start of the war. The Manifesto disavowed Germany's responsibility for starting the war, or for violating the neutrality of Belgium, or for harming its citizens. (Born had not sign. He was too junior.) It enraged British and American scientists at the time. And some six years later the fury still gripped them. Contacting them was asking for an insult. Born learned this lesson when he reached out to Harvard chemist and Nobel Prize winner Theodore Richards.

A Swedish friend wrote to Richards on Born's behalf to ask if he would collaborate on measuring the compressibility of a specific list of salts—to aid in investigating Bohr's planar concept of the structure of matter. Richards answered that he was willing to help Born since he had not signed the Manifesto. Otherwise, he was unwilling to shake the hand of a German scientist and called for Germans to repent. Born replied through his friend, his answer alluding to the devastation of the blockade:

> "You write him that *we* will *not* forgive as long as the body politic exists which holds, both as a party and a judge, for its right to impose starvation on an entire people" [4].

Shortly after, Born established cordial relations with old friend Irving Langmuir at GE in Schenectady NY. He wrote Langmuir of his frustration.

> "Several days ago, Niels Bohr was in Berlin. I discussed much with him about quantum theory and atomic structure. It is a pity that American scholars cannot participate in the discussion; otherwise one would reach agreement quickly" [5].

Although Born's main research in Frankfurt was not quantum theory, by lecturing on it and studying Bohr's articles and Sommerfeld's *Atomic Structure and Spectral Lines*, he stayed in the discussion.

Always hoovering in the background were repercussions from the Treaty of Versailles. In 1921, Born and Einstein corresponded about its impact. Einstein, painting an optimistic picture of world events perhaps to avoid anxiety, wrote,

> "You *need* not be so depressed by the political situation. The huge reparation payments and the threats are only a kind of moral nutrition for the dear public in France, to make the situation appear rosier to them. The more *impossible* the conditions, the more certain it is that they are not going to be put into practice" [6].

Einstein believed that the slovenliness of the French and the growing disunity among the Allies would undermine the intent and impact. Born saw it differently:

> "I can *see* the *effect* of this power politics on the minds of the people; it is a wholly irreversible accumulation of ugly feelings of anger, revenge, and hatred ... It seems to *me* that new catastrophes will inevitably result from all this. The world is not ruled by reason; even less by love" [7].

Against this political and economic backdrop Born was called to the chair of the theoretical physics department at the University of Göttingen. On leaving Frankfurt in 1921, he proposed Otto Stern as his successor, which included his promotion from lecturer to assistant professor. The University—so called the "Jewish University" because its original benefactors had been Jewish and its bylaws stipulated no religious discrimination—and the government's education department refused to make the appointment. A professor there said to Born: "I think very highly of Stern, but he has such an analytical Jewish intellect" [8]. Born felt the decision personally.

An ultimatum that Born placed on accepting the chair in Göttingen was Franck's appointment to a chair in experimental physics. He also wanted another assistant professor. When Frankfurt rejected Stern, Born wanted him to come to Göttingen. He got Franck, but not Stern. Part of the reason was the budget, the other was anti-Semitism. The unofficial Jewish quota on the Göttingen faculty was already oversubscribed.

3. – Quantum mechanics (the 1920s)

By the time Born settled in Göttingen, the infamous hyper-inflation had begun. In November 1921, there were 330 marks to one American dollar. Two years later the rate was over 4 trillion marks. Starving students became a constant concern. In addition to worries about their general survival, Born found them too cold and too hungry to study. He raised funds to support them both academically and personally and wrote scores of letters to find them teaching positions until he had exhausted all possibilities and himself.

The inflation ended in January 1924, just over a year before the formulation of quantum mechanics tumbled forth. During the inflationary period, much of the basic framework of QM was developed—and much of it in Göttingen by Born and his assistants.

Born's ambition was "to bring Göttingen physics to further heights" [9]. Wolfgang Pauli was part of the plan. Discontented with Göttingen and Born's mathematical approach, he only lasted for about six months. Following his departure he and Born continued their collaboration on applying perturbation theory to the helium atom. At first they

thought that their results confirmed an old claim by Niels Bohr about the orientation of the axes of its two electrons. But ultimately they found that the results conflicted with experimental data.

Born would soon discuss these results with Bohr. In 1922, Bohr came to Göttingen for—the *Bohr Festspiele*, as it became known. Bohr lectured for two weeks describing his theory, which he considered unfinished, with its focus on spectral lines and their connection with the elements of the periodic table. He proposed the Correspondence Principle as a means to address a troublesome problem with this theory, namely that the *orbital* frequency of an electron did not correspond with the transition frequency (emission/absorption). The Correspondence Principle ensured that, although the two frequencies were never equal, they converged in the *limit* of high quantum numbers.

The lectures often went past the dinner hour. Friedrich Hund observed that they were scientifically nourished. But given the economic conditions, physically famished. For the next three years, nothing stopped the quantum fervor there.

The Festspiele, organized by Born and David Hilbert, signaled to British and French scientists that their refusal to admit German scientists to their conferences and their boycott of German ones would not marginalize them. Born's trip to the US three years later, where he introduced unaware American physicists to the new quantum mechanics, shows that the lack of contact hardly compromised the Germans' scientific prowess. Perhaps the opposite occurred.

The rupture was fully healed in September, 1927—nine years after the armistice—with the Volta Conference at Como, Italy, one of the first international gatherings to freely invite German scientists. Max Planck did not accept the invitation until he was sure that was really the case.

The Volta Conference, held to commemorate the centennial of Alessandro Volta's death, was in part a precursor to the Solvay conference a month later. It included more than 70 physicists from 13 countries. As the conference took place shortly after Heisenberg discovered the uncertainty principle, quantum physics was the highlight. So was the new political order. Several speakers praised Benito Mussolini and among the banners hanging high in the hall, Russia's new flag with the hammer and sickle was prominent. Born delivered a lecture on the statistical interpretation of the wave function; Bohr gave one on "The Quantum Postulate and the Recent Development of Atomic Theory"; Heisenberg did not present but used the comment period to discuss uncertainty.

Bohr's ideas, although still in process, heralded the beginning of the Copenhagen interpretation on quantum physics. It was the formal introduction to his new overarching concept of complementarity where waves and particles were mutually exclusive but absolutely necessary to each other. In spite of that the conference was, in Born's words, "dull". His most significant event turned out to be bumping into Ernest Rutherford and Francis Aston as those two also escaped from the boredom. The relationship forged with Rutherford that day was instrumental in Born's receiving a position at the University of Cambridge when he fled Germany in 1933.

After Born's triumphal trip to the US in 1925-26 spreading the quantum gospel, Göttingen was awash with eager students and learned professors who, drunk with quan-

tum spirits, continued to explore and expand the theory. Copenhagen with Bohr was much the same. And for a few years there was stability, but then, on October 29, 1929, the stock market crashed. Some students, such as Edward Teller and John von Neumann, still followed the path to the quantum fount in Göttingen as repressive Eastern European regimes squeezed them out. The science continued, but this time the political climate rocked by high unemployment took a sharp turn to the political right rather than the left.

4. – Exile (1933)

Starting in April 1933, life in Göttingen and in other parts of Germany gradually came to an end for most Jewish professors. The Nazi regime dismissed them from their positions and barred them from entering classrooms. On May 10, the day of the book bonfires, the Borns took a train to the South Tyrol—Selva Gardena—leaving Göttingen behind to await judgment on his future. Who would have him?

A letter from Born to physicist Paul Ehrenfest in Leiden conveys some small piece of the impact of this moment.

> "This is a completely new and wonderful experience — the high mountains in the spring. I feel like thanking the powers in Berlin for making this possible. ... We forget all the evil we have experienced and take a deep breath at night and breathe the good air. In the morning I walk for a couple of hours and have time to ponder the meaning of what we've been through. Anger and bitterness slowly disappear except at night. They come back in nightmares and in long hours of wakefulness" [10].

Clearly trying to put the situation in perspective and weigh his worth, Born soon wrote to Ehrenfest again:

> "At the beginning of quantum mechanics I had the experience that bold things occur to me (*e.g.* the matrices, the confidence relationships, the perturbation calculations, the transformation SUS^{-1}, etc., the shares of Heisenberg, Jordan, and me are rather equal, the formulas about 90% mine)" [11].

Perhaps this recounting helped Born emotionally, but Ehrenfest, although living safely in the Netherlands, was beyond words. He committed suicide two months later.

In June, Heisenberg sent Born a letter that Born described as the "well-formulated attitude of well-meaning German colleagues" [12]. Heisenberg wrote that he was shocked to learn that Born did not want to return to Göttingen. He felt that Born must simply wait until the situation sorted out [13]. He clung to a future scenario where their lives were not affected adversely. "Surely, with time the ugly will be separated from the beautiful", and that "in the new political situation there are those who are well worth the wait" [14]. Heisenberg envisioned Born, Franck, and mathematician Richard Courant, returning to

a Göttingen unaffected by the political changes. Although some would certainly be affected, he said, it would only be a few.

For Born the tumult outweighed the science, but there was some. Since the essentials for his research were a pencil and paper, Born used them to delve into physics and push the outside world away. Reaching back to his research as a graduate student, when the electromagnetic mass of the electron stirred his imagination, he decided to apply quantum theory to the electromagnetic field and create a new field theory. His first foray was to investigate Maxwell's equations to determine their compatibility to quantum theory. In the midst of this, two students somehow managed to find him in Selva Gardena—one from Cambridge and one from South Africa. With them he created the "Selva University", and the threesome worked daily on the porch of the cabin or under the trees.

Later in the summer, much more of the scientific community found its way to Selva. Born assessed it as "nearly a mathematical physics conference". Visitors were:

- Hermann Weyl, mathematician, Göttingen (who was about to leave),
- Anny Schrödinger (Weyl's companion and Erwin Schrödinger's wife),
- Wolfgang Pauli, Zurich, and his sister,
- Max and Marga Planck, Berlin,
- Joe Mayer, physicist, Baltimore, (Maria Goeppert's husband),
- Arnold Eucken, physical chemist, Göttingen,
- Hermann Mark, physical chemist, Vienna,
- Frederick Lindemann, physicist, Oxford (later Lord Cherwell and Churchill's advisor)
- Artur Schnabel and Therese Behr, pianist and singer, Tremezzo, Lake Como.

Quite a diverse collection for the rustic village: Jew, non-Jew, German, Austrian, British, American, chemist, physicist, mathematician, pianist, singer. All were safe by virtue of nationality or academic position, except for Mark, a Jew, who later drove out of Austria flying a Nazi flag on his car and secreting his wealth in coat hangers made from platinum that he had created.

One can only guess at the conversation. Perhaps for the first time with such an august gathering of physicists, science was on the back burner. The real work was finding safe positions for young colleagues. The sequela was that many of Germany's most promising scientists would leave.

Lindemann had come from England specifically to pluck talented young physicists. He traveled from university to university in his chauffeured touring car. Earlier Born had written him about finding positions for his Göttingen assistants and students — as he said those "under his wing" [15]—such as Edward Teller, Lothar Nordheim, Fritz London, and Walter Heitler. Born was especially concerned for Edward Teller, writing

earlier to Ehrenfest, "Teller is an unusually talented person who deserves every support, and he is already weighted down enough by being a Hungarian Jew" [16]. To Lindemann he wrote, "take Teller's name to heart" [17]. Lindemann did find temporary positions for Fritz London and Edward Teller in England. Heitler went to Bristol when Born persuaded an anti-Nazi "Aryan" student, Martin Stobbe, to give him his position. (Born then had to place Stobbe). Nordheim, the last one standing, finally found a temporary position in Holland. Except for Heitler, who ended up in Dublin, the others eventually received offers from universities in the US. Almost all the physicists who initially worked in England moved to positions in the US. The British had almost no ability or, perhaps, desire to retain them.

Born himself received a number of offers and chose a position at Cambridge University limited to three years. Correspondence with Ralph Fowler, the chair of theoretical physics at the Cavendish Library, suggests that the position was created specifically for Born.

Once in Cambridge, Born continued his research on nonlinear electrodynamics with the young Polish physicist, Leopold Infeld. However, the plight of German-Jews absorbed much of his time. He became the physics representative for the Swiss-based Emergency Committee for German Scientists in Exile and for the British Society for the Protection of Science and Learning. Sometimes these duties and his emotions overwhelmed him. He wrote to Einstein, "Almost every week some unfortunate wretch approaches me personally, and every day I receive letters from people left stranded" [18]. He continued to search for any opportunity—Turkey, Russia, China, India, anywhere—for the younger generation of scientists and he placed many of them.

Besides the expulsion from Göttingen and worry about the safety of family and friends, the year 1933 further reinforced Born's sense of loss. Heisenberg, Schrödinger and Dirac won Nobel Prizes for their contributions to quantum theory.

Heisenberg wrote Born a conciliatory letter, which he posted from Switzerland.

> "The fact that I am to receive the Nobel Prize alone, for work done in Göttingen in collaboration — you, Jordan and I — this fact depresses me and I hardly know what to write to you. I am, of course, glad that our common efforts are now appreciated, and I enjoy the recollection of the beautiful time of collaboration. I also believe that all good physicists know how great was your and Jordan's contribution to the structure of quantum mechanics — and this remains unchanged by a wrong decision from outside" [19].

Born later wrote to Michael Polanyi about the issue of priority, noting that he was,

> "the first to write down a real quantum-mechanical formula, not only $pq-qp = h/2\pi i$ but also $q' = \partial H/\partial p$, $p' = \partial H/\partial q$ as matrix equations. [...] I think you will not mind if I inform you privately. For we are both in the about the same position, without the backing of a big nation" [20].

Born did not blame Heisenberg for the confusion. Polanyi did.

5. – The atom bomb (1945)

When the atom bomb was dropped, Born wrote to his son: "What have they done to my beautiful science?" Science had given Born both a sense of optimism and a sense of order. He believed all of his life that "Science was a noble pursuit like philosophy and art and true religion". He acknowledged to his son that this belief was not easy to sustain when science was, as he said, "perverted to destruction and hatred" [21]. One of the ironies for Born was that many of the top researchers in Los Alamos were his former students and collaborators, including Edward Teller, Eugene Wigner, John von Neumann, Klaus Fuchs, and Robert Oppenheimer, who earned his Ph.D. with Born in Göttingen in 1927 and had been a close collaborator while there.

The Americans and British had great respect for the tremendous talent in Germany and understood that the bomb had to be developed before the Germans did so. Many of those working in the Manhattan project were German and only a few years earlier had been close friends and colleagues with those working on Hitler's bomb project. Born begrudgingly knew that it had to be done. In thinking back, though, he chastised himself for not having stressed to his students the moral responsibility involved in scientific research.

6. – The Nobel Prize (1954)

Born ultimately received the Nobel Prize for his pioneering work in quantum theory, specifically for the discovery of the statistical interpretation of Schrödinger's wave function — a wave that was not a continuous cloud of electric charges as Schrödinger had argued but rather, as Born theorized, a cloud of the probability of finding a particle in a certain state after a collision. Born was particularly pleased that the Royal Swedish Academy honored this discovery because it was one that he had found completely on his own.

7. – Conclusion (1970)

Born's contemporaries and his students spun helplessly with him in the same dramatic web. Over the course of more than 30 years, they all took different routes to keep their lives together and continue their science. At the end of his life, Born summed up his philosophy on science and humanity:

> "I am convinced that ideas such as absolute certitude, absolute exactness, final truth, and so on are figments of the imagination which should *not* be admissible in any field of science. On the other hand, any assertion of probability is either right or wrong from the standpoint of the theory on which it is based. *This loosing of thinking* seems to me the greatest blessing which modern science has given us. For the belief in a single truth and in being the possessor thereof is the root cause of all evil in the world" [22].

* * *

I would like to thank Joseph D. Martin, Ph.D. at Michigan State University, for his review of this lecture and his very helpful comments.

APPENDIX A.

A few of the scientists saved by Max Born during war and the aftermath with some of their scientific accomplishments

1. **Erwin Madelung**: quantum theory/ Madelung Constant, Madelung Equations, Madelung Rule

2. **Alfred Landé**: quantum theory/ Landé g-factor; Landé interval rule

3. **Rudolf Minkowski**: astronomer/ supernovae: Apollo asteroid 1620 Geographos; Planetary Nebula M2-9; Minkowski (moon crater)

4. **Fritz Reiche**: quantum theory/Thomas-Reiche-Kuhn sum rule; supersonic flow

5. **Otto Stern**: (Nobel Prize) quantum theory/ Stern-Gerlach effect; atomic magnetic moment; molecular beam epitaxy

6. **Max Wertheimer**: psychologist/co-founder of Gestalt therapy

7. **Erich von Hornbostel**: ethnomusicologist/pioneer in this field; timing used in sound ranging

8. **Paul Epstein**: quantum theory/interpretation of the Stark effect and the Zeeman effect

9. **Edward Teller**: nuclear and molecular theory/Gamow-Teller Effect, Jahn-Teller Effect, Renner-Teller Effect, and many more

10. **Fritz London**: physical chemistry-quantum theory/London equations, London moment, London dispersion forces

11. **Lothar Nordheim**: quantum theory/Fowler-Nordheim tunneling, Fowler-Nordheim-type equations, field electron emission

12. **Martin Stobbe**: He may have died during WWII

13. **Walter Heitler**: quantum theory/quantum electrodynamics, quantum field theory

Appendix B.

German scientists whose research during these turbulent years earned Nobel Prizes.

1. James Franck and Gustav Hertz(1925): discovery of the laws governing the impact of an electron upon an atom
2. Max Born (1954): statistical studies of atomic wave functions
3. Walther Bothe (1954): invention of the coincidence method
4. Werner Heisenberg* (1932): creation of quantum mechanics
5. Erwin Schrödinger (1933): introduction of wave equations in quantum mechanics
6. Otto Stern* (1943): discovery of the magnetic moment of the proton
7. Wolfgang Pauli* (1945): discovery of the exclusion principle of electrons
8. Eugene Wigner* (1963): principles governing interaction of protons and neutrons in the nucleus (Maria Goeppert Mayer* also received the Nobel Prize in 1963 for a different discovery. She was a student of Born's with Wigner. She did her Nobel Prize work in the 1940s.)

*Students of Born.

REFERENCES

[1] Forman P., *Hist. Stud. Phys. Sci.*, **3** (1971) 1.
[2] Born M., *Experiment and Theory in Physics* (Dover, New York) 1956, p. 109.
[3] Einstein A. to Born M., in *The Born-Einstein Letters*, translated by Newton-John I. (Walker, New York) 1971, p. 21.
[4] Born M. to Oseen C., 15 May 1919, Oseen Archives, Center for the History of Science, Royal Swedish Academy of Sciences, Stockholm.
[5] Born M. to Langmuir I., 11 May 1920, Irving Langmuir Papers, Library of Congress, Washington DC.
[6] Einstein A., in *The Born-Einstein Letters*, translated by Newton-John I. (Walker, New York) 1971, p. 50.
[7] Born M. to Einstein A., in *The Born-Einstein Letters*, translated by Newton-John I. (Walker, New York) 1971, p. 54.
[8] Born M. to Einstein A., in *The Born-Einstein Letters*, translated by Newton-John I. (Walker, New York) 1971, p. 31.
[9] Born M. to Still C., 8 December 1920, Carl Still Papers, Still family private collection.
[10] Born M. to Ehrenfest P., 19 May 1933, Archive for the History of Quantum Physics, American Institute of Physics, College Park MD.
[11] Born M. to Ehrenfest P., 20 Jul 1933, Archive for the History of Quantum Physics, American Institute of Physics, College Park MD.
[12] Born M. to Ehrenfest P., 11 June 1933, Archive for the History of Quantum Physics, American Institute of Physics, College Park MD.

[13] BORN M. to JAMES FRANCK, 27 May 1933, Archive for the History of Quantum Physics, American Institute of Physics, College Park MD.
[14] HEISENBERG W. to BORN M., enclosure in Born M. to Ehrenfest P., 11 Jun 1933, Archive for the History of Quantum Physics, American Institute of Physics, College Park MD.
[15] BORN M. to EHRENFEST P., 10 May 1933, Archive for the History of Quantum Physics, American Institute of Physics, College Park MD.
[16] BORN M. to EHRENFEST P., ibid.
[17] BORN M. to LINDEMANN F., 2 June 1933, Frederick Lindemann Papers, Nuffield College, Oxford University.
[18] BORN M. to EINSTEIN A., 8 March 1934, in *The Born-Einstein Letters* translated by NEWTON-JOHN I. (Walker, New York) 1971, p. 119.
[19] BORN M., *My Life: Recollections of a Nobel Laureate* (Charles Scribner's Sons, New York) 1978, p. 220.
[20] BORN M. to POLANYI M., 30 January 1942, Michael Polanyi Papers, University of Chicago Library.
[21] BORN M. to BORN G., 2 September 1945, Private papers and letters of the Born family.
[22] BORN M., *Physics in My Generation* (Springer-Verlag, New York) 1969, pp. 182-183.

Proceedings of the International School of Physics "Enrico Fermi"
Course 197 "Foundations of Quantum Theory", edited by E. M. Rasel, W. P. Schleich and S. Wölk
(IOS, Amsterdam; SIF, Bologna) 2019
DOI 10.3254/978-1-61499-937-9-15

But God does play dice: The path to quantum mechanics

Nancy Thorndike Greenspan(*)

Washington DC, USA

Summary. — Quantum mechanics has many fathers. The contributions of some have been lost to its story in part because of political events, the personalities of the fathers, and the overarching Copenhagen Interpretation that highlighted the ideas of Niels Bohr and Werner Heisenberg. One physicist who made fundamental contributions but who is little acknowledged is Max Born. His mathematical formulations provided much of the basis for the solution as well as the interpretation and completion of Heisenberg's mathematical insight.

Introduction

Max Born was one of the seminal thinkers in quantum theory. Much of what he created became part of the Copenhagen Interpretation of quantum mechanics. Because of this, his singular importance has sometimes been overlooked. But Born spent 15 years, 1912-27, searching for a quantum theory and was instrumental in finding one. Here is his story.

(*) Nancy Thorndike Greenspan is a writer based in Washington DC. Her most recent book is *The End of the Certain World: The Life and Science of Max Born.*
E-mail: nancy@nancygreenspan.com

1. – Breslau, Germany (now Wroclaw, Poland)

A bit of background. . .Born (1882 to 1970) grew up in Breslau, Germany, the grandson of a wealthy textile manufacturer. As a child, Born was surrounded by music and art. His mother's family was integral to the city's cultural life and generously funded the arts. Whenever musicians were in town—Johannes Brahms, Richard Strauss, Klara Schumann, Joseph Joachim—they stayed with and entertained family members. His father Gustav was a beloved professor of embryology at the University of Breslau, who before his early death hypothesized accurately about the role of the corpus luteum. Although both families were Jewish by birth, they were as assimilated as possible.

In his school days Born was a very average student. His bad memory hindered his academic interests because the academics, including mathematics, relied on rote learning. Consequently math became a strong dislike. After graduating from the Gymnasium, he went to the University of Breslau, where first he studied philosophy. He found it a bad fit. A friend persuaded him to try mathematics. Much to his surprise, he discovered that the abstractions and analytical thought, so distinct from the earlier exercises, excited him. He had found his strength.

2. – Göttingen

In 1907, the University of Göttingen awarded Born a Ph.D. in mathematics. Initially his ambition was to be a mathematician in the mold of his mentors David Hilbert and Hermann Minkowski, two of the great mathematicians at the turn of the 20^{th} century. But early on he failed to work out the proof of the transcendental character of the roots of the wavelike Bessel function([1]), and his seeming lack of creativity in pure mathematics confronted him. For his dissertation and thereafter, he refocused on applied mathematics and the physical world, but he never forgot Hilbert's dictum: "a perfect formulation to a problem is already half its solution" [1]. For the rest of his career, Born stripped the problem of nonessentials, simplifying without losing the core, and solved. First principles were his base.

The mystery of quanta was not Born's first scientific quest. His early work centered on crystal dynamics, a research area as important to a future quantum theory as that of spectral lines [2].

His first inkling that the life of quanta needed further exploration came in 1912 when he and his housemate Theodor von Kármán—later famous for characterizing airflows in aerodynamics—wrote a theoretical paper on the specific heat of solids. Born created a detailed, repeating three-dimensional model of a solid structure that included all the basic concepts of lattice dynamics. (The real inner structure of a solid was not yet known.) The theoretical results fit experimental ones in all temperature ranges for the first time. In the conclusion Born wrote that he could not "suppress certain doubts" [3]

([1]) That no algebraic equation with integral coefficients exists with π as its root.

about the assumptions underlying the quantum theory of radiation. He knew that there must be something deeper — exactly what was not clear to him.

This doubt was one of two that drew Born's attention to quantum theory. The second grew out of another paper on crystal dynamics. In this research Born realized that the Bohr-Sommerfeld ring model of the atom, visually similar to the Solar System, was not correct. He wrote, "The planar electron orbits are insufficient, atoms are seemingly spatial objects" [4]. He needed a space-filling, three-dimensional model such as a cube. Later Born recalled that that result was the "second hint given by lattice dynamics that in atomic dimensions we ought to face the fact that quite a new mechanics was needed" [5].

3. – Frankfurt

After WWI, the University of Frankfurt called Born to the chair of theoretical physics. He did not involve himself in quantum research there, but he did keep up on the field, giving a seminar on it and studying Arnold Sommerfeld's book *Atomic Structure and Spectral Lines*. However, even though he respected Sommerfeld's accomplishment, he did not like his method. Born considered it guessing. In Hilbertian fashion, Born wanted to develop a general theoretical framework, asking questions that led to mathematical solutions as he had done in crystal dynamics.

In 1919, Born invited Wolfgang Pauli to work with him in Frankfurt. Pauli was then in Munich studying with Arnold Sommerfeld. In his letter Born wrote:

> "You regard the application of the continuum theory to the interior of the electron as meaningless because it is principally not a question of observable things. I have pursued just these thoughts for a long time, certainly until now without positive success, namely that the path out of the quantum difficulty has to be sought from quite principled points: one is not allowed to transfer the concepts of space and time as a 4-dimensional continuum from the macroscopic experience to the atomic world which demands obviously another type of number manifold for an adequate picture" [6].

Born harkened back to the inconsistencies of Bohr's model and reiterated his own approach: to throw out "superfluous elements and describe as simply as possible" [7]. He knew that the macroscopic continuum of the classical world did not describe the microscopic world of the electron.

On November 27, 1920, Born wrote to the US chemist Gilbert Lewis:

> "One thing shines out of the chaos: the classical relationship between the motion of charged particles (nuclei and electrons) and radiation is wrong and must be described by a *Mittlewert Erscheinung*, the true laws are quantum laws that unfortunately we know only little about" [8].

4. – Göttingen again

In 1921, Born moved from Frankfurt to Göttingen to become head of the Theoretical Physics Institute at the University. He again invited Pauli to work with him, and this time Pauli accepted. He was an ideal partner for furthering Born's goal of exploring basic quantum principles. In his recently completed dissertation under Sommerfeld, Pauli had applied Bohr's quantum conditions to the hydrogen molecule ion — a three-body problem. His results did not match experimental ones given by spectral lines. Born and Pauli attacked another three-body problem, applying perturbation theory to the helium atom. Again, it was unsuccessful.

Pauli's assistantship with Born lasted only about 6 months. Rural Göttingen was just too quiet for city-loving Pauli. But probably as limiting was Pauli's dislike of Born's mathematical approach. Pauli's own was more physical and intuitive. In later years, Pauli never missed an opportunity to deliver a jab at Born's proclivity.

A few months later, in June 1922, Niels Bohr came to Göttingen for two weeks of lectures on his quantum theory. History remembers it as the *Bohr Festspiele*. About 100 German physicists attended, one of whom was Arnold Sommerfeld with his young protégé Werner Heisenberg. Quickly assessing Heisenberg's talents, Born arranged for him to work in Göttingen while Sommerfeld was on sabbatical. Niels Bohr saw the same qualities in Heisenberg, but his invitation to Copenhagen came too late.

Stimulated by those two weeks of intense thought and discussion, Born immediately wrote a short article, "On the Model of the Hydrogen Molecule". Its two-pages boldly signaled his future research intentions on quantum theory.

> "The time is perhaps past when the imagination of the investigator was given free rein to devise atomic and molecular models at will. Rather, we are now in a position to construct models with some certainty, although still by no means complete certainty, based on the quantum rules" [9].

Clarity, energy, and a view to the future stamped Born's remarks.

That fall, with Heisenberg, Born re-launched his plan to systematically determine where Bohr's quantum theory could predict experimental results. His goal was to examine a version of the three-body problem more sophisticated than the one with Pauli. He wanted to use the excited helium atom to explore the behavior of electrons in aperiodic motion([2]).

([2]) Born M., *Problems of Atomic Dynamics* (Frederick Ungar Publishing Co., New York) 1960, p. xiii-xiv. Born's terminology of "aperiodic" motions or processes may be unfamiliar to modern-day physicists. He is referring simply to unbound situations, such as the scattering problem, or the continuum states above $E = 0$ for the hydrogen atom. These are "aperiodic" in the sense that there is no Bohr quantization condition based on angular momentum. They were crucial in leading Born to the statistical interpretation of the wave function amplitude as a probability density.

To expand his own knowledge of perturbation theory and to introduce his students to it, Born started a Monday night seminar to study Poincaré's celestial mechanics. Then he and Heisenberg applied Poincaré's theory to confirm Bohr's buildup of the periodic table. Finally they turned to systematically exploring the excited helium atom. Conclusion: They found no agreement with experimental results. A further confirmation for Born on the need for a new quantum theory.

In the summer of 1923, Born wrote to Einstein, "As always I am thinking hopelessly about quantum theory, trying to find a recipe for calculating helium and the other atoms; but I am not succeeding in this either" [10].

A few months later, Born offered a new course, "Perturbation Theory Applied to Atom Mechanics". These lectures later became his first book on quantum theory. True to his word about the demise of the old quantum theory, he began the lectures by proposing a major new approach — to incorporate the electron's transitions from one stationary state to the next by mathematically acknowledging the discrete nature of the jump. Rather than using differential equations, which assume continuity, he replaced them with difference formulas for discontinuities. He made the jumps an integral part of the theory.

There was probably some cross-fertilization here. At the same time as Born's seminar, his good friend and colleague Richard Courant, head of mathematics at the University, was holding a seminar on difference equations. The idea of continuity *versus* discontinuity had also regained focus with Arthur Holly Compton's research on the wave-particle duality of light. Heisenberg reported to Pauli that he was incorporating Born's new mathematical formulation into his own quantum model.

In Copenhagen a few months later, Niels Bohr and his colleague Dutch physicist Hendrik Kramers, along with the young American post-doctoral student John Slater, were working out a new theory. Within the trio, disagreement brewed. Slater proposed that energy and momentum are conserved in each absorption and emission process. Bohr insisted that they were only conserved statistically over a large number of processes. This assumption allowed him to avoid accepting light quanta, for which he had distaste. (By this time it was strongly implied by the Compton Effect.)

The ultimate importance of the Bohr-Kramers-Slater theory lay in Slater's concept of a virtual radiation field originating from virtual oscillators. Kramers capitalized on this idea by incorporating it into his dispersion theory. He relied on atomic transitions rather than the non-observable orbits of electrons. Born followed up on these ideas by publishing "On Quantum Mechanics" in which he laid out a quantum discretizing rule using difference equations. Optimism overtook Born as he wrote to a friend that they would soon "get to the bottom of it" [11].

Heisenberg went to Copenhagen to work with Bohr for the winter of 1924-25. Ph.D. student Pascual Jordan became Born's new assistant. Their immediate research objective was to introduce the concept of transition quantities into Planck's original 1900 quantum theory. From there, they planned to investigate collisions.

They were just finishing their article, "Quantum theory of Aperiodic Processes", when Heisenberg returned from Copenhagen with some new ideas. These, he did not share.

But discussions did ensue among the three on Born's and Jordan's new realization. They had discovered that the number of quantum jumps (the transition probabilities) related to the observed spectral line intensities, the squares of their amplitudes. Born also proposed that one could formulate transition amplitudes (later interpreted as the matrix elements of the position and momentum operators). He thought this idea might be central to a new theory. Observability was also a striking feature of the Born/Jordan article:

> "A postulate of great reach and fruitfulness states that only such quantities should enter into the true laws of nature which are in principle observable and measurable" [12].

Observability was not a new idea in quantum theory. Pauli had talked about it for years, as had Born, and in a dispersion paper, Kramers had included only observable quantities. But Born and Jordan made it explicit for the first time.

All of this was intellectual fodder for Heisenberg. Working on his own, he came up with a strange multiplication rule. Again he did not share. Instead in late-May, he left for Helgoland to rest, recover from hay fever, and clear his thoughts. While there, he was inspired to use his multiplication rule to derive the energies of the quantum stationary states and compare them with the observed spectral line intensities. Amazing to him, the two agreed.

When Heisenberg returned to Göttingen in early July, he handed his article to Born and asked him to review it. He explained that he was not sure what it all meant and whether it was nonsense or not. He needed Born's advice and insight.

Born read the article and puzzled over the mathematical configurations. He recognized that Heisenberg had created multiplication rules for the transition amplitudes. He did not immediately recognize the mathematical technique. Restless and not sleeping, he continued to ponder. Tracing the logic and rewriting Heisenberg's version of Bohr's quantum conditions for p (momentum) and q (position), Born realized that the product of pq did not equal qp. They did not commute. Heisenberg had used matrix multiplication unknowingly. At that time he did not what matrices were.

Born then left for a conference where he saw Pauli. He asked Pauli to collaborate on understanding Heisenberg's formulation. Pauli declined characteristically, saying, "I know you are fond of tedious and complicated formalism. You are only going to spoil Heisenberg's physical ideas by your futile mathematics" [13].

Back in Göttingen, Born noted the assignments for Jordan and his other assistants in the July 23 entry of his Daybook. He specified for himself "Heisenberg's Quantum Mechanics". Under "new problems", he wrote: "Connection between de Broglie's theory and the Duane-Compton derivation of [x-ray] interference" [14]. With that entry, he stopped writing in his Daybook — and he stopped thinking about de Broglie's wave theory. For about the next four days, Heisenberg's article with its symbolic multiplication of transition amplitudes became his intense focus. Then he sent it off to the *Zeitschrift für Physik*.

From there, he did not look back. Using Heisenberg's formula, he found that the value of the diagonal elements $\mathbf{pq} - \mathbf{qp}$ equaled $h/2\pi i$. As for the off-diagonal elements,

he arrived at the conclusion that they were zero. The fundamental commutation law of quantum mechanics was, therefore, $\mathbf{pq} - \mathbf{qp} = h/2\pi iI$. Born later considered his discovery of this formula to be the "climax of my research" [15].

Quickly Born worked to expand Heisenberg's insight into a more comprehensive treatment. The result was a joint article with Pascual Jordan, who had worked at Born's side throughout. In late October, the two reunited with Heisenberg and launched into expanding the features of quantum mechanics in their so-called "three-man" paper.

But Jordan and Heisenberg had to finish the article without Born. During the summer, not foreseeing that the answer to the quantum puzzle was quite so near at hand, Born had scheduled a trip to America for the fall of 1925. Leaving the research behind was wrenching for him.

Just before departing, he wrote to Niels Bohr and expressed his pleasure at Heisenberg's discovery:

> "I am so glad that Heisenberg's idea pleases you. I believe with complete certainty that it signifies great progress and that the form, which Jordan and I have given it, is in a certain sense somewhat final, so far as one can say that at all in physics. For me, the possibility of this formulation has an entirely personal charm. Since my student times, I suffer with an idée fixe, that is to say, that all significant laws of physics must find their adequate formulation as invariants of linear substitutes. [...] Everything I did myself, *e.g.* the contribution on lattice theory of the crystal, always ran out of this" [16].

5. – America

Born arrived in New York City on November 11, 1925. During the next four months and covering 6000 miles, he crisscrossed the US spreading the new quantum gospel to important centers of physics. Thousands of people, scientists and non-scientists, were introduced to the new theory from Göttingen. (See table I.)

This exposure explains in part why so many young American physicists soon travelled to Born and Göttingen. It became, as Karl Compton later dubbed it, "the fount of quantum wisdom" [17].

On Saturday, November 14, 1925, Born arrived in Cambridge, MA and MIT. During his two-month stay he gave 30 lectures—the first 10 titled "The Lattice Theory of Rigid Bodies" and the next 20 "The Structure of the Atom". Born's excitement about the theory growing, he described the lectures as "successful, almost sensational" [18]. In April 1926, MIT published them as *The Problems of Atomic Dynamics*. It was the first book to present the new quantum theory.

Besides the lectures at MIT and others at Harvard, intense discussions about the new theory at lunch, dinner, or on a walk filled Born's days. Ideas bubbled up. When he met the young MIT mathematician Norbert Wiener (now famous as the father of cybernetics), he thought that a more precise method of Fourier analysis developed by Wiener could be applied to the problems of the continuous spectrum, such as collision theory. Collision

TABLE I. – *Itinerary for 1925/1926 US trip.*

Arrival	*Places*	*Faculty/hosts*	
11/11/25	New York NY	Henry Goldman	
11/14/25- 1/22/26	***Massachusetts Institute of Technology***, Cambridge MA	Paul Heymans Norbert Wiener Pres. S. W. Stratton	Hans Mueller C. L. Norton
	Harvard University, Cambridge, MA	Theodore Lyman John C. Slater	Edwin C. Kemble Percy Bridgeman
12/24/25	***Rockefeller Foundation***, New York NY	Wickliffe Rose	
01/23/26	***General Electric Co.***, Schenectady NY	Willis Whitney David Coolidge Albert Hull Burt L. Newkirk	Irving Langmuir Saul Dushman W. H. Mott-Smith
01/26/26	***Cornell University***, Ithaca NY	Ernest Merritt Earle Kennard W. A. Hurwitz Frederick Bedell	Floyd Richtmyer Livingston Ferrand Virgil Snyder R. Clifton Gibbs
02/01/26	***University of Buffalo***, Buffalo NY	Cooke	Edward Moore
02/02/26	***University of Chicago***, Chicago IL	Henry Gale Arthur Dempster Albert A. Michelson	A. H. Compton Harvey B. Lemon
02/10/26	***California Institute of Technology*** Pasadena CA	Robert Millikan Svein Rosseland George Glocker Roscoe Dickinson Richard Tolman Edwin Hubble Wayne B. Hales	Paul Epstein Fritz Zwicky Joseph Mattauch Harry Bateman Dayton Clarence Miller Ernest Charles Watson Harold Delos Babcock
02/24/26	***University of California*** Berkeley CA	Gilbert N. Lewis Hertha Sponer Edward London	Leonard Loeb Frederick Brackett Pres. W. W. Campbell
03/07/26	***University of Wisconsin***, Madison WI	Max Dresden John Van Vleck John Roebuck Pres. Glenn Frank	Alexander Mieklejohn Leonard Ingersoll Harnack (nephew of Adolf) Hagans
03/12/26	***Columbia University***, New York NY	George Pegram Bergen Davis Harold Webb	Kronig Wills
	Rockefeller Foundation, New York NY	Wickcliffe Rose	
	General Electric Co., New York NY	Pres. Gerard Swope	
03/15/26	***Princeton University***, Princeton NJ	Karl Compton Dean West Henry Russel	Oswald Veblen James W. Alexander Sebastian Karrer
03/20/26	***Philosophical Society*** Washington DC	Gregory Breit Sebastian Karrer	Otto LaPorte

experiments in the lab of his colleague James Franck had triggered Born's interest in the interaction of matter with matter as well as of matter with radiation. This problem had seemed impossible to handle with matrices. In the three-man paper, Born had outlined the problem and an approach for replacing an element of a matrix referring to two discrete states by a function of two continuous variables. With the rules for the transition unknown, he could go no further.

By the end of November, Born and Wiener began to generalize matrices as linear operators using Wiener's method. They replaced the mathematics of discontinuous energy spectra with more general systems with continuous spectra. The article was ready by the Christmas holiday. Born was pleased with the results.

Then came a great surprise. In December 1925, Born received a reprint of the article "The fundamental equations of quantum mechanics", from the *Proceedings of the Royal Society*, written by someone he had never heard of, one Paul Dirac. Here was a formulation of quantum mechanics similar to the one that he, Jordan, and Heisenberg had just completed. How had Dirac done this? It turned out that in the fall Cambridge University physicist Ralph Fowler had received the galley proofs of Heisenberg's article on the new quantum formulation and sent them on to Dirac, asking for his opinion. Dirac's formulation was the answer.

Born gave his last lecture at MIT on January 22, 1926. One thousand people attended.

The General Electric Research Lab in Schenectady NY was Born's next stop. Greatly impressed after his three-day stay, he deemed it "the center of American physics; more even than Harvard" [19]. In the following nine days, Born lectured at Cornell, the University of Buffalo, and the University of Chicago. On February 10, he arrived in Pasadena CA for two weeks at the California Institute of Technology. Carl Eckart's recollection echoed those of many others:

> "I was not greatly attracted to this [the new quantum theory], but in the late winter of 1925 or early 1926 Born came to Pasadena, and his lucid lectures aroused my interest[...]. He had conferred with Norbert Wiener on operator calculus, and the interpretation of the commutative law for p and q as identical with the differentiation operator was very strongly emphasized in his lectures. The result was that I spent the spring of 1926 working rather intensively with this operator formulation and was completely familiar with what is now known as the Schrödinger operator (the energy operator) before Schrödinger's appeared in Pasadena [20].

At the end of the Pasadena lectures Born traveled north to the University of California at Berkeley to deliver the Hitchcock lectures. Then he began his trip back to the East coast. His one stop was the University of Minnesota where he dined with John Van Vleck, who traveled from Wisconsin, and his father, a mathematician there. Van Vleck was one of the few American physicists actively pursuing quantum physics in early 1926. He and Born had already been corresponding.

Back on the East Coast, there were more lectures at Columbia University and an important visit with Wickcliffe Rose of the Rockefeller Foundation, The Foundation was

just beginning a program to support the exchange of young physicists. The meeting with Born solidified funding for American students to study with him and his colleagues in Göttingen.

A March 19, 1926 article in *The New York Times*, "His Dynamic Theory Rivals Einstein's", heralded Born's next stop in Princeton. It quoted Karl Compton: "Born [...] has been particularly successful in stimulating and developing his younger colleagues and students, who have made Göttingen a great centre of activity in modern physics" [21].

By the time Born reached his final stop of Washington DC, he was exhausted and sick with the flu. He managed to lecture at the Philosophical Society but never made it to the Bureau of Standards. He sailed from New York on March 24, 1926. In his wake American science was turning upside down: he had lectured on the new quantum theory at twelve universities and research centers; he had inspired ideas in hundreds of physicist, young and old, and he had published the first book on the new quantum theory. E. C. Kemble, then a junior professor at Harvard, reviewed the book. He began with the statement: "it would perhaps be rash to say that the year 1925 marks the beginning of a new era in physics". He followed this miscalculation by describing the tremendous value of Born's lectures to American physicists [22].

Over the next few years many of those in Born's audiences made their way to Göttingen. Norbert Wiener from MIT, E. C. Kemble from Harvard, Earle Kennard at Cornell, E. U. Condon from Berkeley were but a few who arrived in Goettingen for the 1926 winter term. (Robert Oppenheimer shared this same time in Goettingen, but his inducement was Born's visit to Cambridge, England not Cambridge, Massachusetts.) Other physicists, for instance Karl Compton from Princeton, visited later in the year as did those from closer at hand, such as Paul Dirac, who, after the startling introduction with his article, became a close and lasting friend.

6. – Göttingen

Born weathered the surprise of the Dirac article only to be hit with another big one once back in Göttingen: Schrödinger's article on wave mechanics. First, it created anxiety in the advocates of quantum mechanics. Which formulation—waves or particles—held for the quantum world? Fortunately, in May, Schrödinger quickly resolved the confusion by mathematically demonstrating that the two forms were equivalent.

Then there was Born's deep frustration when he saw that he and Wiener had stopped just short of discovering wave mechanics. Using the complicated and nontransparent operator theory, they had developed a formula equivalent to Schrödinger's commutation law but applied it only to the variable time instead of to both time and momentum. Seeing his short-sightedness, Born considered this lapse as "the most outstanding example of my being quite close to an important discovery and letting it slip by" [23].

In hindsight the work with Wiener does not seem for naught. It must have sharpened Born's curiosity about waves. In any case, he recovered from his trip and from his lapse and started thinking about Schrödinger's wave equation. He was inspired to examine the result of a free particle, such as an electron, colliding with an atom. He saw that

Schrödinger's equation allowed a description of stationary states as well as quantum jumps. And he found a probabilistic outcome to the collision. In his article in June 1926, he wrote: "One gets no answer to the question", what is the state after the collision", but only to the question, "how probable is a specific outcome of the collision". His initial thinking about a probabilistic outcome? That it was, "a philosophical question for which physical arguments alone are not decisive" [24].

In his second paper, however, he warmed to the idea of an indeterminate outcome and expanded his thoughts. He explicitly contradicted Schrödinger's interpretation that electron waves were clouds of electricity by theorizing that they were the probability of finding a particle in a certain place after a collision. Inspired by Einstein's earlier representation of waves as guides for light quanta, he envisioned "ghost fields" [25]. And ghostly it was. Born's statistical interpretation was the death knell of causality. It ended the reign of determinism in physics—for most. But not for all.

Einstein and Schrödinger, in particular, never accepted this new reality. Schrödinger was so distressed with Born's interpretation that at one point he wished that he had never written his original article.

In March 1926, prior to the publication of Born's article on Schrödinger's wave function, Einstein commented on quantum mechanics in a letter to Born's wife: "The Heisenberg-Born concepts leave us all breathless, and have made a deep impression on all theoretically oriented people" [26]. It is Einstein's only comment on the initial stages of quantum mechanics in letters with the Borns. After Born's statistical interpretation, Einstein called on a higher order to buttress his objections. In his now famous quote he wrote to Born that the statistical interpretation was:

> "certainly imposing. But an inner voice tells me that it s not yet the real thing. The theory says a lot, but does not really bring us closer to the secret of the 'Old One'. I, at any rate, am convinced that *He* does not play dice" [27].

Born had anticipated that some physicists would argue for a deeper physical level that maintained cause and effect. In his second paper, he wrote that some would "assume that there are other parameters, not given in the theory, that determine the individual event" [28]. Given his own reliance on Einstein's early insight in his creating this theory, he may not have thought his good friend would be the one.

The controversy over the statistical nature of quantum mechanics came to a head in October 1927 at the Solvay Conference in Brussels. Born and Heisenberg made a joint presentation, reviewing its basic features. Their pronouncements were unequivocal.

> "We regard quantum mechanics as a complete theory for which the fundamental physical and mathematical hypotheses are no longer susceptible of modification... Our fundamental hypothesis of essential indeterminism is in accord with experiment. The subsequent development of the theory of radiation will change nothing in this state of affairs" [29].

The exchanges between the two main combatants, Einstein and Bohr, became mythic

in the physics community but they did not create unity. Both left standing, neither bowed.

The Solvay Conference became the springboard for the Copenhagen Interpretation, a theory prominently emphasizing Bohr's complementarity and Heisenberg's uncertainty principle. It cemented the importance of discontinuities embodied in matrix mechanics and the observer's role in quantum measurement. Believers, especially Heisenberg and Pauli, proselytized under the rubric of The Copenhagen Interpretation, sweeping in all components including Born's. His contribution of the statistical interpretation of the wave function, which established the basis for uncertainty, became almost subsumed—even after the Nobel Prize committee recognized his achievement in 1954.

Einstein, for one, did not forget its origins nor consider it trivial. At the time of the award, Einstein tried to be gracious without capitulating. He wrote,

"it was your [...] statistical interpretation of the description (of quantum theory) which has decisively clarified our thinking. It seems to me that there is no doubt about this at all, in spite of our inconclusive correspondence on the subject" [30].

No uncertainty existed for Born. On his tombstone is chiseled:

$$\mathbf{pq} - \mathbf{qp} = h/2\pi i.$$

* * *

I would like to thank Joseph D. Martin, Ph.D., Michigan State University, for his review of this lecture and his very helpful comments.

REFERENCES

[1] COURANT R., "Reminiscences from Hilbert's Göttingen", transcript from a tape made at Yale University (1964) p. 12, Richard Courant Papers, Bobst Library, New York University.
[2] JEREMIAH J. and CHRISTIAN J., *Hist. Stud. Nat. Sci. 45*, **5** (2015) 641.
[3] BORN M. and VON KARMAN T., *Phys. Zei.*, (1912) 297.
[4] BORN M. and LANDÉ A., *Ver. deut. Phys. Gesell.*, **20** (1918) 210.
[5] BORN M., *Lattice Dynamics: Proceedings of the International Conference Held at Copenhagen, Denmark August 5-9*, 1964, (Pergamon Press, New York) pp. 1-7.
[6] BORN M. to PAULI W., *Wolfgang Pauli: Scientific Correspondence with Bohr, Einstein, Heisenberg, A. O.* ***1***, edited by HERMANN A. *et al.* (Springer-Verlag, New York) 1979, p. 10.
[7] Landé A. interview on 3 July 1962, Archive for the History of Quantum Physics, American Institute of Physics, College Park MD, p. 25.
[8] BORN M. to LEWIS G., 27 November 1920, Gilbert Lewis Papers, Bancroft Library, University of California, Berkeley.
[9] BORN M. *Die Naturwissen.*, **10** (1922) 677, as translated in CASSIDY D., *Uncertainty: The Life and Science of Werner Heisenberg* (W.H. Freeman, New York) 1992, p. 146.
[10] BORN M., in *The Born-Einstein Letters*, translated by NEWTON-JOHN I. (Walker, New York) 1971, p. 81.

[11] Born M. to Landé A., 14 July 1924, Archive for the History of Quantum Physics, American Institute of Physics, College Park MD.
[12] Born M. and Jordan P., *Z. Phys.*, **33** (1925) 479.
[13] Born M., *My Life: Recollections of a Nobel Laureate* (Charles Scribner's Sons, New York) 1978, p. 218.
[14] Born M., *Personal Daybook*, 23 July 1925, Private papers and letters of the Born family.
[15] Born M., Application for membership to the Royal Society, London.
[16] Born M. to Bohr N., in *Niels Bohr Collected Works, Volume 5: The Emergence of Quantum Mechanics (mainly 1924–1926)*, edited by Stolzenburg K. (Elsevier, Amsterdam) 2008, pp. 311-313.
[17] Karl Compton, *Nature*, **139** (1937) 238, quoted in Cassidy D., *Uncertainty: The Life and Science of Werner Heisenberg* (W.H. Freeman, New York) 1992, p. 263.
[18] Born M., Outline for chapter in *My Life: Recollections of a Nobel Laureate*, Private papers and letters of the Born family.
[19] Born M. to Born H., 25 January 1926, Private papers and letters of the Born family.
[20] Eckart C. interview with Heilbron J., 1962, Archive for the History of Quantum Physics, American Institute of Physics, College Park MD.
[21] "His Dynamic Theory Rivals Einstein's", *New York Times*, 20 March 1926.
[22] Kemble E. C., *Phys. Rev.*, **28** (1926) 423.
[23] Born M., *My Life: Recollections of a Nobel Laureate* (Charles Scribner's Sons, New York) 1978, p. 227.
[24] Born M., *Zei. Phys.*, **37** (1925) 863, as translated in *Quantum Theory and Measurement*, edited by Wheeler J. and Zurek W. H. (Princeton University Press, Princeton) 1983, pp. 52-61.
[25] Born M. to Einstein A., 30 November 1926, quoted in Pais A., *Science*, **218** (1982) 443.
[26] Einstein A. to Born H., in *The Born-Einstein Letters*, translated by Newton-John I. (Walker, New York) 1971, p. 88.
[27] Einstein A. to Born M., in *The Born-Einstein Letters*, translated by Newton-John I. (Walker, New York) 1971, p. 91.
[28] Born M., *Zei. Phys.*, **38** (1926) 803, as translated in Ludwig G., *Wave Mechanics* (Pergamon Press, New York) 1968, pp. 206-226.
[29] Cassidy D., *Uncertainty: The Life and Science of Werner Heisenberg* (W.H. Freeman, New York) 1992, p. 250.
[30] Einstein A. to Born M., in *The Born-Einstein Letters*, translated by Newton-John I. (Walker, New York) 1971, p. 229.

This page intentionally left blank

Proceedings of the International School of Physics "Enrico Fermi"
Course 197 "Foundations of Quantum Theory", edited by E. M. Rasel, W. P. Schleich and S. Wölk
(IOS, Amsterdam; SIF, Bologna) 2019
DOI 10.3254/978-1-61499-937-9-29

From the Bohr model to Heisenberg's quantum mechanics

M. Kleber(*)

Department of Physics, TU München - Germany

Summary. — The development of quantum mechanics is inseparably connected with Niels Bohr and Werner Heisenberg. We review the period during which the Bohr model was developed, reached its limits, and was finally replaced by Heisenberg's quantum mechanics. We show how the theory was obtained by a team of brilliant scientists. In this lecture we bring together historical aspects and mathematical details.

1. – Introduction

This is a short story of the development of quantum mechanics. The theory was worked out in the first half of the last century and came into existence through intensive work of the world's brightest scientists who were part of a scientific network. It was developed and kept alive through mutual scientific visits and exchange of knowledge through (surface) mail. Niels Bohr and Werner Heisenberg belonged to this team of highly competent theoretical physicists who all worked with the same mission: Understand the

(*) E-mail: mkleber@tum.de

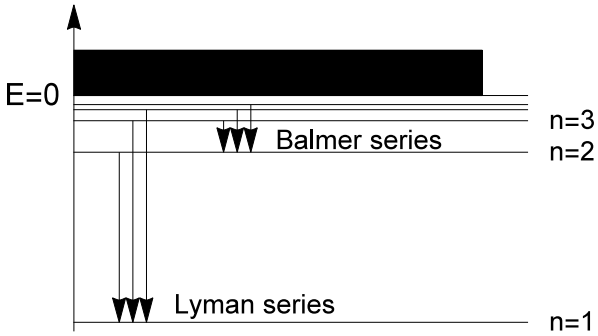


Fig. 1. – Modernized sketch of the Balmer (1885) and Lyman (1906) line series.

world of atoms! The story begins with the Bohr model and ends with Heisenberg's quantum mechanics — just before the appearance of Schrödinger's wave mechanics.

2. – From Balmer to Bohr

The paradigm shift from classical physics to quantum physics started practically unnoticed with Balmer's *Wunderformel* (wonder formula) [1]. Johann J. Balmer was a Swiss mathematician and physicist who in 1885 studied the position of the four(!) visible spectral lines of hydrogen atoms obtained with high precision by Anders J. Ångström who had made careful studies of the spectrum of the Sun, in particular the Fraunhofer lines. Balmer fitted their position with an empirical formula:

$$\lambda_n = 3645,6\frac{n^2}{n^2-4}\ \text{Å} \qquad (n=3,4,5,6), \tag{1}$$

where n runs — as we know today — from $n = 3$ through integer numbers to infinity (see fig. 1). The ensemble of the wavelengths λ_n forms the Balmer series. Balmer's formula (1) describes an amazing regularity of the spectral lines emitted by hydrogen atoms (in vacuum). However, in 1885 it was impossible to relate the formula to the underlying atomic structure. After all the electron was discovered only in 1897 by Joseph J. Thomson.

As is well known, it was Max Planck who in 1900 had to postulate a strange property of the so-called blackbody radiation (*Hohlraumstrahlung*) which stands for the radiation coming out of a hot box. In order to find a correct mathematical description for the intensities of the frequency distribution of the blackbody radiation, Planck had to assume that the radiation leaves the box discontinuously in tiny portions of energy $E = h\nu$, with ν being one of the many frequencies emitted by the box. Planck's constant h would soon turn out to play a decisive role in the upcoming theory of quanta.

For his 1905 explanation of the photoelectric effect Einstein came up with the idea of a "needle radiation" (*Nadelstrahlung*), meaning that light will *always* come in small

pieces which were later called photons. The photoelectric effect did not, however, give a clue how positive and negative charges are distributed in an atom. At that time, many scientists believed in the so-called plum pudding model introduced by J. J. Thomson who thought that the positive charge of an atom should be uniformly distributed over the volume of the atom. This uniform background forms the "pudding" with the electrons being the "plums". They are positioned at fixed points about which they can vibrate and emit and absorb radiation according to classical electrodynamics.

The situation changed when one of Thomson's former students, Ernest Rutherford together with his students Hans W. Geiger and Ernest Marsden measured the scattering of α-particles by a thin metal foil. They found a small but finite probability for backscattering of the α-particles. Some time in 1911 Rutherford concluded from an analysis of the data that the atom must contain a very small nucleus, probably positively charged. Rutherford's model for the nucleus eventually triggered Bohr's model for the atom.

Niels Bohr (1885-1962) [2] had studied physics at the University of Copenhagen and submitted his doctoral thesis (in Danish) on "Studies on the electron theory in metals" in the year 1911. In 1912 he began his postdoctoral time at the University of Cambridge to work with J. J. Thomson on the binding problem of electrons in metals. Cambridge and Manchester were at that time the leading places for theoretical and experimental atomic physics. But it seems that Bohr was not happy with Thomson's plum pudding model of the atom. By some lucky coincidence he got acquainted in the same year with Rutherford's α-particle experiments in Manchester.

Rutherford's research group shared its knowledge and results willingly with Bohr. In particular Georg von Hevesy and Charles G. Darwin made him familiar with the new atomic and nuclear physics. In order to learn more about the atom, Darwin had studied the energy loss of α-particles passing through matter. The data suggested that the energy loss for small scattering angles of the α-particles was caused by electrons. Bohr concluded from the experimental results that the atomic electrons had to be bound to the nucleus. He was flexible enough not to insist that this binding had to be harmonic. Without going into details we can safely say that it was Rutherford's discovery of the nucleus which made Bohr picturize the atom as an almost pointlike nucleus surrounded by moving electrons. The motion was governed by electric forces between all particles. Stability of the molecules (and atoms) had to be achieved through the motion of the electrons. In fact, in the presence of a central heavy point charge a ring of electrons can only be stable if it rotates in such a way that the attractive Coulomb force is counterbalanced by the centrifugal force.

In 1912 Bohr wrote a letter to Rutherford, known as *Rutherford Memorandum* in which he explained the binding of small molecules by electrons that move on discrete orbits. He chose the orbits through intelligent guessing. The most interesting candidate was the H_2-molecule. He described its spatial configuration by two nuclei with fixed distance and two electrons rotating synchronously on opposite sides of a circular orbit with diameter of about 10^{-8} cm and with its center half way between the two nuclei. Bohr however did not think that he had got a final answer to the binding problem. In a letter [3] to his brother Harald he considered the model to be "perhaps a little bit

of reality". He was intrigued by the question what the electrons are doing in an atom. When he went back to Denmark for a break of his postdoc work, Bohr was still heavily occupied with this problem. He came much closer to its solution when he was reminded of the Balmer formula (1) by one of his colleagues. In 1888 the Swedish physicist Johannes Rydberg had cast this formula in the form

$$\frac{1}{\lambda} = R\left(\frac{1}{n'^2} - \frac{1}{n^2}\right) \qquad (n = n'+1, n'+2, n'+3, \ldots,), \tag{2}$$

with $R = 1.097\ 10^7\,\mathrm{m}^{-1}$ being the Rydberg constant. The Balmer series is obtained for $n' = 2$. In 1906 a second spectral series with $n' = 1$ had been discovered in the ultraviolet by Theodore Lyman (see fig. 1). Equation (2) would be useful for Bohr because it characterizes the observed spectral lines of hydrogen by their frequency $\nu = \frac{c}{\lambda}$ instead of their wavelength λ.

Bohr was bold enough to assume that, in contradiction to classical physics a bound electron in a hydrogen can move only on one out of a discrete set of stationary orbits. Transitions between the orbits n and n' with energies E_n and $E_{n'}$ are possible through emission ($E_n > E_{n'}$) or absorption ($E_n < E_{n'}$) of a light quantum with frequency $\nu_{n',n}$. The *Einstein-Bohr frequency condition*

$$E_n - E_{n'} = h\,\nu_{n',n} \tag{3}$$

reflects energy conservation.

Motivated by Planck's hypothesis, $E = h\nu$, Bohr considered $\nu = \omega/(2\pi)$ to be the *mechanical* frequency of the electron in a circular orbit with radius r. The kinetic orbital energy E_{kin} should be half the total energy E in analogy to a harmonic oscillator. Discrete orbits have discrete kinetic energies which Bohr quantized by making the ansatz

$$E_{kin} = \frac{1}{2}\,nh\nu \ \left(= \frac{1}{2}\,n\hbar\omega\right) \quad (n = 1, 2, 3, \ldots) \tag{4}$$

with $\hbar = h/(2\pi)$. At this point Bohr introduced the (realistic) Coulomb force between electron and positive nuclear core. Stability of the atom was guaranteed through the balance between centrifugal force and the attractive electric Coulomb force. This balance requires the potential energy E_{pot} to be twice the negative of the kinetic orbital energy,

$$E_{kin} = \frac{n}{2}\,\hbar\omega = \frac{Ze^2}{2r} = -\frac{1}{2}E_{pot}. \tag{5}$$

(In SI-units Ze^2 has to be replaced by $Ze^2/(4\pi\epsilon_0)$.) By making use of $v = r\omega$ the discrete energies $E_n = E_{kin} + E_{pot}$ follow immediately:

$$E_n = -\frac{\mu}{2}\left(\frac{Ze^2}{\hbar}\right)^2 \frac{1}{n^2} \quad (n = 1, 2, 3, \ldots), \tag{6}$$

with μ being the (reduced) mass of the electron. From the last equation and by using (2) both the wave lengths and the frequencies of the emitted radiation could be calculated. Balmer's Wunderformel was demystified. But Bohr had also found the radius r of each allowed electron orbit together with the corresponding orbital angular momentum L,

$$L = n\hbar \tag{7}$$

which is also quantized. This beautiful result of the Bohr model should be valid in atoms and in molecules.

In 1913 Bohr submitted three papers to the Philosophical Magazine after showing them to Rutherford. They were entitled "On the constitution of atoms and molecules". In the first paper [4] which was devoted to the hydrogen atom he presented his world-famous model of the atom. Bohr used four principles to explain his model:

- The electron revolves around the nucleus in the Coulomb field of the proton. The motion on all (closed) orbits occurs according to the laws of classical mechanics. Therefore the formalism of classical mechanics remains valid which includes the possibility to use the canonical equations without modifications.
- The orbital motion of the electron is free of energy losses. No radiation is emitted as one would expect from classical electrodynamics.
- The bound electron can move on discrete stationary orbits only. The allowed orbits are obtained from constraints imposed by quantum effects.
- In the limit of large masses and orbits (large quantum numbers) the new theory has to pass over into classical mechanics. This is the correspondence principle.

3. – The Bohr model between success and failure

Bohr's formulation of his "quantum theory" introduces the concept of quantization but leaves the formalism of classical mechanics untouched. It was quickly realized that the quantization conditions could be most simply stated within the Hamilton-Jacobi equation by using action-angle variables. The important quantity is the action integral (or action function) J_k defined as the constant area enclosed by closed stream lines in (q_k, p_k) phase space. As an example we consider the quantized action integral (see fig. 2)

$$J = \oint \vec{p}\,\mathrm{d}\vec{l} = n\,h, \tag{8}$$

for the hydrogen problem. The quantization condition on the r.h.s. of (8) will determine the discrete energy eigenvalues of hydrogen. For the calculation of the line integral along the elliptical orbit we need the classical energy E of an electron in the Coulomb field of the proton

$$E = \frac{\vec{p}^{\,2}}{2\mu} - \frac{e^2}{r} \tag{9}$$

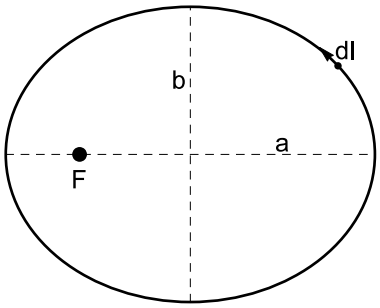


Fig. 2. – Elliptical orbit: Planetary model of an electron in the field of a proton.

with μ being again the reduced electron mass. The analytic expression for J depends only on the semimajor axis a of the ellipse:

$$J = 2\pi \sqrt{\mu\, a e^2} = nh. \tag{10}$$

For an attractive $1/r^2$ force law there is a well-known relation between a and the binding energy $|E|$:

$$a = \frac{e^2}{2\,|E|}\,. \tag{11}$$

Knowledge of the energy determines the semimajor axis of the ellipse a but leaves the value of the semiminor axis $b \leq a$ arbitrary.

From the last two equations one obtains the quantized binding energies of the electron in hydrogen:

$$E_n = -\frac{\mu}{2}\left(\frac{e^2}{\hbar}\right)^2 \frac{1}{n^2}\,. \tag{12}$$

Bohr had used circular orbits to obtain the binding energies in hydrogen. For circular motion the action integral J is simply the orbital angular momentum multiplied by 2π: $J = 2\pi L$.

In 1916 Arnold Sommerfeld extended the Bohr model to elliptical orbits. The Bohr-Sommerfeld model, as it was usually named from then on, used the fact that radial and angular motion are *separable* for the classical Coulomb (or Kepler) problem. Sommerfeld showed that the elliptic motion can be characterized by two quantum numbers which, from a classical point of view determine size and shape of the ellipse. Actually, there is a third quantum number n_ϕ that specifies the projection of the orbital angular momentum onto the z-axis (or some other fixed axis). The three internal quantum numbers n_r, n_ϕ and n_θ add up to the principal quantum number n

$$n = n_r + n_\phi + n_\theta, \quad (n = 1, 2, 3, \ldots). \tag{13}$$

The Bohr-Sommerfeld model was the basis for explaining the splitting of spectral lines in external electric and magnetic fields. The sum $l = n_\phi + n_\theta$ is the quantum number

$l = 0, 1, 2, \ldots$ of the angular momentum whereas $m = n_\phi$ is the magnetic quantum number.

Paul Epstein who had joined Sommerfeld in Munich for his PhD work studied the effect of a static electric field on the quantum numbers. He could show that the linear Stark effect in hydrogen can be solved exactly for the hydrogen atom within the Bohr-Sommerfeld model. For this problem the agreement with the experiment was perfect. Sommerfeld was also successful in explaining the normal Zeeman effect. And, amazingly, his relativistiv extension of the model could also explain the (relativistic) hyperfine level splitting of hydrogen. The scientists seemed to be on the right track.

Einstein however wrote a note of caution [5]: "... it remains unsatisfactory to depend on the separation of variables which has probably nothing to do with the problem of quantization". Einstein suggested a coordinate-invariant principle for quantizing separable systems as well as non-separable systems. Only for separable systems Einsteins principle reduces to the Bohr-Sommerfeld quantization rule. In this context we should point out that Wolfgang Pauli was not aware of Einstein's paper when he came to Sommerfeld in 1918 to work for his doctoral thesis on the properties of the hydrogen molecule ion H_2^+. So he used the unmodified quantization conditions of Bohr and Sommerfeld which for his case were too restrictive and therefore gave a positive ground state energy in contradiction to experiment [6, 7].

Einstein's important work remained unnoticed for several decades. Only Louis de Broglie [8] referred to the paper in his seminal thesis on the wave properties of matter.

In 1920, after finishing Gymnasium (an advanced high school), Werner Heisenberg (1901-1976) [9,10] enrolled in the University of Munich to study physics with Sommerfeld who immediately recognized his student's great scientific pontential. One year before, Sommerfeld had finished his famous book on "Atomic Structure and Spectral Lines" [11]. Sommerfeld was a great teacher and Heisenberg a highly intelligent student. Within one year's time Heisenberg could already publish a paper on the anomalous Zeeman effect. The electron spin was unknown at that time. Therefore, a convincing explanation of the anomalous Zeeman effect within the Bohr-Sommerfeld model was impossible. Questions about the validity of the Bohr-Sommerfeld theory were brought up. It was a fortunate coincidence that Heisenberg met Wolfgang Pauli [12] in Munich which marked the beginning of a lifelong friendship. Discussing physics and understanding the mysterious world of atoms was their common ambition.

Sommerfeld supported his talented students. In the summer of 1922 he travelled with Heisenberg to Göttingen in order to attend a workshop on urgent quantum problems. The workshop was called the "Bohr -Festival" because Bohr was the main speaker. Göttingen's famous mathematicians David Hilbert, Felix Klein, Carl Runge and Richard Courant were in the audience. And of course Max Born [13], Göttingen's leading theoretician with his co-workers Friedrich Hund and Pascual Jordan. Also Pauli from Hamburg and other prominent scientists like Paul Ehrenfest and Oskar Klein. The Bohr-Sommerfeld model was not a final theory. There was growing conflict with spectroscopic data. Bohr complained in one of his Göttingen talks "... how incomplete and uncertain everything still is".

Heisenberg's active participation in the "Bohr-Festival" resulted in an invitation to work with Born in Göttingen and also one to work with Bohr in Copenhagen. But first he had to get his Ph.D. in Munich. He received it in 1923 with a work on turbulence, another research field of Sommerfeld. Thereafter Heisenberg came to Göttingen to work with Born as his assistant. One of their first projects was to understand the spectrum of helium. However, this "simple" two-electron problem could not be solved with the Bohr-Sommerfeld theory. Some success was achieved by implementing other ideas and methods into the theory, including Pauli's exclusion principle. We do not discuss them here because they are not essential for Heisenberg's groundbreaking work.

One had to find a way out of the serious difficulties of the Bohr-Sommerfeld theory. In this almost hopeless situation Heisenberg was lucky to obtain a grant to join Bohr's group in Copenhagen. He stayed at Bohr's institute from September 1924 to April 1925. There he met another young physicist of exceptional quality, Hendrik Kramers [14], who was working on the connection between classical mechanics and quantum theory. Kramers and Heisenberg were both studying the interaction between matter and radiation, and they got along well with each other. Their joint paper [15] on the refraction of radiation by atoms uses for the first time *Fourier series for the amplitudes* of the scattered radiation. The calculational methods were taken from Born's paper on "Quantum mechanics" [16](1) (see also [17] which contains a very useful collection of relevant papers from Bohr to Heisenberg in English).

When Heisenberg returned to Göttingen he had not solved the problems with the Bohr-Sommerfeld model. But influenced by Bohr and Kramers and convinced that this model could not be saved, Heisenberg was "fabricating quantum mechanics" as he wrote in a letter to Pauli dated June 21, 1925. A new quantum mechanics had to replace the old Bohr-Sommerfeld quantum mechanics. It did not take Heisenberg long to write the famous paper entitled "Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen" [18] ("Quantum-theoretical reinterpretation of kinematic and mechanical relations" [17]). In the abstract of the paper Heisenberg announces that he tries to "establish a basis for a theoretical quantum mechanics that is exclusively based on relationships between quantities which are in principle observable". He wanted to get rid of the time-dependent position and the orbital period of an electron which he considered to be unobservable. As Dirac explains [19]: "Heisenberg puts forward a new theory which suggests that it is not the equations of classical mechanics that are in any way at fault, but the mathematical operations by which physical results are deduced from them require modifications".

4. – Heisenberg's path from classical physics to quantum mechanics

The main problem in understanding Heisenberg's paper is that it contains so much information that "one cannot see the forest for the trees". Fortunately Heisenberg wrote

(1) The name QUANTUM MECHANICS appears in the literature here for the first time.

two letters, one to Ralph Kronig in Copenhagen and a second one, on June 24, to Pauli where he clearly explained his main new ideas. The letter to Pauli is contained in [17]. The difficulty for today's reader lies in understanding the reinterpretation of the classical action. It is the quantized action integral J which plays a key role in Heisenberg's desire to obtain an adequate description of observable atomic properties. In the following we work out the three steps that lead from classical to quantum mechanics.

- The classical action is analyzed in Fourier space.
- The classical Fourier coefficients are transformed into amplitudes which depend on two "co-ordinates".
- The equation of motion for the amplitudes is used to determine algebraically the energy spectrum of the quantum system.

4˙1. *Action integral in Fourier space.* – Like in the old quantum mechanics of Bohr and Sommerfeld, the important variable is the action integral. In the absence of friction or radiation a stationary bound state in one dimension is always periodic in time,

$$J = \oint_{period} P\,\mathrm{d}Q = \oint_{period} P\frac{\mathrm{d}Q}{\mathrm{d}t}\,\mathrm{d}t = \int_0^T P\,\frac{\mathrm{d}Q}{\mathrm{d}t}\mathrm{d}t. \tag{14}$$

As usual T denotes the period of the motion. Here the particle's time-dependent position $Q(t)$ and momentum $P(t)$ are denoted with capital letters. Small letters are reserved for their corresponding Fourier expansion coefficients $x(\tau)$ and $p(\tau)$:

$$Q(t) = \sum_{\tau=-\infty}^{\infty} x(\tau)e^{i\tau\omega t} \quad \text{with} \quad \omega = \frac{2\pi}{T} = 2\pi\,\nu. \tag{15}$$

An analogous expansion holds true for $P(t)$:

$$P(t) = \sum_{\tau=-\infty}^{\infty} p(\tau)e^{i\tau\omega t}. \tag{16}$$

Because of

$$\int_0^T p(\tau')e^{i\tau'\omega t}\; x(\tau)e^{i\tau\omega t}\,\mathrm{d}t = T\,p(\tau')x(\tau)\delta_{\tau',-\tau} \tag{17}$$

the action variable J can be easily expressed through the Fourier coefficients

$$J = 2\pi i \sum_{\tau=-\infty}^{\infty} \tau\; p(-\tau)\,x(\tau). \tag{18}$$

For the following it is helpful if we arrange the Fourier coefficients $x(\tau)$ and $p(\tau)$ according to their argument τ along the τ-axis (see fig. 3). The action integral is a sum of bilinear

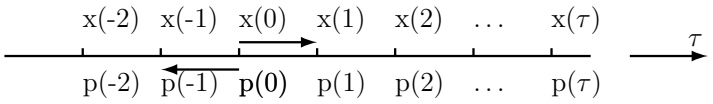


Fig. 3. – Geometric illustration of the bilinear terms in (18) by two equally long arrows pointing in opposite directions.

terms, $x(\tau)\,p(\tau')$ with $\tau' = -\tau$. This condition on τ' makes J independent of time as it must be the case for stationary states. We can picturize this condition by using two equally long arrows in fig. 3 which point into opposite directions, one from 0 to τ and the other from 0 to $-\tau$. For *charged particles* such an arrow indicates the possibility to emit or absorb harmonic radiation. Figure 3 indicates a radiative transition between ground state and first overtone separated by $\Delta\tau = \pm 1$ which corresponds to $\Delta E = \pm h\nu$. Higher-harmonic transitions with $|\tau| > 1$ can be piturized accordingly. The length of an arrow is proportional to the corresponding transition energy between the two states which it connects. One-dimensional periodic motion exhibits only one basic frequency, $\nu = 1/T$, with equidistant overtones as shown in the figure.

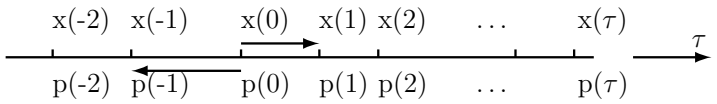


Fig. 4. – Same as fig. 3 for a non-equidistant frequency spectrum.

4˙2. *Extension to an arbitrary frequency spectrum.* – The frequency spectrum of hydrogen is not equidistant. A schematic plot of this new situation is shown in fig. 4. An expansion of the action variable as before does not work because the corresponding arrows have different lengths. We have seen before that a stationary contribution to the action integral J requires the two phases in the bilinear products (17) to cancel. Opposite phases are generally only possible if we draw closed loops as shown in fig. 5 The index 0 characterizes the average property of the periodic motion. Each of the periodic stationary bound states in the hydrogen atom has its own quantum number n and its own action function $J = J_n$. By replacing the index 0 with n we arrive at the numbering given in fig. 6. The arrows are now conveniently described by their starting point and by their endpoint. In the example of fig. 6, the Fourier coefficient $x(n, n-1)$ denotes a transition amplitude from state $n-1$ to state n (read the indices in $x(n, n-1)$ from right to left). It was Heisenberg's brilliant idea to replace the Fourier coefficients in (18)

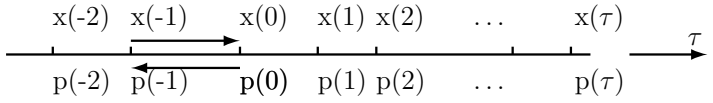


Fig. 5. – Arrows with the same length but pointing in opposite directions.

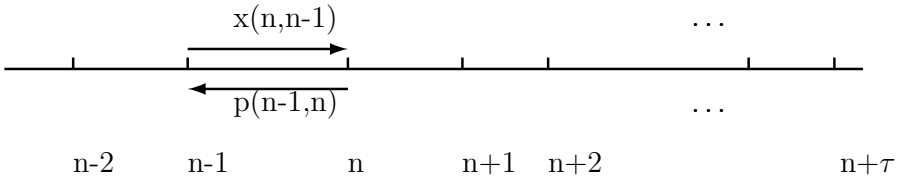


Fig. 6. – Fourier coefficients that depend on two states n and m.

with quantities $x(n,m)$ (in Heisenberg's notation $a(n,m)$) and the frequencies with the transition frequencies from state m to state n. He had "fabricated" the correspondence scheme for replacing classical Fourier coefficients by quantum transition amplitudes:

$$x(1)p(-1) \longrightarrow x(n,n-1)p(n-1,n), \tag{19}$$

$$x(\tau)p(-\tau) \longrightarrow x(n,n-\tau)p(n-\tau,n). \tag{20}$$

Heisenberg knew from Einstein's paper on the quantum theory of radiation [20] and from the many discussions with Kramers that the (observable) intensity of radiation for a transition from state m to state n was proportional to the square of the transition amplitudes. The square of the amplitudes (or their absolute values) should be observable quantities which should be linked to theory. Making use of (19) and (20) the quantized version of (18) now reads

$$J = 2\pi i \sum_{\tau=-\infty}^{\infty} \tau \; x(n,n-\tau)p(n-\tau,n) = n\,h = J(n). \tag{21}$$

In this form the action depends only on transition amplitudes and is no longer restricted to periodic motion – a constraint given up earlier by Born and Jordan [21].

4˙3. *The appearance of non-commuting quantities*. – Heisenberg differentiated the action (21) with respect to n:

$$\frac{\partial}{\partial n} J(n) = 2\pi i \sum_{\tau=-\infty}^{\infty} \tau \; \frac{\partial}{\partial n} \Lambda_\tau(n) = h \tag{22}$$

with

$$\Lambda_\tau(n) := x(n,n-\tau)\, p(n-\tau,n). \tag{23}$$

A similar expression was discussed by Born in his already mentioned important paper on quantum mechanics [16] which preceded Heisenberg's paper. There Born argues by

invoking the correspondence principle that a classical expression of the form $\tau \frac{\partial}{\partial n}\Lambda_\tau(n)$, where n is discrete, should be replaced by the corresponding difference form([2]):

$$\tau \frac{\partial}{\partial n}\Lambda_\tau(n) \longrightarrow \Lambda_\tau(n+\tau) - \Lambda_\tau(n), \tag{24}$$

which means that

$$\tau \frac{\partial}{\partial n}\Lambda_\tau(n) \longrightarrow p(n,n+\tau)\; x(n+\tau,n) - x(n,n-\tau)\, p(n-\tau,n). \tag{25}$$

As a result of this substitution rule we now obtain from (22)

$$2\pi i \sum_{\tau=-\infty}^{\infty} p(n,n+\tau)\; x(n+\tau,n) - x(n,n-\tau)\, p(n-\tau,n) = h. \tag{26}$$

Replacing the summation index τ with the summation indices $l = n+\tau$ and $l = n-\tau$, respectively yields the famous commutation relation between x and p:

$$\sum_{l=-\infty}^{\infty} p(n,l)\; x(l,n) - x(n,l)\, p(l,n) = \frac{\hbar}{i}\,. \tag{27}$$

Heisenberg did not write down this fundamental commutation relation between momentum coordinate and space coordinate because he had replaced the momentum by the velocity, $P(t) = m\frac{\mathrm{d}}{\mathrm{d}t}Q(t)$ (see also next section). *Equation* (27) *is due to Born* who realized that Heisenberg's new theory was based on matrix calculation. In 1925 Heisenberg was not familiar with matrix calculation although his multiplication law for transition amplitudes in eqs. (7) and (8) of his paper is written in matrix notation. One page later Heisenberg remarks that "Whereas in classical theory $x(t)y(t)$ is always equal to $y(t)x(t)$ this is not necessarily the case in quantum theory". The new quantum mechanics got the name "matrix mechanics". Its complete formulation was accomplished in refs. [22,19,23]. We should point out that it was Pascual Jordan who showed [22] that Heisenberg's theory satisfies energy conservation and the Einstein-Bohr frequency relation (3).

To appreciate what Heisenberg had achieved we follow his calculation of the quantized energy levels of a harmonic oscillator.

5. – Quantization of the linear harmonic oscillator

In his paper [18,17] (see also [24,25]) Heisenberg considers space-dependent forces. He takes over the classical equation of motion $\ddot{x} + f(x) = 0$ (here in one spatial dimension).

([2]) The replacement rule is not quite unique. With the same reasoning one could also substitute $\tau \frac{\partial}{\partial n}\Lambda_\tau(n) \longrightarrow \Lambda_\tau(n) - \Lambda_\tau(n-\tau)$.

To avoid confusion with the Fourier coefficients we replace $x(t)$ as we have done before by $Q(t)$:

$$\ddot{Q}(t) + f(Q) = 0. \tag{28}$$

We now show how Heisenberg quantizes the linear harmonic oscillator as is explained in his letter to Pauli (see [17]). Writing

$$\ddot{Q}_n(t) + \omega^2 Q_n(t) = 0, \tag{29}$$

where Q_n is the n-th periodic solution of the oscillator and going over to Fourier space as before we have

$$Q_n(t) = \sum_{\tau=-\infty}^{\infty} x_n(\tau) \exp(i\tau\omega_n t) = Q_n^*(t), \tag{30}$$

because $Q_n(t)$ is real. The frequencies ω_n must satisfy the Einstein-Bohr rule (3). A harmonic oscillator lacks higher harmonics; therefore $\omega_n = \omega$. By combining the last two equations we obtain

$$\sum_{\tau=-\infty}^{\infty} x_n(\tau)[-\tau^2\omega^2 + \omega^2] \exp(i\tau\omega t) = 0. \tag{31}$$

It follows that all coefficients $x_n(\tau)$ must be zero except $x_n(\tau = 1)$ which oscilates with frequency $\omega(n+1, n) = \omega$ and $x_n(\tau = -1)$ which oscillates with frequency $\omega(n-1, n) = -\omega$([3]). Heisenberg used the substitution rule

$$\begin{aligned} x_n(1) &\longrightarrow x(n, n-1) = x(n-1, n)^*, \\ x_n(-1) &\longrightarrow x(n, n+1) = x(n+1, n)^* \end{aligned}$$

and expressed the momentum by the velocity $\dot{Q}_n$

$$P_n(t) = \mu\, \dot{Q}_n(t) = i\mu\,\omega\, [x_n(1) \exp(i\omega t) - x_n(-1) \exp(-i\omega t)]. \tag{32}$$

Heisenberg was then in a position to evaluate (27). From the two contributing terms $l = n + 1$ and $l = n - 1$ he readily obtained

$$2\mu\,\omega\, \left[\, |x(n, n+1)|^2 - |x(n, n-1)|^2 \,\right] = \hbar. \tag{33}$$

([3]) This is a special case of the Ritz combination principle, also used by Heisenberg: $x_n(\tau) \exp[i\tau\omega_n t]) \longrightarrow x(n+\tau, n) \exp[i\omega(n+\tau, n)t]$ with $\omega(n, l) + \omega(l, m) = \omega(n, m)$.

The equation defines a recursion relation:

$$|x(n+1,n)|^2 = |x(n,n-1)|^2 + \frac{\hbar}{2\mu\,\omega} \tag{34}$$

Negative values of n are not possible. Therefore matrix elements with negative indices must be zero. Using $n = 0$ in (34) one obtains

$$|x(1,0)|^2 = |x(0,1)|^2 = \frac{\hbar}{2\mu\,\omega}\,. \tag{35}$$

and from there

$$|x(n,n+1)|^2 = |x(n+1,n)|^2 = (n+1)\frac{\hbar}{2\mu\,\omega}\,. \tag{36}$$

In a last step Heisenberg used this result to calculate the quantized energies

$$E_n = \frac{\mu}{2}\dot{Q}_n^2(t) + \frac{\mu}{2}\omega^2 Q_n^2(t) \tag{37}$$

of the harmonic oscillator. He did this during a stay in Helgoland where he cured a heavy attack of hay fever. Using the matrix elements (36) which determine both $Q_n(t)$ and $\dot{Q}_n(t)$ he obtained the (by now) well-known result

$$E_n = \hbar\,\omega\left(n + \frac{1}{2}\right). \tag{38}$$

The nonvanishing (dipole) transition matrix elements to and from state n are indicated in fig. 7 together with the equidistant energies E_n. From the figure we infer that the Einstein-Bohr frequency condition (3) is satisfied since we have as mentioned before $\omega(n+1,n) = \omega$, and $\omega(n-1,n) = -\omega$. The zero-point energy of the ground state had been anticipated before from Planck's blackbody radiation formula. But Heisenberg had for the first time developed a mathematical method that does not use Bohr's quantization scheme. He obtained the whole spectrum of the oscillator, the ground state included from which obviously no radiation can be emitted.

The transition amplitudes shown in the figure are needed in the scattering of light by atoms (or molecules) when internal degrees of freedom of the atom are excited. Such type of inelastic scattering is the nowadays well known Raman scattering. Let us point out that Heisenberg spent much effort in his paper to quantize the *anharmonic* oscillator. There the frequencies are not known a priori. Heisenberg calculated them in second order in the anharmonicity together with the quantized energies which satisfy the Einstein-Bohr frequency condition (3). The details of this lengthy calculation can be found in [26].

Finally, let us mention that Pauli obtained [27] the hydrogen spectrum from an intricate calculation within Heisenberg's theory.

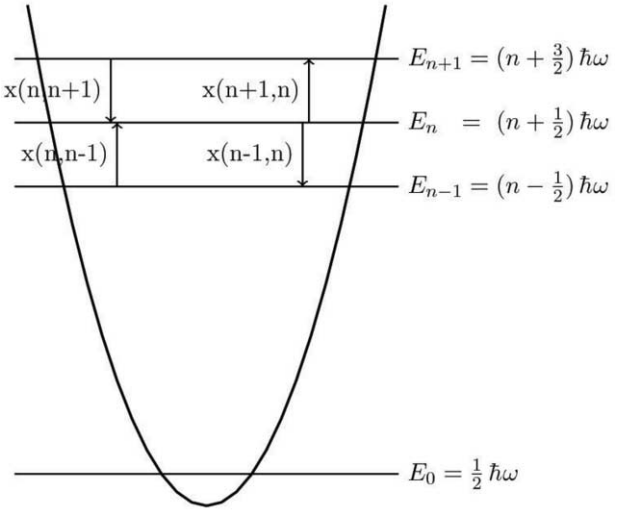


Fig. 7. – Energies and matrix elements for the linear harmonic oscillator. $x(b, a)$ is the matrix element of the operator (matrix) x for a transition from state a to state b.

6. – Light at the end of the tunnel

It is here where we terminate our story of the discovery of quantum mechanics. Heisenberg's contribution to modern quantum mechanics cannot be overestimated. The same is true for Bohr's ground breaking ideas when he introduced his model. The two scientists reached the light of new scientific insight by crossing a dark tunnel which links classical physics to quantum physics.

Some final remarks should be in order. Most textbooks on quantum mechanics do not spend much effort on Heisenberg's matrix approach to quantum mechanics because Schrödinger's wave method yields a much easier access. Schrödinger submitted the first of his three papers on quantum mechanics in January 1926 [28, 29] (see also [30]), a few days after Pauli had submitted his sophisticated calculation of the Balmer formula using Heisenberg's matrix mechanics. Schrödinger's wave function method simplified the calculation of quantum properties considerably. Although his wave mechanics looked different from Heisenberg's matrix mechanics, Schrödinger was able to show that the two approaches to quantum mechanics are identical.

In 1926 Heisenberg was again in Copenhagen where he taught theoretical physics in Danish. During his stay in Denmark Heisenberg provided further evidence of his extraordinary scientific abilities when he solved the long-standing (two-electron) helium problem by making use of the Schrödinger equation in combination with the Pauli principle. Heisenberg's uncertainty principle followed and then in 1932 he got the Noble Prize, ten years after Bohr had gotten this prestigeous award.

The spirit of the Bohr model is still alive because in its extended and improved semiclassical version it can be a surprisingly good approximation to exact quantum-

mechnical results. For example, the low-lying electronic states of H_2^+ which Pauli failed to calculate can be quite accurately reproduced in semiclassical quantization [31]. Another important application is the semiclassical helium atom [32]. The Bohr model also serves as a beautiful example for dimensional scaling [33]. Finally, the coordinate-invariant quantization scheme developed by Einstein on the basis of the Bohr-Sommerfeld model is of great use to understand and to deal with the difficulties in quantizing classically chaotic motion [34, 32].

* * *

I would like to thank my colleagues Nancy Thorndike Greenspan, Wolfgang Schleich, Kurt Schönhammer, Marlan O. Scully and Herbert Wagner for helpful discussions.

REFERENCES

The list of references should be useful but it is certainly not complete.

[1] BALMER J. J., *Notiz über die Spectrallinien des Wasserstoffs*, *Ann. d. Phys. und Chem.*, **25** (1885) 80.
[2] PAIS A., *Niels Bohr's times* (Clarendon Press, Oxford) 1991.
[3] SCHIRRMACHER A., *Der realistische Bohr*, *Physik Journal*, **12** (2013) 43.
[4] BOHR N., *On the constitution of atoms and molecules*, *Philos. Mag. Ser. 6*, **26** (1913) 1.
[5] EINSTEIN A., *Zum Quantensatz von Sommerfeld und Epstein*, *Deutsche Physikalische Gesellschaft, Verhandlungen*, **19** (1917) 82.
[6] PAULI W., *Über das Modell des Wasserstoffmolekülions*, *Ann. Phys. (Leipzig)*, **68** (1922) 177.
[7] GUTZWILLER M., *Chaos in classical and quantum mechanics* (Springer, New York) 1990.
[8] DE BROGLIE L., *Recherches sur la théorie des quanta*, *Ann. Phys. (Paris)*, 10^e série, t.III (1925), Engl. Translation by KRACKLAUER A. F.: *On the theory of quanta.* `http://aflb.ensmp.fr/LDB-oeuvres/De_Broglie_Kracklauer.pdf`.
[9] RECHENBERG W., *Werner Heisenberg – Die Sprache der Atome* (Springer Verlag, Berlin) 2010.
[10] MEHRA J. and RECHENBERG M., *The historical development of quantum theory*, Vol. **2** (Springer Verlag, New York) 1982.
[11] SOMMERFELD A., *Atombau und Spektrallinien* (Friedr. Vieweg & Sohn, Braunschweig) 1919.
[12] ENZ CH., *No Time to be Brief* (Oxford University Press, Oxford) 2002.
[13] GREENSPAN N. TH., this volume, p. 15.
[14] TER HAAR D., *Master of modern Physics: The scientific contributiond of H. A. Kramers.* (Princeton University Press, Princeton) 1998.
[15] KRAMERS H. A. and HEISENBERG W., *Über die Streuung von Strahlen durch Atome*, *Z. Phys.*, **31** (1925) 681, received Jan. 5, 1925.
[16] BORN M., *Über Quantenmechanik.*, *Z. Phys.*, **26** (1924) 379.
[17] VAN DER WAERDEN B. L., *Sources of Quantum Mechanics* (Dover Books on Physics) Reprint of the 1968 Dover edition.
[18] HEISENBERG W., *Über quantenmechanische Umdeutung kinematischer und mechanischer Beziehungen*, *Z. Phys.*, **33** (1925) 879, received July 29, 1925.

[19] Dirac P. A. M., *The Fundamental Equations of Quantum Mechanics*, *Proc. R. Soc. A*, **109** (1926) 642, received November 7, 1925.
[20] Einstein A., *Zur Quantentheorie der Strahlung*, *Phys. Zeit.*, **18** (1917) 121.
[21] Born M. and Jordan P., *Zur Quantentheorie aperodischer Vorgänge*, *Z. Phys.*, **33** (1925) 479.
[22] Born M. and Jordan P., *Zur Quantenmechanik*, *Z. Phys.*, **34** (1925) 858, received Sept. 27, 1925.
[23] Born M., Heisenberg W. and Jordan P., *Zur Quantenmechanik II*, *Z. Phys.*, **35** (1926) 557, received Nov. 16, 1925.
[24] Park D., *Classical Dynamics and its Quantum Analogues* (Springer Verlag, Berlin) 1990.
[25] Razavy M., *Heisenberg's Quantum Mechanics* (World Scientific, Singapore) 2011.
[26] Aitchison I. A. R., MacManus D. A. and Snyder T. M., *Understanding Heisenberg's "magical" paper of July 1925: A new look at the calculational details*, *Am. J. Phys.*, **72** (2004) 1370.
[27] Pauli W., *Über das Wasserstoffspektrum vom Standpunkt der neuen Quantenmechanik*, *Z. Phys.*, **36** (1926) 336, received Jan. 17, 1926.
[28] Schrödinger E., *Quantisierung als Eigenwertproblem*, *Ann. Phys. (Leipzig)*, **79** (1926) 361, received Jan. 27, 1926.
[29] Schrödinger E., *Collected Papers on Wave Mechanics* (AMS Chelsea Publishing, Providence, Rhode Island) 2003.
[30] Scully M. O., Greenberger D. and Schleich W. P., this volume, p. 119.
[31] Strand M. P. and Reinhardt W. P., *Semiclassical quantization of the low lying states of* H_2^+, *J. Chem. Phys.*, **71** (1979) 3812.
[32] Wintgen D., Richter K. and Tanner G., *The semiclassical helium atom*, *Chaos*, **2** (1992) 19.
[33] Herschbach D. R., Scully M. O. and Svidzinsky A. A., *Bohr's Comeback*, *Physik Journal*, **12** (2013) 37.
[34] Stone A. D., *Einstein's unknown insight and the problem of quantizing chaos*, *Phys. Today*, **58**, 8 (2005) 37.

This page intentionally left blank

Proceedings of the International School of Physics "Enrico Fermi"
Course 197 "Foundations of Quantum Theory", edited by E. M. Rasel, W. P. Schleich and S. Wölk
(IOS, Amsterdam; SIF, Bologna) 2019
DOI 10.3254/978-1-61499-937-9-47

The linearity of quantum mechanics and the birth of the Schrödinger equation

W. P. Schleich

Institut für Quantenphysik and Center for Integrated Quantum Science and Technology (IQ^{ST}) Universität Ulm - D-89069 Ulm, Germany

Department of Physics, University of North Texas - Denton, Texas 76205-1427, USA

Hagler Institute for Advanced Study at Texas A&M University, Institute for Quantum Science and Engineering (IQSE), and Texas A&M AgriLife Research, Texas A&M University College Station - Texas 77843-4242, USA

D. M. Greenberger

City College of the City University of New York, New York - New York 10031, USA

Institut für Quantenphysik and Center for Integrated Quantum Science and Technology (IQ^{ST}) Universität Ulm - D-89069 Ulm, Germany

D. H. Kobe

Department of Physics, University of North Texas - Denton, Texas 76205-1427, USA

M. O. Scully

Institute for Quantum Science and Engineering (IQSE) and Department of Physics and Astronomy, Texas A&M University College Station - Texas 77843-4242, USA

Princeton University - Princeton, New Jersey 08544, USA

Baylor University - Waco, Texas 97096, USA

Summary. — We obtain the Schrödinger equation from the Hamilton-Jacobi equation of classical mechanics together with the law of conservation of matter. It is the quantum current in this continuity equation which ensures the linearity of quantum mechanics.

1. – Introduction

The quantum world [1] displays many mind-boggling phenomena. Schrödinger cats [2, 3], Einstein-Podolsky-Rosen correlations [4] and complementarity [5] are only a few of the many manifestations of the micro-cosmos which are alien to the classical way of thinking. Ultimately, the origin of these features is the linearity of quantum theory [6]. Indeed, the Schrödinger equation is a *linear* equation, and the superposition of two solutions is, therefore, again a solution.

In order to probe the linearity of quantum mechanics, non-linear Schrödinger equations of various forms have been suggested [7-9]. However, in every experimental test, for example employing neutron interferometry [10, 11] or rf spectroscopy of laser-cooled stored ions [12], the superposition principle has prevailed.

In contrast to quantum theory, classical mechanics is in general a *non-linear* theory and the superposition principle does not hold. Despite this distinct difference, both theories are intimately connected. Indeed, the Hamilton-Jacobi equation of classical mechanics leads [13-15] to a *non-linear* wave equation which is amazingly close, but not identical to the Schrödinger equation.

To the best of our knowledge there is no procedure how to arrive at the linear Schrödinger equation starting from classical mechanics. In our lecture notes we show that this direct path from classical to quantum mechanics is possible when we allow an additional current in the continuity equation.

1˙1. *Linearization of the non-linear wave equation.* – Despite an extensive literature [16] on the connection between the Hamilton-Jacobi and the Schrödinger equation we have not been able to find the arguments summarized in our lecture notes. The key element of our approach is a mathematical identity which involves the time derivative as well as the Hamiltonian of the Schrödinger equation acting on a wave consisting of an amplitude and a phase. Our relation shows that a *quantum* wave, that is a wave which satisfies the Schrödinger equation, can be interpreted as one whose phase is coupled in a special way to its amplitude.

Indeed, the equation of motion for the amplitude is again a Schrödinger equation but with the momentum operator being replaced by the momentum operator *plus* the gradient of the phase function, and the potential by the potential *plus* the time derivative. This substitution is familiar from the concept of gauge invariance. However, in the present context we use the transformed Schrödinger equation for the *amplitude* to define the familiar Schrödinger equation for the *complete* wave.

For the reader not interested in the mathematical details of the different approaches towards the Schrödinger equation discussed in the present lectures, we have summarized them in fig. 1. In order to distinguish between rigorous derivations, and assumptions based on generalizations of successful concepts, we have introduced in this diagram solid and dashed lines, respectively.

In fig. 1 we compare and contrast four approaches towards the Schrödinger equation: The first one, proposed by Ralph Schiller [13] and shown on the left-hand side of the

Fig. 1. – Four routes connecting the Hamilton-Jacobi equation of non-relativistic classical mechanics with the linear Schrödinger equation and the non-linear wave equation. The Van Vleck determinant D obtained by appropriate differentiations of the classical action $S^{(\mathrm{cl})}$ satisfies the Van Vleck continuity equation which follows from the Hamilton-Jacobi equation. As a result, the super-classical wave $\psi^{(\mathrm{cl})} \equiv D^{1/2}\mathrm{e}^{\mathrm{i}S^{(\mathrm{cl})}/\hbar}$ satisfies a non-linear wave equation similar to the Schrödinger equation, but with a non-linearity given by the potential $Q[|\psi^{((\mathrm{cl}))}|]$. Here we have defined $Q[|\varphi|] \equiv \hbar^2/(2m)\Delta|\varphi|/|\varphi|$ for an arbitrary wave $\varphi = \varphi(\boldsymbol{r}, t)$. When we impose the boundary conditions on $\psi^{(\mathrm{cl})}$ familiar from quantum mechanics, we arrive at *semi-classical* quantum mechanics. For the slightly generalized wave $\psi_R^{(\mathrm{cl})} \equiv A_R^{(\mathrm{cl})}\mathrm{e}^{\mathrm{i}S^{(\mathrm{cl})}/\hbar}$ we have replaced $D^{1/2}$ by a *real-valued amplitude* $A_R^{(\mathrm{cl})}$ but still maintain a continuity equation of the form of Van Vleck. In this case the Hamilton-Jacobi equation together with the continuity equation leads again to the classical non-linear wave equation but now with the non-linearity $Q[|\psi_R^{(\mathrm{cl})}|]$. Here we have to postulate rather than derive the continuity equation as indicated by the dashed line. Nevertheless, conservation of matter is inherent in the two non-linear wave equations, which are equivalent to classical statistical mechanics. Finally we allow a *complex-valued* amplitude R of the quantum wave $\psi^{(\mathrm{q})} \equiv R\,\mathrm{e}^{\mathrm{i}S^{(\mathrm{cl})}/\hbar}$, and generalize the continuity equation to a quantum continuity equation. In this case we arrive at the *linear* Schrödinger equation and at a complete quantum-mechanical description. The linearization of the classical non-linear wave equation is made possible by an additional contribution to the current due to an extra momentum originating from the quantum continuity equation. The wave $\psi^{(\mathrm{q})} \equiv A^{(\mathrm{q})}\mathrm{e}^{\mathrm{i}S^{(\mathrm{q})}/\hbar}$, familiar from the Madelung-Bohm formulation of quantum mechanics with a real-valued amplitude $A^{(\mathrm{q})}$ and a phase $S^{(\mathrm{q})}$ which not only contains $S^{(\mathrm{cl})}$ but also the phase β of R, leads us from the Schrödinger equation to the *quantum* Hamilton-Jacobi equation with the Madelung-Bohm quantum potential $\mathcal{Q}^{(\mathrm{MB})} \equiv -Q[|\psi^{(\mathrm{q})}|]$, and the quantum continuity equation.

figure, defines a wave $\psi^{(\mathrm{cl})} \equiv D^{1/2}\mathrm{e}^{\mathrm{i}S^{(\mathrm{cl})}/\hbar}$ with an amplitude and a phase given by the Van Vleck determinant D and the classical action $S^{(\mathrm{cl})}$, respectively. Since the Hamilton-Jacobi equation implies [17] the Van Vleck continuity equation, our mathematical identity reduces to a non-linear wave equation. Despite its resemblance of quantum mechanics this equation still represents [13] classical statistical mechanics. Only when we impose appropriate boundary conditions on $\psi^{(\mathrm{cl})}$ do we obtain the semi-classical description [18] of quantum mechanics, and $\psi^{(\mathrm{cl})}$ turns into the primitive WKB wave function [18].

In the second approach suggested by Nathan Rosen [14,15], and outlined in the center of fig. 1, we replace the amplitude $D^{1/2}$ of $\psi^{(\mathrm{cl})}$ by an arbitrary real-valued amplitude $A_R^{(\mathrm{cl})}$ which obeys a classical continuity equation *à la* Van Vleck. Again, the wave equation for the classical wave $\psi_R^{(\mathrm{cl})} \equiv A_R^{(\mathrm{cl})}\mathrm{e}^{\mathrm{i}S^{(\mathrm{cl})}/\hbar}$ following from our mathematical identity is non-linear. As indicated by the dashed horizontal line in fig. 1 connecting the Van Vleck continuity equation and the one of Rosen, we have now entered uncharted territory. Indeed, we had to *postulate* rather than could *derive* this conservation law. Needless to say, the classical non-linear wave equation implies the continuity equation.

The third approach depicted on the right-hand side of fig. 1 achieves the linearization of the non-linear wave equation by allowing the amplitude R of the quantum wave $\psi^{(\mathrm{q})} \equiv R\,\mathrm{e}^{\mathrm{i}S^{(\mathrm{cl})}/\hbar}$ to take on complex values. Indeed, both the Schiller as well as the Rosen concept are based on real-valued amplitudes. As a consequence of the complex nature

of the amplitude $R \equiv |R|e^{i\beta}$ the momentum of the particle is not just the classical momentum $\boldsymbol{\nabla} S^{(\mathrm{cl})}$, but also contains the quantum momentum $\hbar\boldsymbol{\nabla}\beta$.

Obviously, we have to include this term in the definition of the current which gives rise to the quantum continuity equation. Again this step can only be motivated but not proven from first principles as indicated in fig. 1 by the dashed line connecting the classical and the quantum continuity equations.

The fourth approach summarized at the bottom right of fig. 1 highlights the connections between the hydrodynamic formulation of quantum mechanics by Erwin Madelung [19] and David Bohm [20], the Hamilton-Jacobi equation of classical mechanics, and our approach towards the Schrödinger equation. Indeed, the decomposition $\psi^{(\mathrm{q})} \equiv A^{(\mathrm{q})}e^{iS^{(\mathrm{q})}/\hbar}$ of the Schrödinger wave $\psi^{(\mathrm{q})}$ into a real-valued amplitude $A^{(\mathrm{q})}$ and a real-valued phase $S^{(\mathrm{q})}$ transforms the linear Schrödinger equation into the Hamilton-Jacobi equation with a non-linear quantum potential, and recovers the quantum continuity equation. Now the quantumness of the Schrödinger wave is included in the phase $S^{(\mathrm{q})}$ rather than the amplitude R. Obviously, the relations $A^{(\mathrm{q})} = |R|$ and $S^{(\mathrm{q})} \equiv \hbar\beta + S^{(\mathrm{cl})}$ connect the amplitude and the action of the Madelung-Bohm wave to the corresponding ones of the third approach.

Finally, we emphasize that the quantumness of the wave increases from the top to the bottom of fig. 1, that is as we make the transition in the amplitude of the wave from the square root of the Van Vleck determinant D, or an arbitrary real-valued amplitude $A_R^{(\mathrm{cl})}$ governed by the classical continuity equation, to the complex-valued R determined by the quantum continuity equation. Whereas the first two approaches only provide a semi-classical description based on WKB waves which satisfy a classical *non-linear* wave equation, the third and fourth one obtain the full quantum theory based on the *linear* Schrödinger equation.

1˙2. *Key ideas of our previous approaches.* – At this point it is worthwhile to put the present discussion into the context of our earlier studies of the origin of the Schrödinger equation. Here we were guided by three principles: i) symmetric coupling [21] between amplitude and phase equations, ii) more freedom in phase [22], and iii) elimination of classical concepts [23].

Indeed, ref. [21] starts from a general mathematical identity for complex-valued functions which relates the operators appearing in the Schrödinger equation to terms reminiscent of the continuity equation and the classical Hamilton-Jacobi equation. The postulate of a conservation law for the intensity of a wave with its phase determined by the classical action leads to a *non-linear* wave equation. In this situation, the equation for the amplitude is driven by the gradient of the phase but the phase equation is free of the amplitude.

We can restore this broken symmetry in the coupling of the two equations by attaching the non-linearity to the equation of motion of the phase rather than the wave. In this case the resulting wave equation is linear, and corresponds to the Schrödinger equation.

In ref. [22] we start from a complex-valued function whose phase is determined in its dynamics by the Hamilton-Jacobi equation, but do not specify yet the time evolution of

the amplitude. As a result, we obtain a master wave equation whose two extreme limits represent the non-linear wave equation of classical statistical mechanics, and the linear Schrödinger equation of quantum mechanics.

In the first case the amplitude remains real for all times. However, in the second one the equation for the amplitude is such that it immediately accumulates phases even if initially the amplitude was purely real.

Hence, the phase of a quantum wave is the sum of the classical action and that of the amplitude. More freedom in phase in quantum than in classical physics therefore constitutes the main theme of ref. [22].

We emphasize that the approach pursued in the present lecture notes is closely related to, but also different from the one in ref. [22]. Indeed, we start again from a general mathematical identity which, in contrast to the one of ref. [21], is at the heart of gauge transformations. Moreover, we make the superposition principle the overarching idea of our lectures and obtain from it the Schrödinger equation.

The elimination of classical mechanics and the emergence of the Schrödinger equation is the guiding principle of ref. [23]. Again, we consider a wave consisting of the product of a complex-valued amplitude and a phase factor. Three assumptions lead us straight to the Schrödinger equation: i) The dynamics of the phase is governed by the classical action, ii) the equation of motion for the amplitude of the wave is determined by the condition that the desired linear wave equation is free of the classical action and its derivatives, and iii) there exists a continuity equation for the intensity of the wave. An additional consequence of this approach is the phase invariance of the Schrödinger equation.

Needless to say, our four approaches towards the Schrödinger equation share a lot of features and are strongly intertwined. They obviously represent different aspects of the same problem.

1˙3. *Outline*. – Our lecture notes are organized as follows: We start in sect. **2** by a historical account describing the developments that took place in the mid 1920's leading to the Schrödinger equation. Here we emphasize especially the application [24] of Hermann Weyl's concept [25] of parallel transport to the electron in an atom. This article [24] by Schrödinger, published already in 1922, and thus even before matrix mechanics, was an important stepping stone towards the Schrödinger equation.

We then dedicate sect. **3** to a brief review of several alternative derivations of the Schrödinger equation. The work [26] closest to ours adds momentum fluctuations governed by the Heisenberg uncertainty relation to the Hamilton-Jacobi equation, and uses a variational principle to obtain from it the Schrödinger equation.

In sect. **4** we establish the mathematical identity which is at the very heart not only of our approach towards the Schrödinger equation but also of gauge transformations. Whereas the latter connects two *wave functions* we use the identity to define *quantum* waves by postulating a specific equation of motion for the *amplitude* of the wave.

We gain more insight into the physical meaning of the amplitude equation in sect. **5** by choosing the classical action expressed in units of Planck's constant as the phase field. In this way we get a rather unusual Schrödinger equation which mixes classical

and quantum mechanics. We emphasize that this equation for the *amplitude* of the wave is completely equivalent to the familiar Schrödinger equation for the *total* wave.

Still the question remains why this particular wave equation for the amplitude should hold true. An answer emerges in sect. **6** where we take advantage of the Hamilton-Jacobi equation to simplify the amplitude equation. This analysis shows that the amplitude of a *quantum* wave has to assume *complex* values, and has to involve Planck's constant. As a consequence of the complex nature of the amplitude a quantum phase and a quantum current emerge. Their existence guarantees the linearity of the Schrödinger equation.

Indeed, *classical* waves given by purely *real* amplitudes can only lead to a non-linear wave equation as shown in sect. **7**. Here the non-linearity is given by the quantum potential of the Madelung-Bohm theory. Fortunately, a special class of classical waves suggests a way to the linear Schrödinger equation. Indeed, waves with an amplitude determined by the Van Vleck determinant are the WKB waves of semi-classical quantum mechanics. Nevertheless, we face a non-linear wave equation. Fortunately it reduces to the linear Schrödinger equation under appropriate conditions.

Finally we dedicate sect. **8** to a summary of our journey from the classical to the quantum domain arriving at the linear Schrödinger equation. We conclude in sect. **9** by summarizing our results and by providing an outlook.

In order to keep our lecture notes self-contained we have included two appendices. In appendix A we rederive the continuity equation satisfied by the Van Vleck determinant, first in one, and then in an arbitrary number of space dimensions. We dedicate appendix B to the proof that the familiar WKB wave function obeys a non-linear wave equation.

2. – Road towards the Schrödinger equation

At the beginning of the last century the discrete energy spectrum of an atom represented a great mystery. Indeed, the early successes of Niels Bohr's and Arnold Sommerfeld's *Atommechanik* [27] which explained this phenomenon by imposing quantization conditions on the classical orbits of the electrons, were soon overcast by serious problems indicating that this approach could not represent the ultimate answer.

It was Werner Heisenberg [28] who in the spring of 1925 recognized in a lonely night on the island of Helgoland that the key to an understanding of the atom is not the motion of the electron on a *single* orbit, but the jump between *two* different ones. As a result, one has to deal with a new type of mechanics where the position and, by the same token, the momentum of a particle are characterized by *two* indices rather than a single one.

Max Born and Pascal Jordan immediately identified [29] these objects as matrices familiar from linear algebra. Matrix mechanics as outlined by these three scientists in the *Drei-Männerarbeit* [30] was born. For a concise summary of matrix mechanics we refer to the lectures of Manfred Kleber in this volume [31].

At the same time Erwin Schrödinger had embarked on an at first sight completely different route. He formulated a wave equation which governs the propagation of a quantity called *Feldskalar* (wave function) describing the electron in the atom. Moreover,

he was also able to resolve the apparent contradiction between the *discontinuous* quantum jumps of Heisenberg, and the *continuous* time evolution of his wave function. Indeed, both approaches represent different, but completely equivalent points of view.

Due to its simplicity the Schrödinger formulation of quantum mechanics quickly superseded the matrix one. Today, the Schrödinger equation is the central tool to address phenomena of the microscopic world.

Despite its success, the origin of the Schrödinger equation is murky. Indeed, Schrödinger in his original papers [32] arrived at it in a rather convoluted way [33]. He was obviously motivated by the ideas of Louis de Broglie [34], who in 1924 had proposed the wave nature of matter.

Indeed, Felix Bloch recalls [35] that when Schrödinger summarized these ideas in a physics colloquium at the request of Peter Debye he was challenged by Debye to find a wave equation. Amazingly only a couple of weeks later, Schrödinger could present in another colloquium his equation, together with the solutions of the hydrogen atom, the harmonic oscillator, and many other elementary quantum systems.

One might wonder [36] how Schrödinger could arrive that quickly at his wave equation and obtain such a wealth of results. The answer to this riddle emerges when we recall that Schrödinger was not only familiar with de Broglie's idea of matter waves but he had also learned from Hermann Weyl to employ the parallel transport of a vector to define a field. This concept was the beginning [37] of gauge theories [38].

In 1918 Weyl [39] proposed a unification of gravitation and electricity by postulating that a vector, when transported along a closed path, does not only change its orientation as a result of the curvature of spacetime induced by the gravitational field, but also its length. This length change expressed by the *Weyl Streckenfaktor* is the exponential of the line integral of the four-vector potential describing the electromagnetic field.

Although this idea immediately attracted criticism from Albert Einstein [40] it was decisive for the development of wave mechanics. Indeed, in 1922 Schrödinger [24] assumed that the electron in the Bohr atom would carry such a vector along its closed orbit. Although he never discussed the exact nature of the vector, which was spelled out only in 1927 by Fritz London [41] *after* the discovery of wave mechanics, he found that in this case the line integral is proportional to an integer multiple of Planck's constant $h = 2\pi\hbar$.

Schrödinger speculated that if the constant of proportionality in the exponential function would involve the imaginary unit "i", and $\hbar$ in the form $\mathrm{i}/\hbar$ the exponential *length* change would turn into a periodic *phase* factor, and the Bohr orbit could be understood as a resonance phenomenon of a wave. Hence, in Schrödinger's article of 1922 phase factors which are central to quantum theory, and manifest themselves for example in the Aharonov-Bohm effect [42, 43], were born.

Schrödinger's article [24] is remarkable since it appeared two years before de Broglie's great insight into the wave nature of matter, and three years before Schrödinger's own development of wave mechanics. However, it was not till 1929 that Weyl [44] recognized that his idea of a length change along a closed path, when translated into phase factors, unifies quantum mechanics and electrodynamics rather than gravity and electrodynamics.

Indeed, in order to preserve the form invariance of the Schrödinger equation under a

phase transformation of the wave function Weyl introduces the electromagnetic potentials which allow to him to absorb the additional terms arising from the transformation. This elemination is made possible by the gauge invariance of the electromagnetic field. It goes without saying that Weyl's seminal paper [44] has given rise to the development of gauge theories [45, 46] which are at the very heart of particle physics.

3. – Comparison with the literature

It is straightforward to start from quantum mechanics and move towards classical mechanics. The most prominent example of such an approach is due to Madelung [19] and Bohm [20], who decompose [47] the wave function into an amplitude and a phase factor. The Schrödinger equation then leads to the continuity equation and the Hamilton-Jacobi equation of classical mechanics with an additional potential which is proportional to the square of Planck's constant. This potential is commonly referred to as the quantum potential [48].

In our lectures notes we move in the opposite direction. We use the Hamilton-Jacobi equation of classical mechanics, that is in the absence of a quantum potential, to derive a wave equation for a wave consisting of an arbitrary amplitude and a phase given by the classical action. A particular choice of this amplitude eliminates in this wave equation all remnants of classical mechanics and the Schrödinger equation emerges. This amplitude which gives birth to quantum mechanics follows from a Schrödinger equation containing the gradient of the gauge field making contact with Schrödinger's paper of 1922.

We are aware of several alternative derivations of the Schrödinger equation and mention here only a few. For an interesting approach using the dispersion relation of matter we refer to the lectures by Gerd Leuchs in this volume [49].

The best known derivation of the Schrödinger equation is due to Richard P. Feynman [50] and is based on his path integral formulation of quantum mechanics. A rather radical idea [51] claims that the time-dependent Schrödinger equation is an approximation, and arises from the elimination of degrees of freedom of the time-independent Schrödinger equation of two coupled systems. The most elementary example is the reduction of the three-dimensional Helmholtz equation in the paraxial wave approximation which leads to a time-dependent Schrödinger equation in two dimensions. An even more extreme position was taken by Willis E. Lamb [52], who asked the question: "Suppose Newton had invented quantum mechanics", or argued in favor of a superclassical theory.

A derivation of the Schrödinger equation based on quantum field theory was outlined in ref. [53]. Moreover, an interesting connection between the path integral and the Hamilton-Jacobi equation was suggested by ref. [54].

The work closest to our approach can be found in ref. [26]. Here momentum fluctuations have been added to the Hamilton-Jacobi equation which when subjected to the Heisenberg uncertainty relation lead to the Schrödinger equation. The connection between ref. [26] and our lecture notes comes to light when we recall that the momentum is determined by the gradient of the phase. Hence, additional contributions to the momentum imply additional phases. Indeed, our approach eliminates the remnants of

classical physics by allowing the phase of the wave function to go beyond the one given by the classical action. Since this quantum phase is determined by another Schrödinger equation, the uncertainty principle is satisfied automatically.

4. – Why zero?

We start our "derivation" of the Schrödinger equation by first establishing a mathematical identity which is at the very heart of gauge invariance. However, the logic of our approach is opposite to the standard one. Indeed, our mathematical relation allows us to *define* for an arbitrary phase field a quantum wave by postulating a specific dynamics of its amplitude.

4˙1. *A curious mathematical identity.* – Throughout these notes we consider waves $\Phi = \Phi(\boldsymbol{r}, t)$ which depend on the three-dimensional position vector $\boldsymbol{r}$ and the time coordinate t, and represent them by the position- and time-dependent amplitude $A = A(\boldsymbol{r}, t)$ and phase $\alpha = \alpha(\boldsymbol{r}, t)$ in the form

$$\Phi = A\mathrm{e}^{\mathrm{i}\alpha}. \tag{1}$$

Here we do not assume that A is real and positive. In the course of these notes, we shall find that quantum mechanics requires complex-valued amplitudes.

We differentiate Φ with respect to time to obtain the relation

$$\mathrm{i}\hbar\,\frac{\partial\Phi}{\partial t} = \mathrm{e}^{\mathrm{i}\alpha}\left[\mathrm{i}\hbar\,\frac{\partial A}{\partial t} - \hbar\,\frac{\partial\alpha}{\partial t}\,A\right], \tag{2}$$

and recall the identity

$$-\hbar^2\Delta\Phi = \mathrm{e}^{\mathrm{i}\alpha}\left(\frac{\hbar}{\mathrm{i}}\,\boldsymbol{\nabla} + \hbar\boldsymbol{\nabla}\alpha\right)^2 A, \tag{3}$$

which follows immediately when we express A in the form

$$A = \mathrm{e}^{-\mathrm{i}\alpha}\Phi. \tag{4}$$

Next we subtract on both sides of eq. (2) the term

$$\hat{H}\Phi \equiv -\frac{\hbar^2}{2m}\,\Delta\Phi + V\Phi, \tag{5}$$

where $V \equiv V(\boldsymbol{r}, t)$ denotes a scalar function which may depend on position and time, and m denotes the mass of the particle.

Together with eq. (3) we arrive at the identity

$$\mathrm{i}\hbar\,\frac{\partial\Phi}{\partial t} - \hat{H}\Phi = \mathrm{e}^{\mathrm{i}\alpha}\left[\mathrm{i}\hbar\,\frac{\partial A}{\partial t} - \frac{1}{2m}\left(\frac{\hbar}{\mathrm{i}}\,\boldsymbol{\nabla} + \hbar\boldsymbol{\nabla}\alpha\right)^2 A - \left(V + \hbar\,\frac{\partial\alpha}{\partial t}\right) A\right]. \tag{6}$$

It is amusing that just by appropriate differentiations of the decomposition, eq. (1), of the wave Φ into an amplitude A and a phase α, we have found a mathematical identity which contains on its left-hand side all the ingredients of the Schrödinger equation, and on the right-hand side a Schrödinger equation for A.

The latter emerges from the one of Φ by the substitutions

$$\frac{\hbar}{\mathrm{i}}\,\boldsymbol{\nabla} \to \frac{\hbar}{\mathrm{i}}\,\boldsymbol{\nabla} + \hbar\boldsymbol{\nabla}\alpha \tag{7}$$

and

$$V \to V + \hbar\,\frac{\partial\alpha}{\partial t}\,, \tag{8}$$

familiar from Weyl's concept [44] of gauge transformations.

4˙2. *Definition of a quantum wave by its amplitude.* – However, we emphasize that the approach of Weyl is the inverse of ours. Indeed, he first assumes that $\psi \equiv \psi(\boldsymbol{r}, t)$ satisfies the Schrödinger equation

$$\mathrm{i}\hbar\frac{\partial\psi}{\partial t} = \hat{H}\psi, \tag{9}$$

and then argues that the wave function

$$\tilde{\psi} \equiv \psi \mathrm{e}^{-\mathrm{i}\Lambda} \tag{10}$$

with the gauge field $\Lambda = \Lambda(\boldsymbol{r}, t)$ obeys the transformed Schrödinger equation

$$\mathrm{i}\hbar\,\frac{\partial\tilde{\psi}}{\partial t} = \frac{1}{2m}\left(\frac{\hbar}{\mathrm{i}}\,\boldsymbol{\nabla} + \hbar\boldsymbol{\nabla}\Lambda\right)^2\tilde{\psi} + \left(V + \hbar\,\frac{\partial\Lambda}{\partial t}\right)\tilde{\psi}. \tag{11}$$

Since, Weyl starts from the Schrödinger equation, eq. (9), for ψ, the left-hand side of our mathematical identity, eq. (6), vanishes. As a consequence, the right-hand side of eq. (6), that is the equation of motion for $\tilde{\psi}$ has to vanish as well, and eq. (11) has to hold.

In contrast, we employ the mathematical identity, eq. (6), to *define* quantum waves

$$\psi \equiv R\mathrm{e}^{\mathrm{i}\alpha} \tag{12}$$

as waves whose amplitudes $R \equiv R(\boldsymbol{r}, t)$ make the right-hand side of eq. (6) vanish, that is they follow from the equation of motion

$$\mathrm{i}\hbar\,\frac{\partial R}{\partial t} = \frac{1}{2m}\left(\frac{\hbar}{\mathrm{i}}\,\boldsymbol{\nabla} + \hbar\boldsymbol{\nabla}\alpha\right)^2 R + \left(V + \hbar\,\frac{\partial\alpha}{\partial t}\right)R. \tag{13}$$

Hence, for a given phase function $\alpha = \alpha(\boldsymbol{r}, t)$ this partial differential equation determines the amplitude R of the wave ψ which by virtue of eq. (6), now satisfies the Schrödinger equation.

4˙3. *Formulation of the problem.* – So far, we have not specified the form of α, nor have we given an argument why eq. (13) should hold true. Indeed, the central question is:

> *Why does the right-hand side of the mathematical identity, and therefore also the left-hand side vanish?*

Hence, the question is not anymore: "Why does the Schrödinger equation contain the operators we are familiar with?", but "Why is the amplitude R of a quantum wave with phase α given by eq. (13)?"

5. – Classical mechanics guides the amplitude of the Schrödinger wave

The answer as to the origin of the vanishing of the mathematical identity must emerge from physics. Indeed, it is the Hamilton-Jacobi formulation of classical mechanics which provides us with insight. In this section we first briefly review this approach towards classical mechanics, and then analyze the consequences of the classical action as a phase field.

5˙1. *Hamilton-Jacobi theory in a nutshell.* – In classical mechanics we are interested in determining the trajectory $\boldsymbol{r} = \boldsymbol{r}(t)$ of a particle of mass m moving in a position- and time-dependent potential $V = V(\boldsymbol{r}, t)$. Many ways of treating this problem ranging from Newton's equation to Hamilton's equations [55] offer themselves.

Most relevant for the present discussion is the formulation of classical mechanics in terms of the action $S^{(\mathrm{cl})} = S^{(\mathrm{cl})}(\boldsymbol{r}, t; \boldsymbol{\varepsilon})$ which satisfies the Hamilton-Jacobi equation [55, 56]

$$-\frac{\partial S^{(\mathrm{cl})}}{\partial t} = \frac{1}{2m}\left(\boldsymbol{\nabla} S^{(\mathrm{cl})}\right)^2 + V. \tag{14}$$

Here we have included the superscript (cl) to indicate that we deal with the *classical* action, and $\boldsymbol{\varepsilon}$ is a time-independent parameter. In particular, $S^{(\mathrm{cl})}$ can take on only real values.

Equation (14) describes the propagation of surfaces of constant phase $S^{(\mathrm{cl})}$ as a function of time. The motion of a particle given by the classical momentum $\boldsymbol{p} \equiv m\mathrm{d}\boldsymbol{r}/\mathrm{d}t$ is orthogonal to these planes and thus $\boldsymbol{p} = \boldsymbol{\nabla} S^{(\mathrm{cl})}$.

5˙2. *Classical action as a phase field.* – Since $S^{(\mathrm{cl})}$ represents wave fronts, it is suggestive to define the wave

$$\psi^{(\mathrm{q})} \equiv R\mathrm{e}^{\mathrm{i}S^{(\mathrm{cl})}/\hbar}, \tag{15}$$

and require it to be a quantum wave.

In this case the phase function α of the mathematical identity, eq. (6), reads

$$\alpha \equiv \frac{1}{\hbar}\, S^{(\mathrm{cl})}, \tag{16}$$

and according to eq. (13) the amplitude R of the quantum wave has to satisfy the equation of motion

$$\mathrm{i}\hbar\, \frac{\partial R}{\partial t} = \frac{1}{2m} \left(\frac{\hbar}{\mathrm{i}}\, \boldsymbol{\nabla} + \boldsymbol{\nabla} S^{(\mathrm{cl})} \right)^2 R + \left(V + \frac{\partial S^{(\mathrm{cl})}}{\partial t} \right) R. \tag{17}$$

This partial differential equation is reminiscent of the Schrödinger equation for a charged particle in an electromagnetic field described by a vector potential and a scalar potential. In the present case however, these roles are played by the classical momentum $\boldsymbol{p} \equiv \boldsymbol{\nabla} S^{(\mathrm{cl})}$ and the energy $\partial S^{(\mathrm{cl})}/\partial t$, respectively.

Our approach of starting from the mathematical identity, eq. (6), of a general wave Φ, and then imposing the amplitude condition, eq. (13), has opened up a new aspect of the Schrödinger equation which is not apparent in the standard formulation of quantum mechanics. Indeed, we now have two different methods to address a given quantum problem.

We can either solve the familiar Schrödinger equation

$$\mathrm{i}\hbar \frac{\partial \psi^{(\mathrm{q})}}{\partial t} = \hat{H} \psi^{(\mathrm{q})} \tag{18}$$

for the wave function $\psi^{(\mathrm{q})}$ with the Hamiltonian, eq. (5). In this formulation the connection to classical mechanics is lost, since we have eliminated [23] the action.

Alternatively, we may solve the Schrödinger equation, eq. (17), for the amplitude R of the wave. Here the classical action $S^{(\mathrm{cl})}$ appears explicitly in the kinetic energy operator as $\boldsymbol{\nabla} S^{(\mathrm{cl})}$, and with the scalar potential V as $\partial S^{(\mathrm{cl})}/\partial t$. Hence, the classical field $S^{(\mathrm{cl})} = S^{(\mathrm{cl})}(\boldsymbol{r}, t; \boldsymbol{\epsilon})$, that is, classical mechanics serves as a guiding field for the amplitude R of the quantum wave $\psi^{(\mathrm{q})}$. We note that this interpretation is reminiscent of the pilot-wave theory of de Broglie.

We emphasize that both $\psi^{(\mathrm{q})}$ and R must be normalizable. Moreover, according to eq. (15) their initial values $\psi^{(\mathrm{q})}(\boldsymbol{r}, t = 0)$ and $R(\boldsymbol{r}, t = 0)$ are connected by the identity

$$\psi^{(\mathrm{q})}(\boldsymbol{r}, t = 0) = R(\boldsymbol{r}, t = 0) \exp\left[\frac{\mathrm{i}}{\hbar}\, S^{(\mathrm{cl})}(\boldsymbol{r}, t = 0; \boldsymbol{\varepsilon}) \right] \tag{19}$$

to the initial action $S^{(\mathrm{cl})}(\boldsymbol{r}, t = 0; \boldsymbol{\varepsilon})$.

Needless to say, we are not aware of a solution of eq. (17) for R in the presence of an arbitrary $S^{(\mathrm{cl})}$. Such an expression would allow us to solve, with the help of the definition eq. (15) of $\psi^{(\mathrm{q})}$ in terms of R and $S^{(\mathrm{cl})}$, the Schrödinger equation in the presence of an arbitrary potential $V = V(\boldsymbol{r}, t)$.

6. – Quantum condition implies linear Schrödinger equation

In the preceding section we have considered the amplitude condition, eq. (13), for the special example $\alpha \equiv S^{(\mathrm{cl})}/\hbar$ of a phase function, and have arrived at the transformed Schrödinger equation, eq. (17) for the amplitude R of the wave. We recall that this equation ensures the Schrödinger equation, eq. (18), for $\psi^{(\mathrm{q})}$.

Thus the origin of the Schrödinger equation is intimately related to the questions: i) What is the meaning of eq. (17), and ii) why should it be satisfied in nature?

6˙1. *Emergence of a quantum phase.* – To answer these questions, we first take advantage of the Hamilton-Jacobi equation (14) to simplify eq. (17). Indeed, the discussion of sect. **5** has not even used the dynamics of $S^{(\mathrm{cl})}$ which now reduces eq. (17) to

$$\frac{\partial R}{\partial t} = \mathrm{i}\,\frac{\hbar}{2m}\,\Delta R - \frac{\boldsymbol{\nabla} S^{(\mathrm{cl})}}{m}\cdot\boldsymbol{\nabla} R - \frac{1}{2}\,\boldsymbol{\nabla}\cdot\left(\frac{\boldsymbol{\nabla} S^{(\mathrm{cl})}}{m}\right) R. \tag{20}$$

We note that the first term on the right-hand side of this equation is proportional to the imaginary unit i and Planck's constant. As a result, in the course of time the amplitude R must assume complex values. Moreover, it must involve $\hbar$ explicitly.

Hence, $\hbar$ enters the wave $\psi^{(\mathrm{q})}$ not only through its phase $S^{(\mathrm{cl})}/\hbar$ but also its amplitude R. For this reason, and since the condition leads us straight to the Schrödinger equation, we refer to eq. (20) as the *quantum condition*.

The quantum condition, eq. (20), enforces a complex-valued amplitude

$$R \equiv |R|\mathrm{e}^{\mathrm{i}\beta} \equiv A^{(\mathrm{q})}\mathrm{e}^{\mathrm{i}\beta} \tag{21}$$

of the wave $\psi^{(\mathrm{q})}$ where the absolute value $|R| \equiv A^{(\mathrm{q})}$, as well as the phase β, depend on $\hbar$. But why is this so seemingly innocent property powerful enough to lead us straight from the Hamilton-Jacobi equation of classical mechanics to the Schrödinger equation summarizing quantum mechanics?

We gain insight into this question when we slightly rewrite the quantum condition, eq. (20), with the decomposition eq. (21). For this purpose we express eq. (20) in real and imaginary parts.

With the help of the identities eqs. (2) and (3), applied to $A^{(\mathrm{q})}$ defined by eq. (21) instead of Φ, eq. (20) turns after minor calculation into the relation

$$\begin{aligned} 2\frac{\partial A^{(\mathrm{q})}}{\partial t} + \mathrm{i}2\,\frac{\partial \beta}{\partial t}A^{(\mathrm{q})} = &-\left\{A^{(\mathrm{q})}\boldsymbol{\nabla}\cdot\left(\frac{1}{m}\,\boldsymbol{\nabla} S^{(\mathrm{q})}\right) + \frac{2}{m}\boldsymbol{\nabla} S^{(\mathrm{q})}\cdot\boldsymbol{\nabla} A^{(\mathrm{q})}\right\} \\ &+\mathrm{i}\left\{\frac{\hbar}{m}\,\Delta A^{(\mathrm{q})} - \frac{\hbar}{m}\,A^{(\mathrm{q})}(\boldsymbol{\nabla}\beta)^2 - 2A^{(\mathrm{q})}\,\frac{\boldsymbol{\nabla} S^{(\mathrm{cl})}}{m}\cdot\boldsymbol{\nabla}\beta\right\}, \end{aligned} \tag{22}$$

where we have introduced the abbreviation

$$S^{(\mathrm{q})} \equiv \hbar\beta + S^{(\mathrm{cl})} \tag{23}$$

for the sum of the quantum and the classical actions, $\hbar\beta$ and $S^{(\mathrm{cl})}$, respectively.

Hence, by taking real and imaginary parts of eq. (22) we find the two equations

$$2\,\frac{\partial A^{(\mathrm{q})}}{\partial t} + A^{(\mathrm{q})}\boldsymbol{\nabla}\cdot\left(\frac{\boldsymbol{\nabla}S^{(\mathrm{q})}}{m}\right) + 2\boldsymbol{\nabla}A^{(\mathrm{q})}\cdot\frac{1}{m}\,\boldsymbol{\nabla}S^{(\mathrm{q})} = 0 \tag{24}$$

and

$$-\frac{\partial\beta}{\partial t} = \frac{\hbar(\boldsymbol{\nabla}\beta)^2}{2m} + \frac{\boldsymbol{\nabla}S^{(\mathrm{cl})}}{m}\cdot\boldsymbol{\nabla}\beta - \frac{\hbar}{2m}\,\frac{\Delta A^{(\mathrm{q})}}{A^{(\mathrm{q})}}\,. \tag{25}$$

Here we have used the fact that the classical action $S^{(\mathrm{cl})}$ is real.

In summary, the mathematical identity, eq. (6), implies that the Schrödinger equation holds for $\psi^{(\mathrm{q})}$ provided the amplitude R satisfies the wave equation, eq. (13). For the special choice of the phase function $\alpha \equiv S^{(\mathrm{cl})}/\hbar$, eq. (13) enforces via eq. (20) the fact that R is complex. The amplitude $|R| \equiv A^{(\mathrm{q})}$ and the phase angle β then need to obey the coupled equations of motion for $A^{(\mathrm{q})}$ and β given by eqs. (24) and (25), respectively.

6˙2. *Continuity equation with quantum current.* – More insight into the physics of the two equations, eqs. (24) and (25), emerges when we multiply eq. (24) by $A^{(\mathrm{q})}$ which yields the continuity equation

$$\frac{\partial}{\partial t}\,\rho + \boldsymbol{\nabla}\cdot\boldsymbol{j} = 0. \tag{26}$$

Here we have introduced the density

$$\rho \equiv |R|^2 = \left(A^{(\mathrm{q})}\right)^2, \tag{27}$$

together with the current

$$\boldsymbol{j} \equiv \left(A^{(\mathrm{q})}\right)^2\,\frac{1}{m}\,\boldsymbol{\nabla}S^{(\mathrm{q})}. \tag{28}$$

When we recall the definition, eq. (23), of $S^{(\mathrm{q})}$, we recognize that two terms contribute to the current: Indeed, the *classical current*

$$j^{(\mathrm{cl})} \equiv \left(A^{(\mathrm{q})}\right)^2\,\frac{\boldsymbol{\nabla}S^{(\mathrm{cl})}}{m} \tag{29}$$

is determined mainly by the classical action $S^{(\mathrm{cl})}$.

In contrast, the *quantum current*

$$\boldsymbol{j}^{(\mathrm{q})} \equiv \left(A^{(\mathrm{q})}\right)^2\,\frac{\boldsymbol{\nabla}(\hbar\beta)}{m} \tag{30}$$

involves $\hbar$ through the phase β as well as $A^{(\mathrm{q})}$.

This decomposition of $\boldsymbol{j}$ allows us to answer the question concerning the special role of a complex-valued amplitude R enforced by the quantum condition. Whereas, for a real-valued amplitude R we have $\beta = 0$ and thus $\boldsymbol{j}^{(\mathrm{q})} = 0$, the quantum condition enforces a non-vanishing quantum current.

6˙3. *Quantum Hamilton-Jacobi equation.* – We conclude by briefly addressing the equation of motion, eq. (25), for β. For this purpose we multiply both sides by $\hbar$ which provides us with the equation of motion

$$-\frac{\partial(\hbar\beta)}{\partial t} = \frac{(\boldsymbol{\nabla}(\hbar\beta))^2}{2m} + \frac{\boldsymbol{\nabla} S^{(\mathrm{cl})}}{m} \cdot \boldsymbol{\nabla}(\hbar\beta) + \mathcal{Q}^{(\mathrm{MB})}[A^{(\mathrm{q})}] \tag{31}$$

for the quantum action $\hbar\beta$. Here we have introduced the Madelung-Bohm quantum potential

$$\mathcal{Q}^{(\mathrm{MB})}[|\varphi|] \equiv -Q[|\varphi|] \equiv -\frac{\hbar^2}{2m}\,\frac{\Delta|\varphi|}{|\varphi|} \tag{32}$$

for a complex-valued function $\varphi = \varphi(\boldsymbol{r}, t)$.

We emphasize that eq. (31) is reminiscent of the *classical* Hamilton-Jacobi equation eq. (14) with two small, but important differences: The appearance of i) a convective derivative $\boldsymbol{\nabla} S^{(\mathrm{cl})} \cdot \boldsymbol{\nabla}(\hbar\beta)/m$, and ii) the Madelung-Bohm quantum potential $\mathcal{Q}^{(\mathrm{MB})}$.

When we add to the equation of motion, eq. (31) for the quantum action $\hbar\beta$ the one for the *classical* action $S^{(\mathrm{cl})}$, that is the classical Hamilton-Jacobi equation, eq. (14), we arrive at the *quantum* Hamilton-Jacobi equation

$$-\frac{\partial S^{(\mathrm{q})}}{\partial t} = \frac{(\boldsymbol{\nabla} S^{(\mathrm{q})})^2}{2m} + V + \mathcal{Q}^{(\mathrm{MB})}[A^{(\mathrm{q})}] \tag{33}$$

for the total action $S^{(\mathrm{q})}$ defined by eq. (23).

This analysis shows that the convective derivative in eq. (31) is a consequence of the quadratic non-linearity of the Hamilton-Jacobi equation, and the total action being the sum of the classical and the quantum action. The Madelung-Bohm potential $\mathcal{Q}^{(\mathrm{MB})}$ reflects the fact that according to eq. (25) the quantum phase β is driven by the amplitude $A^{(\mathrm{q})}$ through the ratio $\Delta A^{(\mathrm{q})}/A^{(\mathrm{q})}$.

7. – Classicality condition implies non-linear wave equation

The quantum current introduced in the preceding section is a consequence of the equation of motion, eq. (20), for the amplitude R of a quantum wave. By virtue of the mathematical identity, eq. (6), it gives rise to the linear Schrödinger equation.

The importance of the quantum current stands out most clearly when we consider for the moment a real-valued amplitude $A_R^{(\mathrm{cl})}$. Here we have included a subscript R in the amplitude as to reflect the fact that Nathan Rosen has advocated this approach. As a

result of $A_R^{(\mathrm{cl})}$ being real, β vanishes, and eq. (25) cannot be satisfied since there are no terms left to compensate the term $\Delta A_R^{(\mathrm{cl})}/A_R^{(\mathrm{cl})}$ driving β.

In the present section we analyze the wave equations resulting from such a classicality condition. Here we consider first the general case of a real-valued amplitude, and then focus on a specific one given by the Van Vleck determinant. The latter is directly connected to the primitive WKB wave functions [18].

7˙1. *General real amplitude.* – According to eq. (25), a wave

$$\psi_R^{(\mathrm{cl})} \equiv A_R^{(\mathrm{cl})} \mathrm{e}^{\mathrm{i}S^{(\mathrm{cl})}/\hbar} \tag{34}$$

with a real amplitude $A_R^{(\mathrm{cl})}$ does not obey the Schrödinger equation. Indeed, the mathematical identity, eq. (6), yields for $\psi_R^{(\mathrm{cl})}$ the relation

$$\left[\mathrm{i}\hbar\,\frac{\partial}{\partial t} - \hat{H}\right]\psi_R^{(\mathrm{cl})} = \left[\frac{\hbar^2}{2m}\,\Delta A_R^{(\mathrm{cl})} + \mathrm{i}\hbar\left(\frac{\partial A_R^{(\mathrm{cl})}}{\partial t} - \boldsymbol{\nabla} A_R^{(\mathrm{cl})}\cdot\frac{\boldsymbol{\nabla} S^{(\mathrm{cl})}}{m} - \frac{1}{2}\,A_R^{(\mathrm{cl})}\boldsymbol{\nabla}\cdot\left(\frac{\boldsymbol{\nabla} S^{(\mathrm{cl})}}{m}\right)\right)\right]\mathrm{e}^{\mathrm{i}S^{(\mathrm{cl})}/\hbar}, \tag{35}$$

where we have also made use of the classical Hamilton-Jacobi equation, eq. (14).

Since we require $A_R^{(\mathrm{cl})}$ to be real, we have to subject it to the condition

$$\frac{\partial A_R^{(\mathrm{cl})}}{\partial t} = \boldsymbol{\nabla} A_R^{(\mathrm{cl})}\cdot\frac{\boldsymbol{\nabla} S^{(\mathrm{cl})}}{m} + \frac{1}{2}\,A_R^{(\mathrm{cl})}\boldsymbol{\nabla}\cdot\left(\frac{\boldsymbol{\nabla} S^{(\mathrm{cl})}}{m}\right). \tag{36}$$

When we multiply this equation by $A_R^{(\mathrm{cl})}$ we find the classical continuity equation

$$\frac{\partial}{\partial t}\,\rho^{(\mathrm{cl})} + \boldsymbol{\nabla}\cdot\boldsymbol{j}^{(\mathrm{cl})} = 0, \tag{37}$$

which contains, apart from the density

$$\rho^{(\mathrm{cl})} \equiv \left(A_R^{(\mathrm{cl})}\right)^2 \tag{38}$$

determined by the amplitude $A_R^{(\mathrm{cl})}$, only the classical current

$$\boldsymbol{j}^{(\mathrm{cl})} \equiv \left(A_R^{(\mathrm{cl})}\right)^2 \frac{\boldsymbol{\nabla} S^{(\mathrm{cl})}}{m}\,, \tag{39}$$

in complete agreement with eq. (29).

As a result, the amplitudes $A_R^{(\mathrm{cl})}$ following from the condition eq. (36) are independent of $\hbar$. For this reason we refer to it as the *classicality condition*.

According to eq. (35) a classical wave of the form eq. (34) with the amplitude $A_R^{(\mathrm{cl})}$ given by the classicality condition, eq. (36), satisfies the non-linear wave equation [13-15].

$$\mathrm{i}\hbar\frac{\partial\psi_R^{(\mathrm{cl})}}{\partial t} = \hat{H}\psi_R^{(\mathrm{cl})} + Q[|\psi_R^{(\mathrm{cl})}|]\psi_R^{(\mathrm{cl})}. \tag{40}$$

Here we have recalled the definitions, eqs. (32) and (34) of Q and $\psi_R^{(\mathrm{cl})}$, and have made use of the fact that $|\psi_R^{(\mathrm{cl})}| = A_R^{(\mathrm{cl})}$.

Hence, the classicality condition, eq. (36), corresponding to a *vanishing* quantum current connects the Hamilton-Jacobi equation, eq. (14), of classical mechanics with the *non-linear* wave equation, eq. (40). In contrast, the quantum condition, eq. (20) which enforces a *non-vanishing* quantum current takes us straight to the *linear* Schrödinger equation, eq. (18). Obviously the linearization of the non-linear wave equation is due to the existence of the quantum current.

7˙2. *Amplitude given by Van Vleck determinant.* – So far we have not provided an explicit expression for the amplitude of the wave. Indeed, our arguments have rested exclusively on the corresponding equations of motion, eqs. (20) and (36) representing the quantum and classical conditions, or the associated continuity equations, eqs. (26) and (37), respectively. We now give an explicit expression for a classical wave which we identify as the primitive WKB function [18]. This identification will guide us to the Schrödinger equation.

7˙2.1. Super-classical waves. In appendix A we recall that the Van Vleck determinant [17]

$$D \equiv \left|\frac{\partial^2 S^{(\mathrm{cl})}}{\partial x_k \partial \varepsilon_l}\right|, \tag{41}$$

formed by the second derivatives of the classical action $S^{(\mathrm{cl})}$ with respect to the Cartesian coordinates x_k and the constants of motion ε_l with $(k,l) = 1,2,3$, satisfies [17, 13] the continuity equation

$$\frac{\partial}{\partial t}D + \boldsymbol{\nabla}\cdot\left(D\frac{1}{m}\boldsymbol{\nabla}S^{(\mathrm{cl})}\right) = 0. \tag{42}$$

Indeed, as shown in appendix A this conservation law is a consequence of the classical Hamilton-Jacobi equation.

We can now define a wave

$$\psi^{(\mathrm{cl})} \equiv D^{1/2}\mathrm{e}^{\mathrm{i}S^{(\mathrm{cl})}/\hbar}, \tag{43}$$

whose amplitude is given by the square root of the Van Vleck determinant.

Since $D^{1/2}$ is real-valued, and we have the continuity equation, eq. (42), the wave $\psi^{(\mathrm{cl})}$ enjoys the classicality condition, eq. (36). Therefore, $\psi^{(\mathrm{cl})}$ obeys the non-linear wave equation

$$\mathrm{i}\hbar\frac{\partial\psi^{(\mathrm{cl})}}{\partial t} = \hat{H}\psi^{(\mathrm{cl})} + Q[|\psi^{(\mathrm{cl})}|]\psi^{(\mathrm{cl})}. \tag{44}$$

We emphasize that despite its similarity to the linear Schrödinger equation the non-linear wave equation, eq. (44), for $\psi^{(\mathrm{cl})}$ still only represents classical mechanics. Indeed, the Hamilton-Jacobi equation, eq. (14), together with the Van Vleck continuity equation, eq. (42), neither of which involve $\hbar$, is equivalent to the non-linear wave equations, eq. (44). For this reason, and to distinguish them from the classical waves of Rosen we refer to $\psi^{(\mathrm{cl})}$ as super-classical waves.

7˙2.2. Super-classical waves are WKB waves. Nevertheless, we are already rather close to quantum mechanics. This fact stands out most clearly when we restrict ourselves for the moment to one space dimension with coordinate x.

In this case we find from appendix A the explicit expression

$$D = \frac{m}{p} \tag{45}$$

for the Van Vleck determinant D in terms of the classical momentum

$$p(x;E) = \sqrt{2m[E - V(x)]} \tag{46}$$

of the particle of mass m and energy E in a potential V.

When we recall the corresponding classical action

$$S^{(\mathrm{cl})}(x,t;E) \equiv \int_{x_0}^{x} \mathrm{d}\tilde{x}\, p(\tilde{x};E) - Et, \tag{47}$$

where x_0 is an arbitrary position, the super-classical wave $\psi^{(\mathrm{cl})}$ defined by eq. (43) reads

$$\psi^{(\mathrm{cl})}(x,t) \equiv \mathcal{N}\,\frac{1}{\sqrt{p}}\exp\left[\frac{\mathrm{i}}{\hbar}\int_{x_0}^{x}\mathrm{d}\tilde{x}\, p(\tilde{x},E)\right]\exp\left[-\frac{\mathrm{i}}{\hbar}\,Et\right], \tag{48}$$

where $\mathcal{N}$ is a normalization constant, and thus represents the familiar primitive WKB wave function [18].

As argued above, $\psi^{(\mathrm{cl})}$ satisfies [18] the non-linear wave equation, eq. (44). However, it is instructive to rederive this wave equation directly from the explicit expression, eq. (48), of $\psi^{(\mathrm{cl})}$.

In appendix B we perform this calculation and demonstrate that in this case Q takes the explicit form

$$Q \equiv \frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(\frac{1}{\sqrt{p}}\right)\sqrt{p}. \tag{49}$$

Needless to say, this expression also follows immediately from the definition, eq. (32), of Q together with the relation

$$\left|\psi^{(\mathrm{cl})}\right| = \frac{|\mathcal{N}|}{\sqrt{p}}\,, \tag{50}$$

which is a consequence of eq. (48).

We note that $\psi^{(\mathrm{cl})}$ is an excellent approximation of the exact solution of the Schrödinger equation in situations when we can neglect Q compared to the effect of $\hat{H}$ on $\psi^{(\mathrm{cl})}$. The explicit expression, eq. (49), of Q in terms of the second derivative of the square root of p allows us to argue [18] that this condition is satisfied when the particle is appropriately away from the turning points of the classical motion where p vanishes, and for potentials V which vary sufficiently slowly.

Hence, for the special class of classical waves represented by WKB waves, and specific domains of space, the non-linear wave equation is almost indistinguishable from the linear Schrödinger equation. In this sense we have achieved the desired linearization.

8. – From Van Vleck via Rosen to Schrödinger

We have now reached the point where we can follow the path from the classical to the quantum world as promised in the introduction of these lecture notes, and outlined in fig. 1. This approach brings out most clearly that it is the linearity of quantum mechanics which gives birth to the Schrödinger equation.

We have come a long way since we have embarked on our journey in search for the origin of the Schrödinger equation. On our trip we have met several wave functions summarized in table I. We now analyze them one more time. However, in contrast to the discussion of the preceding sections which traversed fig. 1 from the bottom right, corresponding to the quantum world, to the top left representing the classical world, we presently move in the opposite direction, starting from the super-classical wave $\psi^{(\mathrm{cl})}$ and finally arriving at the quantum wave $\psi^{(\mathrm{q})}$.

8˙1. *The need for linearity.* – The super-classical wave $\psi^{(\mathrm{cl})}$ consisting of an amplitude given by the square root of the Van Vleck determinant, and a phase determined by the classical action corresponds to classical physics. We emphasize that neither the Hamilton-Jacobi equation nor the continuity equation contain Planck's constant. Nevertheless, we are close to, but not quite at quantum theory. Although $\psi^{(\mathrm{cl})}$ corresponds to a primitive WKB wave function, it does not satisfy the *linear* Schrödinger equation but a *non-linear* wave equation due to the presence of the classicality enforcing potential Q.

Only under appropriate conditions such as being away from the turning points, and for slowly varying potentials can we neglect the non-linearity associated with the classicality, and arrive at the linear Schrödinger equation. Obviously in this part of space, $\psi^{(\mathrm{cl})}$ contains a grain of quantum mechanics, and we marvel at the fact that a quantity constructed exclusively out of classical building blocks, such as the Van Vleck determinant and the classical action, can be so close to quantum theory.

TABLE I. – *Summary of the building blocks such as amplitude and phase, their dynamics and the resulting wave equations of the four major wave functions employed in these lecture notes to analyze the origin of the Schrödinger equation. For this reason we have not included the wave functions ψ and $\tilde{\psi}$ which made their appearance in the discussion of the gauge principle. Nor have we mentioned the function φ used in the definition of the potential Q.*

Wave function	Dynamics of phase	Dynamics of amplitude	Wave equation
$\psi^{(\mathrm{cl})} \equiv D^{1/2}\mathrm{e}^{\mathrm{i}S^{(\mathrm{cl})}/\hbar}$ $D \equiv$ Van Vleck determinant is real $S^{(\mathrm{cl})} \equiv$ classical action	classical Hamilton-Jacobi equation for $S^{(\mathrm{cl})}$	continuity equation for $\rho \equiv D$ and $\rho\boldsymbol{\nabla}S^{(\mathrm{cl})}/m$	non-linear
$\psi_R^{(\mathrm{cl})} \equiv A_R^{(\mathrm{cl})}\mathrm{e}^{\mathrm{i}S^{(\mathrm{cl})}/\hbar}$ $A_R^{(\mathrm{cl})}$ is real $S^{(\mathrm{cl})} \equiv$ classical action	classical Hamilton-Jacobi equation for $S^{(\mathrm{cl})}$	continuity equation for $\rho \equiv (A_R^{(\mathrm{cl})})^2$ and $\rho\boldsymbol{\nabla}S^{(\mathrm{cl})}/m$	non-linear
$\psi^{(\mathrm{q})} \equiv R\mathrm{e}^{\mathrm{i}S^{(\mathrm{cl})}/\hbar}$ $R \equiv A^{(\mathrm{q})}\mathrm{e}^{\mathrm{i}\beta}$ R is complex $A^{(\mathrm{q})}$ and β are real	classical Hamilton-Jacobi equation for $S^{(\mathrm{cl})}$ and quantum equation for β	continuity equation for $\rho \equiv \lvert R\rvert^2$ and $\rho[\boldsymbol{\nabla}(\hbar\beta) + \boldsymbol{\nabla}S^{(\mathrm{cl})}]/m$	linear Schrödinger equation
$\psi^{(\mathrm{q})} \equiv A^{(\mathrm{q})}\mathrm{e}^{\mathrm{i}S^{(\mathrm{q})}/\hbar}$ $S^{(\mathrm{q})} \equiv \hbar\beta + S^{(\mathrm{cl})}$ $A^{(\mathrm{q})}$ and $S^{(\mathrm{q})}$ are real	quantum Hamilton-Jacobi equation for $S^{(\mathrm{q})}$	continuity equation for $\rho \equiv (A^{(\mathrm{q})})^2$ and $\rho\boldsymbol{\nabla}S^{(\mathrm{q})}/m$	linear Schrödinger equation

The form of $\psi^{(\mathrm{cl})}$ is one part. However, the boundary conditions on $\psi^{(\mathrm{cl})}$ are another. Indeed, standing matter waves are an experimental fact. They can be described in their most elementary form by a *superposition* of a right- and a left-running wave, eq. (48), with identical amplitudes.

Likewise, a scattering situation with a wave approaching from the left requires reflected and transmitted waves. Again these secondary waves are of the form of eq. (48).

However, both boundary conditions, corresponding either to a bound state, or a scattering state are intimately connected to the superposition principle, and thus to the linearity of the Schrödinger equation.

8˙2. *Linearization due to quantum current.* – Neglecting the classicality enforcing potential at least in some domain of space in order to obey the superposition principle points us in the right direction. However, we need a more rigorous argument.

For this purpose we now go a step further and slightly generalize the derivation of the non-linear wave equation, eq. (44), based on the classical Hamilton-Jacobi equation and the continuity equation of Van Vleck by allowing [14,15] a wider class of amplitudes, namely real-valued amplitudes $A_R^{(\mathrm{cl})}$. Unless $A_R^{(\mathrm{cl})} = D^{1/2}$, we cannot *derive* a classical continuity equation, eq. (37), which would ensure the classicality condition, eq. (36).

However, we may *postulate* it, and it is indeed reasonable to assume a conservation law for matter. Unfortunately, for a classical wave $\psi_R^{(\mathrm{cl})}$ defined in this way, we arrive again [14, 15] at the non-linear Schrödinger equation.

The next and final step in the generalization is to allow complex-valued amplitudes R of the wave $\psi^{(\mathrm{q})}$. Again, we cannot verify a quantum continuity equation, but *infer* it from conservation of matter. In this case, we arrive at the linear Schrödinger equation.

The linearization, which was already alluded to by the WKB wave, and is now made rigorous, originates from the introduction of a complex-valued amplitude which manifests itself in a phase and a current in addition to the classical ones. Since the corresponding action depends explicitly on Planck's constant we find a quantum phase and a quantum current in addition to the classical phase and the classical current originating from the classical action.

We emphasize that the linear Schrödinger equation and the non-linear wave equation resulting from the quantum and the classical current lead again to continuity equations. Therefore, the dynamics of the wave is consistent with the assumed continuity equations.

9. – Summary and outlook

We are now in the position to summarize our main results. In order to gain some insight into the theoretical underpinning of the Schrödinger equation, and its connection to classical mechanics, we have started from a mathematical identity which is at the heart of gauge transformations. We have then considered a general wave whose phase is governed in its dynamics by the classical Hamilton-Jacobi equation. The equations of motion of the amplitudes of the waves follow from the mathematical identity when apply either the classicality or the quantum condition. Moreover, the corresponding wave equations give rise to continuity equations with a density and a current. For the quantum condition an additional current appears, and the equation for the total wave is the Schrödinger equation. In contrast, a non-linear wave equation emerges from the classicality condition.

Needless to say, both the non-linear wave equation as well as the linear Schrödinger equation following from the classically and the quantum condition, respectively, imply continuity equations identical to the ones following from the equations of motion of the amplitudes of the corresponding waves. Therefore, the classicality and the quantum conditions and the resulting wave equations are consistent with each other.

Moreover, our approach has brought to light two completely equivalent descriptions of quantum mechanics: i) the Schrödinger equation for ψ which is free of any remnants of classical mechanics such as $\boldsymbol{\nabla} S^{(\mathrm{cl})}$, and ii) the Schrödinger equation for R which does remember its roots in the classical world, and contains $\boldsymbol{\nabla} S^{(\mathrm{cl})}$ and $\partial S^{(\mathrm{cl})}/\partial t$ in a form reminiscent of a vector potential and a scalar potential.

We can trace this absence of classical concepts in the familiar Schrödinger equation back to the quantum condition. However, one might also argue that a theory describing the microscopic world should be furthest away from classical mechanics. For this reason, it should be free of derivatives of $S^{(\mathrm{cl})}$. Indeed, the quantum condition just achieves this

goal since it eliminates the remnants of classical mechanics, and at the same time gives birth to the linear Schrödinger equation.

Hence, we are tempted to pursue an approach analogous to the one that has led Heisenberg to the matrix mechanics: Eliminate classical concepts that cannot be observed in principle, and derive in this way the Schrödinger equation! Unfortunately, the pursuit of this intriguing idea goes beyond the scope of our lecture notes and has to be postponed to a future publication [23].

* * *

On December 3, 2013 our friend Donald H. Kobe passed away after a long life dedicated to science and religion. It is with great pleasure that we acknowledge his decisive influence on us. Indeed, it was Don who had triggered our interest in the deeper origin of the Schrödinger equation which initiated our series of articles on this topic. In this adventure we have immensely enjoyed the heated exchange of arguments supporting different points of view which ultimately led to many insights. Unfortunately, we were not able to publish all of our results before Don's untimely death. The present lecture notes fall into this category and emerged from a manuscript prepared originally in close collaboration with Don. For this reason, we feel it is appropriate to still include him in the list of coauthors. We have also profited from numerous fruitful discussions with L. Cohen, J. P. Dahl, J. Dalibard, W. D. Deering, M. Fleischhauer, R. F. O'Connell, H. Paul, E. Sadurni, A. A. Svidzinsky and W. H. Zurek. D. M. G. is grateful to the Alexander von Humboldt-Stiftung for a Wiedereinladung which made this work possible, and to Ulm University and, in particular to IQ^{ST} for the gracious hospitality. M. O. S. acknowledges the support of the Office of Naval Research (Award No. N00014-16-1-3054), the Air Force Office of Scientific Research (Award No. FA9550-18-1-0141) and the Robert A. Welch Foundation (Grant No. A-1261). Moreover, W. P. S. thanks the Hagler Institute of Advanced Study at Texas A&M University for a Faculty Fellowship and Texas A& M AgriLife for the support of this work. The research of the IQ^{ST} is financially supported by the Ministry of Science, Research and Arts Baden-Württemberg.

Appendix A.

Van Vleck continuity equation

In order to keep our lecture notes self-contained we rederive in this appendix the Van Vleck continuity equation for two cases: The first one is for a *single* degree of freedom, and brings out the essential ingredients of the calculation. The second one addresses an *arbitrary* number of degrees. Here, the derivation is slightly more involved and hides the inner workings of the continuity equation behind an opaque curtain of mathematics.

Although these calculations can be found in the literature, for example in the lecture notes of Wolfgang Pauli on field quantization [57], or in the first article [13] by Ralph Schiller we have decided to include them here for the sake of completeness. In particular, we rederive the Jacobi formula [58] for the differentiation of a determinant which is at the heart of the Van Vleck continuity equation.

Moreover, these derivations contain important ingredients of the WKB approximation, and thereby provide a foundation for appendix B where we show that the wave equation of the primitive WKB wave function [18] is identical to the non-linear wave equation of classical mechanics.

A˙1. *One-dimensional case.* – In this section we first rederive the equation of motion for the one-component version of the Van Vleck determinant in the presence of a position- and time-dependent potential. We then illustrate the consequences of the so-obtained Van Vleck continuity equation for a time-independent potential, and focus on the interplay between density and current.

A˙1.1. Derivation of continuity equation. We start from the definition of the quantity

$$D \equiv \frac{\partial^2 S}{\partial x\, \partial \varepsilon} \tag{A.1}$$

as the second derivative of the classical action $S \equiv S(x,t;\varepsilon)$ with respect to the coordinate x and the constant ε of integration. For the sake of simplicity in notation we suppress throughout this appendix the superscript (cl) indicating that we are dealing with the *classical* action.

Indeed, S is the solution of the classical Hamilton-Jacobi equation

$$-\frac{\partial S}{\partial t} = \frac{1}{2m}\left(\frac{\partial S}{\partial x}\right)^2 + V \tag{A.2}$$

for a particle of mass m moving in a position- and time-dependent potential $V = V(x,t)$.

From the definition, eq. (A.1), of D we find interchanging the order of the derivatives the identity

$$\frac{\partial}{\partial t} D = \frac{\partial}{\partial t}\left(\frac{\partial^2 S}{\partial x\, \partial \varepsilon}\right) = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial \varepsilon}\left(\frac{\partial S}{\partial t}\right)\right), \tag{A.3}$$

which with the help of the Hamilton-Jacobi equation, eq. (A.2), reduces to

$$\frac{\partial}{\partial t} D = -\frac{\partial}{\partial x}\left(\frac{\partial}{\partial \varepsilon}\left[\frac{1}{2m}\left(\frac{\partial S}{\partial x}\right)^2 + V\right]\right). \tag{A.4}$$

Since the potential V does not depend on ε, the corresponding derivative with respect to ε vanishes, and we arrive at the relation

$$\frac{\partial}{\partial t} D = -\frac{\partial}{\partial x}\left[\frac{\partial}{\partial \varepsilon}\left(\frac{1}{2m}\left(\frac{\partial S}{\partial x}\right)^2\right)\right] = -\frac{\partial}{\partial x}\left(\frac{1}{m}\frac{\partial S}{\partial x}\frac{\partial}{\partial \varepsilon}\left(\frac{\partial S}{\partial x}\right)\right), \tag{A.5}$$

or

$$\frac{\partial}{\partial t} D = -\frac{\partial}{\partial x}\left[D\frac{1}{m}\left(\frac{\partial S}{\partial x}\right)\right]. \tag{A.6}$$

In the last step we have again interchanged the order of the differentiations with respect to x and ε, and have recalled the definition, eq. (A.1), of D.

Hence, we have verified that the Hamilton-Jacobi equation, eq. (A.2), implies for the quantity D defined by eq. (A.1) the continuity equation

$$\frac{\partial}{\partial t}D + \frac{\partial}{\partial x}\left[D\frac{1}{m}\left(\frac{\partial S}{\partial x}\right)\right] = 0, \tag{A.7}$$

where D represents a density ρ and

$$j_x \equiv D\frac{1}{m}\frac{\partial S}{\partial x} \tag{A.8}$$

is a current.

A˙1.2. Explicit expressions for density and current from action. We now use the solution

$$S(x,t;\varepsilon) \equiv \int_{x_0}^{x} \mathrm{d}\tilde{x}\, p(\tilde{x};\varepsilon) - \varepsilon t \tag{A.9}$$

of the Hamilton-Jacobi equation, eq. (A.2), for a time-independent potential $V = V(x)$ to derive an explicit expression for D, and find the current.

For this purpose we substitute the ansatz, eq. (A.9), into the Hamilton-Jacobi equation, eq. (A.2), which yields with

$$\frac{\partial S}{\partial x} = p \quad \text{and} \quad \frac{\partial S}{\partial t} = -\varepsilon \tag{A.10}$$

the identity

$$\varepsilon = \frac{p^2}{2m} + V. \tag{A.11}$$

Hence, we can identify for the case of a time-independent potential the constant ε of integration with the energy E of the particle, and find the explicit expression

$$p^2(x;E) = 2m[E - V(x)] \tag{A.12}$$

for the square of the momentum p.

Next we obtain D from its definition, eq. (A.1), together with the explicit form, eq. (A.9), of S. Indeed, interchanging the order of differentiations yields with eq. (A.10) the identity

$$D = \frac{\partial^2 S}{\partial x\, \partial E} = \frac{\partial}{\partial E}\left(\frac{\partial S}{\partial x}\right) = \frac{\partial p}{\partial E}\,. \tag{A.13}$$

From the conservation of energy expressed by eq. (A.12) we arrive by differentiation at the relation

$$\frac{\partial}{\partial E}p^2 = 2p\frac{\partial p}{\partial E} = 2m, \tag{A.14}$$

or

$$\frac{\partial p}{\partial E} = \frac{m}{p}\,, \tag{A.15}$$

and eq. (A.13) reduces to

$$D \equiv \rho = \frac{m}{p}\,. \tag{A.16}$$

Hence, the one-dimensional version of the Van Vleck determinant, that is the one-dimensional density is inversely proportional to the momentum. This result is in complete accordance with the corresponding expression obtained from statistical mechanics.

We conclude by substituting eq. (A.10) into eq. (A.8) to obtain the explicit expression

$$j_x = \rho \frac{p}{m} \equiv \rho \cdot v, \tag{A.17}$$

for the current familiar from statistical mechanics.

A˙1.3. Density and current from continuity equation. Needless to say, the continuity equation (A.7) provides us with the same expressions. Indeed, when we recall that the present discussion deals with a stationary system, and thus

$$\frac{\partial}{\partial t} D = 0, \tag{A.18}$$

eq. (A.7) immediately yields

$$0 = \frac{\partial}{\partial x}\left(D \frac{1}{m} \frac{\partial S}{\partial x}\right) = \frac{\partial}{\partial x}\left(D \frac{p}{m}\right), \tag{A.19}$$

where in the last step we have used eq. (A.10).

Obviously, the relation

$$D \frac{p}{m} = \text{constant} \tag{A.20}$$

implies, up to the constant, the expressions eqs. (A.16) and (A.17) for D and j_x, respectively.

A˙2. *Multi-dimensional case.* – Next we turn to the multi-dimensional case where we have N degrees of freedom, that is N Cartesian coordinates x_k which in turn imply N constants ε_l of integration. In this case the classical action $S \equiv S(x_1, \ldots, x_N, t; \varepsilon_1, \ldots, \varepsilon_N)$ depends on time, but most importantly on N coordinates x_k, as well as on N constants ε_l.

Thus we consider the Van Vleck determinant

$$D \equiv \left| \frac{\partial^2 S}{\partial x_k \, \partial \varepsilon_l} \right| \tag{A.21}$$

formed by all second derivatives of the action S with respect to x_k and ε_l.

The derivation of the corresponding continuity equation relies on two ingredients: (i) the Jacobi formula [58]

$$\mathrm{d}\mathcal{M} = \mathcal{M}\,\mathrm{Tr}(\mathbf{M}^{-1}\mathrm{d}\mathbf{M}) \equiv \mathcal{M}\,(\mathbf{M}^{-1})_{lk}(\mathrm{d}\mathbf{M})_{kl} \tag{A.22}$$

for the differential $\mathrm{d}\mathcal{M}$ of a determinant $\mathcal{M}$ of a matrix $\mathbf{M}$ revisited in the next section, and (ii) the Hamilton-Jacobi equation

$$-\frac{\partial S}{\partial t} = \frac{1}{2m}\left(\frac{\partial S}{\partial x_j}\right)\left(\frac{\partial S}{\partial x_j}\right) + V. \tag{A.23}$$

Throughout this part of the appendix we adhere to the Einstein summation convention, and sum over indices which appear twice. Moreover, we allow the potential V to display a position as well as a time dependence.

We first apply the Jacobi formula, eq. (A.22), to the matrix $\mathbf{M}$ defined by the matrix elements

$$M_{kl} \equiv \frac{\partial^2 S}{\partial x_k\,\partial \varepsilon_l} \tag{A.24}$$

and find the relation

$$\frac{\partial}{\partial t}D = D(\mathbf{M}^{-1})_{lk}\frac{\partial}{\partial t}M_{kl}, \tag{A.25}$$

or

$$\frac{\partial}{\partial t}D = D(\mathbf{M}^{-1})_{lk}\frac{\partial}{\partial x_k}\left[\frac{\partial}{\partial \varepsilon_l}\left(\frac{\partial S}{\partial t}\right)\right], \tag{A.26}$$

where we have interchanged the order of differentiations.

Moreover, we obtain from the Hamilton-Jacobi equation, eq. (A.23), the expression

$$\frac{\partial}{\partial \varepsilon_l}\left(\frac{\partial S}{\partial t}\right) = -\frac{1}{2m}\frac{\partial}{\partial \varepsilon_l}\left[\left(\frac{\partial S}{\partial x_j}\right)\left(\frac{\partial S}{\partial x_j}\right)\right] = -\frac{1}{m}\frac{\partial S}{\partial x_j}\frac{\partial^2 S}{\partial x_j\,\partial \varepsilon_l} = -\frac{1}{m}\frac{\partial S}{\partial x_j}M_{jl}, \tag{A.27}$$

where in the first step we have made use of the fact that the potential V does not depend on the constants ε_l of integration, and have recalled in the last step the definition, eq. (A.24) of the matrix elements M_{jl}.

When we substitute this result into eq. (A.26), and employ the product rule we arrive at the formula

$$\frac{\partial}{\partial t}D = -D(\mathbf{M}^{-1})_{lk}\frac{1}{m}\frac{\partial}{\partial x_k}\left(\frac{\partial S}{\partial x_j}\right)M_{jl} - D(\mathbf{M}^{-1})_{lk}\frac{1}{m}\frac{\partial S}{\partial x_j}\frac{\partial M_{jl}}{\partial x_k}\,. \tag{A.28}$$

Next we note the symmetry relation

$$\frac{\partial M_{jl}}{\partial x_k} = \frac{\partial}{\partial x_k}\left(\frac{\partial^2 S}{\partial x_j\,\partial \varepsilon_l}\right) = \frac{\partial}{\partial x_j}\left(\frac{\partial^2 S}{\partial x_k\,\partial \varepsilon_l}\right) = \frac{\partial M_{kl}}{\partial x_j}\,, \tag{A.29}$$

which follows from the definition eq. (A.24) of the matrix elements M_{jl} by interchanging the order of the two differentiations with respect to the coordinates x_k and x_j, and slightly rearrange the order of terms in eq. (A.28) which leads us to

$$\frac{\partial}{\partial t}D = -\frac{D}{m}\frac{\partial}{\partial x_k}\left(\frac{\partial S}{\partial x_j}\right)M_{jl}\left(\mathbf{M}^{-1}\right)_{lk} - \frac{1}{m}\frac{\partial S}{\partial x_j}D\left(\mathbf{M}^{-1}\right)_{lk}\frac{\partial M_{kl}}{\partial x_j}\,. \tag{A.30}$$

At this point we recall the identity

$$M_{jl}\left(\mathbf{M}^{-1}\right)_{lk} = \delta_{jk} \tag{A.31}$$

corresponding to $\mathbf{M}\mathbf{M}^{-1} = \mathbb{1}$, and use the Jacobi formula, eq. (A.22), to establish the relation

$$\frac{\partial}{\partial x_j}D = D\left(\mathbf{M}^{-1}\right)_{lk}\frac{\partial M_{kl}}{\partial x_j} \tag{A.32}$$

for the differentiation of D with respect to x_j.

With the help of these formulae eq. (A.30) reduces to

$$\frac{\partial}{\partial t}D = -\frac{D}{m}\frac{\partial}{\partial x_j}\left(\frac{\partial S}{\partial x_j}\right) - \frac{1}{m}\frac{\partial S}{\partial x_j}\frac{\partial D}{\partial x_j}\,, \tag{A.33}$$

or

$$\frac{\partial}{\partial t}D + \frac{\partial}{\partial x_j}\left[D\frac{1}{m}\left(\frac{\partial S}{\partial x_j}\right)\right] = 0, \tag{A.34}$$

which constitutes the Van Vleck continuity equation.

A˙3. *Differential of a determinant.* – For the sake of completeness we provide in this section the essential ideas needed in the derivation of the Jacobi formula [58]

$$\mathrm{d}\mathcal{M} = \mathcal{M}\,\mathrm{Tr}(\mathbf{M}^{-1}\mathrm{d}\mathbf{M}) \equiv \mathcal{M}(\mathbf{M}^{-1})_{lk}(\mathrm{d}\mathbf{M})_{kl} \tag{A.35}$$

for the differential $\mathrm{d}\mathcal{M}$ of the determinant $\mathcal{M}$ corresponding to the matrix $\mathbf{M}$ with the elements $(\mathbf{M})_{ij} \equiv M_{ij}$. Here $\mathbf{M}^{-1}$ is the inverse of $\mathbf{M}$ and Tr denotes the trace. Throughout this section we assume that we deal with a $N \times N$ matrix where N is an integer.

Since $\mathcal{M} = \mathcal{M}(M_{11}, M_{12}, \ldots, M_{NN})$ is a function of all N^2 elements M_{ij} we find for the differential the expression

$$\mathrm{d}\mathcal{M} = \frac{\partial \mathcal{M}}{\partial M_{ij}}\,\mathrm{d}M_{ij}, \tag{A.36}$$

where we follow again the Einstein summation convention and sum over i as well as j.

Next we recall the Laplace expansion

$$\mathcal{M} = \sum_{r=1}^{N} M_{ir}\, m_{ir} \tag{A.37}$$

of the determinant $\mathcal{M}$ by elements of the i-th row where the co-factor $m_{ir} \equiv (-1)^{i+r}\mathcal{M}_{ir}$ contains the determinant $\mathcal{M}_{ir}$ of the matrix obtained from $\mathbf{M}$ by deleting the i-th row and the r-th column. Since here we only sum over r we have denoted this summation explicitly.

We emphasize that the result for $\mathcal{M}$ is independent of the row we have used for the expansion. Moreover, an expansion by columns is possible as well.

From the Laplace expansion, eq. (A.37), we find immediately the relation

$$\frac{\partial \mathcal{M}}{\partial M_{ij}} = \sum_{r=1}^{N} \frac{\partial M_{ir}}{\partial M_{ij}} m_{ir}, \tag{A.38}$$

where we have made use of the fact that the co-factors m_{ir} are independent of M_{ij} since we have deleted the i-th row.

With the help of the identity

$$\frac{\partial M_{ir}}{\partial M_{ij}} = \delta_{rj}, \tag{A.39}$$

eq. (A.38) reduces to

$$\frac{\partial \mathcal{M}}{\partial M_{ij}} = m_{ij}, \tag{A.40}$$

and eq. (A.36) reads

$$\mathrm{d}\mathcal{M} = m_{ij}\mathrm{d}M_{ij}, \tag{A.41}$$

where now we sum again over both indices.

Provided the determinant $\mathcal{M}$ is non-vanishing the inverse $\mathbf{M}^{-1}$ of $\mathbf{M}$ is given by

$$\left(\mathbf{M}^{-1}\right)_{ij} = \mathcal{M}^{-1} m_{ji}, \tag{A.42}$$

or

$$m_{ij} = \mathcal{M}\left(\mathbf{M}^{-1}\right)_{ji}, \tag{A.43}$$

and thus eq. (A.41) turns into

$$\mathrm{d}\mathcal{M} = \mathcal{M}\left(\mathbf{M}^{-1}\right)_{ji} \mathrm{d}M_{ij}, \tag{A.44}$$

that is the Jacobi formula, eq. (A.35).

Appendix B.

Non-linear wave equation for WKB wave

In this appendix we rederive [18] the wave equation for the primitive WKB wave

$$\phi \equiv \frac{\mathcal{N}}{\sqrt{p}} \mathrm{e}^{\mathrm{i}S/\hbar}, \tag{B.1}$$

where $\mathcal{N}$ is a normalization constant, and

$$p(x;E) \equiv \sqrt{2m[E-V(x)]} \tag{B.2}$$

denotes the classical momentum of the particle of mass m, and energy E in the time-independent potential $V = V(x)$.

In complete analogy to appendix A we suppress again for the sake of simplicity in notation the superscript (cl) in the classical action

$$S(x,t;E) \equiv \int_{x_0}^{x} \mathrm{d}\tilde{x}\; p(\tilde{x};E) - Et, \tag{B.3}$$

where x_0 is an arbitrary coordinate.

Our calculation shows that ϕ defined by eq. (B.1) can be interpreted as a solution of the non-linear wave equation discussed in sect. **7**. Indeed, the non-linearity is the origin of the deviation of the WKB wave from the exact solution of the Schrödinger equation.

When we differentiate ϕ with respect to time and coordinate we find the relations

$$\mathrm{i}\hbar \frac{\partial}{\partial t}\phi = E\phi \tag{B.4}$$

and

$$\frac{\partial}{\partial x}\phi \equiv \phi' = \mathcal{N}\left[\left(\frac{1}{\sqrt{p}}\right)' \mathrm{e}^{\mathrm{i}S/\hbar} + \frac{1}{\sqrt{p}}\mathrm{e}^{\mathrm{i}S/\hbar}\left(\frac{\mathrm{i}}{\hbar}\frac{\partial S}{\partial x}\right)\right], \tag{B.5}$$

or

$$\phi' = \mathcal{N}\left[\left(\frac{1}{\sqrt{p}}\right)' \mathrm{e}^{\mathrm{i}S/\hbar} + \frac{1}{\sqrt{p}}\mathrm{e}^{\mathrm{i}S/\hbar}\left(\frac{\mathrm{i}}{\hbar}p\right)\right]. \tag{B.6}$$

In the last step we have used the relation

$$\frac{\partial S}{\partial x} = p \tag{B.7}$$

following from the definition, eq. (B.3), of S. Moreover, throughout the remainder of this appendix prime denotes differentiation with respect to x.

One more differentiation yields

$$\phi'' = \mathcal{N}\left\{\left(\frac{1}{\sqrt{p}}\right)'' \mathrm{e}^{\mathrm{i}S/\hbar} + \frac{\mathrm{i}}{\hbar}\left[2\left(\frac{1}{\sqrt{p}}\right)' p + \frac{1}{\sqrt{p}}p'\right]\mathrm{e}^{\mathrm{i}S/\hbar} + \left(\frac{\mathrm{i}}{\hbar}p\right)^2 \frac{1}{\sqrt{p}}\mathrm{e}^{\mathrm{i}S/\hbar}\right\}, \tag{B.8}$$

where we have used again eq. (B.7).

We can simplify this expression when we recall the identity

$$2\left(\frac{1}{\sqrt{p}}\right)' = -p^{-3/2}p', \tag{B.9}$$

which makes the terms in the square bracket of eq. (B.8) cancel each other.

Next, we multiply eq. (B.8) by $\hbar^2/(2m)$, and arrive with the definition, eq. (B.2), of p at the formula

$$\frac{\hbar^2}{2m}\phi'' = \frac{\hbar^2}{2m}\left[\sqrt{p}\left(\frac{1}{\sqrt{p}}\right)''\right]\phi - (E-V)\phi. \tag{B.10}$$

Finally, we eliminate the term $E\phi$ with the identity eq. (B.4), and obtain the result

$$\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\phi = -\mathrm{i}\hbar\frac{\partial}{\partial t}\phi + V\phi + \frac{\hbar^2}{2m}\left[\sqrt{p}\left(\frac{1}{\sqrt{p}}\right)''\right]\phi, \tag{B.11}$$

that is

$$\mathrm{i}\hbar\frac{\partial}{\partial t}\phi = \hat{H}\phi + \frac{\hbar^2}{2m}\left[\sqrt{p}\left(\frac{1}{\sqrt{p}}\right)''\right]\phi. \tag{B.12}$$

Here we have recalled the familiar definition

$$\hat{H} \equiv -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x) \tag{B.13}$$

of the Hamiltonian.

From the definition eq. (B.1) of ϕ we obtain the relation

$$|\phi| = \frac{|\mathcal{N}|}{\sqrt{p}}\,, \tag{B.14}$$

which allows us to identify the last term in eq. (B.12) as the potential

$$Q[|\phi|] \equiv \frac{\hbar^2}{2m}\frac{|\phi|''}{|\phi|}\,, \tag{B.15}$$

defined by eq. (32).

As a result, we find that the primitive WKB wave ϕ defined by eq. (B.1) satisfies the non-linear wave equation

$$\mathrm{i}\hbar\frac{\partial}{\partial t}\phi = \hat{H}\phi + Q\left[|\phi|\right]\phi \tag{B.16}$$

of classical statistical mechanics.

REFERENCES

[1] Zurek W. H., *Phys. Today*, **44** (1991) 36.
[2] Leibfried D., Knill E., Seidelin S., Britton J., Blakestad R. B., Chiaverini J., Hume D. B., Itano W. M., Jost J. D., Langer C., Ozeri R., Reichle R. and Wineland D. J., *Nature*, **438** (2005) 639.
[3] Deléglise S., Dotsenko I., Sayrin C., Bernu J., Brune M., Raimond J. and Haroche S., *Nature*, **455** (2008) 510.
[4] Einstein A., Podolsky B. and Rosen N., *Phys. Rev.*, **47** (1935) 777.
[5] Scully M. O., Englert B.-G. and Walther H., *Nature*, **351** (1991) 111.
[6] Bohm D., *Quantum Theory* (Dover Publications, New York) 1989.
[7] Bialynicki-Birula I. and Mycielski J., *Ann. Phys. (N.Y.)*, **100** (1976) 62.
[8] Shimony A., *Phys. Rev. A*, **20** (1979) 394.
[9] Weinberg S., *Ann. Phys. (N.Y.)*, **194** (1989) 336.
[10] Shull C. G., Atwood D. K., Arthur J. and Horne M. A., *Phys. Rev. Lett.*, **44** (1980) 765.
[11] Gähler R., Klein A. G. and Zeilinger A., *Phys. Rev. A*, **23** (1981) 1611.
[12] Bollinger J. J., Heinzen D. J., Itano W. M., Gilbert S. L. and Wineland D. J., *Phys. Rev. Lett.*, **63** (1989) 1031.
[13] Schiller R., *Phys. Rev.*, **125** (1962) 1100; *ibid*, **125** (1962) 1109; *ibid*, **125** (1962) 1116.
[14] Rosen N., *Am. J. Phys.*, **32** (1964) 597.
[15] Rosen N., *Found. Phys.*, **16** (1986) 687.
[16] Jammer M., *The Philosophy of Quantum Mechanics* (Wiley, New York) 1974.
[17] van Vleck J. H., *Proc. Natl. Acad. Sci. U.S.A.*, **14** (1928) 178.
[18] Schleich W. P., *Quantum Optics in Phase Space* (Wiley-VCH, Berlin) 2001.
[19] Madelung E., *Z. Phys.*, **40** (1927) 322.
[20] Bohm D., *Phys. Rev.*, **85** (1952) 166; *ibid*, **85** (1952) 180.
[21] Schleich W. P., Greenberger D. M., Kobe D. H. and Scully M. O., *Proc. Natl. Acad. Sci. U.S.A.*, **110** (2013) 5374.
[22] Schleich W. P., Greenberger D. M., Kobe D. H. and Scully M. O., *Phys. Scr.*, **90** (2015) 108009.
[23] Schleich W. P., Greenberger D. M., Kobe D. H. and Scully M. O., *History and derivation of the Schrödinger equation* (to be published).
[24] Schrödinger E., *Z. Phys.*, **12** (1922) 13.
[25] Weyl H., *Raum, Zeit, Materie* (Springer, Berlin) 1921.
[26] Hall M. J. W. and Reginatto M., *J. Phys. A*, **35** (2002) 3289.
[27] Born M., *Vorlesung über Atommechanik* (Springer, Berlin) 1925.
[28] Heisenberg W., *Z. Phys.*, **33** (1925) 879.
[29] Born M. and Jordan P., *Z. Phys.*, **34** (1925) 858.
[30] Born M., Heisenberg W. and Jordan P., *Z. Phys.*, **35** (1926) 557.

[31] KLEBER M., this volume, p. 29.
[32] SCHRÖDINGER E., *Ann. Phys. (Leipzig)*, **79** (1926) 361; *ibid*, **79** (1926) 489; *ibid*, **80** (1926) 437; *ibid*, **81** (1926) 109; see also his *Collected Papers on Wave Mechanics* (Chelsea Publishing Company, New York, 1982).
[33] YANG C. N., in *Schrödinger: Centenary Celebration of a Polymath*, edited by KILMISTER C. W. (Cambridge University Press, Cambridge) 1987, p. 53.
[34] HASLETT J. W., *Am. J. Phys.*, **40** (1972) 1315.
[35] BLOCH F., *Phys. Today*, **29** (1976) 23.
[36] RAMAN V. V. and FORMAN P., *Hist. Studies Phys. Sci.*, **1** (1969) 291.
[37] YANG C. N., in *Quantum Coherence and Decoherence*, edited by FUJIKAWA K. and ONO Y. (Elsevier, Amsterdam) 1996, p. 30.
[38] WEYL H., *The Theory of Groups and Quantum Mechanics* (E. P. Dutton, New York) 1931.
[39] WEYL H., *Sitzungsber. Preuss. Akad. Wiss.*, (1918) 465.
[40] EINSTEIN A., *Sitzungsber. Preuss. Akad. Wiss.*, (1918) 478, see also the response by WEYL H., *ibid*, (1918) 478.
[41] LONDON F., *Z. Phys.*, **42** (1927) 375.
[42] EHRENBERG W. and SIDAY R., *Proc. Phys. Soc. London B*, **62** (1949) 8.
[43] AHARANOV Y. and BOHM D., *Phys. Rev.*, **115** (1959) 485.
[44] WEYL H., *Z. Phys.*, **56** (1929) 330.
[45] YOUNG C. N., *Phys. Rev. Lett.*, **33** (1974) 445.
[46] WU T. and YOUNG C., *Phys. Rev. D*, **12** (1975) 3845.
[47] MERZBACHER E., *Quantum Mechanics* (Wiley, New York) 1998.
[48] HOLLAND P. R., *The Quantum Theory of Motion: An Account of the de Broglie-Bohm Causal Interpretation of Quantum Mechanics* (Cambridge University Press, Cambridge) 1993.
[49] LEUCHS G., this volume, p. 81.
[50] FEYNMAN R. P., *Rev. Mod. Phys.*, **20** (1948) 367.
[51] BRIGGS J. S., BOONCHUI S. and KHEMMANI S., *J. Phys. A: Math. Theor.*, **40** (2007) 1289.
[52] LAMB W. E., *Am. J. Phys.*, **62** (1994) 201; *ibid*, **69** (2001) 413.
[53] SCULLY M. O., *J. Phys. Conf. Ser.*, **99** (2008) 012019.
[54] FIELD J., *Eur. J. Phys.*, **32** (2011) 63.
[55] GOLDSTEIN H., *Classical Mechanics* (Addison-Wesley, Reading) 1980.
[56] RUND H., *Hamilton-Jacobi Theory in the Calculus of Variations: Its Role in Mathematics and Physics* (Nostrand, Amsterdam) 1966.
[57] PAULI W., *Ausgewählte Kapitel aus der Feldquantisierung* (Akad. Buchgenossenschaft, Zürich) 1957.
[58] MAGNUS J. R. and NEUDECKER H., *Matrix Differential Calculus with Applications in Statistics and Econometrics* (Wiley, New York) 1999.

This page intentionally left blank

Proceedings of the International School of Physics "Enrico Fermi"
Course 197 "Foundations of Quantum Theory", edited by E. M. Rasel, W. P. Schleich and S. Wölk
(IOS, Amsterdam; SIF, Bologna) 2019
DOI 10.3254/978-1-61499-937-9-81

Wave phenomena and wave equations

G. Leuchs

Department Physik, Universität Erlangen-Nürnberg - Erlangen, Germany

Max-Planck-Institut für die Physik des Lichts - Erlangen, Germany

Physics Department, University of Ottawa - Ottawa, Canada

Summary. — For any kind of wave phenomenon one can find ways to derive the respective dispersion relation from experimental observations and measurements. This dispersion relation determines the structure of the wave equation and thus characterizes the dynamics of the respective wave. Different wave phenomena are thus governed by different differential equations. Here we want to emphasize the experimental approach to matter waves, but before doing so we will discuss and test the procedure for other types of waves, in particular water waves.

1. – Preludium

There are a number of ways how one can motivate or even derive the wave equation for matter waves from theoretical principles. In two recent papers Schleich, Greenberger, Kobe and Scully revisit the Schrödinger equation and point out the importance of the phase dynamics and of the strong coupling between phase and amplitude [1, 2]. In this lecture I want to show to what extent one can obtain the Schrödinger equation starting from specific experimental observations. I will not simply "postulate the classical-to-quantum rules" [1], but show in a consistent way how one can construct a wave equation from experimental observations of a wave phenomenon. In principle this approach works for any wave phenomenon and some examples are given.

Any linear partial differential equation of spatio-temporal variables is solved by a plane wave ansatz

$$f(t,\vec{x}) = e^{-i\omega t + i\sum_{l=1}^{D} k_l x_l}, \tag{1}$$

with ω denoting the temporal frequency and $\vec{k} = \sum_{l=1}^{D} k_l \vec{x_l}^{(0)}$ denoting the wave vector or spatial frequencies. The dimension D can be one, two or three, and the $\vec{x_l}^{(0)}$ represent the D unit vectors in position space. Inserting eq. (1) in any linear partial differential equation yields a relation between ω and k_l, $l = 1$ to D. In the standard cases this results in a relation between ω and $|\vec{k}| = k$ called the dispersion relation for this type of wave phenomenon. In this process each temporal derivative $\frac{\partial}{\partial t}$ yields $-i\omega$ and each spatial derivative $\frac{\partial}{\partial x_l}$ yields ik_l.

Conversely, if you have the characteristic dispersion relation for a particular type of wave phenomenon $g(\omega, k_1, k_2, k_3) = 0$, then you get to the wave equation by making the following substitutions in $g(\omega, k_1, k_2, k_3)\psi(\vec{x}, t) = 0$: substitute $i\frac{\partial}{\partial t}$ for ω and $-i\frac{\partial}{\partial x_l}$ for each k_l. $\psi(\vec{x}, t)$ is the function describing the wave motion provided it is a solution to the equation

$$g\left(i\frac{\partial}{\partial t}, -i\frac{\partial}{\partial x_1}, -i\frac{\partial}{\partial x_2}, -i\frac{\partial}{\partial x_3}\right)\psi(\vec{x}, t) = 0. \tag{2}$$

Note that, while this procedure tells us the mathematical structure of the differential equation determining the wave dynamics, called the wave equation, it tells us nothing about the physical nature of "the wave". To illustrate this point, let us take light in vacuum and apply the procedure. First we need experimental evidence for the dispersion relation. This we get from astronomical observations: objects on the sky, changing with time, do not change their colour, such as a moon reappearing from behind Jupiter, or the varying star Algol in Perseus or a supernova such as SN 1987A [3]. This means that the speed of light is the same for all frequencies or $\omega = c|\vec{k}|$. Strictly speaking it is the group velocity, which is constant. This would imply that the frequency is linearly dependent on the modulus of $\vec{k}$ including a constant offset. Here we assume that this offset is zero. For the experimental evidence one has to measure both the frequency and the wavelength. The former is facilitated by the frequency comb technique developed by Hänsch [4] and Hall [5] and the latter is routinely done with a spectrometer. Without some coherence of the light waves we would not have this information.

Before we can apply the transformation used to obtain eq. (2), we have to get rid of the modulus of $\vec{k}$. There are two possibilities.

1) We restrict the discussion to one spatial dimension and consider only waves propagation in one direction. This reduces the dispersion relation to $\omega = k_x$. This choice, however, does not give us any interesting insight. The more interesting optical phenomenon in vacuum is diffraction requiring at least two spatial dimensions. This brings us to the second possibility.

2) We square $\omega = c|\vec{k}|$ yielding $\omega^2 = c^2(k_x^2 + k_y^2 + k_z^2)$. Applying the above transformation, we obtain the familiar wave equation for light

$$\left(\frac{\partial^2}{\partial t^2} - c^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\right)\Psi(t,x,y,z) = 0. \tag{3}$$

In the case of light we cannot easily "see" the physical quantity, which oscillates. As a result, in the history of optics one had to make assumptions about some fluidum called ether, the substance assumed to oscillate and carrying the energy. Needless to emphasise that this ether had to uniformly fill the whole universe. After James Clark Maxwell's stroke of genius, making the connection between light and electromagnetism, the electric field and the magnetic field were associated with this ether. At the time scientists were thinking in terms of the Galileo transformation when relating mutually moving coordinate systems and expected to be able to see an effect of the lab and the Earth moving through this ether. Using a specially developed interferometer Albert Abraham Michelson showed first alone and later with Edward Williams Morley [6] and with higher precision that there was no evidence for this relative motion. In order to adjust the theory to cope with this null result, George Francis FitzGerald [7] proposed to replace the Galileo transformation with a new transformation which a few years later was also derived by Hendrik Antoon Lorentz [8]. The Lorentz transformation, as it is often called, introduced a Lorentz-invariant ether. Maxwell's equations for empty space survived because they were curiously enough invariant under the Lorentz transformation. It was when Einstein showed that the Lorentz transformation is a direct consequence of assuming that the speed of light in empty space is the same in all inertial co-ordinate systems and independent of their relative velocity, that a paradigm shift happened and the concept of the ether was abandoned. Since then the name ether has this old-fashioned and backward connotation. However, our modern concept of the vacuum is not that of empty space, but rather of a space filled with vacuum fluctuations. The vacuum —acting much like a dielectric would do— screens the bare point charge of *e.g.* the electron [9,10], an effect called vacuum polarization, to give rise to the elementary charge observed in low-energy experiments. At higher energy collisions between point charge particles the screening reduces [11, 12] giving rise to a modification of the effective charge and equivalently to the so-called running of the fine-structure constant (for all this, see also the "Further readings" section, second item). As a result, one might today think again of ether, but now it is a modern Lorentz-invariant quantum ether the properties of which are determined by the different types of charged elementary particles responsible for the vacuum fluctuations. And these properties determine the parameters appearing in Maxwell's equations giving rise to an interesting connection between high-energy physics and low-energy optics (see "Further readings" section). Anyway, the purpose of the present section was to underline and emphasize that dispersion relations tell us little if anything about the physical nature of a wave, they tell us only about the mathematical structure of the wave equation and the dynamics of the solution.

2. – Water waves

Water waves are an interesting example, as we will see, because of their added complexity. The types we will discuss here are the deep water waves, not affected by a finite water depth, and waves of wavelength longer than a few centimeters, for which surface tension can be neglected. The properties of this special type of water wave are governed by gravitation.

We begin by experimentally determining the dispersion relation. A straightforward experiment uses a wave machine. A movable wall in a large enough basin moves forward and backward with a time period of T. The medium responds by developing a wave with a particular wavelength λ. The experimental finding is that [13]

$$\lambda \propto T^2. \tag{4}$$

This result is independent of amplitude as long as it is not too large. The physical explanation is that both the inertial property and the gravitational restoring force are proportional to the same mass in a single hump of the wave.

There are other ways to obtain the same result. You can also stand at a pond and through stones into the water watching the pattern develop [14] and possibly even taking a video of the waves. For a particular size and weight of the stone you will get a specific pair of values for the temporal and the spatial period. Using stones of different size and weight you can get different pairs of values. Looking closely one sees that the circular wave pattern created after a single throw of a stone develops a chirp, further out the periods become longer. Thus the video taken after a single throw with one particular stone will give you different pairs of values. All combined one again retrieves $\lambda \propto T^2$.

The last way I want to mention briefly involves analysing the wave pattern behind any object moving fast enough through water. A duck at typical travelling speed is fast enough. Behind these objects a conical wave pattern develops, the opening angle of which is surprisingly independent of the speed (fig. 1). Already William Thomson, better known as Lord Kelvin, determined this opening angle to be $\theta = 2\arcsin(1/3)$ or approximately 38.94 degrees [15]. The group velocity being half as large as the phase velocity in the case of water determines this mathematical relationship [11]:

$$\frac{\partial \omega}{\partial k} = \omega/2k. \tag{5}$$

Integrating this equation yields again the relation shown in eq. (4). This corresponds to the dispersion relationship $\omega^2 \propto k$. The proportionality constant must have the dimension of acceleration. Considering that these waves are driven by gravitation one might guess that it should be the gravitational acceleration on Earth, which is g. Quantitative measurements in the first two experimental possibilities discussed above to obtain the dispersion relation yield also this proportionality factor. The full dispersion relation is [11]

$$\omega^2 = gk. \tag{6}$$

Fig. 1. – Wave pattern behind a duck traveling on water.

2˙1. *Wave equation for water waves.* – Here again we have the complication that the modulus of the wave vector $\vec{k}$ appears. Unlike in the case of light discussed in the preludium it makes sense to restrict the discussion to one dimension of $\vec{k}$. This already reveals some non-trivial properties of these special types of waves. The corresponding wave equation is [16]

$$\frac{\partial^2}{\partial t^2}\Psi(t,x) = gi\frac{\partial}{\partial x}\Psi(t,x). \tag{7}$$

We note two points: first of all in this particular case we notice that we know from observation the physical nature of the wave. It is the height of the water surface that oscillates. Secondly the differential equation contains the imaginary unit. This would imply that the solution consists of complex function values, but how can this be related to the height of the water surface? Let us first address this latter point. We can get rid of the imaginary unit by writing the wave function as the sum of a real and an imaginary part $\Psi = \Psi_r + i\Psi_i$. Inserting this into eq. (7) the real terms on the left hand side have to match the real term on the right and the same holds for the imaginary terms. This yields two real-valued coupled differential equations

$$\begin{cases} \dfrac{\partial^2}{\partial t^2}\Psi_r(t,x) = -g\dfrac{\partial}{\partial x}\Psi_i(t,x), \\ \dfrac{\partial^2}{\partial t^2}\Psi_i(t,x) = g\dfrac{\partial}{\partial x}\Psi_r(t,x). \end{cases} \tag{8}$$

We see that the imaginary unit in eq. (7) simply means that the differential equation describes the combined dynamics of two real functions. In the two equations in (8) the roles of Ψ_r and Ψ_i are interchanged and there is a difference in sign. Using the dispersion relation $\omega^2 = gk$, the set of eqs. (8) is solved by

$$\begin{cases} \Psi_r = a\sin(\omega t - kx), \\ \Psi_i = a\cos(\omega t - kx). \end{cases} \tag{9}$$

For positive k eq. (9) describes a wave propagating in the positive x direction. We already mentioned that observation reveals that the height of the surface is oscillating.

But the height is only one real variable. What is then the other oscillating variable, *i.e.* the other one of the two functions? Taking a closer look at water waves we notice that the surface is not only going up and down but also to and fro. We notice this whenever something is floating in the water and the motion is up and to and down and fro.

A particular point on the surface is obviously making a circular motion properly described by eqs. (9). So we found that the two functions Ψ_r and Ψ_i describe the vertical and longitudinal periodic motions of the positions of surface molecules. However, our approach does not tell us whether the height of the surface is described by the real or the imaginary part of the wave function. So we have two choices, resulting in two different wave patterns. In fig. 2a the water molecules at the crest move against the propagation direction of the crest and the crest is wider than the trough([1]). This solution is, however, not compatible with the observation in nature. Figure 2b shows the other possibility compatible with eq. (9), the water molecules move along with the crest and the crest is more peaked [17]. Inspection of water waves in nature shows that fig. 2b is the proper description. For a full mathematical treatment see *e.g.* Horace Lamb [11].

Everybody enjoying the waves near a beach has experienced the power of the wave crest when one is hit and pushed to the beach. The trick is to dive under the crest —where the velocity of the water molecules is much smaller than at the surface and one swiftly reappears in the next valley. One can thus exploit this special wave pattern in both directions: diving under the crest when going into the sea and body surfing on the slope preceding the crest when coming back. It is quite remarkable that all this is made possible by the occurrence of the imaginary unit in eq. (7), giving rise to the circular motion.

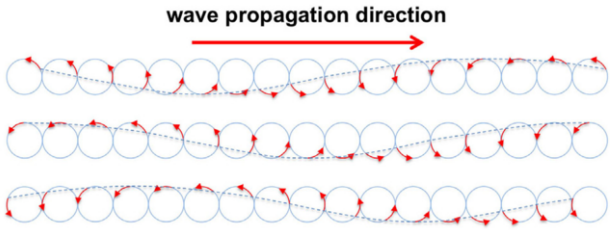
Note that recently, when analysing the wake field of fast boats, Rabaud and Moisy [18] reported on some discrepancies to Lord Kelvin's postulate that the wake field angle is constant. The observed reduction of the wakefield angle is attributed to fast enough small boats not necessarily exciting any water waves with a phase velocity at or close to the boat velocity.

3. – Matter wave

With this digression to water waves we are well prepared to tackle the problem of matter waves. Again we start by looking for some experimental evidence for the dispersion relation, *i.e.* for a wave phenomenon in nature associated with matter waves hoping to extract the dispersion relation for matter waves. The obvious showcases are interference patterns, which we recognize as such, when a matter wave is diffracted off a small structure. It goes without saying that this structure must have a dimension somewhat comparable to the wavelength of the matter waves in order to see a pronounced effect.

([1]) This is a valid mathematical model compatible with eq. (9), `http://demonstrations.wolfram.com/MotionOfParticlesInOceanWaves/`. However, Yuliya Troitskaya of IAP-RAS pointed out to me that for the solution shown in fig. 2a the amplitude of the circular motion grows exponentially with distance from the surface, which is unphysical.

a)



b)

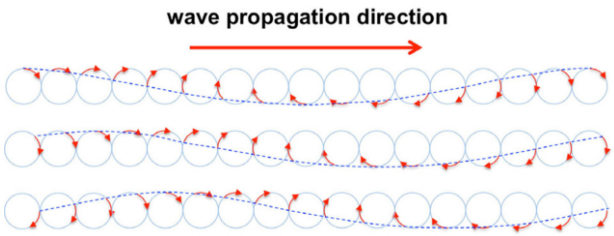


Fig. 2. – (a) Propagation of deep-water surface waves is the result of a circular motion of the surface molecules. Variant 1: the individual molecules in the crest are moving in the opposite direction to the crest. The same is true for the molecules below the surface (details of the collective motion not shown). (b) Propagation of deep-water surface waves is the result of a circular motion of the surface molecules. Variant 2: the individual molecules in the crest are moving along with the crest. The same is true for the molecules below the surface (details of the collective motion not shown).

We choose diffraction of electrons by randomly oriented graphite crystals [19] (fig. 3). In a vacuum chamber electrons emitted by a hot wire at high negative potential $-U_0$ accelerate towards a grounded metal aperture. Electrons passing through the central aperture hole emerge as an expanding electron beam with an average velocity given by

$$u^2 = \frac{2eU_0}{m}\,. \tag{10}$$

The mass m and charge $-e$ of the electron are fixed quantities in the non-relativistic regime. The electron beam hits a thin target made of graphite powder scattering the electrons towards a fluorescent screen. The fluorescence intensity at a particular point on the screen indicates the flux of electrons impinging at this point. What one observes is a pattern of concentric rings reminiscent of the rings around a bright yellow streetlight when looking through the fogged window of a bus at wintertime. The halo occasionally seen around the moon is the same optical phenomenon only that a mono disperse layer of ice crystals in the upper atmosphere takes over the role of the fogged glass. Likewise, the effect observed with fogged glass is also the more pronounced the more mono disperse the size distribution of the fine droplets on the glass is. The angle under which a ring is

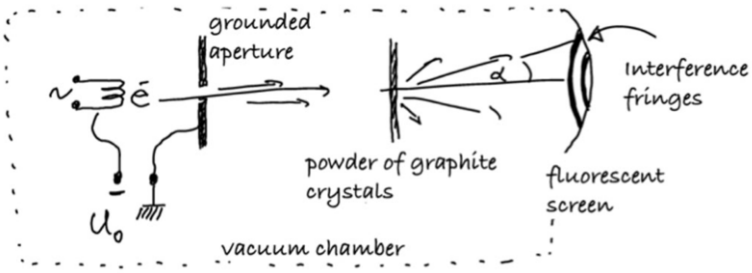


Fig. 3. – Sketch of an apparatus for studying the diffraction of electrons after impinging on graphite. The angle α is proportional to the wavelength, as we know from analogous scenarios in optics, such as a coloured halo of the moon. This angle changes when the electron acceleration voltage U_0 is varied.

observed is proportional to the wavelength —red light appearing further out than blue light. We do not see the matter wave directly, but we see something, which is very close to an interference pattern we know. Hence we proceed with the assumption that electrons behave like waves under this condition. Based on this analogy we conclude that also here the angle α is proportional to the wavelength of the electrons for small angles for which the sine function can be replaced by its argument. Next we vary the acceleration voltage U_0 and find that this changes also the angle under which the rings appear on the screen. A more quantitative measurement yields the following relation:

$$\alpha \propto U_0^{-1/2} = \left(u\sqrt{m/2e}\right)^{-1/2}. \tag{11}$$

When thinking of electrons as waves we cannot help but thinking of more or less localised wave packets. In wave language the group velocity $u = \frac{\partial\omega}{\partial k}$ gives the speed of the wave packet. Using in addition $\alpha \propto \lambda$, we can transform eq. (11) into

$$\lambda \propto \left(\frac{\partial\omega}{\partial k}\right)^{-1}, \qquad \text{or} \qquad \frac{\partial\omega}{\partial k} \propto Ak, \tag{12}$$

where the constant A turns out to be $\hbar/m$, see footnote([2]). This translates to the integral form

$$\int \mathrm{d}\omega = \frac{\hbar}{m}\int k\mathrm{d}k, \qquad \text{and to} \qquad \omega = \frac{\hbar}{2m}k^2 + C. \tag{13}$$

([2]) When repeating the experiment with particles of a different mass, one notices that the wavelength depends not only on the velocity but also on the mass, such that the wavelength is inversely proportional to the linear momentum of the particle. This proportionality constant is a fundamental constant: Planck's constant h. It is in principle possible to determine $h/2\pi$ from diffraction experiments of this kind.

This is the dispersion relation for matter waves we were looking for. C is the integration constant, *i.e.* it is not a function of k or ω but may well be a function of other parameters such as position and time.

3˙1. *Wave equation for matter wave.* – In eq. (13), it is the square of the wave vector, which enters. Therefore we can directly proceed with transforming k and ω into the corresponding differential operators, yielding

$$i\frac{\partial}{\partial t}\Psi(t,\vec{x}) = -\frac{\hbar}{2m}\Delta\Psi(t,\vec{x}) + C(t,\vec{x})\Psi(t,\vec{x}). \tag{14}$$

Multiplying eq. (14) by $\hbar$ yields the Schrödinger equation (see p. 112 in [20]) and C allows for accommodating the potential $V(t,x)$, which may depend on the spatial and also on the temporal co-ordinates,

$$i\hbar\frac{\partial}{\partial t}\Psi(t,\vec{x}) = -\frac{\hbar^2}{2m}\Delta\Psi(t,\vec{x}) + V(t,\vec{x})\Psi(t,\vec{x}). \tag{15}$$

4. – Final remark

We thus achieved our goal to derive the wave equation for matter waves and could stop here. But the journey through the empirical derivation of the various wave equations for the different types of wave phenomena in physics gives also room for speculation. In the case of the water waves the asymmetry between the temporal and spatial differential operators gave rise to the imaginary unit in the wave equation, which in turn meant that this wave equation could be written as two coupled real differential equations. And indeed we found that the water waves are associated with a 2D motion. Now, in the case of matter waves, we have the same situation and could write eq. (15) as

$$\begin{cases} \hbar\dfrac{\partial}{\partial t}\Psi_r(t,\vec{x}) = -\dfrac{\hbar^2}{2m}\Delta\Psi_i(t,\vec{x}) + V(t,\vec{x})\Psi_i(t,\vec{x}), \\ \hbar\dfrac{\partial}{\partial t}\Psi_i(t,\vec{x}) = \dfrac{\hbar^2}{2m}\Delta\Psi_r(t,\vec{x}) - V(t,\vec{x})\Psi_r(t,\vec{x}), \end{cases} \tag{16}$$

with real functions Ψ_r and Ψ_i, with $\Psi = \Psi_r + i\Psi_i$. So far we have a physical interpretation only for the squared modulus of the full wave function $|\Psi|^2 = \Psi_r^2 + \Psi_i^2$ and one might speculate whether new evidence will bring to light more physical insight. Recently, in a most remarkable paper, Schleich, Greenberger, Kobe, and Scully also address this very question. But they followed a different approach. Instead of writing $\Psi = \Psi_r + i\Psi_i$ they used an amplitude and a phase function, $\Psi(\vec{r},t) = A(\vec{r},t)e^{i\theta(\vec{r},t)}$, also resulting in two coupled real valued differential equations. The square of the amplitude represents the spatial density (or rather the position) and the gradient of the phase represents the momentum. This provides novel insight into the classical quantum transition.

5. – Further readings

Regarding the two other lectures I gave, I would like to point attention to manuscripts available on arXiv.org:

– "Time reversal symmetry - a powerful tool in quantum optics"
arXiv:1205.1374 and arXiv:1309.6167

– "Evidence at low energy for high energy particles"
arXiv:1612.03394 and arXiv:1612.06263

* * *

I thank Shadi Sorayya and Ankan Bag for helping preparing the manuscript.

REFERENCES

[1] SCHLEICH W. P., GREENBERGER D. M., KOBE D. H. and SCULLY M. O., *Proc. Natl. Acad. Sci. U.S.A.*, **110** (2013) 5374.
[2] SCHLEICH W. P., GREENBERGER D. M., KOBE D. H. and SCULLY M. O., *Phys. Scr.*, **90** (2015) 108009.
[3] This argument is attributed to Dominique François Jean Arago. see p. 225 in LOMMEL E. *The nature of light with a general account of physical optics* (Henry S. King & Co., London) 1875.
[4] HÄNSCH T. W., see Nobel lecture `http://www.nobelprize.org/nobel_prizes/physics/laureates/2005/hansch-lecture.pdf`.
[5] HALL J. L., see Nobel lecture `http://www.nobelprize.org/nobel_prizes/physics/laureates/2005/hall-lecture.pdf`.
[6] MICHELSON A. A. and MORLEY E. W., *Am. J. Sci.*, **34** (1881) 333.
[7] FITZGERALD G. F., *Science*, **13** (1889) 390.
[8] LORENTZ H. A., *Zittingsverlag Akad. v. Wet.*, **1** (1892) 74.
[9] GOTTFRIED K. and WEISSKOPF V. F., *Concepts of particle physics*, Vol. II (Oxford University Press, New York) 1986.
[10] HAWTON M., SANCHEZ-SOTO L. L. and LEUCHS G., *The linear optical response of the quantum vacuum*, arXiv:1612:06263 (2016).
[11] HOGAN C. J., *Rev. Mod. Phys.*, **72** (2000) 1149.
[12] HAWTON M., SANCHEZ-SOTO L. L. and LEUCHS G., *Quantum field theory and classical optics: determining the fine structure constant*, arXiv:1612.03394 (2016).
[13] LAMB H., *Hydrodynamics* (Dover publications, New York) 1945 - originally published by Cambridge University Press, 6th edition 1932.
[14] See p. 74 in LEUCHS G. and DEISSINGER E. *P.M. Magazin*, **9** (1991) 70.
[15] THOMSON W., *Inst. Mech. Eng. Proc.*, **38** (1887) 409.
[16] WEISMAN D., FU S., GONÇALVES M., SHEMER L., ZHOU J., SCHLEICH W. P. and ARIE A., *Phys. Rev. Lett.*, **118** (2017) 154301.
[17] For a video see `https://youtu.be/NShUBfJQEHk`.
[18] RABAUD M. and MOISY F., *Phys. Rev. Lett.*, **110** (2013) 214503.
[19] For early work on diffraction from polycrystalline substance see THOMSON G. P., *Proc. R. Soc. London, ser. A*, **117** (1928) 600. For first diffraction from single crystal see DAVISSON C. and GERMER L., *Phys. Rev.*, **30** (1927) 705.
[20] SCHRÖDINGER E., *Ann. Phys. (Leipzig)*, **81** (1929) 109.

Proceedings of the International School of Physics "Enrico Fermi"
Course 197 "Foundations of Quantum Theory", edited by E. M. Rasel, W. P. Schleich and S. Wölk
(IOS, Amsterdam; SIF, Bologna) 2019
DOI 10.3254/978-1-61499-937-9-91

History leading to Bell's inequality and experiments

Edward S. Fry

Department of Physics & Astronomy, Texas A&M University
College Station, TX 77843-4242, USA

Summary. — A brief history of the beginnings of quantum mechanics will be presented together with arguments regarding its interpretation; including the very important, but ignored, argument by Grete Hermann. The early discussions were basically philosophical and it was not until Bell produced his inequality that an experimental test became a real possibility. Nevertheless, in the two decades following Bell's development of his inequality, there was a negative attitude by many physicists towards questioning quantum mechanics and a disdain for doing experiments to test a Bell inequality. Nevertheless, four experiments were done in the 1970's by young physicists at the beginning of their careers; those initial experiments and their results will be briefly described. It should be noted that by the 1980's and with the completion of the experiments led by the young Alain Aspect, the culture had begun to change and many experiments have since been done; these culminate in the three recent tests of Bell inequalities that for the first time simultaneously closed both the detection and locality loopholes.

1. – Introduction

Quantum entanglement reaches to the heart of the foundations of quantum mechanics and is the essential feature that makes quantum information processing (including quantum teleportation, quantum computation and quantum cryptography) different from classical information processing. Entanglement produces a special quantum state with a

subtle kind of correlation that has no classical equivalent and suggests a "spooky action at a distance".

Studies of entanglement have focused on tests of Bell inequalities; twenty-seven of them are listed in these references [1-27]; all but one [2] of these experimental tests have satisfied the Bell inequality and violated what were thought to be its quantum-mechanical predictions (it was just the second of all these Bell inequality experiments). The first experimental tests [1-6] of the Bell inequalities were done with photons and are considered to contain loopholes. The two main loopholes were "enforcing locality" and "detection efficiency". In 1982, Aspect, *et al.* [7] made the first major step towards closing the locality loophole. Then, in 1998 the locality condition was fully enforced in an experiment by Zeilinger's group at Innsbruck [14]. Their experiment employed truly random analyzer settings that were outside each other's light cone; that experiment gave a clear violation of a Bell inequality.

Since massive particles present a different approach to entanglement, several proposals for atom-based experiments with high detection efficiency have been advanced [28-31]. The first such experiment by Rowe *et al.* [16] used two massive particles ($^9Be^+$ ions in an ion trap) to provide a conclusive test; their detection efficiency is considered to be essentially 100%. But, there are concerns due to the time required for multiple transitions in the detection process, and due to the close proximity of the $^9Be^+$ ions that leads to the possibility for an "outcome dependence" loophole [32]. While some of the previous experiments have closed either the "locality" or the "detection efficiency" loophole, it is only recently that both loopholes have been closed simultaneously in a single experiment; this was actually achieved in three different experiments by independent groups in the fall of 2015 [25-27]. One of these three was also the first and only test of a Bell inequality to use fermions [25]; all of the other Bell inequality tests have used bosons.

2. – Early history

Near the end of the 19th century, classical mechanics (Newton) together with classical electrodynamics (Maxwell) appeared to provide a firm and final foundation for all of science. There was a sense that only the details were left. As examples, in 1894 Albert A. Michelson (Michelson-Morley experiment) said "*[...] it seems probable that most of the grand underlying principles have been firmly established and that further advances are to be sought chiefly in the rigorous application of these principles [...]*" [33, 34]. And, in 1874 Philipp von Jolly who was Max Planck's physics professor at the University of Munich, advised Planck against going into physics. He told Planck that, "*in this field, almost everything is already discovered, and all that remains is to fill a few holes*" [35]. But, Planck persisted and actually did the only experimental work of his career under Jolly's guidance before turning to theoretical physics.

However, the limits of classical physics began to manifest themselves. Discoveries of fascinating phenomena that did not fit the classical picture inspired radical conjectures and led to the development of Quantum Mechanics. Examples are:

i) *1901 Radiation energy distribution (Planck).* Measurements of the intensity of blackbody radiation as a function of wavelength disagreed dramatically with the classical Rayleigh-Jeans prediction. By analyzing energy and entropy, Planck concluded that only integer multiples of a quantized amount of energy ($E = h\nu$) can be absorbed or emitted, where h is now known as Planck's constant and ν is the frequency of the radiation [36]. With this restriction there was excellent agreement between theory and experiment. Thus we have the first appearance of a quantum and the beginnings of quantum mechanics. And, this fundamental discovery was from someone who was advised against even going into physics [35]!

ii) *1905 Photoelectric effect (Einstein).* When light is incident on a metal surface, electrons can be emitted. Classically, this is thought to be a transfer of energy from the light to the electron. But it was experimentally shown that the kinetic energy of the emitted electrons is independent of the light intensity and that the kinetic energy of emitted electrons depends on the frequency (wavelength) of the light. In fact, if the wavelength of the light is greater than some specific value (depending on the metal), no electrons are emitted. Einstein provided a solution by proposing that, as Planck had argued, light actually consists of discrete packets (photons) of energy $E = h\nu$. If E_0 is the energy required to remove an electron from the surface, then if $h\nu$ is greater than E_0, electrons will be emitted; if not, no electrons will be emitted. And, the amount by which $h\nu$ is greater than E_0 determines the kinetic energy of emitted electrons. Einstein's proposal was confirmed experimentally by Millikan in 1914 [37].

iii) *1913 Bohr model of the H atom (Bohr).* The Rydberg formula for spectral lines of an atom worked well for hydrogen and hydrogen-like atoms, but there was no theoretical justification. Bohr proposed that the electron would orbit the nucleus in circular orbits [38], but only in orbits of a radius r for which the electron's orbital angular momentum L was quantized in integral multiples of Planck's constant h divided by 2π or simply $\hbar$. Specifically, Bohr's model sets $L = mvr = n\hbar$, where m is the mass of the electron, v is the speed of the electron, and n is an integer. Setting the centripetal force on the electron equal to the coulomb attraction, Bohr then showed that the electron in an orbit with this quantized angular momentum would have a total energy (kinetic plus potential) proportional to $1/n^2$. The change in the energy of an electron when it moves from one orbit to another with smaller (larger) radius would be the energy of the photon emitted (absorbed) in the process. This, model successfully reproduced the wavelengths of the spectral lines of hydrogen. Of course, there was still the question of why the circular orbits would be stable because classically an electron undergoing such centripetal acceleration should radiate energy.

iv) *1922 Electron spin (Stern-Gerlach).* Classically, if a particle with a magnet moment passes through an inhomogeneous magnetic field, it will be deflected from its straight path by an amount depending on the angle between the directions of the magnetic moment and the field gradient. Consequently, if particles in a beam have a classical magnetic moment based on their spin angular momentum, they will be deflected into a continuous distribution.

Otto Stern and Walther Gerlach did the experiment with silver atoms using appara-

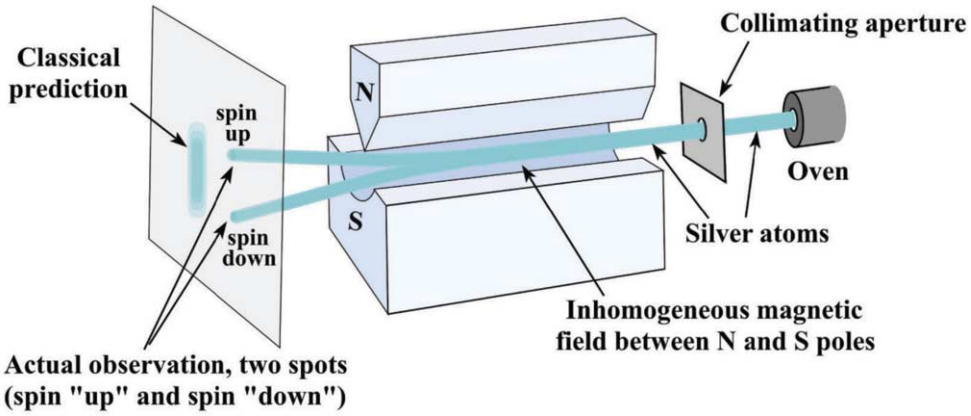


Fig. 1. – Stern-Gerlach apparatus to observe the quantization of angular momentum.

tus illustrated in fig. 1. Rather than the classical prediction of a continuous distribution, the beam of silver atoms were either deflected up or down into two spots [39]. This was a clear demonstration that not only is the spatial orientation of the angular momentum quantized, but that the orientation of the measurement apparatus determines the quantum-mechanical spatial orientation.

v) *1923 Compton Scattering (Compton).* Compton observed that when light is scattered by an electron, the wavelength of the light increases, this increase is called the Compton shift; at low intensities it cannot be explained in the classical theory of electromagnetic waves. The important consequence is that Compton scattering is a demonstration that light can be like a stream of particles (photons) that have energy hv; the directions and energies of the scattered photons and electrons just follow from the conservation of energy and momentum. For the photon momentum, Compton used Einstein's mass energy relation $E = mc^2$ to write $mc^2 = hv$ where m is the "mass" of the photon; and since the photon is traveling with speed c, its momentum is $p = mc$. From these two relations he finds the photon momentum $p = hv/c = h/\lambda$. Compton not only derived his scattering formula, he verified it experimentally [40].

vi) *1924 Wave properties of matter (De Broglie).* De Broglie argued that since light has both wave and particle like properties, electrons (which have particle properties) should also have wave properties. From Einstein's mass energy relation, Compton had argued that the photon momentum is $p = h/\lambda$. De Broglie just reversed that relation to argue that a particle of mass m, speed v and hence momentum $p = mv$ must have a wavelength $\lambda = h/p = h/mv$ [41].

vii) *1927 Electron matter waves (Davisson-Germer).* Clinton J. Davisson and Lester H. Germer observed electron diffraction from a nickel crystal and thereby made the first experimental observation of the wavelength of a moving electron [42]. The lattice spacing in the nickel was $d = 0.215\,\text{nm}$ and a peak in the number of scattered electrons (constructive interference) was observed at $\theta = 50°$. From the grating equation, the wave-

length required to produce constructive interference would be $\lambda = d \sin\theta = 0.165\,\text{nm}$. Since the energy of the incident electrons was 54 eV, their wavelength according to de Broglie would be $\lambda = 0.167\,\text{nm}$, in good agreement with the experimental result.

viii) *1927 Uncertainty Principle (Heisenberg).* If a particle acts like a wave, then its position is described by a wave function that only gives the quantum-mechanical probability of observing the particle at any specific position. To better constrain the position requires a superposition of plane waves; the more plane waves the better localized the quantum-mechanical probability of observing the particle at a specific location. But, since each plane wave represents a different particle momentum, more waves used to provide better position localization means a wider distribution of momentum possibilities. A result of these analyses is the Heisenberg Uncertainty Principle for momentum and position, $\sigma_x \sigma_p \geq \hbar/2$ [43].

3. – The beginnings of quantum mechanics

All these beautiful experimental results led to a broad interest in the development of quantum mechanics and to a euphoria in physics that was widespread by the end of the 1920's. And, in the mid to late 1920's Bohr and Heisenberg were largely responsible for the development of what was eventually called the Copenhagen Interpretation [44] of quantum mechanics; its objective was an attempt to provide a foundation for understanding the quantum world of atoms and waves. Basically, quantum mechanics predicts probabilities for the specific values that measurements can produce. In the Copenhagen Interpretation physical systems typically do not have definite properties prior to being measured; the act of measurement affects the system, causing the set of probabilities to reduce to only one of the specific quantum-mechanical values that are possible (wave function collapse).

But, there were a few physicists who had concerns; perhaps foremost among them was Einstein. Einstein was uneasy about the implications of quantum mechanics even though he had made seminal contributions to its development. In a letter to E. Schrödinger dated May 31 (1928), Einstein wrote: "The Heisenberg-Bohr tranquilizing philosophy — or religion? — is so delicately contrived that, for the time being, it provides a gentle pillow for the true believer from which he cannot very easily be aroused. So let him lie there. But this religion has so damned little effect on me [...]" [45]. In a letter to Max Born dated November 7, 1944, Einstein wrote: "You believe in God playing dice and I in perfect laws in the world of things existing as real objects [...]" [46]. Basically, Einstein was concerned about quantum mechanics because it could only give probabilities (*i.e.* a throw of the dice) for the outcome of a measurement; he felt nature should be deterministic. Consequently, he liked to make comments to the effect that God does not play dice; and at one point Bohr may have responded: "Einstein, stop telling God what to do!" [47].

Clearly, quantum mechanics presents some conceptual and interpretive problems. These are clearly expressed by several other quotes: Niels Bohr wrote: "If quantum

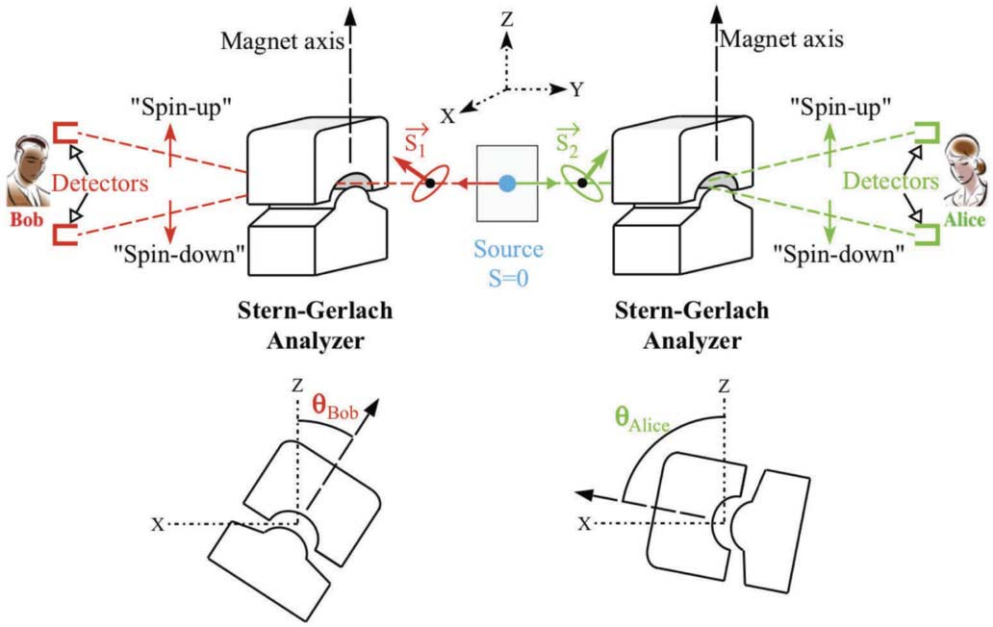


Fig. 2. – Bohm's version of the EPR argument.

mechanics hasn't profoundly shocked you, you haven't understood it yet" [48]. And, Richard Feynman wrote: "No, you're not going to be able to understand it [quantum mechanics] [...]. It is my task to convince you not to turn away because you don't understand it. You see my physics students don't understand it either. That is because I don't understand it. Nobody does" [49]. Although quantum mechanics remains counter-intuitive and unsettling for some, the big thing going for it is that it is wildly successful — its predictions have *never* been found to be incorrect!

By the mid 1930s, Einstein had accepted quantum mechanics as a consistent theory for the statistics of the behavior of atoms. He recognized that quantum mechanics was "the most successful physical theory of our time" [50]. But nevertheless, he felt it was an "incomplete theory". His concerns were most successfully presented in the famous EPR paper "Can quantum-mechanical description of physical reality be considered complete?" [51]. The EPR discussion was based on non-commuting operators such as position and momentum, but Bohm has presented their argument in an especially beautiful form based on noncommuting measurements in different directions of the spin of two spin-1/2 particles that were originally in a spin-zero singlet state [52, 53].

Bohm's approach is depicted in fig. 2; it uses the Stern-Gerlach magnets to observe whether a spin–one-half particle is deflected up or down along the direction of the magnet axis (as in fig. 1). A quantum-mechanical representation for the singlet state of the two

spin-1/2 particles is

$$|\Psi\rangle = \frac{1}{\sqrt{2}}\left\{|\uparrow\rangle_{\text{Bob}}|\downarrow\rangle_{\text{Alice}} - |\downarrow\rangle_{\text{Bob}}|\uparrow\rangle_{\text{Alice}}\right\}. \tag{1}$$

Initially, both Stern-Gerlach magnets are oriented so as to measure the spin in the Z direction: spin up $(+Z)$ or down $(-Z)$. Now, if Bob measures spin up $(+Z)$, then Alice will always observe spin down $(-Z)$, and vice-versa. But, suppose Bob is free to rotate his Stern-Gerlach so that it measures spin along either X or Z, while Alice only measures along Z. If Bob rotates to X, then whenever he measures spin up $(+X)$, Alice will only observe spin down $(-Z)$ half the time, *i.e.* 50% probability. But if Bob rotates to Z, then whenever he measures spin up $(+Z)$, Alice will observe spin down $(-Z)$ every time, 100% probability. How can Bob's decision to measure in the X or Z direction be affecting Alice's measurements?

To summarize, if Bob's measurement of the spin of particle #1 in the Z-direction has outcome "spin-up" $(+Z)$, then one can predict with certainty that Alice's measurement of the spin of particle #2 in the Z-direction will have the outcome "spin-down" $(-Z)$. And, Einstein would therefore argue that if it is certain the measurement will give spin down $(-Z)$, there must be something "real" about the spin of particle #2 in the Z-direction. Specifically, there must be some additional parameters (*i.e.* "hidden variables") that if known would specify the spin in the Z-direction. Similarly, if Bob's measurement of the spin of particle #1 in the X-direction has outcome "spin-up" $(+X)$, then one can predict with certainty that Alice's measurement of the spin of particle #2 in the X-direction would have the outcome "spin-down" $(-X)$. And, Einstein would therefore argue that there is something "real" about the spin of particle #2 in the X-direction. Since quantum mechanics does not simultaneously encompass two components of the spin, Einstein concludes that quantum mechanics is "incomplete".

Perhaps the impact on locality was also a big concern to Einstein. Specifically, if the spin direction of particle #2 in the Z-direction were not a "real" property of particle #2, then the measurement on particle #2 would appear to depend non-locally on the orientation of Stern-Gerlach analyzer #1.

Three years prior to the EPR paper, John von Neumann had already supposedly proved you could not supplement quantum mechanics in the sense that Einstein wanted. This was in his book on the foundations for quantum mechanics [54,55], a pioneering work that had an overpowering influence on the physics community. Of course, Von Neumann was an extraordinary mathematician and computer scientist; he was regarded as one of the foremost mathematicians of the 20th century, and he made many major contributions to a number of scientific fields, *e.g.* mathematics, computer science, statistics, physics, and economics.

Von Neumann's book provided the basis for a broad acceptance and understanding of his approach to the new quantum mechanics. Jammer's historical studies into the response of the scientific community concluded that, "he was hailed by his followers and credited even by his opponents as having succeeded in bringing the foremost

methodological and interpretative problem of quantum mechanics down from the realm of speculation into the reach of mathematical analysis and empirical decision" [56]. In the words of Louis Caruana, "von Neumann provided what seemed to be a bulwark protecting the Copenhagen Interpretation against the claim that determinism could be recovered, and in doing so, he was satisfying the needs of the physics community. He legitimised mathematically what the great majority of physicists had quite peacefully accepted as the "constraint" of the new very powerful theory. It should be remembered that this "impossibility proof" was just one part of a book which puts the whole of quantum mechanics on a 'proper' mathematical and axiomatic base" [57].

For 20 years, scientists believed it was "impossible" to complete quantum mechanics. The 1932 proof by von Neumann was a dominant factor. There were hostile and bitter arguments. But, it was all philosophy; there were no experimental proofs. Then, in 1952 David Bohm did the "impossible"; he developed a model of a deterministic "hidden variable" theory that reproduced the predictions of quantum mechanics [58]. However, Bohm's approach did not get much traction in the mainstream physics community, mainly because it was inherently non-local. In a letter to Max Born dated May 12, 1952, Einstein wrote "Have you noticed that Bohm believes (as deBroglie did, by the way, 25 years ago) that he is able to interpret the quantum theory in deterministic terms? That way seems too cheap to me. But you, of course, can judge this better than I" [59].

John Stewart Bell noted that Bohm's "hidden variable" model was non-local and went on to develop one of the most important theorems of quantum physics, Bell's Theorem [60]. He also showed that von Neumann's "impossibility" proof was circular; von Neumann's basic axioms for quantum mechanics inherently excluded deterministic predictions! David Mermin has even stated that "von Neumann's argument was silly" [61]. For perspective, Jeffrey Bub has argued that Bell misconstrued von Neumann's proof and that the physics community as a whole had misinterpreted his proof [62].

Actually, the flaw in von Neumann's proof had been previously discovered by Grete Hermann in 1935 [63]; but, her work basically went unnoticed until after Bell had rediscovered the flaw almost 30 years later. Von Neumann's proof and the fact that Hermann's critique of it was not on the radar screen of the physics community for decades probably had a strong influence on the development of quantum mechanics. In fact, "Some have posited that had her critique not remained nearly unknown for decades, the historical development of quantum mechanics may have been greatly affected; in particular, it has been speculated that a wider awareness of her work would have put in question the unequivocal acceptance of the Copenhagen interpretation of quantum mechanics, by providing a credible basis for the further development of nonlocal hidden variable theories" [64].

This history is discussed in a superb book, "*The Age of Entanglement*" by Louisa Gilder [65]. On page 161, she points out that at Princeton around 1938 "Einstein proceeded to open von Neumann's tome and point to the assumption that, unbeknownst to him, Grete Hermann had criticized a few years earlier. "Why should we believe in that?" he asked [...]" And on page 195, she emphasizes that "Von Neumann was one of the greatest minds of the twentieth century, and few aside from the distracted Einstein

and the ignored Grete Hermann wondered if it could perhaps be von Neumann who was mistaken". Basically, von Neumann had made such important scientific contributions and was so highly regarded that whatever he said was accepted.

So, what was von Neumann's "impossibility" proof? And, what did Einstein point to in von Neumann's book when he said "Why should we believe in that?". Basically, von Neumann's objective was to provide axioms to create a rigorous axiomatic foundation for quantum mechanics. The problem that Hermann immediately identified and that Einstein, and then eventually Bell saw was that one of von Neumann's axioms basically precluded "hidden variables". Clearly, if the basic assumptions of a theory preclude "hidden variables", then that theory can be used to prove their non-existence. The axiom in question is

$$\langle \boldsymbol{\varphi}|\mathbf{A}|\boldsymbol{\varphi}\rangle + \langle \boldsymbol{\varphi}|\mathbf{B}|\boldsymbol{\varphi}\rangle = \langle \boldsymbol{\varphi}|\mathbf{A}+\mathbf{B}|\boldsymbol{\varphi}\rangle, \tag{2}$$

where A and B are operators and each quantity is an expectation value. This is rigorously true in QM, regardless of whether the operators A and B commute. But, suppose i) there are "hidden variables" that prescribe the result of each measurement and ii) that A and B do not commute. Then, for any single system in state $|\boldsymbol{\varphi}\rangle$, the "hidden variables" tell us what the result for each measurement would be; those results cannot be expected to satisfy this axiom.

As an example: Suppose A and B are operators for spin angular momentum. For simplicity, drop the factor $\hbar/2$ everywhere, then A and B are just Pauli matrices with eigenvalues ± 1. Specifically, take

$$\mathbf{A} = \boldsymbol{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \mathbf{B} = \boldsymbol{\sigma}_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \tag{3}$$

$$\boldsymbol{\sigma}_{45} = \frac{1}{\sqrt{2}}(\boldsymbol{\sigma}_x + \boldsymbol{\sigma}_y) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1-i \\ 1+i & 0 \end{pmatrix},$$

where σ_{45} is the operator for measurement of spin in the x-y plane at an angle of 45° to the x and y axes. All three operators ($\boldsymbol{\sigma}_x$, $\boldsymbol{\sigma}_y$, and $\boldsymbol{\sigma}_{45}$) have eigenvalues ± 1.

Now von Neumann's axiom states

$$\langle \boldsymbol{\sigma}_x \rangle + \langle \boldsymbol{\sigma}_y \rangle = \frac{1}{\sqrt{2}} \langle \boldsymbol{\sigma}_x + \boldsymbol{\sigma}_y \rangle. \tag{4}$$

Again, in QM this is rigorously accurate for the average over an ensemble of systems, but to impose it on every member of the ensemble is "silly" (as Mermin and Bell have both said). In particular, for each member of the ensemble the eigenvalue is ± 1, hence for each member the axiom requires the impossible result,

$$(\pm \mathbf{1}) + (\pm \mathbf{1}) = \frac{\mathbf{1}}{\sqrt{\mathbf{2}}} (\pm \mathbf{1}). \tag{5}$$

The axiom is simply invalid for a hidden variable theory.

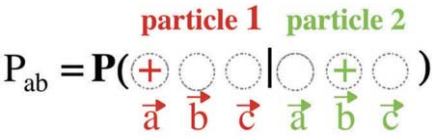


Fig. 3. – P_{ab}, the probability that particle 1 deflects in the $+\vec{a}$ direction and particle 2 deflects in the $+\vec{b}$ direction.

4. – Bell Inequalities

John Stewart Bell considered an EPR-type experiment [60] and assumed i) locality, *i.e.* two spatially separated systems can affect each other only after a time delay greater than the time it takes light to travel from one system to the other; and ii) the "completion" of quantum mechanics, *i.e.* hidden variables. He showed that with these assumptions, the statistical predictions of any local theory that "completes" quantum mechanics in the sense of Einstein must satisfy an *inequality* and that under appropriate experimental conditions the statistical predictions of quantum mechanics will violate that inequality. The incredible significance of this work is that after decades of philosophical arguments, a definitive laboratory experiment was possible. The ultimate goal of any theory in physics or science in general is to find experimental verification; Bell provided a way to do that!

A derivation for a simple example of a Bell inequality was provided by Eugene Wigner [66]. He used Bohm's version of the EPR argument (fig. 2) with two spin-1/2 particles in a total spin zero (singlet) state and chose three different directions $\vec{a}, \vec{b}, \vec{c}$ in the x-z plane (fig. 2) for possible orientations of the two Stern-Gerlach magnets. Now consider, for example, the probability P_{ab} of observing spin up for particle 1 in direction $\vec{a}$ and spin up for particle 2 in direction $\vec{b}$, similarly for P_{bc} and P_{ac}. For the notation to describe this, P_{ab} is written as in fig. 3.

Now if there are hidden variables that specify all components of the spin, then we can add some information to the expression for P_{ab} in fig. 3. Specifically, since particle 1 deflected in the $+\vec{a}$ direction, then the hidden variables tell us that particle 2 would deflect in the $-\vec{a}$ direction; similarly since particle 2 deflected in the $+\vec{b}$ direction, then the hidden variables tell us that particle 1 would deflect in the $-\vec{b}$ direction. So P_{ab} is rewritten as in fig. 4.

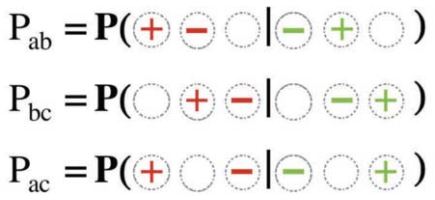


Fig. 4. – Expressions for P_{ab}, P_{bc}, and P_{ac} including some hidden-variable information.

Now, P_{ab} does not involve any measurements in the direction $\vec{c}$; but if there are hidden variables, the two particles will always deflect in opposite directions. Considering all possibilities for results in direction $\vec{c}$, P_{ab} can be written as two terms. One term has plus and minus in direction $\vec{c}$ for particles 1 and 2, respectively; the second term has minus and plus,

$$P_{ab} = \mathbf{P}(+-+|-+-) + \mathbf{P}(+--|-++). \tag{6}$$

In a similar fashion, the expressions for P_{bc} and P_{ac} are

$$P_{bc} = \mathbf{P}(++-|--+) + \mathbf{P}(-+-|+-+), \tag{7}$$

$$P_{ac} = \mathbf{P}(++-|--+) + \mathbf{P}(+--|-++). \tag{8}$$

Now note that the second terms in P_{ab} and P_{ac} are identical, as also are the first term P_{bc} and P_{ac}. It follows immediately that

$$P_{ab} + P_{bc} = P_{ac} + \mathbf{P}(+-+|-+-) + \mathbf{P}(-+-|+-+). \tag{9}$$

Since probabilities are always positive, it follows that

$$P_{ab} + P_{bc} \geq P_{ac}. \tag{10}$$

This is a Bell inequality that will be satisfied if there are hidden variables. What about the quantum-mechanical predictions? In quantum mechanics the probability of two particles having spin "up" in directions θ_a and θ_b is

$$P_{ab} = \frac{1}{4}\{1 - \cos(\theta_a - \theta_b)\}. \tag{11}$$

Take $\theta_a = 0°$, $\theta_b = 45°$, $\theta_c = 90°$, then

$$P_{ab} = P_{bc} = \frac{2-\sqrt{2}}{8}\,; \qquad P_{ac} = \frac{1}{4}. \tag{12}$$

Inserting these values in the Bell inequality, eq. (10), gives

$$\frac{2-\sqrt{2}}{4} \geq \frac{1}{4}\,, \tag{13}$$

which is clearly not valid; *i.e.* the quantum-mechanical predictions do not satisfy the Bell inequality. So now it is critical to do an experiment to determine if the quantum predictions are correct or if the Bell inequality is satisfied.

Physically, Bell's result shows that any hidden-variable theory that also satisfies locality will restrict the strength of the statistical correlations; there is an upper limit on

their magnitude. Quantum mechanics predicts very strong correlations that can violate that restriction.

It was argued that quantum mechanics is an incomplete theory. Specifically, since its predictions for spatially separated systems were incompatible with locality, there must be something more, *e.g.* hidden variables. Basically, the argument is that locality implies quantum mechanics is "incomplete". The irony is that Bell showed just the reverse — if locality is retained, then the quantum mechanical predictions for spatially separated systems cannot be reproduced by any theory that "completes" quantum mechanics.

5. – Initial experiments

A famous paper by Clauser, Horne, Shimony, and Holt laid the ground work for all the initial experiments; they involved polarization correlations between photons in an atomic cascade [67]. But, the environment in the physics community for doing such experiments to test quantum mechanics was caustic at best. In Louisa Gilder's book [65] there are several quotes that provide a good sense of that environment:

1) "Clauser was trying to get a job. "I must have applied to at least a dozen different places, and at all of them I was totally rejected". Universities were uneasy about hiring a professor who would encourage the next generation to question the foundations of quantum theory".

2) "Fry himself had better luck with academia". Realizing that the tenure committee was about to reject the Bell experimenter, one of Fry's friends asked Pipkin (Holt's advisor at Harvard) to come to College Station, TX." ... Pipkin's renown in atomic Physics won over the skeptical committee". Fry got tenure!

3) After Aspect visited with John Bell and completed a presentation on his planned experiment, "[...] Bell asked his first question with a trace of irony: "Have you a permanent position?" Aspect was only a graduate student, but — because of the uniqueness of the French system, and in drastic contrast to his counterparts in America — his position at the École Normale Supérieure was actually permanent. Even with this advantage, it was not easy. There will be serious fights, Bell warned him".

The first five experiments will be briefly reviewed in chronological order. The data from the second experiment satisfied the Bell Inequality and disagreed with the quantum mechanical predictions, all the others violated the Bell inequality and agreed with QM. All these experiments employ a two photon cascade, and the coincidence rates are measured as a function of the angle between the polarizer transmission axes. The Bell inequality for these experiments was

$$\delta = \left| \frac{R(67.5^\circ)}{R_0} - \frac{R(22.5^\circ)}{R_0} \right| \leq \frac{1}{4}\,, \tag{14}$$

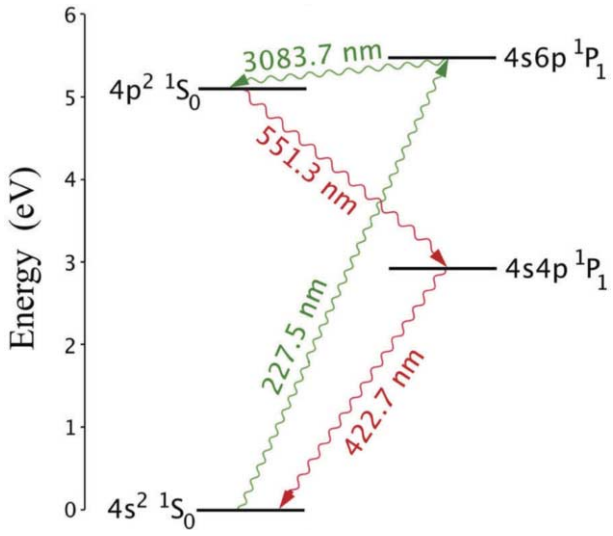


Fig. 5. – The 551.3 nm and 422.7 nm cascade in the Freedman/Clauser experiment.

where $R(67.5°)$ and $R(22.5°)$ are the measured coincidence rates with angles of 67.5° and 22.5° between the polarizer transmission axes and R_0 is the coincident rate with polarizers removed.

The first experiment was completed by Freedman and Clauser at Berkeley in 1972 [1]. They used the two photon cascade in Ca shown in fig. 5. In their experiment a beam of Ca atoms was produced in a vacuum system. At the observation point, a D_2 arc lamp was used to excite the Ca atoms to the $4s6p\,^1P_1$ state. Approximately 7% of the excited atoms decay to the initial state, $4p^2\,^1S_0$, of the cascade. The two cascade photons, 551.3 nm and 422.7 nm, pass through pile-of-plates polarizers and are detected in coincidence as a function of the angle between the polarizer transmission axes. They observed $\delta = 0.300 \pm 0.008$ in violation of the Bell inequality, eq. (14), and agreement with the quantum-mechanical prediction $\delta_{\mathrm{QM}} = 0.301$. They collected data for 200 hours for these results.

The second experiment was completed by Holt and Pipkin at Harvard in 1973 [2]. They used the two photon cascade in Hg shown in fig. 6. They used the zero nuclear spin isotope of ^{198}Hg in a cell and pumped the atoms into the 9^1P_1 state with a 100 eV electron beam. The polarization coincidence rates were measured with calcite polarizers. They observed $\delta = 0.216 \pm 0.013$ in agreement with the Bell inequality and disagreement with the quantum-mechanical prediction $\delta_{\mathrm{QM}} = 0.266$. Nothing appears to have been published as to what might have gone wrong with this experiment

The third experiment was published by Clauser at Berkeley in 1976 [3]. It was a repeat version of the Holt/Pipkin experiment, (fig. 6), but using a cell in which only 91% of the Hg was the zero nuclear spin isotope, ^{202}Hg. The atoms were pumped into the 9^1P_1 state with a 135 eV electron beam. The polarization coincidence rates were

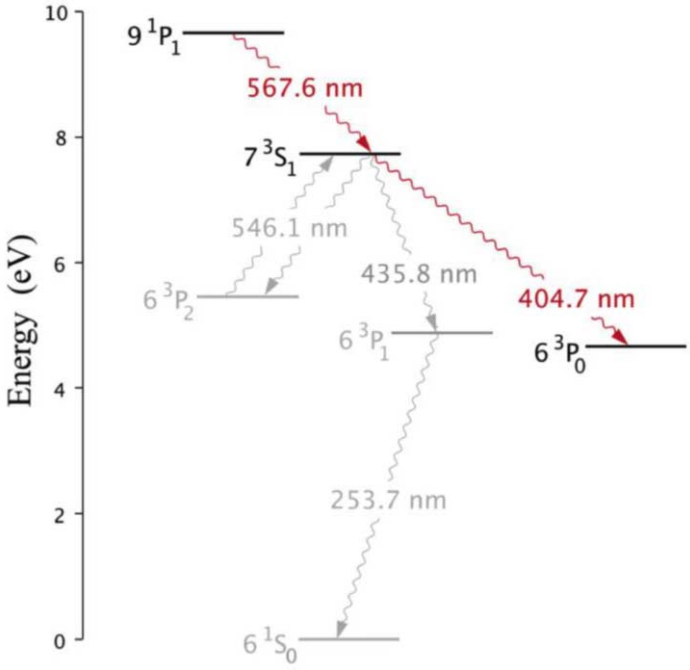


Fig. 6. – The 567.6 nm and 404.7 nm cascade in the Holt/Pipkin experiment.

measured with calcite polarizers. He observed $\delta = 0.2885 \pm 0.0093$ in violation of the Bell inequality and agreement with the quantum-mechanical prediction $\delta_{\mathrm{QM}} = 0.2848$. Data were collected for 412 hours for these results.

The fourth experiment was published a few months later in 1976 by Fry/Thompson at Texas A&M [4]. They used the two photon cascade in ^{202}Hg shown in fig. 7. A beam of natural Hg isotopic abundance was passed through a solenoid electron gun where Hg atoms were excited to the metastable 6^3P_2 state. Further downstream only ^{200}Hg atoms were excited to the 7^3S_1 initial state of the cascade by using a narrow linewidth laser at 546.1 nm; so that only cascade photons from the zero nuclear spin isotope ^{200}Hg were observed. This was the first experiment to use a laser to populate the initial state of the cascade. Polarization correlations between the cascade photons at 435.8 nm and 253.7 nm were observed using pile of plates polarizers. The initial state of the cascade had angular momentum $J = 1$ (previous experiments had $J = 0$ in the initial state). The quantum-mechanical predictions required measurement of the relative populations in the $m_J = 0, \pm 1$ substates. They observed $\delta = 0.296 \pm 0.04$ in violation of the Bell inequality and agreement with the quantum-mechanical prediction $\delta_{\mathrm{QM}} = 0.294 \pm 0.007$. Data were collected for 80 minutes for these results.

The fifth experiment was completed by Aspect/Grangier/Roger in Paris in 1981 [5]. They did exactly the same experiment as the 1972 Freedman/Clauser experiment

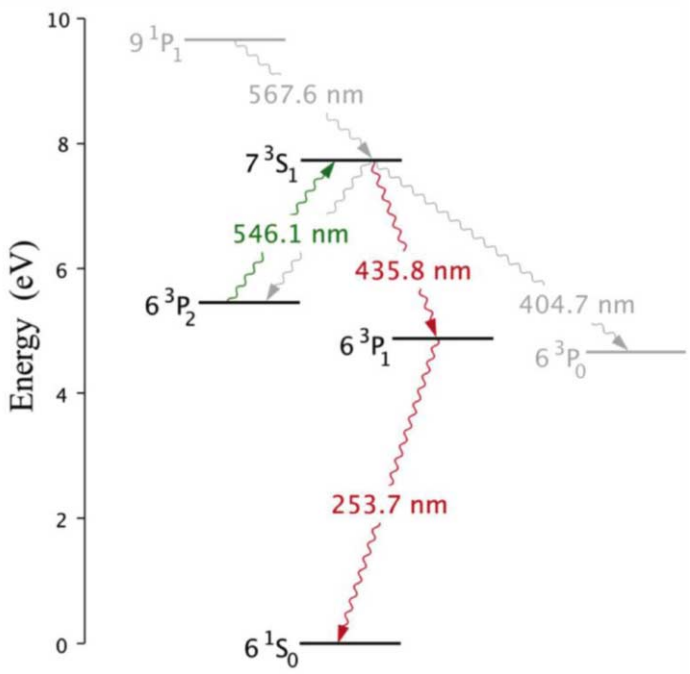


Fig. 7. – The 435.8 nm and 253.7 nm cascade in the Fry/Thompson experiment.

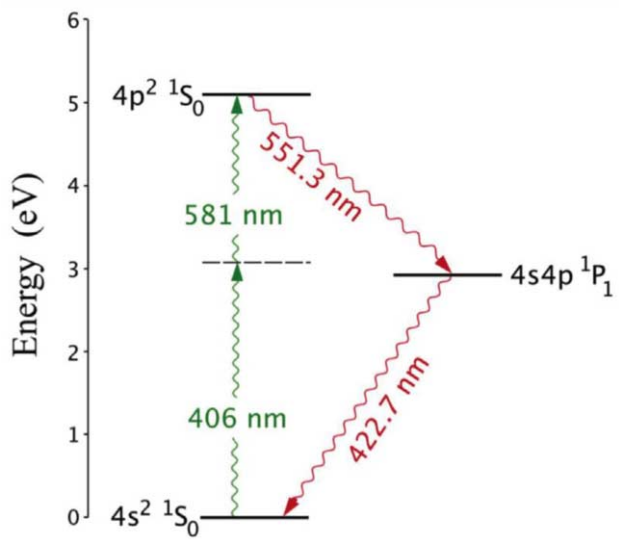


Fig. 8. – The 551.3 nm and 422.7 nm cascade in the Aspect/Grangier/Roger experiment.

(including pile of plates polarizers), except rather than using an arc lamp to excite a state that partially decayed to the initial state of the cascade, they directly excited the initial state of the cascade via a two-photon process. They used 406.7 nm from a single-mode krypton ion laser and 581 nm from a tunable single-mode rhodamine 6G dye laser, see fig. 8. They observed $\delta = 0.3072 \pm 0.0043$ in violation of the Bell inequality and agreement with the quantum-mechanical prediction $\delta_{\mathrm{QM}} = 0.308 \pm 0.002$. This was a great scheme; they observed 150 coincidences per second.

These five experiments all had two important loopholes, detection efficiency and enforcing locality. The detection efficiency concern is that if some particles are not detected, maybe nature will arrange to only detect those particles that give the quantum-mechanical result. The locality loophole concerns the possibility that information will be communicated between the particles if the experimental decisions are not made in a time less than the time it takes light to travel between the particles and the measurement apparatus. Shortly after this fifth experiment, Aspect's group performed an experiment with dual channel polarizers [6] to make the test more robust and another experiment [7] that was the first to address locality concerns. Since then there have been many Bell inequality experiments as discussed in the introduction, including the recent three experiments [25-27] that were the first to simultaneously close both the detection efficiency and locality loopholes and the first and only experiment with fermions [25].

For the future, an especially interesting experiment would be one that is an exact experimental realization of Bohm's version of the EPR experiment with two spin one-half particles. Specifically, the experiment with two ^{199}Hg atom in a nuclear spin singlet state [30]. It would also close both loopholes and be the second test with fermions, but it is especially attractive because of its simplicity and exact analogy to Bohm's approach.

REFERENCES

[1] FREEDMAN S. J. and CLAUSER J. F., *Phys. Rev. Lett.*, **28** (1972) 938.
[2] HOLT R. A., PhD thesis, *Atomic Cascade Experiments* (Harvard University) 1973.
[3] CLAUSER J. F., *Phys. Rev. Lett.*, **36** (1976) 1223.
[4] FRY E. S. and THOMPSON R. C., *Phys. Rev. Lett.*, **37** (1976) 465.
[5] ASPECT A., GRANGIER P. and ROGER G., *Phys. Rev. Lett.*, **47** (1981) 460.
[6] ASPECT A., GRANGIER P. and ROGER G., *Phys. Rev. Lett.*, **49** (1982) 91.
[7] ASPECT A., DALIBARD J. and ROGER G., *Phys. Rev. Lett.*, **49** (1982) 1804.
[8] OU Z. Y. and MANDEL L., *Phys. Rev. Lett.*, **61** (1988) 50.
[9] SHIH Y. H. and ALLEY C. O., *Phys. Rev. Lett.*, **61** (1988) 2921.
[10] RARITY J. G. and TAPSTER P. R., *Phys. Rev. Lett.*, **64** (1990) 2495.
[11] KIESS T. E., SHIH Y. H., SERGIENKO A. V. and ALLEY' C. O., *Phys. Rev. Lett.*, **71** (1993) 3893.
[12] KWIAT P. G., MATTLE K., WEINFURTER H., ZEILINGER A., SERGIENKO A. V. and SHIH Y., *Phys. Rev. Lett.*, **75** (1995) 4337.
[13] TITTEL W., BRENDEL J., ZBINDEN H. and GISIN N., *Phys. Rev. Lett.*, **81** (1998) 3563.
[14] WEIHS G., JENNEWEIN T., SIMON C., WEINFURTER H. and ZEILINGER A., *Phys. Rev. Lett.*, **81** (1998) 5039.

[15] Pan J.-W., Bouwmeester D., Daniell M., Weinfurter H. and Zeilinger A., *Nature*, **403** (2000) 515.
[16] Rowe M. A., Kielpinski D., Meyer V., Sackett C. A., Itano W. M., Monroe C. and Wineland D. J., *Nature*, **409** (2001) 791.
[17] Matsukevich D. N., Chanelière T., Jenkins S. D., Lan S.-Y., Kennedy T. A. B. and Kuzmich A., *Phys. Rev. Lett.*, **96** (2006) 030405.
[18] Ursin R., Tiefenbacher F., Schmitt-Manderbach T., Weier H., Scheidl T., Lindenthal M., Blauensteiner B., Jennewein T., Perdigues J., Trojek P., Ömer B., Fürst M., Meyenburg M., Rarity J., Sodnik Z., Barbieri C., Weinfurter H. and Zeilinger A., *Nat. Phys.*, **3** (2007) 481.
[19] Matsukevich D. N., Maunz P., Moehring D. L., Olmschenk S. and Monroe C., *Phys. Rev. Lett.*, **100** (2008) 150404.
[20] Ansmann M., Wang H., Bialczak R., Hofheinz M., Lucero E., Neeley M., O'Connell A., Sank D., Weides M., Wenner J., Cleland A. and Martinis J., *Nature*, **461** (2009) 504.
[21] Hofmann J., Krug M., Ortegel N., Gérard L., Weber M., Rosenfeld W. and Weinfurter H., *Science*, **337** (2012) 72.
[22] Vermeyden L., Bonsma M., Noel C., Donohue J. M., Wolfe E. and Resch K. J., *Phys. Rev. A*, **87** (2013) 032105.
[23] Giustina M., Mech A., Ramelow S., Wittmann B., Kofler J., Beyer J., Lita A., Calkins B., Gerrits T., Nam S. W., Ursin R. and Zeilinger A., *Nature*, **497** (2013) 227.
[24] Christensen B. G., McCusker K. T., Altepeter J. B., Calkins B., Gerrits T., Lita A. E., Miller A., Shalm L. K., Zhang Y., Nam S. W., Brunner N., Lim C. C. W., Gisin N. and Kwiat P. G., *Phys. Rev. Lett.*, **111** (2013) 130406.
[25] Hensen B., Bernien H., Dréau A. E., Reiserer A., Kalb N., Blok M. S., Ruitenberg J., Vermeulen R. F. L., Schouten R. N., Abellán C., Amaya W., Pruneri V., Mitchell M. W., Markham M., Twitchen D. J., Elkouss D., Wehner S., Taminiau T. H. and Hanson R., *Nature*, **526** (2015) 682.
[26] Giustina M., Versteegh M. A. M., Wengerowsky S., Handsteiner J., Hochrainer A., Phelan K., Steinlechner F., Kofler J., Larsson J.-Å., Abellán C., Amaya W., Valerio Pruneri, Mitchell M. W., Beyer J., Gerrits T., Lita A. E., Shalm L. K., Sae Woo Nam, Scheidl T., Ursin R., Wittmann B. and Zeilinger A., *Phys. Rev. Lett.*, **115** (2015) 250401.
[27] Shalm L. K., Meyer-Scott E., Christensen B. G., Bierhorst P., Wayne M. A., Stevens M. J., Gerrits T., Glancy S., Hamel D. R., Allman M. S., Coakley K. J., Dyer S. D., Hodge C., Lita A. E., Verma V. B., Lambrocco C., Tortorici E., Migdall A. L., Zhang Y., Kumor D. R., Farr W. H., Marsili F., Shaw M. D., Stern J. A., Abellán C., Amaya W., Pruneri V., Thomas Jennewein, Mitchell M. W., Kwiat P. G., Bienfang J. C., Mirin R. P., Knill E. and Nam S. W., *Phys. Rev. Lett.*, **115** (2015) 250402.
[28] Beige A., Munro W. J. and Knight P. L., *Phys. Rev. A*, **62** (2000) 1.
[29] Freyberger M., Aravind P. K., Horne M. A. and Shimony A., *Phys. Rev. A*, **53** (1996) 1232.
[30] Fry E. S., Walther T. and Li S., *Phys. Rev. A*, **52** (1995) 4381.
[31] Oliver B. J. and Stroud J. C. R., *J. Opt. Soc. Am. B*, **4** (1987) 1426.
[32] Vaidman L., *Phys. Lett. A*, **286** (2001) 241.
[33] Michelson A. A., *University of Chicago, Annual Register* 1894-1895 (University of Chicago Press) 1896, pp. 159-162.
[34] Michelson A. A., *Elect. Eng.*, **XXI** (1896) 7.

[35] LIGHTMAN A., *The Discoveries - Great Breakthroughs in 20th-Century Science, Including the Original Papers* (Alfred A. Knopf, Canada) 2005, p. 8.
[36] PLANCK M., *Ann. Phys. (Leipzig)*, **309** (1901) 553.
[37] MILLIKAN R. A., *Phys. Rev.*, **4** (1914) 73.
[38] BOHR N., *Philos. Mag.*, **26** (1913) 1.
[39] GERLACH W. and STERN O., *Z. Phys.*, **9** (1922) 353.
[40] COMPTON A. H., *Phys. Rev.*, **21** (1923) 483.
[41] DE BROGLIE L., thesis, *Recherches sur la théorie des quanta (Research on quantum theory)* (Université de Paris) 1924.
[42] DAVISSON C. and GERMER L. H., *Phys. Rev.*, **30** (1927) 705.
[43] HEISENBERG W., *Z. Phys.*, **43** (1927) 172.
[44] HOWARD D., *Philos. Sci.*, **71** (2004) 669.
[45] EINSTEIN A., SCHROEDINGER E., PLANCK M. and LORENTZ H. A., in *Letters on Wave Mechanics: Schrodinger • Planck • Einstein • Lorentz*, edited by K. PRZIBRAM (Philosophical Library) 1967, p. 75.
[46] STUERMANN W. E., *J. Bible Relig.*, **28** (1960) 399.
[47] RAE A., *Quantum Physics: Illusion or Reality* (Cambridge University Press) 1994 p. 22.
[48] BOHR N., *Atomic Physics and Human Knowledge* (Wiley) 1958.
[49] FEYNMAN R. P., *QED: The Strange Theory of Light and Matter* (Princeton University Press) 1985.
[50] Einstein: Image and Impact. `http://www.aip.org/history/einstein/`; `http://history.aip.org/exhibits/einstein/quantum1.htm`.
[51] EINSTEIN A., PODOLSKY B. and ROSEN N., *Phys. Rev.*, **47** (1935) 777.
[52] BOHM D., *Quantum Theory* (Prentice Hall) 1951.
[53] BOHM D. and AHARONOV Y., *Phys. Rev.*, **108** (1957) 1070.
[54] VON NEUMANN J., *Mathematische Grundlagen der Quanten-mechanik* (Springer) 1932.
[55] VON NEUMANN J., translated by R. T. BEYER, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press) 1955.
[56] JAMMER M., *The Philosophy of Quantum Mechanics* (Wiley) 1974.
[57] CARUANA L., *Physis*, **32** (1995) 109.
[58] BOHM D., *Phys. Rev.*, **85** (1952) 166.
[59] EINSTEIN A., BORN M. and BORN H., *The Born-Einstein letters* (Macmillan) 1971.
[60] BELL J. S., *Physics*, **1** (1964) 195.
[61] MERMIN N. D., *Rev. Mod. Phys.*, **65** (1993) 803.
[62] BUB J., *Found. Phys.*, **40** (2010) 1333.
[63] HERMANN G., *Abhandlung. Fries'schen Schule*, **6** (1935) 75.
[64] Wikipedia, Grete Hermann. `https://en.wikipedia.org/wiki/Grete_Hermann`.
[65] GILDER L., *The Age of Entanglement* (Alfred A. Knopf, Canada) 2008.
[66] WIGNER E. P., *Am. J. Phys.*, **38** (1970) 1005.
[67] CLAUSER J. F., HORNE M. A., SHIMONY A. and HOLT R. A., *Phys. Rev. Lett.*, **23** (1969) 880.

Proceedings of the International School of Physics "Enrico Fermi"
Course 197 "Foundations of Quantum Theory", edited by E. M. Rasel, W. P. Schleich and S. Wölk
(IOS, Amsterdam; SIF, Bologna) 2019
DOI 10.3254/978-1-61499-937-9-109

Sewing Greenberger-Horne-Zeilinger states with a quantum zipper

Da-Wei Wang

Institute of Quantum Science and Engineering, Texas A&M University
College Station, TX 77843, USA

Summary. — A mechanism of a chiral spin wave rotation is introduced to systematically generate mesoscopic Greenberger-Horne-Zeilinger states.

1. – Introduction

One of the astounding principles of quantum mechanics is that a quantum object can exist in a superposition state, *e.g.*, a spin in a superposition of up and down states, $|\psi\rangle = (|\uparrow\rangle + |\downarrow\rangle)/\sqrt{2}$. Applying this principle to two quantum particles, a superposition state such as a singlet state $|\psi\rangle = (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)/\sqrt{2}$ allows the instantaneous determination of the quantum state of the second particle after one measures the quantum state of the first particle, as revealed by Einstein, Podolsky and Rosen in 1935 [1]. This instantaneous determination seems like a violation of a principle of the special relativity, *i.e.*, nothing can travel faster than light, leading to the so-called EPR paradox. The incompleteness of quantum mechanics and the existence of hidden variables were therefore suspected.

In a seminal paper by Bell in 1964 [2], an equality was proposed to show that the statistical results of quantum mechanics is incompatible with any local hidden-variable theories. Subsequent experiments overwhelmingly support the predictions of quantum mechanics. In particular, any local hidden-variable theories are excluded by loophole-free experiments in 2015 [3].

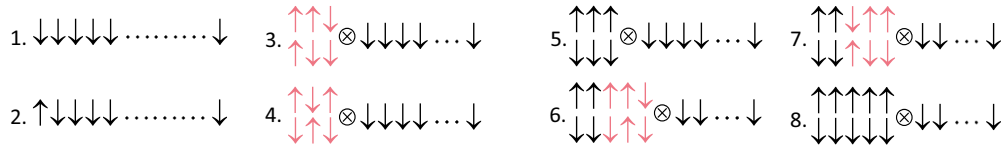


Fig. 1. – Basic steps of a quantum zipper sewing GHZ states. 1, initialize an all-down state. 2, flip the first spin. 3, prepare the second spin in a superposition of the up and down states. 4, introduce an interaction among the red spins, which leads to chiral spin wave rotations as shown in eq. (1). 5, send a π pulse to flip the second spin. 6, send in a π pulse to flip the fourth spin. 7, introduce the same interaction as in step 4 among the red spins until achieving the state shown in this step. 8, send a π pulse to flip the third spin. Repeat the steps 6-8 until zipping all spins in the GHZ state, each time adding two spins to the GHZ chain. For a GHZ state with $2n+1$ spins, we need in total $2n$ π pulses and one $\pi/2$ pulse.

In the exploration of EPR paradox, the Greenberger-Horne-Zeilinger (GHZ) states [4] play an important role. It concerns with a three-particle entangled state, $|\psi\rangle = (|\uparrow\uparrow\uparrow\rangle + |\downarrow\downarrow\downarrow\rangle)/\sqrt{2}$, which allows a direct contradiction between the predictions of quantum mechanics and local hidden-variable theories with a single test, in contrast to the statistical nature of Bell's inequality. GHZ states have also been found useful in the Heisenberg limit metrology [5]. Mesoscopic GHZ states with M particles $|\psi\rangle = (|\uparrow\rangle^{\otimes M} + |\downarrow\rangle^{\otimes M})/\sqrt{2}$ can be generated by using the Mølmer-Sørensen approach [6] in ion traps, with M up to 14 in experiments [7]. To generate GHZ states in a wide range of quantum systems, new mechanisms are highly wanted. Here we introduce a "quantum zipper" to sew GHZ states in a systematic way.

The basic process for generating GHZ states is shown in fig. 1. We start with an all-down state and send a π pulse to the first spin to flip it up. Then we send a $\pi/2$ pulse to prepare the second spin in a superposition of the up and down states. The following step 4 is crucial. We introduce a special interaction between the first three spins, which undergo opposite chiral spin wave rotations for $|\uparrow\downarrow\downarrow\rangle$ and $|\downarrow\uparrow\uparrow\rangle$ states,

$$|\uparrow\downarrow\downarrow\rangle \to |\downarrow\uparrow\downarrow\rangle \to |\downarrow\downarrow\uparrow\rangle \to |\uparrow\downarrow\downarrow\rangle, \tag{1a}$$

$$|\downarrow\uparrow\uparrow\rangle \to |\uparrow\uparrow\downarrow\rangle \to |\uparrow\downarrow\uparrow\rangle \to |\downarrow\uparrow\uparrow\rangle. \tag{1b}$$

The spin states move to the right for the states containing one up spin while they move to the left for the states containing two up spins. The step 4 corresponds to the rotations $|\uparrow\uparrow\downarrow\rangle \to |\uparrow\downarrow\uparrow\rangle$ and $|\uparrow\downarrow\downarrow\rangle \to |\downarrow\uparrow\downarrow\rangle$ in eq. (1). We then send a π pulse to the second spin and prepare the first three spins in a GHZ state, as shown in the step 5. Following and repeating the steps 6-8 (see fig. 1 caption), we can zip the following spins into a GHZ state.

2. – Mechanism

In a previous work, we studied chiral photon rotations among three cavities by generating a synthetic magnetic field [8]. The process in eq. (1) can be realized in a similar

way by the following Hamiltonian:

$$H = i\hbar\kappa \sum_{j=1}^{3} \sigma_{j+1}^{+}\sigma_{j}^{-} + \text{H.c.}, \tag{2}$$

where σ_j^+ and σ_j^- are the raising and lowering operators for the jth spin and the summation over j is cyclic. The Hamiltonian in eq. (2) commutes with $\sum \sigma_j^z$ and the number of up spins is conserved. We first investigate the dynamics in the subspace expanded by $|\uparrow\downarrow\downarrow\rangle$, $|\downarrow\uparrow\downarrow\rangle$ and $|\downarrow\downarrow\uparrow\rangle$, in which the eigenfrequencies are $\lambda_1 = 0$, $\lambda_2 = \sqrt{3}\kappa$ and $\lambda_3 = -\sqrt{3}\kappa$. The corresponding eigenstates are

$$|\psi_1\rangle = \frac{1}{\sqrt{3}}(|\uparrow\downarrow\downarrow\rangle + |\downarrow\uparrow\downarrow\rangle + |\downarrow\downarrow\uparrow\rangle), \tag{3a}$$

$$|\psi_2\rangle = \frac{1}{\sqrt{3}}(|\uparrow\downarrow\downarrow\rangle + e^{i2\pi/3}|\downarrow\uparrow\downarrow\rangle + e^{i4\pi/3}|\downarrow\downarrow\uparrow\rangle), \tag{3b}$$

$$|\psi_3\rangle = \frac{1}{\sqrt{3}}(|\uparrow\downarrow\downarrow\rangle + e^{i4\pi/3}|\downarrow\uparrow\downarrow\rangle + e^{i2\pi/3}|\downarrow\downarrow\uparrow\rangle). \tag{3c}$$

The evolution of the initial state $|\Psi(0)\rangle = |\uparrow\downarrow\downarrow\rangle = (|\psi_1\rangle + |\psi_2\rangle + |\psi_3\rangle)/\sqrt{3}$ is

$$\begin{aligned} |\Psi(t)\rangle = \frac{1}{\sqrt{3}}\sum_{j=1}^{3} e^{-i\lambda_j t}|\psi_j\rangle = \frac{1}{3}[&(1 + 2\cos(\sqrt{3}\kappa t))|\uparrow\downarrow\downarrow\rangle \\ &+(1 + 2\cos(\sqrt{3}\kappa t - 2\pi/3))|\downarrow\uparrow\downarrow\rangle + (1 + 2\cos(\sqrt{3}\kappa t + 2\pi/3))|\downarrow\downarrow\uparrow\rangle]. \end{aligned} \tag{4}$$

It is clear that at time $t = T \equiv 2\pi/(3\sqrt{3}\kappa)$, $|\Psi(T)\rangle = |\downarrow\uparrow\downarrow\rangle$, and at time $t = 2T$, $|\Psi(2T)\rangle = |\downarrow\downarrow\uparrow\rangle$. We obtain the chiral spin wave rotation in eq. (1*a*).

We can follow the same procedure to calculate the dynamics in the subspace with two up spins, $|\downarrow\uparrow\uparrow\rangle$, $|\uparrow\downarrow\uparrow\rangle$ and $|\uparrow\uparrow\downarrow\rangle$. The spin states move to the left, as shown in eq. (1*b*). To understand this surprising result, we try to know how $|\downarrow\uparrow\uparrow\rangle$ evolves based on the knowledge that the state $|\uparrow\downarrow\downarrow\rangle$ rotates to the right. By reversing our definition of up and down, the state $|\downarrow\uparrow\uparrow\rangle$ becomes $|\uparrow\downarrow\downarrow\rangle$ in the new basis. To express the Hamiltonian in this upside down basis, we need to make the replacement $\sigma_j^+ \to \sigma_j^-$ and $\sigma_j^- \to \sigma_j^+$, which results in $H \to -H$. The state evolves backward in time, *i.e.*, the spin wave moves to the left.

3. – Implementation

The key feature of the Hamiltonian in eq. (2) is the imaginary interaction strength between the spins. Previously we proposed the synthetic magnetic field for photons by oscillating the frequencies of three cavities that are coupled to the same spin [8]. Here we consider three spins with oscillating frequencies coupled to the same cavity mode. Let

us first consider two spins with frequencies being modulated with different phases. The interaction Hamiltonian can be written as

$$H_I = \hbar g a^\dagger [\sigma_1^- e^{if\cos(\nu_d t+\phi_1)} + \sigma_2^- e^{if\cos(\nu_d t+\phi_2)}] + \text{H.c.}, \tag{5}$$

where a is the annihilation operator of the cavity, f, ν_d and ϕ_j are the modulation amplitude, frequency and phase for the j-th spin. We assume the central frequencies of the spins are the same as that of the cavity. Since $e^{if\cos(\nu_d t+\phi_j)} = \sum_{n=-\infty}^{\infty} i^n J_n(f) e^{in(\nu_d t+\phi_j)}$ where $J_n(f)$ is the nth order Bessel function of the first kind, we can expand the interaction Hamiltonian $H_I = \sum_n h_n e^{in\nu_d t}$, where

$$h_n = i^n J_n(f)\hbar g a^\dagger [\sigma_1^- e^{in\phi_1} + \sigma_2^- e^{in\phi_2}] + i^n J_n(-f)\hbar g[\sigma_1^+ e^{in\phi_1} + \sigma_2^+ e^{in\phi_2}]a, \tag{6}$$

The effective Hamiltonian is [8]

$$H_e = h_0 + \sum_{n=1}^{\infty} \frac{1}{n\hbar\nu_d}[h_n, h_{-n}] = h_0 + i\frac{\hbar g^2}{\nu_d}\eta(\sigma_2^+\sigma_1^- - \sigma_1^+\sigma_2^-), \tag{7}$$

where $\eta = 2\sum_{n=1}^{\infty} J_n^2(f)\sin[n(\phi_2-\phi_1)]/n$. When $f = 2.4$, $J_0(f) = 0$ and $h_0 = 0$, we obtain the imaginary interaction strength between the two spins. We introduce the third spin with the same modulation amplitude f and frequency ν_d but a different phase ϕ_3, and we set $\phi_j = 2j\pi/3$. We obtain $H_e = H$ in eq. (2) with $\kappa = g^2\eta/\nu_d$ and $\eta = 0.307$.

In the experiments, the transition frequencies of the spins can be modulated by the dynamic Stark shift of a detuned light field with modulated intensities. The spins can be atoms coupled to a one-dimensional waveguide that can mediate long range interactions between atoms [9].

REFERENCES

[1] EINSTEIN A., PODOLSKY B. and ROSEN N., *Phys. Rev.*, **47** (1935) 777.
[2] BELL J. S., *Physics*, **1** (1964) 195.
[3] HENSEN B. *et al.*, *Nature*, **526** (2015) 682; GIUSTINA M. *et al.*, *Phys. Rev. Lett.*, **115** (2015) 250401; SHALM L. K. *et al.*, *Phys. Rev. Lett.*, **115** (2015) 250402.
[4] GREENBERGER D. M., HORNE M. A., SHIMONY A. and ZEILINGER A., *Am. J. Phys.*, **58** (1990) 1131.
[5] LEIBFRIED D., BARRETT M. D., SCHAETZ T., BRITTON J., CHIAVERINI J., ITANO W. M., JOST J. D., LANGER C. and WINELAND D. J., *Science*, **304** (2004) 1476.
[6] MØLMER K. and SØRENSEN A., *Phys. Rev. Lett.*, **82** (1999) 1835.
[7] MONZ T., SCHINDLER P., BARREIRO J. T., CHWALLA M., NIGG D., COISH W. A., HARLANDER M., HANSEL W., HENNRICH M. and BLATT R., *Phys. Rev. Lett.*, **106** (2011) 130506.
[8] WANG D.-W., CAI H., LIU R.-B. and SCULLY M. O., *Phys. Rev. Lett.*, **116** (2016) 220502.
[9] CHANG D. E., CIRAC J. I. and KIMBLE H. J., *Phys. Rev. Lett.*, **110** (2013) 113606.

Proceedings of the International School of Physics "Enrico Fermi"
Course 197 "Foundations of Quantum Theory", edited by E. M. Rasel, W. P. Schleich and S. Wölk
(IOS, Amsterdam; SIF, Bologna) 2019
DOI 10.3254/978-1-61499-937-9-113

Quantum state generation via frequency combs

P. Roztocki

INRS-EMT - Varennes, Canada

M. Kues

INRS-EMT - Varennes, Canada

University of Glasgow - Glasgow, UK

C. Reimer

INRS-EMT - Varennes, Canada

R. Morandotti

INRS-EMT - Varennes, Canada

University of Electronic Science and Technology of China - Chengdu, China

ITMO University - St. Petersburg, Russia

Summary. — In the classical regime, frequency comb sources have revolutionized high-precision metrology and spectroscopy; in this paper, we discuss recent developments, which are extending their use to scalable quantum state generation.

1. – Optical quantum state preparation

Optical quantum state generation and control is, besides its importance towards answering fundamental questions in quantum mechanics, a basis for quantum information applications. In particular, specially-prepared (*e.g.* single-, entangled-, and multi-photon)

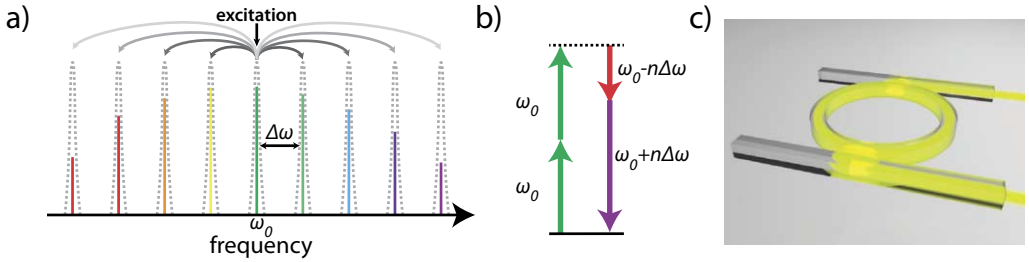


Fig. 1. – a) A frequency comb generated through b) spontaneous four-wave mixing, the nonlinear optical process occurring when a resonance of a c) $\chi^{(3)}$ microring resonator is optically-excited.

quantum states constitute the underlying resources for secure quantum communications [1] and powerful quantum computing [2]. The platforms chosen for quantum state generation must thus point towards "out-of-lab" portability, while, most importantly, allowing for scalable access to complex states. Although photons experience an exceptionally low decoherence that allows them to preserve their state for a long time, large intricate optical quantum states nevertheless remain difficult to prepare and entangle.

Frequency combs (broadband optical sources that have equidistantly-spaced, coherent spectral lines, fig. 1a) operated in the classical regime have revolutionized high-precision metrology/spectroscopy. Recently, their properties have begun to be explored in the context of quantum state generation. This approach brings many benefits, especially for the creation of large quantum states. Frequency combs offer many experimentally-accessible frequency modes in a single spatial mode, giving high dimensionality at a low footprint cost. This enables many simultaneously-operating channels for quantum state generation and the manageable scaling of the setup footprint with state complexity.

2. – Quantum frequency combs from bulk-based systems

The first investigations of quantum frequency combs, based on bulk free-space setups, quickly revealed their potential for the generation of large quantum states. This approach exploited the effect of spontaneous parametric down-conversion in below-threshold–operated optical parametric oscillators to generate continuous-variable non-classical states (*i.e.* squeezed vacuum). In particular, each cavity mode of such a frequency comb can be described by a quantum harmonic oscillator and, analogously to the position and momentum observables, the field's continuous-variable Hilbert space can be described by its amplitude- or phase-quadrature observables. Through the tailored superposition of nonlinear optical processes to link these comb modes, complex states [3] and even quantum networks [4] can be generated. Furthermore, their properties can be verified by observing the squeezed (quantum-noise reduced) field quadratures via homodyne detection characterization.

Quantum state preparation efforts using this and related approaches have been remarkably successful, especially for the generation of many complex states with *e.g.* the

simultaneous generation of 15 quadripartite entangled quantum states using 60 cavity modes [3] or the generation of one 60-mode and two 30-mode copies of a dual-rail quantum-wire state [5]. Richer excitation spectra and more tailored nonlinear optical interactions are predicted to enable larger states [3], and even in the latter experiment the mode number was limited by measurement, not generation, with the maximum number of entangled modes predicted to be at least 6700 [5].

However, the use of such bulk-based approaches coincides with the necessity of using large, expensive, very complex, and actively-stabilized setups (unsuitable for "out-of-lab", portable use). As well, thus far, the continuous-variable non-classical states that have been demonstrated with this approach have not yet achieved the quality (amount of squeezing) required for optical quantum computation.

3. – Integrated quantum frequency combs

In recent years, integrated (on-chip) photonics has stood out as a promising platform for quantum optics, combining mass-producibility (exploiting existing and established silicon chip technologies), portability, and operational stability with a diverse variety of functionalities for quantum state engineering highlighted by demonstrations of on-chip photon sources, path-entangled quantum states, and basic algorithms [6]. However, realizations of on-chip quantum states and quantum gates have mainly focused on the use of polarization- or path-entangled photons, wherein each state dimension corresponds to a waveguide mode. Such architectures are intrinsically limited, as the state/gate complexity directly scales with the quantum circuit footprint and intricacy.

The on-chip generation of classical frequency combs is a very active research field, and many of its principles were assimilated into the first demonstrations of on-chip quantum combs. As materials used for on-chip integration usually exhibit third-order optical nonlinearities ($\chi^{(3)}$), spontaneous four-wave mixing (SFWM) is used for the generation of on-chip quantum frequency combs. The nonlinearity mediates the annihilation of two photons from an excitation field and the simultaneous generation of two daughter photons named signal and idler (the sum of the signal and idler energies is equal to the total energy of the annihilated photons, fig. 1b). To increase the efficiency of this process and also provide the equidistant-line feature of combs, cavity enhancement in on-chip resonant structures is used, particularly in high-Q microring resonators (fig. 1c).

By optically-exciting a single microring resonance, SFWM symmetrically populates neighboring resonances with photon pairs, creating a stable source of several-channel heralded single photons (wherein the signal heralds the presence of the idler, and vice versa) [7]. The double-pulse excitation of a single resonance ($\Psi_{pulses} = |1\rangle + \exp(i\theta)|2\rangle$), was used to demonstrate the generation of time-bin entangled photon pairs (where the generated photon pair is in a coherent superposition of the first and second time bin, $\Psi_{photons} = |1\rangle_{signal}|1\rangle_{idler} + \exp(i\theta)|2\rangle_{signal}|2\rangle_{idler}$, see [8] for further details) over the entire frequency comb spectrum [9]. Quantum state tomography confirmed the high fidelity (0.96) and purity (1.0) of this entangled comb source. The frequency comb nature, enabled by the resonance characteristics of the microring, allows the construction

of more complex states. As the generated photons and the excitation field have equal resonance bandwidths in the ring cavity, the coherence length of different photon pairs is the same and matches the excitation field coherence time. This enables the multiplication of Bell states and the generation of a four-photon time-bin entangled state comb. The realization of this four-photon entangled state was confirmed through four-photon quantum interference with a measured visibility of 89% (without background correction) and through quantum state tomography (exhibiting a fidelity of 64%) [9]. More recently, on-chip resonators have also been used to generate high-dimensional entangled photon pairs [10].

This integrated approach, besides its on-chip portability, presents several benefits. The quantum comb spans the S, C, and L telecommunications bands, with photon frequencies corresponding to standard telecommunication channels spaced by 200 GHz (this large spacing arises from the small cavity size, making the modes easily addressable via the standard filters typically used in telecommunications). A measured photon bandwidth of 110 MHz signifies compatibility with state-of-the-art quantum memories, and the generation process is also very efficient due to cavity enhancement. However, not all components in the photonic circuit are presently integrated; future work may thus be directed towards the implementation of emerging technologies, such as on-chip high-isolation filters and single-photon detectors.

4. – Conclusion

Quantum frequency combs, both in bulk and integrated setups, are a scalable platform for quantum state generation and control. Their versatility, demonstrated in their use as single-photon, entangled-photon, and multi-photon state sources, suggests they will play an important role in future practical implementations of quantum technologies.

* * *

PR and CR acknowledge support from the NSERC Vanier Canada Graduate Scholarship program. MK acknowledges support from the FQRNT-MELS and Marie Skłodowska-Curie grant programs. RM acknowledges support from the Discovery, Strategic, Steacie, Acceleration, and CRC Programs (NSERC); the MESI PSR-SIIRI Initiative (Quebec); the ITMO Fellowship and Professorship Program (#074-U 01, Russian Federation); and the 1000 Talents Sichuan Program (China).

REFERENCES

[1] Kimble H. J., *Nature*, **453** (2008) 1023.
[2] Walther P., Resch K. J., Rudolph T., Schenck E., Weinfurter H., Vedral V., Aspelmeyer M. and Zeilinger A., *Nature*, **434** (2005) 169.
[3] Pysher M., Miwa Y., Shahrokhshahi R., Bloomer R. and Pfister O., *Phys. Rev. Lett.*, **107** (2011) 030505.
[4] Roslund J., de Arajo R. M., Jiang S., Fabre C. and Treps N., *Nat. Photon.*, **8** (2014) 109.

[5] CHEN M., MENICUCCI N. C. and PFISTER O., *Phys. Rev. Lett.*, **112** (2014) 120505.
[6] BONNEAU D., SILVERSTONE J. W. and THOMPSON M. G., *Silicon Quantum Photonics*, in *Silicon Photonics III*, edited by PAVESI L. and LOCKWOOD D. J. (Springer International Publishing AG) 2016, pp. 41–82.
[7] REIMER C., CASPANI L., CLERICI M., FERRERA M., KUES M., PECCIANTI M., PASQUAZI A., RAZZARI L., LITTLE B. E., CHU S. T., MOSS D. J. and MORANDOTTI R., *Opt. Express*, **22** (2014) 6535.
[8] BRENDEL J., GISIN N., TITTEL W. and ZBINDEN H., *Phys. Rev. Lett.*, **82** (1999) 2594.
[9] REIMER C., KUES M., ROZTOCKI P., WETZEL B., GRAZIOSO F., LITTLE B. E., CHU S. T., JOHNSTON T., BROMBERG Y., CASPANI L., MOSS D. J. and MORANDOTTI R., *Science*, **351** (2016) 1176.
[10] KUES M., REIMER C., ROZTOCKI P., ROMERO CORTÉS L., SCIARA S., WETZEL B., ZHANG Y., CINO A., CHU S., LITTLE B., MOSS D., CASPANI L., AZAÑA J. and MORANDOTTI R., *Nature*, **546** (2017) 622.

This page intentionally left blank

Proceedings of the International School of Physics "Enrico Fermi"
Course 197 "Foundations of Quantum Theory", edited by E. M. Rasel, W. P. Schleich and S. Wölk
(IOS, Amsterdam; SIF, Bologna) 2019
DOI 10.3254/978-1-61499-937-9-119

Time after time: From EPR to Wigner's friend and quantum eraser

Marlan O. Scully

Texas A & M University - College Station, TX 77843, USA

Baylor University - Waco, TX, 76798, USA

Daniel M. Greenberger

City University of New York - New York, NY 10031, USA

Wolfgang P. Schleich

Texas A & M University - College Station, TX 77843, USA

Ulm University - Ulm, 89069, Germany

Summary. — The famous Einstein-Podolsky-Rosen (EPR) "paradox" is a good example of how "before" and "after" arguments can be misleading. Wigner's friend is an even better case asking questions such as: when and how does the observer change or "reduce" the state vector? Perhaps the best and most insightful Lehrbeispiel of how to think about before and after issues comes from the quantum eraser. In this case Baysian logic helps clear up before and after confusion via detailed, but simple, calculations.

1. – Introduction

During the present Enrico Fermi School on the foundation of quantum physics it became clear that there was some diversity of opinion (confusion) concerning *e.g.* the famous Einstein-Podolsky-Rosen (EPR) [1, 2] "paradox" and the role of the observer in quantum measurement. In many such discussions the issue of time before and after measurement came in. Our lectures were thus revised to address these issues and the three of us felt that the present notes might help clarify the key role of timing in quantum measurement theory.

The strange nature of time in quantum mechanics was first unveiled by Einstein *et al.* [1] who tried to infer the state of a system "just before" a measurement event, by what the state was after the observation [2]. Wigner [3, 4] took things further by asking when state reduction takes place: at the time the apparatus (or friend) registers an event, or only "when some information" (the yes or no of my friend) enters my consciousness? Schwinger *et al.* [5-7] resolve the Wigner conundrum showing that the incorporation of the observer is enough to produce reduction-type decoherence [8]. The friend never needs to communicate with us. More recently timing in atom two-slit quantum eraser [9, 10] scenarios has been discussed. Here "which slit" measurements can be made and/or erased before or after the atoms are registered on a screen. Many thought that in order to "see" an interference pattern the "which way" (which slit) information has to be erased *before* the atom hits the screen [11]. This reasonable conclusion is incorrect, as we show by careful calculation.

The order and tenor of the presentation of these three topics is as follows:

1. The Einstein *et al.* logic is presented using the spin-$\frac{1}{2}$ approach of Bohm and the before-after paradoxical EPR arguments are resolved by using the reduced density matrix.

2. The role of the observer, *i.e.* Wigner's friend, and the process of observation are illustrated by Wigner's study of coherence and decoherence. Schwinger *et al.* [5-7] show that the friend never needs to communicate his measurement result in order to wipe-out coherence. The mere fact that our friend "knows" is enough to destroy coherence – she never needs to tell us.

3. Quantum eraser and the timing of the eraser event provide insight into the effects of timing of which-way (Welcher Weg) local measurements before and after distant observation. Baysian statistical arguments are used to resolve the timing problem.

2. – The Einstein-Podolsky-Rosen (EPR) problem

In the following, we first present the EPR problem in Einstein's own words. His point of view is also summarized in fig. 1. We then present the density matrix approach [2] to resolving the EPR problem.

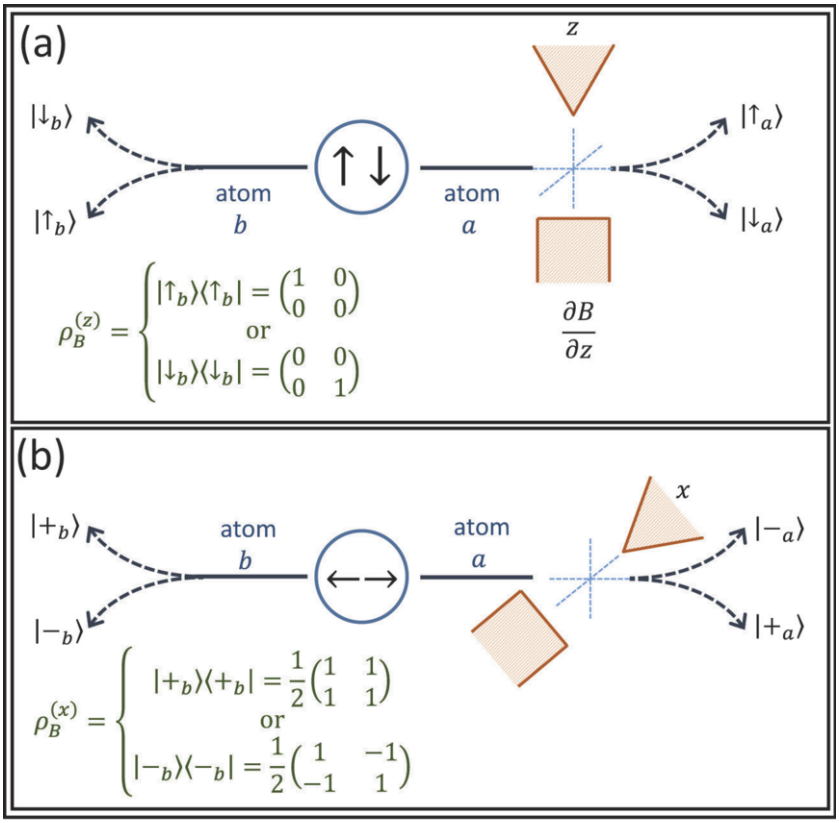


Fig. 1. – (a) **z**-orientation of SGA, (b) **x**-orientation of SGA.

Let us begin by quoting Einstein from his 1936 Franklin Society paper [1] where he says at the bottom of p. 375:

> *"The ψ function does not in any way describe a condition which could be that of a single system; it relates rather to many systems, to 'an ensemble of systems' in the sense of statistical mechanics."*

We will use the statistical or density matrix throughout our discussion. He then goes on to say (p. 376):

> *"Consider a mechanical system constituted of two partial systems A and B which have interaction with each other only during limited time. Let the ψ function before their interaction be given. Then the Schrödinger equation will furnish the ψ function after the interaction has taken place. Let us now determine the physical condition of the partial system A as completely as possible by measurements. Then the quantum mechanics allows us to determine the ψ*

function of the partial system B from the measurements made, and from the ψ function of the total system. This determination, however, gives a result which depends upon which of the determining magnitudes specifying the condition of A has been measured(for instance coordinates or momenta). Since there can be only one physical condition of B after the interaction and which can reasonably not be considered as dependent on the particular measurement we perform on the system A separated from B it may be concluded that the function is not unambiguously coordinated with the physical condition. This coordination of several ψ functions with the same physical condition of system B shows again that the function cannot be interpreted as a (complete) description of a physical condition of a unit system."

Following the Bohm formulation of the EPR problem in terms of the spin-singlet states in the **z**-representation appropriate to a Stern-Gerlach apparatus (SGA) as in fig. 1a, we write

$$|\Psi_{AB}^{(z)}\rangle = \frac{1}{\sqrt{2}}\left[|\uparrow_a\downarrow_b\rangle - |\downarrow_a\uparrow_b\rangle\right], \tag{1}$$

and in the **x**-representation

$$|\Psi_{AB}^{(x)}\rangle = \frac{1}{\sqrt{2}}\left[|+_a -_b\rangle - |-_a +_b\rangle\right]. \tag{2}$$

In fig. 1 we see that the state vector (or density matrix) for the AB system gives different descriptions of the state of spin b depending on what kind of measurement (**z**- or **x**-orientation of the SGA) one envisions, that is

$$\rho_B^{(z)} = \begin{cases} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \\ \text{or} \\ \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \end{cases} \tag{3}$$

on the one hand, or

$$\rho_B^{(x)} = \begin{cases} \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \\ \text{or} \\ \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \end{cases} \tag{4}$$

on the other.

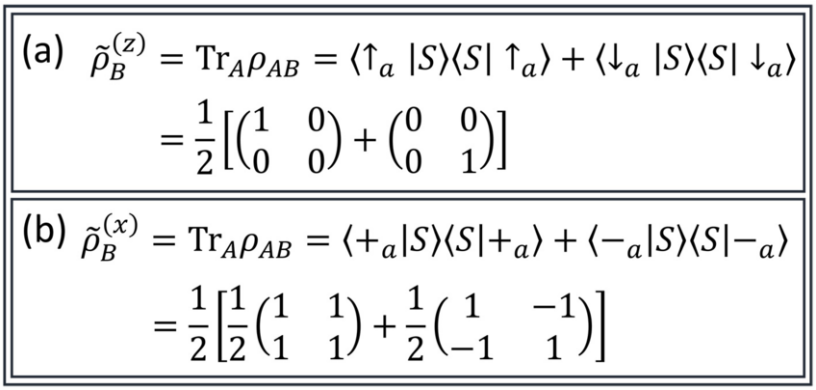


Fig. 2. – (a) Reduced density matrix in **z**-representation (b) Reduced density matrix in **x**-representation, where $|S\rangle$ denotes the spin singlet state of eqs. (1) and (2).

Thus there is more than one density matrix ρ (or state vector ψ) for a single system (spin b). This is what EPR consider "paradoxical".

Next, we show how the resolution of the problem is achieved via the reduced density matrix

$$\tilde{\rho}_B = \mathrm{Tr}_A \rho_{AB}. \tag{5}$$

In fig. 2 we see that the density matrix for the B system (spin b) is the same for both the **z**- and **x**-representation, that is we have the *same* density matrix $\rho_B = \frac{1}{2}\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ for subsystem B (spin b) in both the **z**- and **x**-representations. This resolves the EPR problem; however, the EPR paper points to a deep fact of quantum mechanical life: b spin has neither up and down, nor in and out character before spin a is measured. This is the main take-away lesson.

The before and after character of the time sequence is to be emphasized: Only *after* measuring the spin of atom a, the spin of b is known. Furthermore, it is argued, spin b must be the same before as after. The error in this logic is, as is now well-known, atom b has no state vector *before* measurement. It has only the statistical mixture of fig. 2.

In the next section we follow in the foot steps of Wigner in the study of coherence and observation.

3. – Of Wigner and Wigner's friends

A key aspect of the EPR conundrum is the after and before character of their arguments. In essence they say: After we learn the results of subjecting our spin (atom a) to a **z**-deflection due to a SGA as in fig. 1a, we can infer the spin state of atom b. Then, the argument goes, we "know" the state of the b atom even *before* the measurement because we did nothing to atom b so it must have been in the same state ($|\uparrow_b\rangle$ or $|\downarrow_b\rangle$) before and after the measurement on atom a.

Moreover, we could have chosen to subject atom a to a **x**-oriented SGA and learn the state $|+_b\rangle$ or $|-_b\rangle$ of atom b. Hence, before even measuring the state of atom a we have two state vectors as depicted in fig. 1 for atom b. The "before and after" nature of things is a key part of the EPR logic.

Then comes Wigner with his observer friend. He envisions an EPR singlet-type system in which his friend monitors one system, say atom a, and tells Wigner what he observes so that Wigner then performs the associated state reduction. In Wigner's words:

> *"It is natural to inquire about the situation if one does not make the observation oneself but lets someone else carry it out. What is the wave function if my friend looked at time t? [· · ·] However, even in this case, in which the observation was carried out by someone else, the typical change in the wave function occurred only when some information (the yes or no of my friend) entered my consciousness."*

For example if the friend reports atom a to be in the state $|\uparrow_a\rangle$ then the state reduction to be performed is

$$\frac{1}{\sqrt{2}}[|\uparrow_a\downarrow_b\rangle - |\downarrow_a\uparrow_b\rangle] \rightarrow |\uparrow_a\downarrow_b\rangle. \tag{6}$$

But clever Mr. Wigner asks: When should the state reduction be performed:

– *After* my friend told me the result of his measurement?

– *Before* he told me but when the measurement was actually performed?

Such before and after questions are clarified by designing very specific "experiments" together with the appropriate analysis. The next section on the quantum eraser makes this point in some detail: However, we now consider the observation-state-reduction issue following the Schwinger team [5-7] of Wigner's friends who resolve subtle problems flagged by Wigner.

It is a tenet of the "Copenhagen interpretation" that the act of measuring "reduces" the wave function to a single eigenstate. The problem may be summarized by noting that before its measurement the system of interest is represented by a state vector such as

$$|\psi\rangle = \sum_m a_m |m\rangle, \tag{7a}$$

or the density matrix

$$\rho = \sum_{m,n} a_m^* a_n |n\rangle\langle m|. \tag{7b}$$

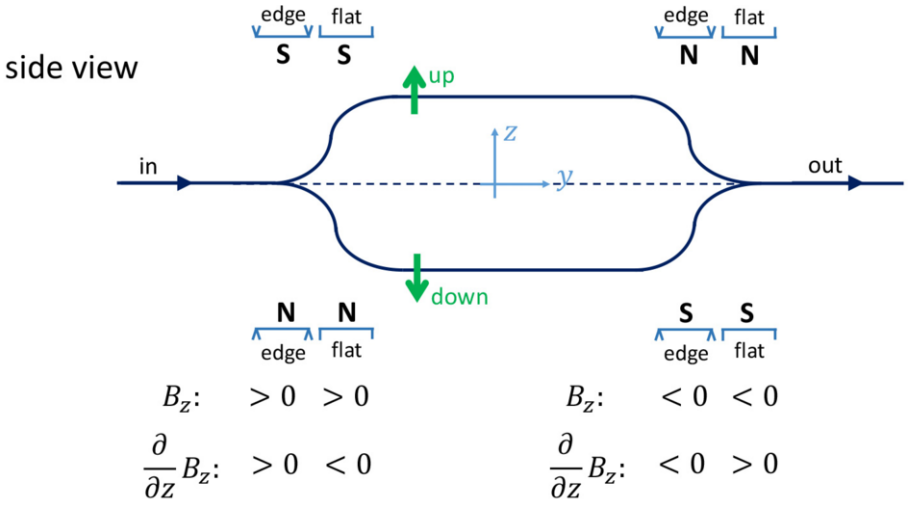


Fig. 3. – Side view of the Stern-Gerlach interferometer (SGI) showing the spins split into two paths and reunite.

But upon "looking" at the system it is found to be in some state

$$|\psi\rangle = |m\rangle, \tag{8a}$$

or equivalently is represented by the density matrix

$$\rho = |m\rangle\langle m|\,. \tag{8b}$$

Please note that it is the off-diagonal nature of the density matrix (7b) that leads to "trouble". As has been observed: the paradoxes associated with quantum theory typically involve the off-diagonality of the density matrix.

Thus, it seems that the simple act of "looking" has changed (reduced) the representation of the system from the general (off-diagonal) density matrix (7b) to the simpler expression (8b). In this regard, Wigner considered a Stern-Gerlach interferometer (SGI) in which a polarized beam of spin-$\frac{1}{2}$ particles is split into two partial beams, and then subsequent Stern-Gerlach deflecting magnets are used to reconstitute these two beams into one, as in fig. 3.

In later work the Schwinger team [5-7] considered the fate of a spin-$\frac{1}{2}$ particle when a detector is present that is sensitive to the passage of particles along one trajectory, but not the other as in fig. 4. It is not surprising that coherence is destroyed as soon as one is able to tell along which path the atom traveled. Most people argue that decoherence arises from dynamical perturbations. However, ref. [7] clearly shows that the loss of coherence in measurements on quantum systems can be traced to the information gleaned by the measuring apparatus ("friend") and the system (spin) being observed. The objective of

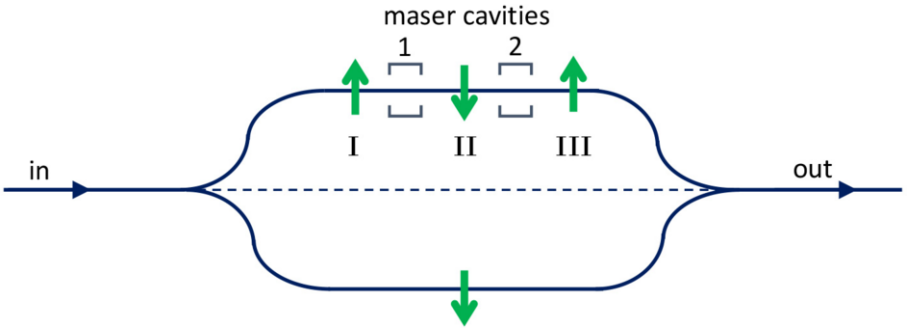


Fig. 4. – SGI with micromaser *welcher Weg* detector in plane along the upper path.

ref. [7] is to understand what happens to spin coherence when a "which path" detector, (*i.e.* a friend) is in place along one arm of the SGI. It is generally agreed *that* spin coherence is destroyed, when one is able to tell along which path the spin-$\frac{1}{2}$ atom traveled through the SGI. But, the question, *how* this loss of coherence arises is controversial. Some argue that the system under observation is affected in a form analogous to the recoil acquired by the scattering of the photon in "Heisenberg's microscope." Others point out that the large number of degrees of freedom in the macroscopic measuring apparatus, leads to an irreversible loss of coherence.

Schwinger *et al.* show that the loss of coherence in measurements on quantum systems can always be traced to information and/or correlation between the (relevant) degrees of freedom of the measuring apparatus and the system being observed.

Following [7] we add two micromaser cavities [8] to the upper path as in fig. 4. The single-mode maser fields are prepared such that the probability that a spin-up atom entering cavity 1 is spin-down between the cavities and spin-up again after leaving cavity 2 is practically equal to unity. Since only the upper partial beam runs through the cavities, the evolution operator describing the SGI with the cavities has the structure

$$U(t_i,t_f) = U_{\rm SGI}(t_i,t_f)\left[\frac{1+\sigma_z(t_i)}{2}U^{(\uparrow)}_{\rm cav}(t_i,t_f) + \frac{1-\sigma_z(t_i)}{2}U^{(\downarrow)}_{\rm cav}(t_i,t_f)\right]. \tag{9}$$

We consider two extreme situations: i) the cavities are prepared in *coherent states* of the maser fields which are classical states of the electromagnetic field. ii) In contrast, the so-called *number states* may be used.

We find that classical coherent maser states preserve spin coherence, but number states destroy it. That is, which-path information is potentially available, provided the properties of the maser fields are changed by the interaction. It is here that the distinction between coherent states and number states enters.

The *welcher Weg* nature of these two extreme situations, that is number *versus* coherent-state preparation, is also clearly displayed upon expressing the state of the spin-photon system in terms of states referring to measurements at the final time t_f. As

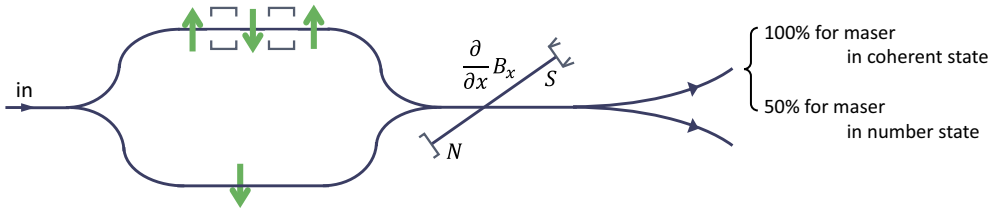


Fig. 5. – SGI with micromasers and SGA for final σ_x measurements, comparing classical coherent states and number states.

shown in [7] at $t = t_f$, the spin⊗detector state reads

$$|\Psi(t_f)\rangle = \frac{1}{\sqrt{2}} \left(|N_1, N_2, \downarrow\rangle + |N_1 + 1, N_2 - 1, \uparrow\rangle\right). \tag{10a}$$

The spin-up and spin-down components are physically distinguished here, inasmuch as there is a one-to-one correspondence between spin down *and* final photon counts N_1, N_2, as well as spin up *and* $N_1 + 1$, $N_2 - 1$. This spin-photon number coupling signifies the correlations established by the interaction. In short, number-state preparation provides us with a good *welcher Weg* detector.

In contrast, cavities prepared in classical coherent states lead to the final state going as

$$|\Psi(t_f)\rangle = \frac{1}{\sqrt{2}} \left(|\alpha_1 e^{-i\beta}, \alpha_2 e^{-i\beta}, \downarrow\rangle + |\alpha_1 e^{-i\beta}, \alpha_2 e^{-i\beta}, \uparrow\rangle\right). \tag{10b}$$

Here the photon and spin degrees of freedom are evidently uncorrelated, no information is available, and therefore we do not have a *welcher Weg* detector.

In order to sharpen the discussion of the SGI with a *cavity-which-way* detector in place we write the density matrix of the spin-$\frac{1}{2}$ system after passing through the cavities and reuniting the spins from the upper and lower arms of the SGI. From ref. [7] the spin density matrix then reads

$$\rho_{\text{spin}}(t_f) = \frac{1}{2} \begin{bmatrix} 1 & \langle a_1^\dagger(t_i)\rangle\langle a_2(t_i)\rangle/\sqrt{N_1 N_2} \\ \langle a_1(t_i)\rangle\langle a_2^\dagger(t_i)\rangle/\sqrt{N_1 N_2} & 1 \end{bmatrix}, \tag{11}$$

where $N_1(N_2)$ is the mean photon number in cavity 1(2) and $a_1^\dagger(a_2)$ is the creation (annihilation) operator for cavity 1(2).

Thus we see that for coherent states in cavities 1 and 2 the spin density matrix is $\left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, while for number states in the cavities $\langle a_1^\dagger\rangle$ and $\langle a_2\rangle$ vanish, and the spin density matrix is $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ as summarized in fig. 5.

Hence, loss of spin coherence (as in state reduction) is caused by and associated with the mere presence of a good detector (friend). We also note that there is no need for a macroscopic number of states in the detector, and no need for information to enter "my

consciouseness". But what if we "erase" the information in the detector? Will coherence then return? We next turn to such questions.

4. – Quantum eraser and the Mohrhoff conundrum

In sect. **2** we have shown how the EPR "two-states-for-one-reality" can be resolved using the reduced density matrix. However, there are many related subtle questions. For example, the problem of Wigner's friend. What if the friend forgets, or has his memory erased? How does the timing of the erasure affect the outcome? Questions such as these are addressed in this section on the quantum eraser.

To lay the ground work for the present problem we recall the abstract from the article summarizing the resolution refutation [12] of the before and after problem posed by Mohrhoff [11]:

> *"Recently, Mohrhoff [Am. J. Phys.* ***64****, 1468–1475 (1996)] has analyzed a thought experiment of ours [Nature (London)* ***351****, 111-116 (1991)] where a double-slit interferometer for atoms is supplemented by a pair of which-way detectors. Owing to the quantum nature of these detectors, the experimenter can choose between acquiring which-way knowledge and observing an interference pattern. The latter option makes use of a procedure called "quantum erasure." Mohrhoff (along with other bright colleagues who have made similar statements) claims erroneously that the experimenter has to make this choice before the atom hits the screen. We readdress this issue here and demonstrate that our original assertion was correct: The experimenter can choose between which-way knowledge and quantum erasure at any time, even after the atom has left its mark on the screen."*

The set-up shown in fig. 6 is a simple physical model of a quantum-eraser experiment following ref. [12]. We consider an array of detectors at the screen at various points $\mathbf{r}_i$. Each detector has a shutter which opens at time t_a for a short time τ_a. The quantum-erasure photodetector is activated by opening its shutters at time t_μ for a time τ_μ. Then, we can easily calculate the state vector for the total system.

At time $t = 0$ the total state vector is

$$\text{which way: } |\Psi^{(i)}(0)\rangle = \frac{1}{\sqrt{2}}\left[|\mu_1\gamma_1\rangle + |\mu_2\gamma_2\rangle\right] \otimes |\beta\rangle|g_i\rangle, \tag{12a}$$

which can also be written as

$$\text{which wave: } |\Psi^{(i)}(0)\rangle = \frac{1}{\sqrt{2}}\left[|\mu_+\gamma_+\rangle + |\mu_-\gamma_-\rangle\right] \otimes |\beta\rangle|g_i\rangle, \tag{12b}$$

where $|\mu_\pm\rangle = \frac{1}{\sqrt{2}}[|\mu_1\rangle \pm |\mu_2\rangle]$ denotes the quantum state of the μ-photon with a similar expression for $|\gamma_\pm\rangle$ describing the γ-photon. The μ-photon detector atom in the ground (excited) state is β (α) and the i-th detector has excited (ground) state e_i (g_i).

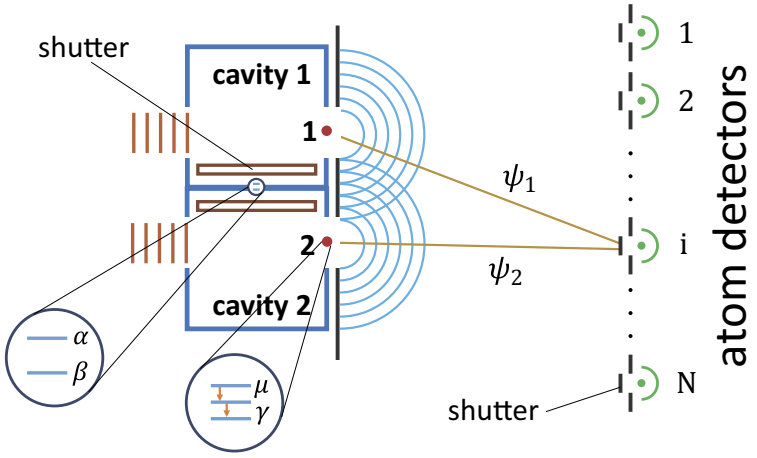


Fig. 6. – Quantum-eraser set-up. The photon detector is represented by a two-level system which is excited upon absorption of a photon. The screen consists of an array of detectors. The microwave (μ) and visible photons (γ) are emitted in cascade from atom 1 or 2 located at the position 1 or 2 of the figure. The quantities ψ_1 and ψ_2 are the γ-photon wave functions as discussed in the text.

Later in time the state $|\Psi^{(i)}(0)\rangle$ in the $\pm$ basis reads

$$|\Psi^{(i)}(t,t_a,t_\mu)\rangle = \Big\{\frac{1}{\sqrt{2}}\Big[\cos\Big(g\frac{\tau_\mu}{2}\Big)|\mu_+\rangle|\beta\rangle - i\sin\Big(g\frac{\tau_\mu}{2}\Big)|0\rangle|\alpha\rangle\Big]|\gamma_+(t)\rangle + \frac{1}{\sqrt{2}}[|\mu_-\rangle|\beta\rangle]\,|\gamma_-(t)\rangle\Big\}\otimes|\mathscr{D}\rangle, \tag{13}$$

where the states $|\gamma_\pm\rangle$ are related to the photon wave function at the i-th detector by

$$\psi_\pm(r_i,t) = \langle 0|\hat{E}(r_i)|\gamma_\pm(t)\rangle = \frac{\mathcal{E}}{r_{1i}}\theta(t-r_{1i}/c)e^{-\gamma(t-r_{1i}/c)}e^{i\omega t} \pm \frac{\mathcal{E}}{r_{2i}}\theta(t-r_{2i}/c)e^{-\gamma(t-r_{2i}/c)}e^{i\omega t}. \tag{14}$$

Here g is the coupling constant between the detector atom and the μ-photon, $|\mathscr{D}(t)\rangle$ denotes the detector array state vector, $\mathcal{E}$ is a constant, and $r_{1i}(r_{2i})$ is the distance from the 1(2) atom to the i-th detector.

Now if we send an atom through the apparatus and find it excites the i-th detector at time t_a; and if we do not activate the quantum eraser then we must use the which-way state (12a) and we can correlate the likelihood of exciting the i-th detector $|\psi_1(r_i,t_a)|^2/2$ and $|\psi_2(r_i,t_a)|^2/2$.

However, if we choose to open the eraser shutter at $t_\mu \approx t_a$, *i.e.*, after the γ-photon has been detected at $\mathbf{r}_i$ on the screen, we can now measure which wave, *i.e.* use state (12b).

Here we do not make any measurements of the type discussed in the previous paragraph. Let us now write the joint count probability that the photodetector is excited at time t_μ, and the i-th screen detector records a count at time t_a as

$$P(e_i;\alpha;t_a,t_\mu)=\langle\Psi^{(i)}(t_a,t_\mu)|\{|e_i\rangle\langle e_i|\otimes|\alpha\rangle\langle\alpha|\}|\Psi^{(i)}(t_a,t_\mu)\rangle, \tag{15}$$

and consulting eq. (13) we see that eq. (15) is just

$$P(e_i;\alpha;t_a,t_\mu)=\frac{1}{2}|\psi_+(r_i)|^2\sin^2\left(g\frac{\tau_\mu}{2}\right). \tag{16a}$$

It is useful to write this expression in terms of the conditional probability $P(e_i|\alpha)$ as

$$P(e_i;\alpha)=P(e_i|\alpha)P(\alpha), \tag{16b}$$

because the probability of finding the i-th detector excited given that the μ-photon detector atom is in the excited state α goes (for perfect detectors) as $P(e_i|\alpha)=\frac{1}{2}|\psi_+(r_i)|^2$. Furthermore $P(\alpha)=\sin^2(g\frac{\tau_\mu}{2})$, so that using (16b) we obtain (16a).

But, Mohrhoff argued, this is only valid if we open the μ-photon shutter *before* the γ-photon hits the detector array. To see that this line of reasoning is *not* correct we turn things around and note that information gleaned from the probability of exciting the i-th detector independent of the state of the μ-detector is $P(e_i)=\frac{1}{2}[|\psi_+(r_i)|^2+|\psi_-(r_i)|^2]$, and the conditional probability of exciting the cavity detector given the i-th atom excited (from eq. (13)) is

$$P(\alpha|e_i)=\frac{|\psi_+(r_i)|^2\sin^2\left(g\frac{\tau_\mu}{2}\right)}{|\psi_+(r_i)|^2+|\psi_-(r_i)|^2}, \tag{17}$$

where the denominator arises from normalization ([1]). Hence, we have the joint probability

$$P(\alpha;e_i;t_\mu,t_a)=\frac{|\psi_+(r_i)|^2\sin^2(g\frac{\tau_\mu}{2})}{|\psi_+(r_i)|^2+|\psi_-(r_i)|^2}\times\frac{1}{2}[|\psi_+(r_i)|^2+|\psi_-(r_i)|^2], \tag{18}$$

which is the same as (16a).

([1]) Perhaps better said: $P(\alpha,e_i)$ is the probability that if the detector i is excited with probability

$$P=\frac{|\psi_+(r_i)|^2}{|\psi_+(r_i)|^2+|\psi_-(r_i)|^2},$$

then the μ-photon detector atom will be in the excited state α with the probalility

$$P\sin^2\left(g\frac{\tau_\mu}{2}\right)=P(\alpha,e_i).$$

.

This resolves the Mohrhoff problem. Do we want particle-like *welcher-weg* information? Then keep the eraser shutters closed and use eq. (12a). Do we want wave-like complementary information? Then open the eraser shutters and use eq. (12b). It does not matter in which order the atomic detection and the photodetection occur. The result is the same.

5. – Summary

EPR logic says that the density matrix of spin B of the AB singlet state has a density matrix determined *after* measuring spin a of the form $\rho_B^{(z)} = |\uparrow\rangle\langle\uparrow|$ or $|\downarrow\rangle\langle\downarrow|$. However a different choice of measurement would yield $\rho_B^{(x)} = |x\rangle\langle x|$ or $|-x\rangle\langle -x|$. Thus EPR argue that the two different density matrices $\rho_B^{(z)}$ and $\rho_B^{(x)}$ describe the B spin before measuring A. In fact the correct density matrix *before* measurement is the reduced density matrix given in fig. 2.

Wigner takes the timing question further by including an observer (friend) who makes a measurement at time t_1 and tells us the results at time t_2. When should the wave function collapse, or equivalently the decoherence of the density matrix take place? Wigner sharpens the decoherence issue by introducing a Stern-Gerlach interferometer (SGI) to study the question of decoherence. Following Schwinger *et al.* we extend the Wigner SGI to include an observation device sensitive to which path the spins take. We show that the density matrix of the spins becomes diagonal when the detector is included in the quantum-mechanical treatment. Questions concerning the timing of collapse or decoherence, *i.e.* state reduction times never arise. Futhermore, macroscopic detectors are not needed to decohere the spins.

The Wigner friend/which-way detector however suggests the interesting question concerning the possibility of which-path information eraser. Indeed, erasing such information makes it possible to regain coherence *e.g.* interference fringes. But do we have to erase which-slit information *before* the γ-photons hit the screen? This is a well-posed question which we can and do answer in the negative. As in the case of EPR and Wigner's SGI, we answer these questions by carrying out detailed calculations on simpler but realistic gedanken experiments with no philosophical discourse.

We conclude that Quantum Theory is a complete and self-contained discipline which can answer any physically sensible question including those involving quantum measurement.

REFERENCES

[1] EINSTEIN A., PODOLSKY B. and ROSEN N., *Phys. Rev.*, **47** (1935) 777; EINSTEIN A., *Physics and Reality* (J. Franklin Institute) 1936, pp. 349-382.

[2] The present treatment follows that of CANTRELL C. and SCULLY M., *Phys. Rep.*, **43** (1978) 499. See also SCULLY M. and ZUBAIRY M., *Quantum Optics* (Cambridge Press) 1997.

[3] GOOD I. J., ed. *The Scientist Speculates* (William Heinemann, Ltd., London) 1961; (Basic Books, Inc., New York) 1962.

[4] WIGNER E., Chapter 13 in *Symmetries and Reflections* (University of Indiana Press) 1967.
[5] ENGLERT B.-G., SCHWINGER J. and SCULLY M., *Found. Phys.*, **18** (1988) 1045.
[6] SCHWINGER J., SCULLY M. and ENGLERT B.-G., *Z. Phys. D*, **10** (1988) 135.
[7] SCULLY M., ENGLERT B.-G. and SCHWINGER J., *Phys. Rev. A*, **40** (1989) 1775.
[8] HAROCHE S. and RAIMOND J.-M., *Exploring the Quantum* (Cambridge University Press) 2006.
[9] SCULLY M. and DRÜHL K., *Phys. Rev. A*, **25** (1982) 2209.
[10] SCULLY M. and WALTHER H., *Found. Phys.*, **28** (1998) 399.
[11] MOHRHOFF U., *Am. J. Phys.*, **64** (1996) 1468.
[12] ENGLERT B.-G., SCULLY M. and WALTHER H., *Am. J. Phys.*, **67** (1999) 325.

Proceedings of the International School of Physics "Enrico Fermi"
Course 197 "Foundations of Quantum Theory", edited by E. M. Rasel, W. P. Schleich and S. Wölk
(IOS, Amsterdam; SIF, Bologna) 2019
DOI 10.3254/978-1-61499-937-9-133

QBism: Quantum theory as a hero's handbook

Christopher A. Fuchs and Blake C. Stacey

Department of Physics, University of Massachusetts Boston - Boston, MA 02125, USA

Summary. — This article situates QBism among other well-known interpretations of quantum mechanics. QBism has its roots in personalist Bayesian probability theory, is crucially dependent upon the tools of quantum information, and in latest developments has set out to investigate whether the physical world might be of a type sketched in the philosophies of pragmatism, pluralism, nonreductionism, and meliorism. Beyond conceptual issues, the technical side of QBism is focused on the mathematically hard problem of finding a *good* representation of quantum mechanics purely in terms of probabilities, without amplitudes or Hilbert-space operators. The best candidate representation involves an entity called a symmetric informationally complete quantum measurement, or SIC. Contemplation of it gives a way to think of the Born rule as an *addition* to the *rules* of probability theory, applicable when an agent considers gambling on the consequences of her actions on the external world, duly taking into account a new universal capacity: namely, Hilbert-space dimension. The article ends by showing that the egocentric elements in QBism represent no impediment to pursuing quantum cosmology and even open up possibilities never dreamt of in the usual formulations.

1. – Introduction

> Chauncey Wright a nearly forgotten philosopher of real merit, taught me when young that I must not say *necessary* about the universe, that we don't know whether anything is necessary or not. So I describe myself as a *bet*tabilitarian. I believe that we can *bet* on the behavior of the universe in its contact with us. [Our underlining; his italics.]
>
> — Oliver Wendell Holmes, Jr.

Quantum physics works. But why?

The mathematical apparatus which we heft onto physics students is astonishingly successful. It guides our understanding of phenomena from the submicroscopic clustering of

quarks to the spectra of quasars, and it underpins technological advances that perfuse throughout our lives. But when we take a moment and ask—whether in a dorm room or the pages of a philosophy journal—what the theory is all about, we find ourselves thrashing about in decades of accumulated murk. How much of the mathematical gadgetry is human convention and historical happenstance, and how much of it truly indicates the character of the natural world that was here before we were and would endure in our absence? Can we take discourse about observership and agency, about what it means to be an agent whose actions have consequences, about the relation between truth and what works in practice, and make honest mathematics of it?

According to the research program of QBism [1-13], the answer is wholeheartedly *Yes*. On one hand, QBism is a way of investing meaning in the abstract structure of quantum theory: It tells us that quantum theory is fundamentally about agency, actions, consequences and expectations. On the other, QBism points out the virtue of reconstructing quantum theory from deep, physical principles. Of all the ideas and theorems that look important, which are *the* ideas, the captivating and compelling seed from which all the formulae would grow given only careful thinking?

We can illustrate the trouble with quantum mechanics by comparing it with other areas of physics in which we have collectively honed our understanding to a high degree of sophistication. Two examples that come to mind are the science of thermodynamics and the special theory of relativity. An old joke has it that the three laws of thermodynamics are "You can't win," "You can't break even," and "You can't get out of the game." To these, we ought to prepend the zero-th law, which we could state as, "At least the scoring is fair." But consider the premise of the joke, which is really rather remarkable: There *are* laws of thermodynamics—a concise list of deep physical principles that underlie and nourish the entire subject. Likewise for special relativity: Inertial observers Alice and Bob can come to agree on the laws of physics, but no experiment they could ever do can establish that one is "really moving" and the other "really standing still"—not even measuring the speed of light. We invest a little mathematics, and then close and careful consideration of these basic principles yields all the details of the formal apparatus, with its nasty square roots or intermingling partial derivatives.

This level of understanding brings many advantages. Having the deep principles set out in explicit form points out how to *test* a theory in the most direct manner. Moreover, it greatly aids us when we *teach* the theory. We do not have to slog through all the confusions that bedeviled the physicists who first developed the subject, to say nothing of the extra confusions created by the fact that "historical" pedagogy is almost inevitably a caricature. In addition, a principled understanding helps us *apply* a theory. As we work our way into a detailed calculation, we can cross-check against the basic postulates. Does our calculation imply that signals travel faster than light? Does our seventeenth equation imply that entropy is flowing the wrong way? We must have made an error! And, when we found our theory upon its deep principles, we have a guide for *extending* our theory, because we know what properties must obtain in order for a new, more general theory to reduce to our old one in a special case.

To our great distress, we must admit that in the matter of quantum mechanics, the physics profession lacks this level of understanding.

Instead, we have a mathematical apparatus and day-to-day experience on how to apply it successfully. But the deep principles remain elusive. What we have in their place is almost a century of "interpretations" cooked up for the theory—campfire stories told to give meaning to the mathematics and say what it is "all about." And what a host of tales have there been!

- Copenhagen Interpretations – Bohr, Heisenberg, Pauli, von Weizsäcker, Peierls, Wheeler (they're actually all different!)
- Nonlocal Hidden Variables – de Broglie, Bohm, Bub (early on), Maudlin, Goldstein, Valentini, Norsen, Hardy
- Stochastic Mechanics – Nelson, Smolin (Lee)
- Modal Interpretations – Kochen, Dieks, van Fraassen, Bub (later on), Healey (for a time), Spekkens & Sipe (briefly)
- Quantum Logics – Birkhoff & von Neumann, Mackey, Jauch, Piron, Finkelstein, Putnam
- Consciousness-Induced Collapse Theories – Wigner, von Neumann
- Objective Collapse Models – Ghirardi-Rimini-Weber, Pearl, Penrose, Frigg, Gisin, Tumulka, Albert
- Consistent Histories – Griffiths, Omnès
- Transactional Interpretation – Cramer, Kastner
- Relational Interpretations – Rovelli, Spekkens (TBA)
- Ensemble Interpretation – Einstein (?), Ballentine
- Informational Interpretations – Zeilinger, Brukner, Bub (presently)
- Superdeterminism – Bell (as a joke), 't Hooft (in full seriousness)
- Many-Worlds Interpretations – Everett, Albert & Loewer, Barbour, Coleman, Deutsch (early version), Deutsch (later version), DeWitt, Gell-Mann & Hartle, Geroch, Graham, Greaves, Lockwood, Papineau, Saunders, Smolin (John), Tegmark, Vaidman, Wallace, Zurek ...

This list is far from exhaustive, and as already hinted abundantly, two authors listed under the same bullet may well disagree on significant issues([1]).

([1]) Over the years, the many-worlders in particular have done a remarkably poor job of agreeing with one another about what specifically there are supposed to be many of.

It is often said that "all interpretations of quantum mechanics make the same predictions, so we cannot tell them apart by experiment." This is false for at least two reasons, a small one and a big one. The small point is that some ideas classified among the "interpretations" are deliberately fashioned to depart from quantum theory in some way. Objective Collapse models are the chief examples of this. But more importantly, it is not clear that all of the interpretations *can be made to yield predictions at all*, when they are thought upon with adequate stringency.

We should also mention the attitude that there is no *point* to "interpreting" quantum theory, because quantum theory will turn out to be wrong anyway. Perhaps the reason why we have yet to fit quantum physics together with gravity is that the quantum side is fundamentally defective, in some fashion that has only manifested when we put the full weight of general relativity upon it. (Feynman even speculated that gravity itself could be the result of quantum mechanics breaking down, in which case "quantizing gravity" would be meaningless [14].) Fair enough! But how, then, do we modify quantum theory in such a way that it still works in all the many places it has worked so far?

So the field of quantum foundations is not unfounded; it is absolutely vital to physics as a whole. But what constitutes "progress" in quantum foundations? How would one know progress if one saw it? Through the years, it seems the most popular strategy has been to remove the observer from the theory just as quickly as possible, and with surgical precision. In practice this has generally meant to keep the *mathematical structure* of quantum theory as it stands (complex Hilbert spaces, operators, tensor products, etc.), but, by hook or crook, find a way to tell a story about the mathematical symbols that involves no observers at all.

In short, the strategy has been to reify or objectify all the mathematical symbols of the theory and then explore whatever comes of the move. All the various interpretations that result see quantum states as physical entities, like a blob of ψ-flavored gelatin, sliding about in accord with its own dynamical laws. Three examples suffice to give a feel: In the de Broglie-Bohm "pilot wave" version of quantum theory, there are no fundamental measurements, only "particles" flying around in a $3N$-dimensional configuration space, pushed around by a wave function regarded as a real physical field in that space. In "spontaneous collapse" versions, systems are endowed with quantum states that generally evolve unitarily, but from time-to-time collapse without any need for measurement. In Everettian or "many-worlds" quantum mechanics, it is only the world as a whole—they call it a multiverse—that is really endowed with an intrinsic quantum state, and that quantum state evolves deterministically, with only an *illusion from the inside* of probabilistic "branching."

The trouble with all these interpretations as quick fixes for quantum strangeness is that they look to be just that, *really quick fixes*. They look to be interpretive strategies hardly compelled by the particular details of the quantum formalism, giving only more or less arbitrary appendages to it. This already explains in part why we have been able to exhibit three such different strategies, but it is worse: Each of these strategies gives rise to its own set of tough-to-swallow ideas. Pilot-wave theories, for instance, give instantaneous action at a distance, but not actions that can be harnessed to send

detectable signals. If so, then what a delicately balanced high-wire act nature presents us with. And how much appeal does the idea of waves pushing particles about really have when it turns out that, in order to be consistent, "position measurements" don't really measure particle positions [15]? Or take the Everettians. Their world purports to have no observers, but then it has no probabilities either. What are we then to do with the Born rule for calculating quantum probabilities? Throw it away and say it never mattered? It is true that quite an effort has been made by the Everettians to rederive the rule by one means or another. But these attempts may have re-imported at least as much vagueness as they claim to eliminate [16-20]. To many in the outside world, it looks like the success of these derivations depends upon where they are assessed—for instance, whether in Oxford [21, 22] or Cambridge [23, 24].

QBists hold that the way forward is to own up to the following lesson. Before there were people using quantum *theory* as a branch of physics, before they were *calculating* neutron-capture cross-sections for uranium and working on all the other practical problems the theory suggests, there were no quantum states. The world may be full of stuff and things of all kinds, but among all the stuff and all the things, there is no unique, observer-independent, *quantum-state kind of stuff*.

The immediate payoff of this strategy is that it eliminates the conundrums arising in the various objectified-state interpretations. A paraphrase of a quote by James Hartle makes the point decisively [25]:

> A quantum-mechanical state being a summary of the observers' information about an individual physical system changes both by dynamical laws, and whenever the observer acquires new information about the system through the process of measurement. The existence of two laws for the evolution of the state vector becomes problematical only if it is believed that the state vector is an objective property of the system. If, however, the state of a system is defined as a list of [experimental] propositions together with their [probabilities of occurrence], it is not surprising that after a measurement the state must be changed to be in accord with [any] new information. The "reduction of the wave packet" does take place in the consciousness of the observer, not because of any unique physical process which takes place there, but only because the state is a construct of the observer and not an objective property of the physical system.

The objective quantum state is the latter-day equivalent of the luminiferous æther. But recognizing this is only the first step of an adventure. Luckily the days for this expedition are ripe, thanks in large part to the development of the field of quantum information theory in the last 25 years—that is, the multidisciplinary field that has brought about quantum cryptography, quantum teleportation, and will one day bring about full-blown quantum computation. Terminology can say it all: A practitioner in this field, whether she has ever thought an ounce about quantum foundations, is just as likely to say "quantum information" as "quantum state" when talking of any $|\psi\rangle$. "What does the quantum

teleportation protocol do?" A now completely standard answer is: "It transfers *quantum information* from Alice's site to Bob's." What we have here is a change of mindset [6].

What the facts and figures, protocols and theorems of quantum information pound home is the idea that quantum states look, act, and feel like information in the technical sense of the word—the sense provided by probability theory and Shannon's information theory. There is no more beautiful demonstration of this than Robert Spekkens's "toy model" for mimicking various features of quantum mechanics [26]. In that model, the "toys" are each equipped with four possible mechanical configurations; but the players, the manipulators of the toys, are consistently impeded from having more than one bit of information about each toy's actual configuration. (Or a total of two bits for each two toys, three bits for each three toys, and so on.) The only things the players can know are their own states of uncertainty about the configurations. The wonderful thing is that these states of uncertainty exhibit many of the characteristics of quantum information: from the no-cloning theorem to analogues of quantum teleportation, quantum key distribution, entanglement monogamy, and even interference in a Mach-Zehnder interferometer. More than two dozen quantum phenomena are reproduced *qualitatively*, and all the while one can always pinpoint the underlying cause of this: The phenomena arise in the uncertainties, never in the mechanical configurations. It is the states of uncertainty that mimic the formal apparatus of quantum theory, not the toys' so-called *ontic states* (states of reality).

What considerations like this tell the ψ-ontologists([2])—*i.e.*, those who attempt to remove the observer too quickly from quantum mechanics by giving quantum states an unfounded ontic status—was well put by Spekkens:

> [A] proponent of the ontic view might argue that the phenomena in question are not mysterious if one abandons certain preconceived notions about physical reality. The challenge we offer to such a person is to present a few simple physical principles by the light of which all of these phenomena become conceptually intuitive (and not merely mathematical consequences of the formalism) within a framework wherein the quantum state is an ontic state. Our impression is that this challenge cannot be met. By contrast, a single information-theoretic principle, which imposes a constraint on the amount of knowledge one can have about any system, is sufficient to derive all of these phenomena in the context of a simple toy theory ...

The point is, far from being an appendage cheaply tacked on to the theory, the idea of quantum states as information has a simple unifying power that goes some way toward explaining why the theory has the very mathematical structure it does([3]). By contrast,

---

([2]) This beautiful word, not to be confused with a practitioner of Scientology, was coined by Christopher Granade [27].

([3]) We say "goes some way toward" because, though the toy model makes about as compelling a case as we have ever seen that quantum states are states of information (an extremely valuable

who could take the many-worlds idea and derive any of the structure of quantum theory out of it? This would be a bit like trying to regrow a lizard from the tip of its chopped-off tail: The Everettian conception never purported to be more than a reaction to the formalism in the first place.

But there are still deep puzzles left outstanding. Above all, there are the old questions of *Whose information?* and *Information about what?*—these certainly must be addressed before any resolution of the quantum mysteries can be declared a success. It must also be settled whether quantum theory is obligated to give a criterion for what counts as an observer. Finally, because no one wants to give up on physics, we must tackle head-on the most crucial question of all: If quantum states are not part of the stuff of the world, then what is? What sort of stuff does quantum mechanics say the world *is* made of?

An understanding of the quantum, like all things worth having, will not come easily. But this much is sure: The glaringly obvious (that a large part of quantum theory, the central part in fact, is about information) should not be abandoned rashly. To do so is to lose grip of the theory as it is applied in practice, with no better grasp of reality in return. If on the other hand, one holds fast to the central point about information, initially frightening though it may be, one may still be able to reconstruct a picture of reality from the unfocused edge of one's vision. Often the best stories come from there anyway.

So, what is the overarching story of QBism, and what does the QB stand for anyway? The Q clearly stands for Quantum, but the B? Initially, we had taken the B to stand for Bayesian, the interpretation of probability from which our efforts grew. This is reflected in much of the literature on the subject—see, for instance, the title of the *Stanford Encyclopedia of Philosophy* article on QBism [13]. However, QBists eventually became dissatisfied with this meaning for the B, as there are just too many varieties of Bayesianism [28]—QBism only represents one very specific strain of it [11]. So, for a time we would say jokingly that the B rather stands for Bruno de Finetti, our hero in probability theory. But eventually and with good reason, we landed on the idea of B for Bettabilitarian! This rolling word, coined by U.S. Supreme Court Justice Oliver Wendell Holmes, Jr. seemed to capture QBism perfectly. Author Louis Menand, in his magisterial book *The Metaphysical Club: A Story of Ideas in America* [29], put it like this:

> "The loss of certainty" is a phrase many intellectual historians have used the characterize the period in which Holmes lived. But the phrase has it backward. It was not the loss of certainty that stimulated the late-nineteenth-century thinkers with whom Holmes associated; it was the discovery of uncertainty. Holmes was, in many respects, a materialist. He believed, as he

---

step forward), it gravely departs from quantum theory in other aspects. For instance, by its nature, it can give no Bell inequality violations or analogues of the Kochen-Specker noncolorability theorems. Later sections of this paper will indicate that the cause of the deficit is that the toy model differs crucially from quantum theory in its answer to the question *Information about what?*

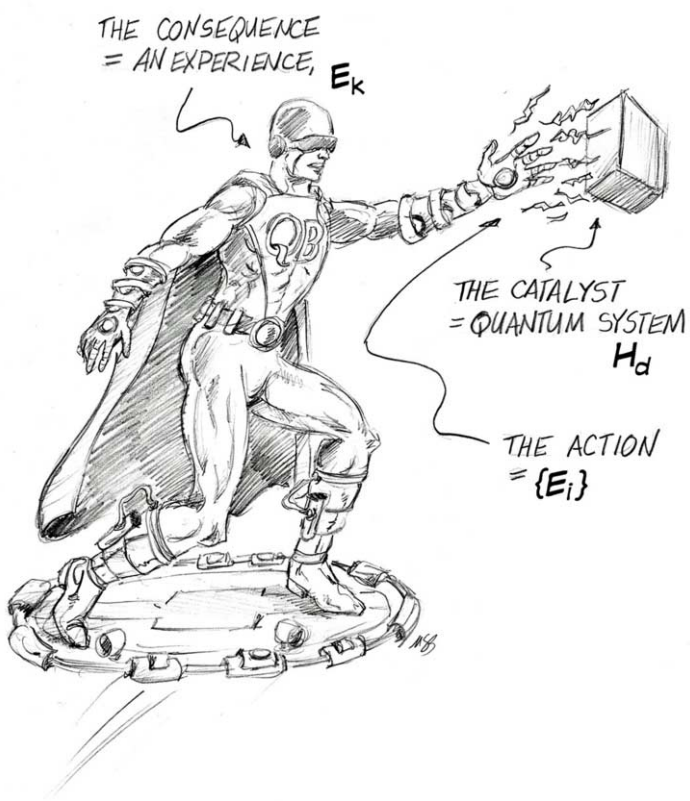


Fig. 1. – "May not the *creatia* of a quantum observer's actions likewise be such additions to the universe as to enhance its total value? And on this view, is not the QBist quantum observer—the agent—a kind of superhero for the universe as a whole, making *extra* things happen wherever, whenever he is called to duty?" (Drawing courtesy of Mark Staff Brandl.)

put it, that "the law of the grub ... is also the law for man." And concerning the hope of social betterment, he was something worse than a pessimist. "I despise," he said, "the upward and onward." But he was not entirely a determinist, because he did not think that the course of human events was fixed .... Complete certainty was an illusion; of that he was certain. There were only greater and lesser degrees of certainty, and that was enough. It was, in fact, better than enough; for although we always want to reduce the degree of uncertainty in our lives, we never want it to disappear entirely, since uncertainty is what puts the play in the joints. Imprecision, the sportiveness, as it were, of the quantum, is what makes life interesting and change possible. Holmes liked to call himself a "bettabilitarian": we cannot know what consequences the universe will attach to our choices, but we can bet on them, and we do it every day.

And that is it—*that* is QBism! To be a QBist is to use quantum theory to be a "better bettabilitarian" in this world in which we are all immersed and which we shape with our every action.

Let us unroll this big idea.

2. – Exactly how quantum states fail to exist

> An experimental physicist usually says that an "experimentally determined" probability has an "error," and writes $P(H) = N_H/N \pm 1/2\sqrt{N}$. There is an implication in such an expression that there is a "true" or "correct" probability which could be computed if we knew enough, and that the observation may be in "error" due to a fluctuation. There is, however, no way to make such thinking logically consistent. It is probably better to realize that the probability concept is in a sense subjective, that it is always based on uncertain knowledge, and that its quantitative evaluation is subject to change as we obtain more information.
>
> — Richard P. Feynman
> *The Feynman Lectures on Physics*

Every area of human endeavor has its bold extremes. Ones that say, "If this is going to be done right, we must go this far. Nothing less will do." In probability theory, the bold extreme is the personalist Bayesian account of probability [30]. It says that probability theory is of the character of formal logic—a set of criteria for testing consistency. In the case of formal logic, the consistency is between truth values of propositions. However logic itself does not have the power to *set* the truth values it manipulates. It can only say if various truth values are consistent or inconsistent; the actual values come from another source. Whenever logic reveals a set of truth values to be inconsistent, one must dip back into the source to find a way to alleviate the discord. But precisely in which way to alleviate it, logic gives no guidance. "Is the truth value for this one isolated proposition correct?". Logic itself is powerless to say.

The key idea of personalist Bayesian probability theory is that it too is a calculus of consistency (or "coherence" as the practitioners call it), but this time for one's decision-making degrees of belief. Probability theory can only say if various degrees of belief are consistent or inconsistent with each other. The actual beliefs come from another source, and there is nowhere to pin their responsibility but on the agent who holds them. Dennis Lindley put it nicely in his book *Understanding Uncertainty* [31]:

> The Bayesian, subjectivist, or coherent, paradigm is egocentric. It is a tale of one person contemplating the world and not wishing to be stupid (technically, incoherent). He realizes that to do this his statements of uncertainty must be probabilistic.

A probability *assignment* is a tool an agent uses to make gambles and decisions—it is a tool she uses for navigating life and responding to her environment. Probability *theory* as a whole, on the other hand, is not about a single isolated belief, but about a whole mesh of them. When a belief in the mesh is found to be incoherent with the others, the theory flags the inconsistency. However, it gives no guidance for how to mend any incoherences it finds. To alleviate the discord, one can only dip back into the source of the assignments—specifically, the agent who attempted to sum up all her history, experience, and expectations with those assignments in the first place. This is the reason for the terminology that a probability is a "degree of belief" rather than a "degree of truth" or "degree of facticity."

To give an example of how mere internal consistency can yield the mathematical rules of probability is to quantify the consequences of acting upon beliefs in terms of costs and benefits. An agent, whom we can call Alice, does business with a bookie, whose goal is to profit by exposing inconsistencies in Alice's mesh of beliefs. Alice's goal is to avoid gambling in a way that forces her into a sure loss. The bookie buys and sells lottery tickets of the form

(1) $\boxed{\text{Worth \$1 if the event } E \text{ occurs.}}$

Based on her own expectations about the event E, Alice assigns a number $p(E)$, which is the price in dollars at which she is willing to buy or to sell a lottery ticket of this form. The normative rule that Alice should avoid a sure loss implies bounds on $p(E)$: If $p(E) < 0$, then Alice would be willing to pay money to have the bookie take the ticket off her hands, whereas if $p(E) > 1$, she would be willing to pay more for a ticket than it could ever be worth. Furthermore, consider two events E and F, which Alice believes to be mutually exclusive. Alice must strive for consistency in her pricing for the following three tickets. First,

(2) $\boxed{\text{Worth \$1 if } E.}$

Second,

(3) $\boxed{\text{Worth \$1 if } F.}$

And finally,

(4) $\boxed{\text{Worth \$1 if } E \text{ or } F.}$

The value of the third ticket should be the sum total value of the first two. If Alice sets her prices such that $p(E \text{ or } F) > p(E) + p(F)$, then the bookie can have a set of transactions that lead her into a sure loss. In the jargon, Alice is vulnerable to a "Dutch book": The bookie sells Alice the third ticket and buys the first two, leaving Alice in debt. Whichever event happens, Alice cannot recoup the loss. Likewise, the bookie

can lead Alice into a sure loss if Alice chooses $p(E \text{ or } F) < p(E) + p(F)$. In this way, striving to satisfy the normative rule of avoiding sure loss, Alice builds up the theory of probabilities: Her personal probability for an event E is simply her fair price for a gamble upon E.

Suppose now that Alice feels compelled to making assignments $p(E)$, $p(F)$, and $p(E \text{ or } F)$ such that $p(E \text{ or } F) \neq p(E) + p(F)$. She detects that something is amiss. What to do to fix the problem? For this, probability theory says nothing: It's now up to Alice to make an adjustment somewhere, somehow, or else she will be vulnerable to a sure loss.

Where personalist Bayesianism breaks away the most from other developments of probability theory is that it says there are no *external* criteria for declaring an isolated probability assignment right or wrong. The only basis for a judgment of adequacy comes from the *inside*, from the greater mesh of beliefs Alice may have the time or energy to access when appraising coherence.

It was not an arbitrary choice of words to title this section *Exactly How Quantum States Fail To Exist*, but a hint of what we must elaborate to develop a perfected vaccine against the fever of quantum interpretations. This is because the phrase has a precursor in a slogan that Bruno de Finetti, the founder of personalist Bayesianism, used to vaccinate probability theory itself. In the preface to his seminal book [32], de Finetti writes, centered in the page and in all capital letters,

PROBABILITY DOES NOT EXIST.

It is a powerful statement, constructed to put a finger on the single most-significant cause of the conceptual problems in pre-Bayesian probability theory. A probability is not a solid object, like a rock or a tree that the agent might bump into, but a feeling, an estimate inside herself.

Previous to Bayesianism, probability was often thought to be a physical property([4])—something objective and having nothing to do with decision-making or agents at all. But when thought so, it could be thought only inconsistently so. And hell hath no fury like an inconsistency scorned. The trouble is always the same in all its varied and complicated forms: If probability is to be a physical property, it had better be a rather ghostly one—one that can be told of in campfire stories, but never quite prodded out of the shadows. Here's a sample dialogue:

Pre-Bayesian: Ridiculous, probabilities are without doubt objective. They can be seen in the relative frequencies they cause.

Bayesian: So if $p = 0.75$ for some event, after 1000 trials we'll see exactly 750 such events?

([4]) Witness Richard von Mises, who even went so far as to write, "Probability calculus is part of *theoretical physics* in the same way as classical mechanics or optics, it is an entirely self-contained theory of certain phenomena [...]" [33].

Pre-Bayesian: You might, but most likely you won't see that exactly. You're just likely to see something close to it.

Bayesian: "Likely"? "Close"? How do you define or quantify these things without making reference to your degrees of belief for what will happen?

Pre-Bayesian: Well, in any case, in the infinite limit the correct frequency will definitely occur.

Bayesian: How would I know? Are you saying that in one billion trials I could not possibly see an "incorrect" frequency? In one trillion?

Pre-Bayesian: OK, you can in principle see an *incorrect* frequency, but it'd be ever less *likely*!

Bayesian: Tell me once again, what does "likely" mean?

This is a cartoon of course, but it captures the essence and the futility of every such debate. It is better to admit at the outset that probability is a degree of belief, and deal with the world on its own terms as it coughs up its objects and events. What do we gain for our theoretical conceptions by saying that along with each actual event there is a ghostly spirit (its "objective probability," its "propensity," its "objective chance") gently nudging it to happen just as it did? Objects and events are enough by themselves.

Similarly for quantum mechanics. Here too, if ghostly spirits are imagined behind the actual events produced in quantum measurements, one is left with conceptual troubles to no end. The defining feature of QBism [1-12] is that it says along the lines of de Finetti, "If this is going to be done right, we must go this far." Specifically, there can be no such thing as a right and true quantum state, if such is thought of as defined by criteria *external* to the agent making the assignment: Quantum states must instead be like personalist, Bayesian probabilities.

The direct connection between the two foundational issues is this. Quantum states, through the Born rule, can be used to calculate probabilities. Conversely, if one assigns probabilities for the outcomes of a well-selected set of measurements, then this is mathematically equivalent to making the quantum-state assignment itself. The two kinds of assignments determine each other uniquely. Just think of a spin-$\frac{1}{2}$ system. If one has elicited one's degrees of belief for the outcomes of a σ_x measurement, and similarly one's degrees of belief for the outcomes of σ_y and σ_z measurements, then this is the same as specifying a quantum state itself: For if one knows the quantum state's projections onto three independent axes, then that uniquely determines a Bloch vector, and hence a quantum state. Something similar is true of all quantum systems of all sizes and dimensionality. There is no mathematical fact embedded in a quantum state ρ that is not embedded in an appropriately chosen set of probabilities([5]). Thus generally, if probabilities are personal in the Bayesian sense, then so too must be quantum states.

What this buys interpretatively, beside airtight consistency with the best understanding of probability theory, is that it gives each quantum state a home. Indeed, a home

([5]) See sect. **7** where this statement is made precise in all dimensions.

localized in space and time—namely, the physical site of the agent who assigns it! By this method, one expels once and for all the fear that quantum mechanics leads to "spooky action at a distance," and expels as well any hint of a problem with "Wigner's friend" [34]. It does this because it removes the very last trace of confusion over whether quantum states might still be objective, agent-independent, physical properties.

The innovation here is that, for most of the history of efforts to take an informational point of view about quantum states, the supporters of the idea have tried to have it both ways: that on the one hand quantum states are not real physical properties, yet on the other there is a right and true quantum state independent of the agent after all. For instance, one hears things like, "The *right* quantum state is the one the agent should adopt if he had all the information." The tension in these two desires leaves their holders open to attack on both flanks and general confusion all around.

Take first instantaneous action at a distance—the horror of this idea is often one of the strongest motivations for those seeking to take an informational stance on quantum states. But, now an opponent can say:

> If there is a *right quantum state*, then why not be done with all this squabbling and call the state a physical fact to begin with? It is surely external to the agent if the agent can be wrong about it. But, once you admit that (and you should admit it), you're sunk: For, now what recourse do you have to declare no action at a distance when a delocalized quantum state changes instantaneously?
>
> Here I am with a physical system right in front of me, and though *my* probabilities for the outcomes of measurements I can do on it might have been adequate a moment ago, there is an objectively better way to gamble *now* because of something that happened far in the distance? (Far in the distance and just now.) How could that not be the signature of action at a distance? You can try to defend yourself by saying "quantum mechanics is all about relations"([6]) or some other feel-good phrase, but I'm talking about measurements right here, in front of me, with outcomes I can see right now. Ones entering my awareness—not outcomes in the mind of God who can see everything and all relations. It is that which I am gambling upon with the

([6]) A typical example is of a woman traveling far from home when her husband divorces her. Instantaneously she becomes unmarried—marriage is a relational property, not something localized at each partner. It seems to be popular to give this example and say, "Quantum mechanics might be like that." The conversation usually stops without elaboration, but let's carry it a little further: Suppose the woman, Carol, is right in front of Alice. Alice has a set of probabilities for what might happen should she buy Carol a celebratory bottle of tequila and congratulate Carol on losing the deadbeat. Would the far-off divorce mean that there is instantaneously a different set of probabilities that Alice could use for weighing the consequences of those actions? Not at all. Alice would have no account to change her probabilities (not due to the divorce, anyway) until Alice became aware of Carol's changed relation, however long it might take that news to get to Alice.

> help of the quantum formalism. An objectively better quantum state would mean that my gambles and actions, though they would have been adequate a moment ago, are now simply wrong in the eyes of the world—they could have been better. How could the quantum system in front of me generate outcomes instantiating that declaration without being privy to what the eyes of the world already see? That's action at a distance, I say, or at least a holism that amounts to the same thing—there's nothing else it could be.

Without the protection of truly personal quantum-state assignments, action at a distance is there as doggedly as it ever was. And things only get worse with "Wigner's friend" if one insists there be a *right* quantum state. As it turns out, the method of mending this conundrum displays one of the most crucial ingredients of QBism. Let us put it in plain sight.

"Wigner's friend" is the story of two agents, Wigner and his friend, and one quantum system—the only deviation we make from a more common presentation([7]) is that we put the story in informational terms. It starts off with the friend and Wigner having a conversation: Suppose they both agree that some quantum state $|\psi\rangle$ captures their mutual beliefs about the quantum system([8]). Furthermore suppose they agree that at a specified time the friend will make a measurement on the system of some observable (outcomes $i = 1, \ldots, d$). Finally, they both note that if the friend gets outcome i, he will (and should) update his beliefs about the system to some new quantum state $|i\rangle$. There the conversation ends and the action begins: Wigner walks away and turns his back to his friend and the supposed measurement. Time passes to some point beyond when the measurement should have taken place.

What now is the "correct" quantum state each agent should have assigned to the quantum system? We have already concurred that the friend will and should assign some $|i\rangle$. But what of Wigner? If he were to consistently dip into his mesh of beliefs, he would very likely treat his friend as a quantum system like any other: one with some initial quantum state ρ capturing his (Wigner's) beliefs of *it* (the friend), along with a linear evolution operator([9]) U to adjust those beliefs with the flow of time([10]). Suppose this quantum state includes Wigner's beliefs about everything he assesses to be interacting with his friend—in old parlance, suppose Wigner treats his friend as an isolated system. From this perspective, before any further interaction between himself and the friend or the other system, the quantum state Wigner would assign for the two together would be $U(\rho \otimes |\psi\rangle\langle\psi|)U^\dagger$—most generally an entangled quantum state. The

---

([7]) For instance, see ref. [35].

([8]) Being Bayesians, of course, they don't have to agree at this stage—for recall $|\psi\rangle$ is not a physical fact for them, only a catalogue of *beliefs*. But suppose they do agree.

([9]) We suppose for the sake of introducing less technicality that U is a unitary operation, rather than the more general completely positive trace-preserving linear maps of quantum information theory [36]. This, however, is not essential to the argument.

([10]) For an explanation of the status of unitary operations from the QBist perspective, as personal judgments directly analogous to quantum states themselves, see sect. **7** and refs. [2, 5, 37].

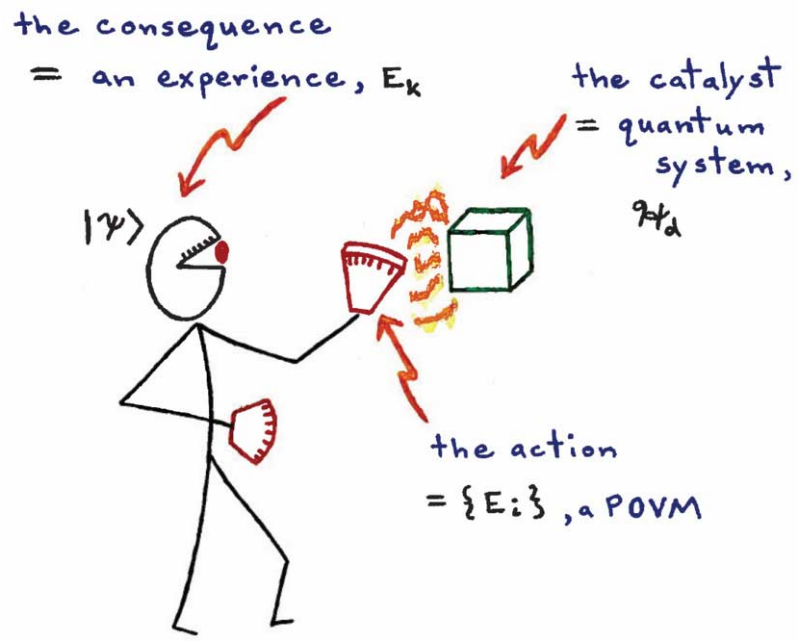


Fig. 2. – In contemplating a quantum measurement, one makes a conceptual split in the world: one part is treated as an agent, and the other as a kind of reagent or catalyst (one that brings about change in the agent itself). The latter is a quantum system of some finite dimension d. A quantum measurement consists first in the agent taking an *action* on the quantum system. The action is represented formally by a set of operators $\{E_i\}$—a positive-operator–valued measure. The action generally leads to an incompletely predictable *consequence* E_i for the agent. The quantum state $|\psi\rangle$ makes no appearance but in the agent's head; for it captures his degrees of belief concerning the consequences of his actions, and, in contrast to the quantum system itself, has no existence in the external world. Measurement devices are depicted as prosthetic hands to make it clear that they should be considered an integral part of the agent. The sparks between the measurement-device hand and the quantum system represent the idea that the consequence of each quantum measurement is a unique creation within the previously existing universe. Two points are decisive in distinguishing this picture of quantum measurement from a kind of solipsism: 1) The conceptual split of agent and external quantum system: If it were not needed, it would not have been made. 2) Once the agent chooses an action $\{E_i\}$ to take, the particular consequence E_k of it is beyond his control—that is, the actual outcome is not a product of his whim and fancy.

state of the system itself for Wigner would be gotten from this larger state by a partial trace operation; in any case, it will not be an $|i\rangle$.

Does this make Wigner's new state assignment incorrect? After all, "if he had all the information" (*i.e.*, all the facts of the world) wouldn't that include knowing the friend's measurement outcome? Since the friend should assign some $|i\rangle$, shouldn't Wigner himself (if he had all the information)? Or is it the friend who is incorrect? For if the friend had "all the information," wouldn't he say that he is neglecting that Wigner could

put the system and himself into the quantum computational equivalent of an iron lung and forcefully reverse the so-called measurement? *I.e.*, Wigner, if he were sufficiently sophisticated, should be able to force

$$U(\rho \otimes |\psi\rangle\langle\psi|)U^\dagger \longrightarrow \rho \otimes |\psi\rangle\langle\psi|. \tag{5}$$

And so the back and forth goes. Who has the *right* state of information? The conundrums simply get too heavy if one tries to hold to an agent-independent notion of correctness for otherwise personalistic quantum states. A QBist dispels these and similar difficulties by being conscientiously forthright. *Whose information?* "Mine!" *Information about what?* "The consequences (for *me*) of *my* actions upon the physical system!" It's all "I-I-me-me mine," as the Beatles sang.

The answer to the first question surely comes as no surprise by now, but why on earth the answer for the second? Why something so egocentric, anthropocentric, psychology-laden, and positivistic (we've heard any number of expletives) as *the consequences (for me) of my actions upon the system*? Why not simply say something neutral like "the outcomes of measurements"? Or, why not fall in line with Wolfgang Pauli and say [38]:

> The objectivity of physics is [...] fully ensured in quantum mechanics in the following sense. Although in principle, according to the theory, it is in general only the statistics of series of experiments that is determined by laws, the observer is unable, even in the unpredictable single case, to influence the result of his observation—as for example the response of a counter at a particular instant of time. *Further, personal qualities of the observer do not come into the theory in any way—the observation can be made by objective registering apparatus, the results of which are objectively available for anyone's inspection.* [Our emphasis.]

To the uninitiated, our answer for *Information about what?* surely appears to be a cowardly, unnecessary retreat from realism. But it is the opposite. The answer we give is the very injunction that keeps the potentially conflicting statements of Wigner and his friend in check, at the same time as giving each agent a hook to the external world in spite of QBism's egocentric quantum states. Pauli's statement certainly wouldn't have done that. Results objectively available for anyone's inspection? This is the whole issue with "Wigner's friend" in the first place. If both agents could just "look" at the counter simultaneously with negligible effect in principle, we would not be having this discussion.

You see, for the QBist, the real world, the one both agents are embedded in—with its objects and events—is taken for granted. What is not taken for granted is each agent's access to the parts of it he has not touched. Wigner holds two thoughts in his head: 1) that his friend interacted with a quantum system, eliciting some consequence of the interaction for himself, and 2) after the specified time, for any of Wigner's own further interactions with his friend or system or both, he ought to gamble upon their consequences according to $U(\rho\otimes|\psi\rangle\langle\psi|)U^\dagger$. One statement refers to the friend's potential experiences, and one refers to Wigner's own. So long as it is explicit that $U(\rho\otimes|\psi\rangle\langle\psi|)U^\dagger$

refers to the latter—*i.e.*, how Wigner should gamble upon the things that might happen to him—making no statement whatsoever about the former, there is no conflict. The world is filled with all the same things it was before quantum theory came along, like each of our experiences, that rock and that tree, and all the other things under the sun; it is just that quantum theory provides a calculus for gambling on each agent's own experiences—it doesn't give anything else than that. It certainly doesn't give one agent the ability to conceptually pierce the other agent's personal experience. It is true that with enough effort Wigner could enact eq. (5), causing him to predict that his friend will have amnesia to any future questions on his old measurement results. But we always knew Wigner could do that—a mallet to the head would have been good enough.

The key point is that quantum theory, from this light, takes nothing away from the usual world of common experience we already know. It only *adds*([11]). At the very least it gives each agent an extra tool with which to navigate the world. More than that, the tool is here for a reason. QBism says that when an agent reaches out and touches a quantum system—when he performs a *quantum measurement*—this process gives rise to birth in a nearly literal sense. With the action of the agent upon the system, the no-go theorems of Bell and Kochen–Specker assert that something new comes into the world that wasn't there previously: It is the "outcome," the unpredictable consequence for the very agent who took the action. John Archibald Wheeler said it this way, and we follow suit, "Each elementary quantum phenomenon is an elementary act of 'fact creation'." [39].

With this much, QBism has a story to tell on both quantum *states* and quantum *measurements*, but what of quantum *theory* as a whole? The answer is found in taking it as a *universal* single-user theory in much the same way that Bayesian probability theory itself is. It is a user's manual that *any* agent can pick up and use to help make wiser decisions in this world of inherent uncertainty([12]). To say it in a more poignant way: In my case, it is a world in which *I* am forced to be uncertain about the consequences of most of *my* actions; and in your case, it is a world in which *you* are forced to be uncertain about the consequences of most of *your* actions. "And what of God's case? What is it for him?" Trying to give *him* a quantum state was what caused this trouble in the first place! In a quantum mechanics with the understanding that each instance of its

([11]) This point will be much elaborated on in sect. **9**.

([12]) Most of the time one sees Bayesian probabilities characterized (even by very prominent Bayesians like Edwin T. Jaynes [40]) as measures of ignorance or imperfect knowledge. But that description carries with it a metaphysical commitment that is not at all necessary for the personalist Bayesian, where probability theory is an extension of logic. Imperfect knowledge? It sounds like something that, at least in imagination, could be perfected, making all probabilities zero or one—one uses probabilities only because one does not know the true, pre-existing state of affairs. Language like this, the reader will notice, is never used in this paper. All that matters for a personalist Bayesian is that there is *uncertainty* for whatever reason. There might be uncertainty because there is ignorance of a true state of affairs, but there might be uncertainty because the world itself does not yet know what it will give—*i.e.*, there is an objective indeterminism. As will be argued in later sections, QBism finds its happiest spot in an unflinching combination of "subjective probability" with "objective indeterminism."

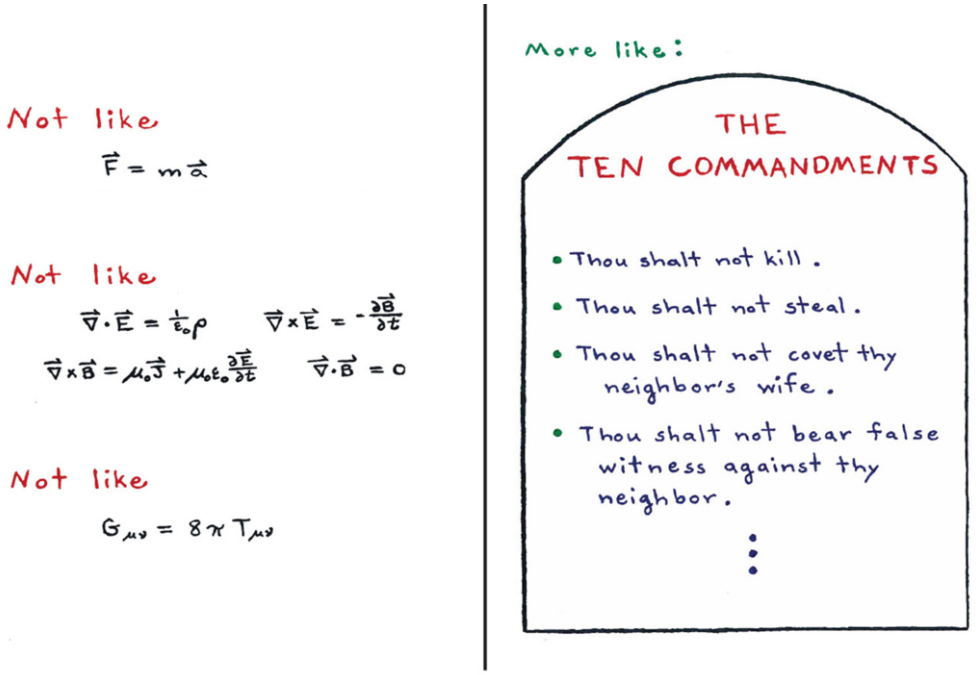


Fig. 3. – The Born rule is not like the other classic laws of physics. Its normative nature means, if anything, it is more like the Biblical Ten Commandments. The classic laws on the left give no choice in their statement: If a field is going to be an electromagnetic field at all, it must satisfy Maxwell's equations; it has no choice. Similarly for the other classic laws. Their statements are intended to be statements concerning nature *just exactly as it is*. But think of the Ten Commandments. "Thou shalt not steal." People steal all the time. The role of the Commandment is to say, "You have the power to steal if you think you can get away with it, but it's probably not in your best interest to do so. Something bad is likely to happen as a result." Similarly for "Thou shalt not kill," and all the rest. It is the worshipper's choice to obey each or not, but if he does not, he ought to count on something potentially bad in return. "And I commend enjoyment," urges Ecclesiastes, "for man has nothing better under the sun than to eat, to drink and to be merry." This is a guideline for behavior—but one conditioned on, and justified in terms of, the character of the natural world. The Born rule guides, "Gamble in such a way that all your probabilities mesh together through me." The agent is free to ignore the advice, but if he does so, he does so at his own peril. Yet, as with the advice of Ecclesiastes, the specifics of the rule can tell us about the character of the world we inhabit.

use is strictly single-user—"My measurement outcomes happen right here, to me, and I am talking about my uncertainty of them."—there is no room for most of the standard, year-after-year quantum mysteries.

The only substantive *conceptual* issue left([13]) before synthesizing a final vaccine is

([13]) Not to worry, there are still plenty of technical ones, as well as plenty more conceptual ones waiting for after the vaccination.

whether quantum mechanics is obligated to derive the notion of agent for whose aid the theory was built in the first place. The answer comes from turning the tables: Thinking of probability theory in the personalist Bayesian way, as an extension of formal logic, would one ever imagine that the notion of an agent, the user of the theory, could be derived out of its conceptual apparatus? Clearly not. How could you possibly get flesh and bones out of a calculus for making wise decisions? The logician and the logic he uses are two different substances—they live in conceptual categories worlds apart. One is in the stuff of the physical world, and one is somewhere nearer to Plato's heaven of ideal forms. Look as one might in a probability textbook for the ingredients to reconstruct the reader herself, one will never find them. So too, the QBist says of quantum theory.

What counts as a "user" of quantum theory? Must the user be conscious? We find that an inopportune way of phrasing things, for it takes the issue too far afield. Instead, we prefer to say it is whatever it takes to be a user of probability theory. Dogs don't collapse wave functions because dogs don't use wave functions. Upon reading this argument, one correspondent immediately protested: "But ants already use probabilities, it has been shown. For the paths they take, one can model the trajectories with an appropriate choice of probabilities and utilities."

To which the QBist involved replied, "No, that's not what I mean. And that is no proof whatsoever that ants use probability theory in the sense I mean it. To use probability theory, I mean one must use it internally, and in a *normative* sense." Probability assignments spring from an attempt to organize one's previous experience for the purpose of future actions. Ants are surely not using it normatively. Modeling agents from the outside (at least in the discussions we've seen so far) never takes into account the normative struggle that is required for any but the most trivial probability assignments.

With this we finally pin down the precise way in which quantum theory is "different in character from any physical theory posed before." For the QBist, quantum theory is not something *outside* probability theory—it is not a picture of the world as it is, as say Einstein's program of a unified field theory hoped to be—but rather it is an *addition* to probability theory itself. As probability theory is a *normative* theory, not saying what one *must* believe, but offering rules of consistency an agent should strive to satisfy within his overall mesh of beliefs, so it is the case with quantum theory. If quantum theory is a user's manual, one cannot forget that the world is its author. And from its writing style, one may still be able to tell something of the author herself. The question is how to tease out the psychology of the style, frame it, and identify the underlying motif.

To take this idea into one's mindset is all the vaccination one needs against the threat that quantum theory carries something viral for theoretical physics as a whole. A healthy body is made healthier still. With this protection, we are for the first time in a position to ask, with eyes wide open to what the answer could not be, *just what after all is the world made of?* Far from being the last word on quantum theory, QBism, we believe, is the start of a great adventure. An adventure full of mystery and danger, with hopes of triumph ... and all the marks of life.

3. – Teleportation

"Teleportation" in the quantum information sense isn't so very much like the Star Trek version as the press always wants to portray it. It's not about getting things from here to there without going in between, but about making your information stop being about *this* and start being about *that* without being about *anything else* in between. In slightly stuffier language: Quantum teleportation is the transference of an agent's predictions about one object onto another object that has never interacted with the first.

In the usual way the teleportation drama is staged, the cast of characters includes an Alice and a Bob who share two systems in a maximally entangled state, and implicitly, a Charlie who prepares a third system in the state of his choice and then hands it off to Alice. Alice then performs a measurement on the two systems in her possession and announces the result of the measurement to Bob. The teleportation process is completed with Bob performing an operation on his system conditioned upon his newly acquired information.

In what sense is it completed? Only in this: If Charlie has the promise that Alice and Bob went through all the actions described above, then he can safely ascribe the same quantum state to Bob's system that he had originally ascribed to the system he handed off to Alice.

The way in which the story is typically told leads one to ask, "How is so much information transmitted?" and "Just how does the information get from Alice to Bob?" The honest answer is that *no* information is transmitted in the process of teleportation (excepting the two bits that tell Bob which action to perform). The only nontrivial thing transferred in the process of teleportation is *reference*. Charlie's information—that is, his compendium of Bayesian degrees of belief—stops being *about* the qubit he just handed off to Alice and starts being *about* Bob's.

Here's a corresponding classical example. In place of entanglement, let us equip Alice and Bob each with a coin (oriented heads or tails) encased in a magical opaque box. These magical opaque boxes have the following properties: 1) though one cannot see how a coin is oriented within it, one can nevertheless reach inside a box and turn the coin over if one wishes, and 2) if one touches two of these boxes together, they will glow green if the coins within them have the same orientations, and they will glow red if they have opposite orientations—the glowing reveals nothing about the actual orientation of either coin, only about their relationship. Finally let us stipulate the following for Alice and Bob: That their opaque boxes contain identically oriented coins, but Alice and Bob (or anyone else for that matter) know nothing more about the coins beyond that. In other words, Alice and Bob possess HH or TT, but they do not know which.

Now, as in quantum teleportation let us introduce a third character, Charlie. Charlie has an opaque box of his own. But let us give him some partial certainty about the orientation of his coin. Particularly, let us suppose he ascribes a probability p for his coin to be heads. This is a real number between 0 and 1, and in principle it might take an arbitrarily huge number of bits to specify.

Here's the protocol. Charlie hands off his coin (encased in a magical opaque box) to Alice. Alice touches her newly acquired box to her old box. The two glow red or green, and she communicates the result to Bob. If the result was green, Bob leaves his coin alone. If the result was red, he reaches into his opaque box and turns the coin over. Meanwhile, Alice randomizes the coins in her possession, *i.e.*, she shuffles them so that Charlie no longer knows which is which. Thus from Charlie's perspective, he now knows nothing about the coin in the original box he gave Alice, and he would write down a 50–50 distribution for heads versus tails. Charlie's original state for his original coin is, in this way, "destroyed". At that point the "teleportation" process is completed.

Again we can ask, "In what sense is it completed?" Only in this: If Charlie has the promise that Alice and Bob went through all the actions described above, then he can safely ascribe the same probability p to the coin in Bob's box (*i.e.*, p that it will be heads) that he had originally ascribed to the coin in his own box. In other words, Charlie has everything it takes to update his *epistemic state* about the orientation of the coin in Bob's box to what he had originally thought of the coin in his own box.

Is this wildly exciting? The stuff that would make headlines in papers all around the world and be called "teleportation"? At the material cost of transferring a single bit from Alice to Bob, has Charlie instantaneously transferred an arbitrarily stupendously big number of bits (in the form of the real number p) between the two sites? Not at all! The only thing that was materially transported from one site to the other was a single bit (that the boxes glowed red or green). The rest was just "conditionalizing" or "updating". And there is no shocking headline in that.

4. – The meaning of no-cloning

It is often underappreciated that taking a stand on the interpretation of quantum states carries with it a fairly distinctive force on one's research. For, the stand one takes implicitly directs the analogies (and disanalogies) one will seek for comparing classical physics to quantum physics. At least this is the case for an information theoretic conception of quantum states.

In 1995, one of the authors (CAF) was quite taken with a point that both Asher Peres and Michael Berry emphasized in their discussions of quantum chaos and more broadly [41]. In making a comparison between quantum mechanics and classical Hamiltonian mechanics, the proper correspondence is not between quantum states and points in phase space, but between quantum states and *Liouville distributions* on the phase space. The key insight is that the points of phase space are meant to represent states of reality, whereas the Liouville distributions are rather explicitly meant to represent one's uncertainty about the true state of reality. In an information theoretic conception of quantum states, a quantum state too should not be a state of reality, but uncertainty about something (maybe not uncertainty of the true state of reality, but nonetheless uncertainty about *something*).

If the analogy worked once, then it should be tested further afield! Furthermore, maybe one could even take the insight the other way around, from quantum to classical.

This prompted the thought that despite all the hoopla over the no-cloning theorem, there was nothing particularly quantum mechanical about it. Classical Liouville evolution preserves phase space volume—this any graduate student versed in Goldstein's classical-mechanics book [42] knows—but a less emphasized consequence of it is that Hamiltonian evolution must preserve the overlap between Liouville distributions as well (as Peres and Berry had been stressing in their quantum chaos work). But then the same argument that drives the proof of the no-cloning theorem for quantum states would also drive it for classical Liouville distributions—for the no-cloning theorem is a nearly immediate consequence of the fact that unitary evolutions preserve Hilbert-space inner products. So, nonorthogonal quantum states cannot be cloned, but neither can nonorthogonal classical Liouville distributions [43].

A historical aside: The issue of no-cloning boils down to an almost immediate consequence of unitarity—inner products cannot decrease. In fact, Wigner's theorem on symmetries [44] even shows that the group of time-continuous, inner-product preserving maps on Hilbert space is strictly equivalent to the unitary group. Therefore, it is an intriguing historical fact that Wigner himself just missed the no-cloning theorem! In a 1961 paper, Wigner took on the question, "How probable is life?" [45]. He did this by identifying the issue of self-reproduction with the existence of the types of maps required for the cloning of quantum states. He didn't tackle the question of cloning for a completely *unknown* quantum state head on, but instead analysed the "fraction" of unitary operators on a tensor-product Hilbert space that can lead to a cloning transformation for at least some states. Nevertheless, he states quite clearly that an arbitrary linear superposition of clonable states ought also to be clonable. But this, of course, cannot be.

The fact that both teleportation and no-cloning arise in classical statistical theories has implications for the project of reconstructing quantum theory. In our search for deep principles, should we try to rederive quantum physics from the postulate that "quantum information can be teleported," or that "quantum information cannot be cloned"? These phenomena being not at all fundamentally quantum makes them feel like poor candidates for the seed from which quantum theory grows. It would be far better to seek that essential DNA in the answer to the question, *Information about what?*

5. – The essence of Bell's theorem, QBism style

It is easy enough to *say* that a quantum system (and hence each piece of the world) is a "seat of possibility." In a spotty way, certain philosophers have been saying similar things for 150 years. What is unique about quantum theory in the history of thought is the way in which its mathematical structure has pushed this upon us to our very surprise. It wasn't that all these grand statements on the philosophical structure of the world were built into the formalism, but that the formalism reached out and shook its users until they opened their eyes. Bell's theorem and all its descendants are examples of that.

So when the users opened their eyes, what did they see? From the look of several recent prominent expositions [46-48], the lesson was indisputably what Tim Maudlin put

so forcefully in [49],

> What Bell proved, and what theoretical physics has not yet properly absorbed, is that the physical world itself is nonlocal.

The world really is full of spooky action at a distance—live with it and love it! But conclusions drawn from even the most rigorous of theorems can only be *additions* to one's prior understanding and beliefs when the theorems do not contradict those beliefs flat out. Such was the case with Bell's theorem. It has just enough room in it to not contradict a misshapen notion of probability, and that is the hook and crook that the sci-fi fans have thrived on. A QBist, however, with a different understanding of probability and a commitment to the idea that quantum measurement outcomes are personal, draws quite a different conclusion from the theorem. In fact it is a conclusion from the far opposite end of the spectrum: It tells of a world unknown to most monist and rationalist philosophies: The universe, far from being one big nonlocal block, should be thought of as a thriving community of connubial, but otherwise autonomous entities. That the world should violate Bell's theorem remains, even for QBism, the deepest statement ever drawn from quantum theory. It says that quantum measurements are *moments of creation*.

This language has already been integral to our presentation, but seeing it come about in a formalism-driven way like Bell's makes the issue particularly vivid. Here we devote some effort to showing that the language of creation is a consequence of three things: 1) the quantum formalism, 2) a personalist Bayesian interpretation of probability, and 3) the elementary notion of what it means to be two objects rather than one. We do not do it however with Bell's theorem precisely, but with an argument that more directly implicates the EPR "criterion of reality" as the source of trouble with quantum theory. The thrust of it is that it is the EPR criterion that should be jettisoned, not locality.

Our starting point is like our previous setup—an agent and a system—but this time we make it two systems: One of them, the left-hand one, is ready. The other, the right-hand one, is waiting. The agent will eventually measure each in turn([14]). Simple enough to say, but things get hung at the start with the issue of what is meant by "two systems?" A passage from a 1948 paper of Einstein [50] captures the essential issue well:

> If one asks what is characteristic of the realm of physical ideas independently of the quantum-theory, then above all the following attracts our attention: the concepts of physics refer to a real external world, *i.e.*, ideas are posited of things that claim a "real existence" independent of the perceiving subject (bodies, fields, etc.), and these ideas are, on the one hand, brought into as secure a relationship as possible with sense impressions. Moreover, it is characteristic of these physical things that they are conceived of as being arranged in a space-time continuum. Further, it appears to be essential for

([14]) It should be noted how we depart from the usual presentation here: There is only the single agent and his two systems. There is no Alice and Bob accompanying the two systems.

> this arrangement of the things introduced in physics that, at a specific time, these things claim an existence independent of one another, insofar as these things "lie in different parts of space." Without such an assumption of the mutually independent existence (the "being-thus") of spatially distant things, an assumption which originates in everyday thought, physical thought in the sense familiar to us would not be possible. Nor does one see how physical laws could be formulated and tested without such a clean separation. [...]
>
> For the relative independence of spatially distant things (A and B), this idea is characteristic: an external influence on A has no *immediate* effect on B; this is known as the "principle of local action," [...]. The complete suspension of this basic principle would make impossible the idea of (quasi-) closed systems and, thereby, the establishment of empirically testable laws in the sense familiar to us.

We hope it is clear to the reader by now that QBism concurs with every bit of this. Quantum states may not be the stuff of the world, but QBists never shudder from positing quantum systems as "real existences" external to the agent. And just as the agent has learned from long, hard experience that he cannot reach out and touch anything but his immediate surroundings, so he imagines of every quantum system, one to the other. What is it that A and B are spatially distant things but that they are causally independent?

This notion, in Einstein's hands([15]), led to one of the nicest, most direct arguments that quantum states cannot be states of reality, but must be something *more like* states of information, knowledge, expectation, or belief [52]. The argument is important—let us repeat the whole thing from Einstein's most thorough version of it [53]. It more than anything sets the stage for a QBist development of a Bell-style contradiction.

> Physics is an attempt conceptually to grasp reality as it is thought independently of its being observed. In this sense on speaks of "physical reality." In pre-quantum physics there was no doubt as to how this was to be understood. In Newton's theory reality was determined by a material point in space and time; in Maxwell's theory, by the field in space and time. In quantum mechanics it is not so easily seen. If one asks: does a ψ-function of the quantum theory represent a real factual situation in the same sense in which this is the case of a material system of points or of an electromagnetic field, one hesitates to reply with a simple "yes" or "no"; why? What the ψ-function (at a definite) time asserts, is this: What is the probability for finding a definite physical magnitude q (or p) in a definitely given interval, if I measure it at time t? The probability is here to be viewed as an empirically determinable, therefore certainly as a "real" quantity which I may determine

([15]) Beware! This is *not* to say in the hands of EPR—Einstein, Podolsky, and Rosen. The present argument is not their argument. For a discussion of Einstein's dissatisfaction with the one appearing in the EPR paper itself, see [51].

if I create the same ψ-function very often and perform a q-measurement each time. But what about the single measured value of q? Did the respective individual system have this q-value even before this measurement? To this question there is no definite answer within the framework of the theory, since the measurement is a process which implies a finite disturbance of the system from the outside; it would therefore be thinkable that the system obtains a definite numerical value for q (or p) the measured numerical value, only through the measurement itself. For the further discussion I shall assume two physicists A and B, who represent a different conception with reference to the real situation as described by the ψ-function.

A. The individual system (before the measurement) has a definite value of q (or p) for all variables of the system, and more specifically, *that* value which is determined by a measurement of this variable. Proceeding from this conception, he will state: The ψ-function is no exhaustive description of the real situation of the system but an incomplete description; it expresses only what we know on the basis of former measurements concerning the system.

B. The individual system (before the measurement) has no definite value of q (or p). The value of the measurement only arises in cooperation with the unique probability which is given to it in view of the ψ-function only through the act of measurement itself. Proceeding from this conception, he will (or, at least, he may) state: The ψ-function is an exhaustive description of the real situation of the system.

We now present to these two physicists the following instance: There is to be a system which at the time t of our observation consists of two partial systems S_1 and S_2, which at this time are spatially separated and (in the sense of classical physics) are without significant reciprocity. The total system is to be completely described through a known ψ-function ψ_{12} in the sense of quantum mechanics. All quantum theoreticians now agree upon the following: If I make a complete measurement of S_1, I get from the results of the measurement and from ψ_{12} an entirely definite ψ-function ψ_2 of the system S_2. The character of ψ_2 then depends upon *what kind* of measurement I undertake on S_1.

Now it appears to me that one may speak of the real factual situation of the partial system S_2. Of this real factual situation, we know to begin with, before the measurement of S_1, even less than we know of a system described by the ψ-function. But on one supposition we should, in my opinion, absolutely hold fast: The real factual situation of the system S_2 is independent of what is done with the system S_1, which is spatially separated from the former. According to the type of measurement which I make of S_1, I get, however, a very different ψ_2 for the second partial system. Now, however, the real situation of S_2 must be independent of what happens to S_1. For the same

> real situation of S_2 it is possible therefore to find, according to one's choice, different types of ψ-function. ...
>
> If now the physicists, A and B, accept this consideration as valid, then B will have to give up his position that the ψ-function constitutes a complete description of a real factual situation. For in this case it would be impossible that two different types of ψ-functions could be coordinated with the identical factual situation of S_2.

Aside from asserting a frequentistic conception of probability, the argument is nearly perfect. It tells us one important reason why we should not be thinking of quantum states as the ψ-ontologists do. Particularly, it is one we should continue to bear in mind as we move to a Bell-type setting: Even there, there is no reason to waiver on its validity. It may be true that Einstein implicitly equated "incomplete description" with "there must exist a hidden-variable account" (though we do not think he did), but the argument as stated neither stands nor falls on this issue.

There is, however, one thing that Einstein does miss in his argument, and this is where the structure of Bell's thinking steps in. Einstein says, "to this question there is no definite answer within the framework of the theory" when speaking of whether quantum measurements are "generative" or simply "revealing" of their outcomes. If we accept everything he said above, then with a little clever combinatorics and geometry one can indeed settle the question.

Let us suppose that the two spatially separated systems in front of the agent are two ququarts (*i.e.*, each system is associated with a four-dimensional Hilbert space $\mathcal{H}_4$), and that the agent ascribes a maximally entangled state to the pair, *i.e.*, a state $|\psi\rangle$ in $\mathcal{H}_4 \otimes \mathcal{H}_4$ of the form,

$$|\psi\rangle = \frac{1}{2}\sum_{i=1}^{4} |i\rangle|i\rangle. \tag{6}$$

Then we know that there exist pairs of measurements, one for each of the separate systems, such that if the outcome of one is known (whatever the outcome), one will thereafter make a probability-one statement concerning the outcome of the other. For instance, if a nondegenerate Hermitian operator H is measured on the left-hand system, then one will thereafter ascribe a probability-one assignment for the appropriate outcome of the transposed operator H^{T} on the right-hand system. What this means for a Bayesian agent is that after performing the first measurement he will bet his life on the outcome of the second.

But how could that be if he has already recognized two systems with no instantaneous causal influence between each other? Mustn't it be that the outcome on the right-hand side is "already there" simply awaiting confirmation or registration? It would seem Einstein's physicist B is already living in a state of contradiction.

Indeed it must be this kind of thinking that led Einstein's collaborators Podolsky and Rosen to their famous sufficient criterion for an "element of [preexistent] reality" [51]:

> If, without in any way disturbing a system, we can predict with certainty (*i.e.*, with probability equal to unity) the value of a physical quantity, then there exists an element of reality corresponding to that quantity.

Without doubt, no personalist Bayesian would ever utter such a notion: Just because he believes something with all his heart and soul and would gamble his life on it, it would not make it necessarily so by the powers of nature—even a probability-one assignment is a state of belief for the personalist Bayesian. But he might still entertain something not unrelated to the EPR criterion of reality. Namely, that believing a particular outcome will be found with certainty on a causally disconnected system entails that one *also* believes the outcome to be "already there" simply awaiting confirmation.

But it is not so, and the QBist has already built this into her story of measurement. Let us show this presently([16]) by combining all the above with a beautifully simple Kochen–Specker style construction discovered by Cabello, Estebaranz, and García-Alcaine (CEGA) [57]. Imagine some measurement H on the left-hand system; we will denote its potential outcomes as a column of letters, like this:

$$\begin{matrix} a \\ b \\ c \\ d \end{matrix} \tag{7}$$

Further, since there is a fixed transformation taking any H on the left-hand system to a corresponding H^{T} on the right-hand one, there is no harm in identifying the notation for the outcomes of both measurements. That is to say, if the agent gets outcome b (to the exclusion of a, c, and d) for H on the left-hand side, he will make a probability-one prediction for b on the right-hand side, even though that measurement strictly speaking is a different one, namely H^{T}. If the agent further subscribes to (our Bayesian variant of) the EPR criterion of reality, he will say that he believes b to be TRUE of the right-hand system as an element of reality.

Now let us consider two possible measurements, H_1 and H_2 for the left-hand side, with potential outcomes

$$\begin{matrix} a \\ b \\ c \\ d \end{matrix} \quad \text{and} \quad \begin{matrix} e \\ f \\ g \\ h \end{matrix} \tag{8}$$

respectively. Both measurements cannot be performed at once, but it might be the case that if the agent gets a specific outcome for H_1, say c particularly, then not only will

([16]) Overall this particular technique has its roots in Stairs [54], and seems to bear some resemblance to the gist of Conway and Kochen's "Free Will Theorem" [55, 56].

he make a probability-one assignment for c in a measurement of H_1^{T} on the right-hand side, but also for e in a measurement of H_2^{T} on it. Similarly, if H_2 were measured on the left, getting an outcome e; then he will make a probability-one prediction for c in a measurement of H_1^{T} on the right. This would come about if H_1 and H_2 (and consequently H_1^{T} and H_2^{T}) share a common eigenvector. Supposing so and that c was actually the outcome for H_1 on the left, what conclusion would the EPR criterion of reality draw? It is that both c and e are elements of reality on the right, and none of a, b, d, f, g, or h are. Particularly, since the right-hand side could not have known whether H_1 or H_2 was measured on the left, whatever c and e stands for, it must be the same thing, the same property. In such a case, we discard the extraneous distinction between c and e in our notation and write

$$\begin{array}{cc} a & \\ b & \\ c & \text{and} \\ d & \end{array} \qquad \begin{array}{c} c \\ f \\ g \\ h \end{array} \tag{9}$$

for the two potential outcome sets for a measurement on the right.

We now have all the notational apparatus we need to have some fun. The genius of CEGA was that they were able to find a set of nine "interlocking" Hermitian operators $H_1, H_2, \ldots, H_9$ for the left, whose set of potential outcomes for the corresponding operators on the right would look like this:

$$\begin{array}{ccccccccc} a & a & h & h & b & i & p & p & q \\ b & e & i & k & e & k & q & r & r \\ c & f & c & g & m & n & d & f & m \\ d & g & j & l & n & o & j & l & o \end{array} \tag{10}$$

Take the second column as an example. It means that if H_2 were measured on the left-hand system, only one of a, e, f, or g would occur—the agent cannot predict which—but if a occurred, he would be absolutely certain of it also occurring in a measurement of H_1^{T} on the right. And if e were to occur on the left, then he would be certain of getting e as well in a measurement of H_5^{T} on the right. And similarly with f and g, with their implications for H_8^{T} and H_4^{T}.

The wonderful thing to note about (10) is that every letter $a, b, c, \ldots, r$ occurs exactly twice in the collection. But the EPR criterion of reality (or our Bayesian variant of it) would require exactly one letter to have the truth value TRUE in each column, with the other three having the value FALSE. In total, nine values of TRUE: A clean contradiction! For if every letter occurs exactly twice in the collection, whatever the total number of TRUE values is, it must be an even number.

To emphasize the point, let us sketch a similar argument for the case where the two halves are both *qutrits*, quantum systems of dimension 3. Again, the argument is a

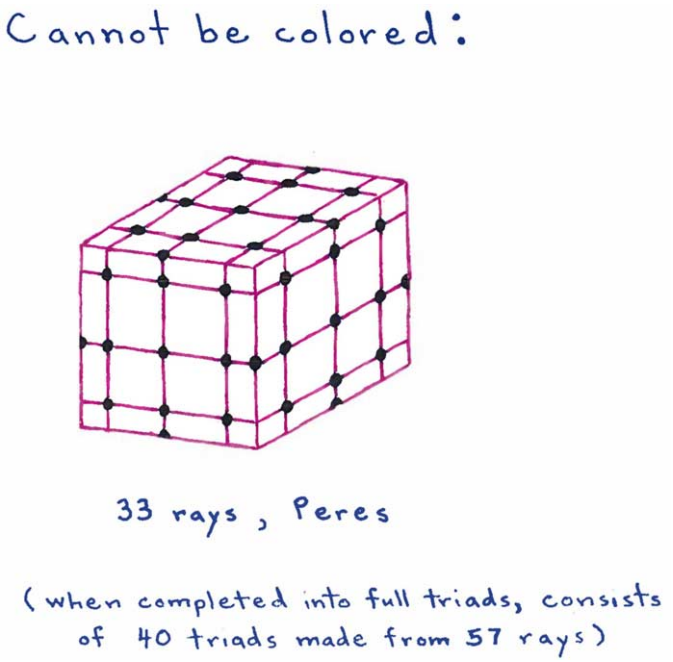


Fig. 4. – Each ray in Asher Peres's qutrit Kochen-Specker construction can be written as a ray in $\mathbb{R}^3$ and illustrated by where those rays intersect a cube.

little variation of the EPR thought experiment. This time, we begin by ascribing to our bipartite system the maximally entangled state

$$|\psi\rangle = \frac{1}{2}\sum_{i=1}^{3} |i\rangle|i\rangle. \tag{11}$$

The experimentalist—let us call her Alice—considers making a measurement on the left-hand particle in some basis. If Alice obtains outcome number 2 on the left side, then she can predict that if she were to make a particular measurement on the right side, she would get outcome number 2. (The bases for the left-hand and right-hand measurements are related by a transpose operation.) So, under the assumption of Einsteinian locality and the EPR criterion, Alice would say, "Aha! It must be the case that there is an element of reality on the right side corresponding to outcome number 2 of that measurement. It's something inherent in that body." But we can play this game with any basis: If Alice were to get outcome i for a measurement on one particle, she would predict with certainty that she would get outcome i for that measurement, transposed, on the other particle.

We're talking about noncommuting variables here; by the EPR criterion and locality, Alice would conclude it must be the case that there were elements of reality associated

with those noncommuting observables. But it is no fun to consider only one basis, or only two. Instead, we proceed to analyze a whole set of them, corresponding to one of the qutrit Kochen-Specker constructions, for instance the one that Asher Peres found [41]. In order to present this construction in a concise way, let us use an abbreviated notation in which a "2" stands for $\sqrt{2}$ and an overline means a minus sign. Thus, "$1\bar{1}2$" stands for the ray $(1, -1, \sqrt{2})^{\mathrm{T}}$. We tabulate the following ten sets of three orthogonal rays apiece:

$$
\begin{array}{ccc}
001 & 100 & 010 \\
101 & \bar{1}01 & 010 \\
011 & 0\bar{1}0 & 100 \\
1\bar{1}2 & \bar{1}12 & 110 \\
102 & \bar{2}01 & 010 \\
211 & 0\bar{1}1 & \bar{2}11 \\
201 & 010 & \bar{1}02 \\
112 & 1\bar{1}0 & \bar{1}\bar{1}2 \\
012 & 100 & 0\bar{2}1 \\
121 & \bar{1}01 & 1\bar{2}1
\end{array}
\tag{12}
$$

The EPR criterion tells us that for each basis, there exists an "element of physical reality" that determines what the outcome of a measurement in that basis will be. This applies to each row above: According to the EPR criterion, in any set of three orthogonal rays, one must be marked TRUE and the other two marked FALSE. But if we start labeling our rays, marking each one orthogonal to a TRUE ray with FALSE, eventually we hit a point where we can't do it consistently. We end up marking a set of three orthogonal rays all FALSE, which is against the rules.

For each of these bases, Alice would have said, "By making a measurement here, I draw an inference about the element of reality over there." She can do it for one, she can do it for another, and another ... but she can't do it for all of them without running into trouble!

Something must give. The quick reaction of most of the quantum foundations community has been to question the causal independence of the two systems under consideration. But if one gives up on the autonomy of one system from the other—after very explicitly assuming it—this surely amounts to saying that there were never two systems there after all; the very idea of separate systems is a broken concept. This first raises a minor conundrum: Why then would the quantum formalism engender us to formulate our description from beginning to end in terms of $\mathcal{H}_3 \otimes \mathcal{H}_3$, rather than simply a raw nine-dimensional space $\mathcal{H}_9$? Why is that separating symbol $\otimes$, apparently marking some kind of conceptual distinction, always hanging around?

Reaching much deeper however, if one is willing to throw away one's belief in systems' autonomy from each other, why would one ever believe in one's own autonomy? All

stringent reason for it gets lost, and indeed as Einstein warns, what now is the meaning of science? As Hans Primas once emphasized [58],

> It is a tacit assumption of all engineering sciences that nature can be *manipulated* and that the initial conditions required by experiments can be brought about by interventions of the world external to the object under investigation. That is, *we assume that the experimenter has a certain freedom of action which is not accounted for by first principles of physics.* Without this freedom of choice, experiments would be impossible. Man's free will implies the ability to carry out actions, it constitutes his essence as an actor. We act under the idea of freedom, but the topic under discussion is neither man's sense of personal freedom as a subjective experience, nor the question whether this idea could be an illusion or not, nor any questions of moral philosophy, but that *the framework of experimental science requires the freedom of action as a constitutive though tacit presupposition.*

If the left-hand system can manipulate the right-hand system, *even when by assumption it cannot*, then who is to say that the right-hand system cannot manipulate the agent herself? To put it still differently: If one is never allowed to assume causal independence between separated systems because of a contradiction in the term, then one can never assume it of oneself either, even with respect to the components of the world that one thinks one is manipulating. It would be a wackier world than even the one QBism entertains.

But QBism's world is not such a bad world, and some of us find its openness to possibility immensely exciting. What gives way in this world is not the idea of reality, but simply the narrow-minded EPR criterion for it. We jettison *both* the idea that a probability-one assignment implies there is a pre-existent outcome (property) "over there" waiting to be revealed and, barring that, that it must have been "over here" pre-existent, waiting to be transferred and then revealed. The solution lies closer to one of John Wheeler's quips, "No question? No answer." A probability-one assignment lays no necessary claim on what the world *is*, but what the agent using it believes with all her heart and soul. In the case of our present example, what the agent believes is that if an outcome b *came about* as a result of her action H on the left-hand system, an outcome b *would come about* if she were to perform the action H^{T} on the right-hand system. But if she does not walk over to the right-hand system and take the action, there is no good sense in which the outcome (or property) b is already there. Measurement is not a passive process, but instead a fundamentally *participatory* one.

At the instigation of a quantum measurement, something new comes into the world that was not there before; and that is about as clear an instance of *creation* as one can imagine. Sometimes one will have no strong beliefs for what will result from the creation (as with the measurement of H), and sometimes one will have very strong beliefs (as with the subsequent measurement of H^{T}), but a free creation of nature it remains.

6. – The quantum de Finetti theorem

> You know how men have always hankered after unlawful magic, and you know what a great part in magic *words* have always played. If you have his name, ... you can control the spirit, genie, afrite, or whatever the power may be. Solomon knew the names of all the spirits, and having their names, he held them subject to his will. So the universe has always appeared to the natural mind as a kind of enigma, of which the key must be sought in the shape of some illuminating or power-bringing word or name. That word names the universe's *principle*, and to possess it is after a fashion to possess the universe itself.
>
> But if you follow the pragmatic method, you cannot look on any such word as closing your quest. You must bring out of each word its practical cash-value, set it at work within the stream of your experience. It appears less as a solution, then, than as a program for more work, and more particularly as an indication of the ways in which existing realities may be *changed*.
>
> *Theories thus become instruments, not answers to enigmas, in which we can rest.* We don't lie back upon them, we move forward, and, on occasion, make nature over again by their aid.
>
> — William James

Since the beginning, those who brought Bayesian probability into quantum physics have been on the run proving technical theorems whenever required to close a gap in their logic or negate an awkwardness induced by their new way of speaking. It was never enough to "lie back upon" the pronouncements: They had to be shown to have substance, something that would drive physics itself forward. A case in point is the *quantum de Finetti theorem* [3, 59].

The term "unknown state" is ubiquitous in quantum information: Unknown quantum states are teleported, protected with quantum error correcting codes, used to check for quantum eavesdropping, and arise in innumerable other applications. For a QBist, though, the phrase can only be an oxymoron: If quantum states are compendia of beliefs, and not states of nature, then the state is known to someone, at the very least the agent who holds it. But if so, then what are experimentalists doing when they say they are performing quantum-state tomography in the laboratory? The very goal of the procedure is to characterize the unknown quantum state a piece of laboratory equipment is repetitively preparing. There is certainly no little agent sitting on the inside of the device devilishly sending out quantum systems representative of his beliefs, and smiling

as the experimenter on the outside slowly homes in on those private thoughts through his experiments.

The quantum de Finetti theorem is a result that allows the story of quantum-state tomography to be told purely in terms of a single agent—namely, the experimentalist in the laboratory. In a nutshell, the theorem is this. Suppose the experimentalist walks into the laboratory with the very minimal belief that, of the systems her device is spitting out (no matter how many), she could interchange any two of them and it would not change the statistics she expects for any measurements she might perform. Then the theorem says that "coherence with this belief alone" requires her to make a quantum-state assignment $\rho^{(n)}$ (for any n of those systems) that can be represented in the form:

$$\rho^{(n)} = \int P(\rho)\, \rho^{\otimes n}\, \mathrm{d}\rho, \tag{13}$$

where $P(\rho)\, \mathrm{d}\rho$ is some probability measure on the space of single-system density operators and $\rho^{\otimes n} = \rho \otimes \cdots \otimes \rho$ represents an n-fold tensor product of identical quantum states.

To put it in words, this theorem licenses the experimenter to act *as if* each individual system has some state ρ unknown to her, with a probability density $P(\rho)$ representing her ignorance of which state is the true one. But it is only *as if*—the only active quantum state in the picture is the one the experimenter actually possesses in her mind, namely $\rho^{(n)}$. When the experimenter performs tomography, all she is doing is gathering data system-by-system and updating, via Bayes rule [60], the state $\rho^{(n)}$ to some new state $\rho^{(k)}$ on a smaller number of remaining systems. Particularly, one can prove that this form of quantum-state assignment leads the agent to expect that with more data, she will make her $P(\rho)$ more and more narrow, and thus she will approach ever more closely a posterior state of the form $\rho^{(k)} \approx \rho^{\otimes k}$. This is the real, underlying reason that excuses the habit of speaking of tomography as revealing "the unknown quantum state."

One important consequence of this theorem is the following. Suppose that Alice is collaborating with Bob. From her perspective, Bob is a physical system. Alice mathematically models Bob as having expectations about the sequence of systems they are studying, expectations that matter to Alice because she can ask Bob questions and get answers. In Alice's mental model of Bob, she writes a de Finetti representation $P_B(\rho)$, satisfying the same general properties as her own expectations, which she encodes into the function $P_A(\rho)$. Alice imagines that she and Bob are receiving the same data. It follows from the quantum de Finetti theorem that if $P_A(\rho)$ and $P_B(\rho)$ initially agree to at least a small extent, then Alice should expect that their expectations will come into greater and greater agreement.

Furthermore, just as there is a de Finetti theorem to make sense of "unknown states," there is a de Finetti theorem to make sense of "unknown measurements" and "unknown processes" [61].

Despite the explicitly foundational motivation, the quantum de Finetti theorem has nonetheless fared like a stand-alone result for quantum information theory. Among other things, it turned out to be useful for proving the security of some quantum key distribu-

tion schemes [62-64], it became an important component in the analysis of entanglement detection [65, 66], and even served in an analysis of the quantum state of propagating laser light [67, 68].

During the effort to prove the quantum de Finetti theorem, Carlton Caves brought attention to a special type of quantum measurement. Up to that point, those of us who were bringing Bayesianism to quantum physics were in the habit of regarding a quantum state ρ as a "catalogue of probabilities." This way of thinking rested on the old way of modeling measurements, that is, the von Neumann tradition, where a measurement corresponds to an orthonormal basis. The Born rule lets us get probabilities out of a density matrix ρ, yet no single von Neumann basis can yield sufficient information to reconstruct ρ; hence, the "catalogue of probabilities" language. But quantum information theory brought a new manner of thinking, in which a measurement can be any collection of positive operators that sum to the identity—a *Positive-Operator–Valued Measure*, or POVM. This change of perspective opened the possibility of an *informationally complete* experiment, *i.e.*, a *single* measurement whose statistics suffice to reconstitute an entire density matrix ρ. Once we have constructed an "IC" POVM, we can then replace any ρ with a probability distribution. One matrix ρ, one vector $\vec{p}$.

At first, Caves believed that proving the quantum de Finetti theorem would require a specific *type* of IC POVM. Luckily, this turned out to be wrong: Finding a construction of an IC POVM in arbitrary finite Hilbert-space dimension was enough. However, the particular kind of IC POVM to which Caves called attention soon took on a life of its own, and it is to that kind which we now turn.

7. – Seeking SICs – The Born rule as fundamental

If quantum theory is so closely allied with probability theory, then why is it not written in a language that starts with probability, rather than a language that ends with it? Why does quantum theory invoke the mathematical apparatus of complex amplitudes, Hilbert spaces, and linear operators? This question quickly brings us to the research frontier.

For, actually there are ways to pose quantum theory purely in terms of probabilities—indeed, there are many ways, each with a somewhat different look and feel [69]. The work of W. K. Wootters is an example, and as he emphasized long ago [70],

> It is obviously *possible* to devise a formulation of quantum mechanics without probability amplitudes. One is never forced to use any quantities in one's theory other than the raw results of measurements. However, there is no reason to expect such a formulation to be anything other than extremely ugly. After all, probability amplitudes were invented for a reason. They are not as directly observable as probabilities, but they make the theory simple. I hope to demonstrate here that one *can* construct a reasonably pretty formulation using only probabilities. It may not be quite as simple as the usual formulation, but it is not much more complicated.

What has happened in the intervening years is that the mathematical structures of

quantum information theory have grown significantly richer than the ones he had based his considerations on—so much so that we may now be able to optimally re-express the theory. What was once "not much more complicated," now has the promise of being downright insightful.

The key ingredient is a hypothetical structure called a "symmetric informationally complete positive-operator-valued measure," or SIC (pronounced "seek") for short. This is a set of d^2 rank-one projection operators $\Pi_i = |\psi_i\rangle\langle\psi_i|$ on a finite d-dimensional Hilbert space such that

$$|\langle\psi_i|\psi_j\rangle|^2 = \frac{1}{d+1} \quad \text{whenever } i \neq j. \tag{14}$$

Because of their extreme symmetry, it turns out that such sets of operators, when they exist, have three very fine-tuned properties: 1) the operators must be linearly independent and span the space of Hermitian operators, 2) there is a sense in which they come as close to an orthonormal basis for operator space as they can under the constraint that all the elements in a basis be positive semi-definite [71], and 3) after rescaling, they form a resolution of the identity operator, $I = \sum_i \frac{1}{d}\Pi_i$.

The symmetry, positive semi-definiteness, and properties 1) and 2) are significant because they imply that an arbitrary quantum state ρ—pure or mixed—can be expressed as a linear combination of the Π_i. Furthermore, the expansion is likely to have some significant features not found in other, more arbitrary expansions. The most significant of these becomes apparent when one takes property 3 into account. Because the operators $H_i = \frac{1}{d}\Pi_i$ are positive semi-definite and form a resolution of the identity, they can be interpreted as labeling the outcomes of a quantum measurement device—not a standard-textbook, von Neumann measurement device whose outcomes correspond to the eigenvalues of some Hermitian operator, but to a measurement device of the most general variety allowed by quantum theory, the POVMs [36,41]. Particularly noteworthy is the smooth relation between the probabilities $P(H_i) = \mathrm{tr}(\rho H_i)$ given by the Born rule for the outcomes of such a measurement([17]) and the expansion coefficients for ρ in terms of the Π_i:

$$\rho = \sum_{i=1}^{d^2} \left((d+1)\, P(H_i) - \frac{1}{d} \right) \Pi_i. \tag{15}$$

There are no other operator bases that give rise to such a simple formula connecting probabilities with density operators.

Before getting to that, however, we should reveal what is so consternating about the SICs: It is the question of when they exist. Despite years of growing effort since the

([17]) There is a slight ambiguity in notation here, as H_i is dually used to denote an operator and an outcome of a measurement. For the sake of simplicity, we hope the reader will forgive this and similar abuses.

definition was first introduced [72-74], no one has been able to show that they exist in completely general dimension. All that is known firmly is that they exist in dimensions 2 through 144 inclusive, as well as in some scattered cases beyond those: 147, 168, 172, 195, 199, 228, 259 and 323. Dimensions 2–21, 24, 28, 30, 31, 35, 37, 39, 43 and 48 are known through direct or computer-assisted analytic proof; the remaining solutions are known through numerical calculation, satisfying eq. (14) to high precision (in some cases, up to 16000 digits accuracy)([18]). For the remainder of the article we will proceed as if they do always exist for finite d. At least this is the conceit of our story. We note in passing, however, that the SIC existence problem is not without wider context: If they do exist, they solve at least four other (more practical, non-foundational) optimality problems in quantum information theory [71, 76-79]. In addition, the results they imply for Lie and Jordan algebras indicate that if SICs didn't always exist, linear algebra itself would have a drastically different character from one dimension to another [80]. It would be a nasty trick if SICs failed to exist!

So suppose they do. Thinking of a quantum state as *literally* an agent's probability assignment for the outcomes of a potential SIC measurement leads to a new way to express the Born rule for the probabilities associated with any *other* quantum measurement. Consider the diagram in fig. 4. It depicts a SIC measurement "in the sky," with outcomes H_i, and any standard von Neumann measurement "on the ground"([19]). For the sake of specificity, let us say the latter has outcomes $D_j = |j\rangle\langle j|$, the vectors $|j\rangle$ representing some orthonormal basis. We conceive of two possibilities (or two "paths") for a given quantum system to get to the measurement on the ground: "Path 1" is that it proceeds directly to the measurement on the ground. "Path 2" is that it proceeds first to the measurement in the sky and only subsequently to the measurement on the ground—the two measurements are cascaded.

Suppose now, we are given the agent's personal probabilities $P(H_i)$ for the outcomes in the sky and his personal conditional probabilities $P(D_j|H_i)$ for the outcomes on the ground subsequent to the sky. *I.e.*, we are given the probabilities the agent would assign on the supposition that the quantum system follows Path 2. Then "coherence alone" (in the Bayesian sense) is enough to tell what probabilities $P(D_j)$ the agent should assign for the outcomes of the measurement on the ground—it is given by the Law of Total Probability applied to these numbers:

$$P(D_j) = \sum_i P(H_i)P(D_j|H_i). \tag{16}$$

([18]) Numerical and some analytical solutions up through dimension 67 are detailed in [75]. Others were reported to us by Andrew Scott and Marcus Appleby. Numerical solutions in dimensions 122 through 142 and in dimension 144 are due to Michael C. Hoang, in collaboration with the authors, using the Chimera supercomputer.

([19]) Do not, however, let the designation "SIC sitting in the sky" make the device seem too exalted to be of any interest. Already, announcements of experimental implementations have been made for qubits [81], qutrits [82], and still higher-dimensional systems [83].

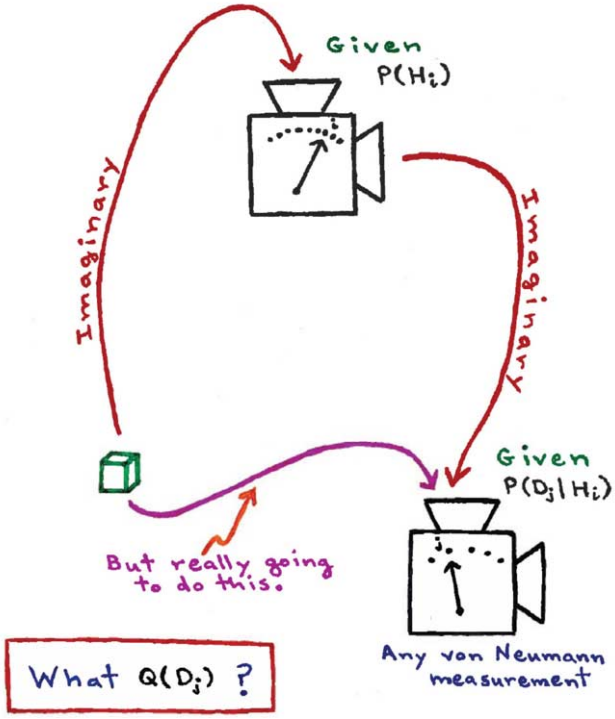


Fig. 5. – Any quantum measurement can be conceptualized in two ways. Suppose an arbitrary von Neumann measurement "on the ground," with outcomes $D_j = 1, \ldots, d$. Its probabilities $P(D_j)$ can be derived by cascading it with a fixed fiducial SIC measurement "in the sky" (of outcomes $H_i = 1, \ldots, d^2$). Let $P(H_i)$ and $P(D_j|H_i)$ represent an agent's probabilities, assuming the measurement in the sky is actually performed. The probability $Q(D_j)$ represents instead the agent's probabilities under the assumption that the measurement in the sky is *not* performed. The Born rule, in this language, says that $P(D_j)$, $P(H_i)$, and $P(D_j|H_i)$ are related by the Bayesian-style eq. (19).

That takes care of Path 2, but what of Path 1? Is this enough information to recover the probability assignment $Q(D_j)$ the agent would assign for the outcomes on Path 1 via a normal application of the Born rule? That is, that

$$Q(D_j) = \mathrm{tr}(\rho D_j) \tag{17}$$

for some quantum state ρ? Maybe, but the answer will clearly not be $P(D_j)$. One has

$$Q(D_j) \neq P(D_j) \tag{18}$$

simply because Path 2 is *not* a coherent process (in the quantum sense!) with respect to Path 1—there is a measurement that takes place in Path 2 that does not take place in Path 1.

What is remarkable about the SIC representation is that it implies that, even though $Q(D_j)$ is not equal to $P(D_j)$, it is still a function of it. Particularly,

$$Q(D_j) = (d+1)P(D_j) - 1 \\ = (d+1)\sum_{i=1}^{d^2} P(H_i)P(D_j|H_i) - 1. \tag{19}$$

The Born rule is nothing but a kind of Quantum Law of Total Probability! No complex amplitudes, no operators—only probabilities in, and probabilities out. Indeed, it is seemingly just a rescaling of the old law, eq. (16). And in a way it is.

Earlier, we stressed the importance of considering quantum measurements in all their generality: The notion of "measurement" should include all POVMs, not just the von Neumann ones. So, what happens if the measurement on the ground is an arbitrary POVM, not necessarily given by an orthonormal basis? Then,

$$Q(D_j) = \sum_{i=1}^{d^2} \left[(d+1)P(H_i) - \frac{1}{d}\right] P(D_j|H_i). \tag{20}$$

This is only slightly more intricate than our previous expression, eq. (19), which is itself a special case of this more general formula. And the general formula is still quite similar to the classical prescription for combining conditional probabilities: We simply stretch the $P(H_i)$ and then shift them to preserve the overall normalization.

But beware: One should not interpret eq. (20) as invalidating probability theory itself in any way! For the old Law of Total Probability has no jurisdiction in the setting of our diagram, which compares a two distinct, mutually exclusive hypothetical scenarios. Path 1 is what Alice intends to do, but she recognizes that she could in principle follow Path 2 instead, and eq. (20) sets the standard of consistency to which Alice should strive when meshing her probabilities together([20]). Indeed as any Bayesian would emphasize, if there is a distinguishing mark in one's considerations—say, the fact of two distinct experiments, not one—then one ought to take that into account in one's probability assignments (at least initially so). Thus there is a hidden, or at least suppressed, condition in our notation: Really we should have been writing the more cumbersome, but honest, expressions $P(H_i|\mathcal{E}_2)$, $P(D_j|H_i,\mathcal{E}_2)$, $P(D_j|\mathcal{E}_2)$, and $Q(D_j|\mathcal{E}_1)$ all along. With this

([20]) This is one place where we can point out a mild historical antecedent to QBism. In the historical study [84], it is pointed out that Born and Heisenberg, *already at the 1927 Solvay conference*, refer to the calculation $|c_n(t)|^2 = |\sum_m S_{mn}(t)c_m(0)|^2$ and say, "it should be noted that this "interference" does not represent a contradiction with the rules of the probability calculus, that is, with the assumption that the $|S_{nk}|^2$ are quite usual probabilities." Their reasons for saying this may have been different from our own, but at least they had come this far.

explicit, it is no surprise that,

$$Q(D_j|\mathcal{E}_1) \neq \sum_i P(H_i|\mathcal{E}_2)P(D_j|H_i,\mathcal{E}_2). \tag{21}$$

The message is that quantum theory supplies something—a new form of "Bayesian coherence," though empirically based (as quantum theory itself is)—that raw probability theory does not. The Born rule in these lights is an addition to Bayesian probability, not in the sense of a supplier of some kind of more-objective probabilities, but in the sense of giving extra normative rules to guide the agent's behavior when she interacts with the physical world.

It is a normative rule for reasoning about the consequences of one's proposed actions in terms of the potential consequences of an alternative action. It is like nothing else physical theory has contemplated before. Seemingly at the heart of quantum mechanics from the QBist view is a statement about the impact of hypotheticals on our expectations for the actual. The impact parameter is metered by a single, significant number associated with each physical system—its Hilbert-space dimension d. The larger the d associated with a system, the more $Q(D_j)$ must deviate from $P(D_j)$. Of course this point must have been implicit in the usual form of the Born rule, eq. (17). What is important from the QBist perspective, however, is how the new form puts the significant parameter front and center, displaying it in a way that one ought to nearly trip over.

Understanding this as the goal helps pinpoint the role of SICs in our considerations. The issue is not that quantum mechanics *must* be rewritten in terms of SICs, but that it *can* be([21]). Certainly no one is going to drop the usual operator formalism and all the standard methods learned in graduate school to do their workaday calculations in SIC language exclusively. It is only that the SICs form an ideal coordinate system for a particular problem (an important one to be sure, but nonetheless a particular one)—the problem of *interpreting* quantum mechanics. The point of all the various representations of quantum mechanics (like the various quasi-probability representations of [69], the Heisenberg and Schrödinger pictures, and even the path-integral formulation) is that they give a means for isolating one or another aspect of the theory that might be called for by a problem at hand. Sometimes it is really important to do so, even for deep conceptual issues and even if all the representations are logically equivalent([22]). In our case, we want to bring into plain view the idea that quantum mechanics is an *addition* to Bayesian probability theory—not a generalization of it [85], not something orthogonal to it altogether [86], but an addition. With this goal in mind, the SIC representation is

([21]) If everything goes right, that is, and the damned things actually exist in all dimensions!

([22]) Just think of the story of Eddington-Finkelstein coordinates in general relativity. Once upon a time it was not known whether a Schwarzschild black hole might have, beside its central singularity, a singularity in the gravitational field at the event horizon. Apparently it was a heated debate, yes or no. The issue was put to rest, however, with the development of the coordinate system. It allowed one to write down a solution to the Einstein equations in a neighborhood of the horizon and check that everything was indeed all right.

a particularly powerful tool. Through it, one sees the Born rule as a replacement for a usage of the Law of Total Probability that one would have made in another context (one mutually exclusive with the first).

Furthermore it is similarly so of unitary time evolution in a SIC picture. To explain what this means, let us change considerations slightly and make the measurement on the ground a unitarily rotated version of the SIC in the sky. In this setting, $D_j = \frac{1}{d} U \Pi_j U^\dagger$, which in turn implies a simplification of eq. (20) to,

$$Q(D_j) = (d+1) \sum_{i=1}^{d^2} P(H_i) P(D_j|H_i) - \frac{1}{d}\,, \tag{22}$$

for the probabilities on the ground. Note what this is saying! As the Born rule is a replacement for the Law of Total Probability, unitary time evolution is a replacement for it as well. For, if we thought in terms of the Schrödinger picture, $P(H_i)$ and $Q(D_j)$ would be the SIC representations for the initial and final quantum states under an evolution given by $U^\dagger$. The similarity is no accident. This is because in both cases the conditional probabilities $P(D_j|H_i)$ completely encode the identity of a measurement on the ground.

Moreover, it makes abundantly clear another point of QBism that has not been addressed so much in the present paper. Since a personalist Bayesian cannot turn his back on the clarification that *all* probabilities are personal judgments, placeholders in a calculus of consistency, he certainly cannot turn his back on the greater lesson eqs. (20) and (22) are trying to scream out. Just as quantum states ρ are personal judgments $P(H_i)$, quantum measurement operators D_j and unitary time evolutions U are personal judgments too—in this case $P(D_j|H_i)$. The only distinction is the technical one, that one expression is an unconditioned probability, while the other is a collection of conditionals. Most importantly, it settles the age-old issue of why there should be two kinds of state evolution at all. When Hartle wrote, "A quantum-mechanical state being a summary of the observers' information about an individual physical system changes both by dynamical laws, and whenever the observer acquires new information about the system through the process of measurement", what is his dynamical law making reference to? There are not two things that a quantum state can do, only one: Strive to be consistent with all the agent's other probabilistic judgments on the consequences of his actions, across all hypothetical scenarios. The SICs emphasize and make this point clear.

In fact, much of the most intense research of the UMass Boston QBism group is currently devoted to seeing how much of the essence of quantum theory is captured by eq. (20). We are frankly quite happy to have an extremely hard problem about the structure of quantum states spaces leading our thinking! (And, as we'll see in the next section, it is a problem that prompts a traveler to question the received wisdom about the boundary between physics and pure mathematics.) For instance, one way to approach this is to take eq. (20) as a fundamental axiom and ask what further assumptions are required to recover all of quantum theory. To give some hint of how a reconstruction of quantum theory might proceed along these lines, note eq. (15) again. What it expresses

is that any quantum state ρ can be reconstructed from the probabilities $P(H_i)$ the state ρ gives rise to. This, however, does not imply that plugging just any probability distribution $P(H_i)$ into the equation will give rise to a valid quantum state. A general probability distribution $P(H_i)$ in the formula will lead to a Hermitian operator of trace one, but it may not lead to an operator with nonnegative eigenvalues. Indeed it takes further restrictions on the $P(H_i)$ to make this true. That being the case, the QBist starts to wonder if these restrictions might arise from the requirement that eq. (20) simply always make sense. For note, if $P(D_j)$ is too small in the special case of eq. (19), $Q(D_j)$ will go negative; and if $P(D_j)$ is too large, $Q(D_j)$ will become larger than 1. So, $P(D_j)$ must be restricted. But that in turn forces the set of valid $P(H_i)$ to be restricted as well. And so the argument goes. We already know how to reconstruct many features of quantum theory in this fashion [5,87-89]. The question now is how to get the whole theory in the most economical way [90]. Should we succeed, we will have a new development of quantum theory, one that puts its beguiling deviation from classicality, as encoded in eq. (20), front and center.

Another exciting development comes from loosening the form of eq. (20) to something more generic:

$$Q(D_j) = \sum_{i=1}^{n} [\alpha P(H_i) - \beta] P(D_j|H_i);, \tag{23}$$

where there is *initially* no assumed relation between α, β, and n as there is in eq. (20). Then, under a few further conditions with only the faintest hint of quantum theory in them—for instance, that there should exist measurements on the ground for which, under appropriate conditions, one can have certainty for their outcomes—one immediately gets a significantly more restricted form for what becomes the analogue of eq. (19):

$$Q(D_j) = \left(\frac{1}{2}qd + 1\right) \sum_{i=1}^{n} P(H_i) P(D_j|H_i) - \frac{1}{2}q. \tag{24}$$

Here, very interestingly, the parameters q and d can only take on integer values, $q = 0, 1, 2, \ldots, \infty$ and $d = 2, 3, 4, \ldots, \infty$, and $n = \frac{1}{2}qd(d-1) + d$.

The $q = 2$ case can be identified with the quantum mechanical one we have seen before. On the other hand, the $q = 0$ case can be identified with the usual vision of the classical world: A world where hypotheticals simply do not matter, for the world just "is." In this case, an agent is well advised to take $Q(D_j) = P(D_j)$, meaning that there is no operational distinction between experiments $\mathcal{E}_1$ and $\mathcal{E}_2$ for her. It should not be forgotten however, that this rule, trivial though it looks, is still an addition to raw probability theory. It is just one that meshes well with what had come to be expected by most classical physicists. To put it yet another way, in the $q = 0$ case, the agent says to herself that the fine details of her actions do not matter. This to some extent authorizes the view that observation is a passive process in principle—again the classical worldview. Finally, the cases $q = 1$ and $q = 4$, though not classical, track still other structures that

have been explored previously: They correspond to what the Born rule would look like if alternate versions of quantum mechanics, those over real [91] and quaternionic [92] vector spaces, were expressed in the equivalent of SIC terms([23]).

Several years ago, Rüdiger Schack gave a talk on this material in Zurich, and Rob Spekkens asked, "Why that particular choice for modification of the Law of Total Probability?" Schack replied, "If one is going to modify it in any way at all, this is the simplest modification one can imagine". The remarkable fact is that the simplest possible modification to the Law of Total Probability carries with it so much interlocking structure.

If all that you desire is a story that you can tell about the current quantum formalism, then all this business about SICs and probabilistic representations might be of little moment. Of our fellow QBists, we know of one who likely doesn't care one way or the other about whether SICs exist. Another would like to see a general proof come to pass, but is willing to believe that QBism can just as well be developed without them—*i.e.*, they are not part of the essential philosophical ideas—and is always quick to make this point. On the other hand, we two are inclined to believe that QBism will become stagnant in the way of *all other* quantum foundations programs without a deliberate effort to rebuild the formalism. SICs might not be the only path to this goal [94], but they engage our attention for the following reason.

Formula (24) from the general setting indicates more strongly than ever that it is the role of *dimension* that is key to distilling the motif of our user's manual. Quantum theory, seen as a normative addition to probability theory, is just one theory (the second rung above classical) along an infinite hierarchy. What distinguishes the levels of this hierarchy is the strength q with which dimension "couples" the two paths in our diagram of fig. 5. It is the strength with which we are compelled to deviate from the Law of Total Probability when we transform our thoughts from the consequences of hypothetical actions upon a d's worth of the world's stuff to the consequences of our actual ones. Settling upon $q = 2$ (*i.e.*, settling upon quantum theory itself) sets the strength of the coupling, but the d variable remains. Different systems, different d, different deviations from a naive application of the Law of Total Probability.

In some way yet to be fully fleshed out, each quantum system seems to be a seat of active creativity and possibility, whose outward effect is as an "agent of change" for the parts of the world that come into contact with it. Observer and system, "agent and

([23]) The equivalent of SICs (*i.e.*, informationally complete sets of equiangular projection operators) certainly do *not* exist in general dimensions for the real-vector-space case—instead these structures only exist in a sparse set of dimensions, $d = 2, 3, 7, 23, \ldots$. With respect to the quaternionic theory, it appears from numerical work that they do not generally exist in that setting either [93]. Complex quantum mechanics, like baby bear's possessions, appears to be just right. This raises the possibility that *if* one had reason to think that the user's manual should be one of those three alternatives—real, complex and quaternionic—demanding the existence of SIC-type structures in all dimensions could narrow down the choice exactly to the complex case.

reagent," might be a way to put it. Perhaps no metaphor is more pregnant for QBism's next move than this: If a quantum system is comparable to a chemical reagent, then d is comparable to a valence. But valence for *what* more exactly?

8. – Mathematical intermezzo: The sporadic SICs

Before we leap off into cosmological speculations, let us take a moment to make a few things more concrete. What does a SIC *look like*, anyway? How do we write one out explicitly? One of the ongoing challenges of SIC research is that the solutions look so complicated: As we go up in the dimension, the vectors soon take many pages of computer printout. Moreover, there are not obvious relations between a SIC vector in one dimension and one in another, so finding one solution doesn't help with finding the next. It is only recently that we have been able to tease out some interconnections happening beneath the surface. In this section, we'll explore the patterns that bind sets of SIC solutions together [95, 96].

A SIC is *group covariant* if it can be constructed by starting with a single vector (the *fiducial*) and acting upon that vector with the elements of some group. All known SICs are group covariant, although since group covariance simplifies the search process, this could be a matter of the light being under the lamppost. Furthermore, in all known cases but one, that group is a *Weyl-Heisenberg group.* Working in dimension d, let $\omega_d = e^{2\pi i/d}$, and define the shift and phase operators

$$X\,|j\rangle = |j+1\rangle \quad \text{and} \quad Z\,|j\rangle = \omega_d^j\,|j\rangle\,, \tag{25}$$

where the shift is modulo d. Products of powers of X and Z, together with dimension-dependent phase factors that we can neglect for the present purposes, define the Weyl-Heisenberg group for dimension d.

A historical aside: This group dates back to the earliest days of quantum physics. Note that the two operators X and Z just fail to commute, doing so up to a phase factor:

$$ZX = \omega XZ. \tag{26}$$

In July of 1925, Max Born had the idea that he could solve one of the equations in Heisenberg's seminal 1925 paper if he made an *ansatz* that the position observable $\hat{q}$ and momentum observable $\hat{p}$ satisfied the commutation relation [97, 98]

$$\hat{q}\hat{p} - \hat{p}\hat{q} = i\hbar\hat{I}. \tag{27}$$

Pascual Jordan later proved that Born's ansatz was the only one that could work, and the paper they wrote up together was received by *Zeitschrift für Physik* on 27 September 1925. It was titled "On Quantum Mechanics." On the same day, Max Born received a letter from Hermann Weyl [99] saying that a previous discussion they had earlier in the month inspired him to generalize Born's relation (27) to eq. (26). Part of what pleased

Weyl was that his generalization was not dependent upon infinite-dimensional spaces—Weyl's relation would always have a solution in the complex matrices, finite and infinite dimensional. And by that circumstance he declared to have a way of "defining a general quantum system"—to each would be associated a "phase space." In the discrete case, the points of the phase space would be associated with the operators

$$(m, n) \longrightarrow X^m Z^n. \tag{28}$$

See Weyl's 1927 textbook [100], as well as Julian Schwinger's later development and extension of the idea in [101]. If a SIC covariant under the Weyl-Heisenberg group always exists, then it would mean that not only could a phase space be associated with a quantum system abstractly, but that the points of the phase space very directly correspond to the outcomes of a potential measurement. It is interesting to see how some of the earliest math in one of the earliest formulations of the theory—Weyl's—came so close to what we are working with today!

In $d = 2$, we can draw a SIC in the Bloch representation. Any qubit SIC forms a tetrahedron inscribed in the Bloch sphere [74]. One such tetrahedron is, in terms of the Pauli matrices,

$$\Pi_{s,r} = \frac{1}{2}\left(I + \frac{1}{\sqrt{3}}(s\sigma_x + r\sigma_y + sr\sigma_z)\right), \tag{29}$$

where the sign variables s and r take the values ± 1. The outcome probabilities for a state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ are given by the Born rule, $p(s, r) = \mathrm{tr}(\Pi_{s,r}|\psi\rangle\langle\psi|)/2$. Explicitly,

$$p(s, r) = \frac{1}{4} + \frac{\sqrt{3}}{12} sr\left(|\alpha|^2 - |\beta|^2\right) + \frac{\sqrt{3}}{6}\,\mathrm{Re}\left[\alpha\beta^*(s + ir)\right]. \tag{30}$$

This expression simplifies in terms of the Cartesian coordinates (x, y, z) of points on the Bloch sphere:

$$p(s, r) = \frac{1}{4} + \frac{\sqrt{3}}{12}\left(sx + ry + srz\right). \tag{31}$$

Two SICs in higher dimensions will be important for our purposes. First is the *Hesse SIC* in $d = 3$, constructed by applying the Weyl-Heisenberg group to the fiducial

$$\left|\psi_0^{(\mathrm{Hesse})}\right\rangle = \frac{1}{\sqrt{2}}(0, 1, -1)^{\mathrm{T}}. \tag{32}$$

Second is the *Hoggar SIC* in $d = 8$. We have multiple choices of fiducial in this case, but they all yield structures that are equivalent up to unitary or antiunitary transformations, so for brevity we speak of "the" Hoggar SIC [102]. One such fiducial [103, 104] is

$$\left|\psi_0^{(\mathrm{Hoggar})}\right\rangle \propto (-1 + 2i, 1, 1, 1, 1, 1, 1, 1)^{\mathrm{T}}. \tag{33}$$

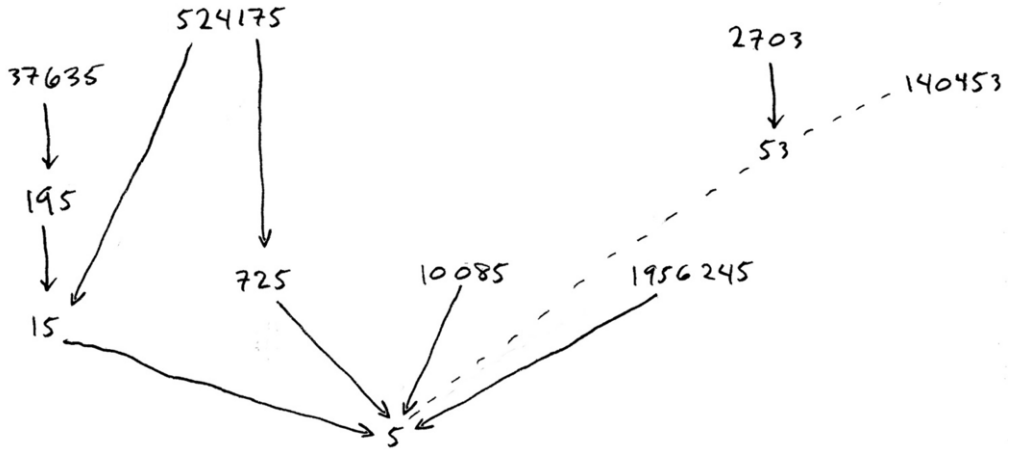


Fig. 6. – The lowest levels of a dimension tower. Conjecturally, Weyl-Heisenberg SICs in these dimensions are related by way of algebraic number theory [95]. The arithmetical meaning of the arrows is left as an exercise to the interested reader.

The Hoggar SIC is the *only* known case where the group that constructs the SIC from the fiducial is not the Weyl-Heisenberg group for d dimensions itself [102]. It is, however, of a related kind: It is the tensor product of three copies of the qubit Weyl-Heisenberg group. The more we poke at the Hoggar SIC, the more odd and unusual things turn up about it [104]. We'll take a look at one of them in this section.

The SICs in dimensions 2 and 3, as well as the Hoggar SIC in dimension 8, stand apart in some respects from the other known solutions [95, 105]. Recently, Appleby *et al.* [95] found a link between SICs and *algebraic number theory.* Their results apply to Weyl-Heisenberg SICs in dimensions 4 and larger. The SICs in dimensions 2 and 3, as well as the Hoggar SIC, fall outside of this category. Either their dimensions are too small, or (in the case of the Hoggar SIC) they have the wrong symmetry group. We can think of them as the *sporadic SICs.*

First, let us sketch the picture for the SICs studied by Appleby *et al.*. The pattern, which is just beginning to come clear, is a story about *number fields.* To a physicist, a "field" means something like the electric field, but to a number theorist, a field is a set of numbers where addition and multiplication can both be done, and where both additive and multiplicative inverses exist, and everything plays together nicely. The real numbers $\mathbb{R}$ constitute a field, as do the rational numbers $\mathbb{Q}$ within them. We can build up a field by starting with some base, like the rationals, and augmenting it with a new element. For example, let us invent a number "$\sqrt{3}$," about which all we know is that it is a positive number that solves the equation $x^2 - 3 = 0$. We then consider all the numbers of the form $a + b\sqrt{3}$, where a and b are rational. This set is a new field, $\mathbb{Q}$ extended by the new ingredient $\sqrt{3}$, which we write as $\mathbb{Q}(\sqrt{3})$.

The connection between SICs and number fields happens when we take the inner-

product conditiont that defines a SIC,

$$|\langle\psi_j|\psi_k\rangle|^2 = \frac{1}{d+1}\,, \tag{34}$$

and we leave off the magnitude-squared step:

$$\langle\psi_j|\psi_k\rangle = \frac{e^{i\theta_{jk}}}{\sqrt{d+1}}\,. \tag{35}$$

The phase factors $e^{i\theta_{jk}}$ turn out to live within very special number fields, and they are particularly special numbers within those fields. They are *units of ray class fields or extensions thereof*—as Bengtsson quips, "These words carry deep meaning for algebraic number theorists" [106]. They mean that we are knocking on the door of *Hilbert's twelfth problem*, one of the last remaining unsolved puzzles on history's most influential list of mathematical challenges [107]. And, remember, we got here because we were trying to find a better way to talk about probability in quantum mechanics!

Part of this still-emerging story [95] is that SICs in different dimensions are related in a hidden way because their number fields are related. The SIC phase factors live in fields that are extensions twice over of the rationals. That is, for a Weyl-Heisenberg SIC in dimension d, the phase factors $e^{i\theta_{jk}}$ make their home in an extension of $\mathbb{Q}(\sqrt{(d-3)(d+1)})$. Because we can factor perfect squares out from under the radical sign, different values of d can yield the same extension of $\mathbb{Q}$. This has led to the image of a *dimension tower*, an infinite sequence $\{d_1, d_2, d_3, \ldots\}$ where each d_j follows neatly from those before, and the number theory tells us how to build a SIC in each d_j. But this remains in the realm of conjecture.

And what of the sporadic SICs? Quite unexpectedly, the qubit SICs, the Hesse SIC and the Hoggar SIC also connect to a subject in pure mathematics that has gone mostly un-utilized in physics. Specifically, their symmetries are linked with a *lattice of integers* in the set of numbers known as the *octonions* [96]. Physics students grow familiar with the complex numbers $\mathbb{C}$ by repeated exposure, internalizing the image of a number plane that extends out on either side of the number line. The quaternions (usually written with an $\mathbb{H}$ for William Hamilton) and the octonions (denoted $\mathbb{O}$) arise when one tries to repeat this dimension-doubling stunt, from two dimensions to four and then to eight. It so happens that familiar properties of arithmetic are lost with each repetition: Multiplication of quaternions is not commutative, but it is still associative. And multiplication of octonions is not even associative! Nineteenth-century physics made much use of quaternions to study 3D rotations, and they still find application in geometry, for example in computer graphics. Octonions are less familiar still, and are perhaps best known for their relations to *exceptional structures* in mathematics. John Baez observed [108],

> Often you can classify some sort of gizmo, and you get a beautiful systematic list, but also some number of exceptions. Nine times out of 10 those exceptions are related to the octonions.

It so transpires that this applies to SICs, too. Moreover, the link with octonionic integers connects the sporadic SICs to the problem of *sphere packing*, that is, the question of how to fit Euclidean spheres of arbitrary dimension together in the most efficient way [109].

A few years ago, one of the authors (BCS) was attending an interdisciplinary workshop and, over lunch, fell into conversation with a mathematician who had that morning lectured on higher-dimensional sphere packing. After a little while, the mathematician asked, "And what are you working on?"

"Too many different things—but one problem has the same feel as sphere packing, because solutions in one dimension don't seem to tell you about solutions in others. I guess the math people know it as the problem of 'complex equiangular lines'."

"Ah! SICs! You know, when I first heard about that conjecture, I thought I could just sit down and solve it. But that didn't quite happen."

As the mathematical properties of SIC dimensionalities grow more intriguing, we are led back to the question of what the dimension of a Hilbert space means physically.

9. – Hilbert-space dimension as a universal capacity

> It is entirely possible to conceive of a world composed of individual atoms, each as different from one another as one organism is from the next.
>
> — John Dupré

A common accusation heard by the QBist([24]) is that the view leads straight away to solipsism, "the belief that all reality is just one's imagining of reality, and that one's self is the only thing that exists"([25]). The accusation goes that, if a quantum state $|\psi\rangle$ only represents the degrees of belief held by some agent—say, the one portrayed in fig. 1—then the agent's beliefs must be the source of the universe. The universe could not exist without him: This being such a ridiculous idea, QBism is dismissed out of hand, *reductio ad absurdum*. It is so hard for the QBist to understand how anyone could think this (it being the antithesis of everything in his worldview) that a little of our own Latin comes to mind: *non sequitur*. See fig. 7.

A fairer-minded assessment is that the accusation springs from our opponents "hearing" much of what we do say, but interpreting it in terms drawn from a particular conception of what physical theories *always ought to be*: Attempts to directly represent (map, picture, copy, correspond to, correlate with) the *universe*—with "universe" here

([24]) For perhaps the loudest, see ref. [110].

([25]) This is the definition of *The American Heritage New Dictionary of Cultural Literacy*, third edition (2005). *Encyclopedia Brittanica* (2008) expands, "in philosophy ... the extreme form of subjective idealism that denies that the human mind has any valid ground for believing in the existence of anything but itself. The British idealist F. H. Bradley, in *Appearance and Reality* (1897), characterized the solipsistic view as follows: 'I cannot transcend experience, and experience is my experience. From this it follows that nothing beyond myself exists; for what is experience is its (the self's) states"'.

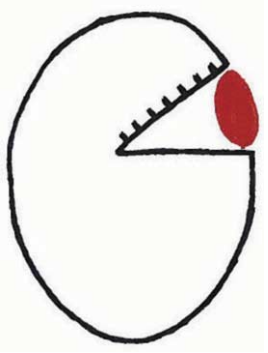


Fig. 7. – *Sarcasm.* In a lecture bottlenecked by repeated accusations of QBism's solipsism, the authors sometimes use the following technique to move things along. Referring to the previous fig. 1, one asks the stubborn accuser, "What about this diagram do you *not* get? It shows an agent and a physical system external to him. It says that a quantum state is a *state of belief* about what will come about as a consequence of his actions upon the system. The quantum state is not a state of nature, but so what? There is an agent with his belief; there is a system that is not part of him; and there is something that really, eventually comes about—it is called *the outcome*. No agent, no outcome for sure, but that's not solipsism: For, no system, no outcome either! A quantum measurement without an external system participating would be like the sound of one hand clapping, a Zen koan. If we were really expressing solipsism, wouldn't a diagram like the one above be more appropriate? A big eyeball surveying nothing. Now there's really no external system and nothing to act upon. *That's* solipsism."

thought of in its totality as a pre-existing, static system; an unchanging, monistic something that just *is*. From such a "representationalist" point of view, *if* a) quantum theory is a proper physical theory, b) its essential theoretical objects are quantum states, and c) quantum states are states of belief, *then* the universe that "just is" corresponds to a state of belief. This chain of deduction is logical and clear, but completely misguided.

QBism sidesteps the poisoned dart, as the previous sections have tried to convey, by asserting that quantum theory is just not a physical theory in the sense the accusers want it to be. Rather it is an addition to personal, Bayesian, normative probability theory. Its normative rules for connecting probabilities (personal judgments) were developed in light of the *character of the world*, but there is no sense in which the quantum state itself represents (pictures, copies, corresponds to, correlates with) a part or a whole of the external world, much less a world that *just is*. In fact the very character of the theory seems to point to the inadequacy of the representationalist program when attempted on the particular world we live in.

There are no lofty philosophical arguments here that representationalism must be wrong always and in all possible worlds (perhaps because of some internal inconsistency([26])). Representationalism may well be true in this or that setting—we take no stand on the matter. We only know that for nearly 90 years quantum theory has been

([26]) As, *e.g.*, Rorty [111] might try to argue.

actively resistant to representationalist efforts on *its* behalf. This suggests that it might be worth exploring some philosophies upon which physics rarely sets foot. Physics of course should never be constrained by any one philosophy (history shows it nearly always lethal), but it does not hurt to get ideas and insights from every source one can. If one were to sweep the philosophical literature for schools of thought representative of what QBism actually is about, it is not solipsism one will find, but nonreductionism [112,113], (radical) metaphysical pluralism [114,115], empiricism [116,117], indeterminism and meliorism([27]) [118], and above all pragmatism [29,119].

A form of nonreductionism can already be seen in play in our answer to whether the notion of agent should be derivable from the quantum formalism itself. We say that it cannot be and it should not be, and to believe otherwise is to misunderstand the subject matter of quantum theory. But nonreductionism also goes hand in hand with the idea that there is real particularity and "interiority" in the world. Think again of the "I-I-me-me mine" feature that shields QBism from inconsistency in the "Wigner's friend" scenario. When Wigner turns his back to his friend's interaction with the system, that piece of reality is hermetically sealed from him. That phenomenon has an inside, a vitality that he takes no part in until he again interacts with one or both relevant pieces of it. *With respect to Wigner*, it is a bit like a universe unto itself.

If one seeks the essence of indeterminism in quantum mechanics, there may be no example more directly illustrative of it than "Wigner's friend." For it expresses to a tee William James's notion of indeterminism [118]:

> [Chance] is a purely negative and relative term, giving us no information about that of which it is predicated, except that it happens to be disconnected with something else—not controlled, secured, or necessitated by other things in advance of its own actual presence. [...] What I say is that it tells us nothing about what a thing may be in itself to call it "chance." [...] All you mean by calling it "chance" is that this is not guaranteed, that it may also fall out otherwise. For the system of other things has no positive hold on the chance-thing. Its origin is in a certain fashion negative: it escapes, and says, Hands off! coming, when it comes, as a free gift, or not at all.

([27]) Strictly speaking, meliorism is the doctrine "that humans can, through their interference with processes that would otherwise be natural, produce an outcome which is an improvement over the aforementioned natural one." But we would be reluctant to take a stand on what "improvement" really means. So said, all we mean in the present essay by meliorism is that the world before the agent is malleable to some extent—that his actions really can change it. Adam said to God, "I want the ability to write messages onto the world." God replied, "You ask much of me. If you want to write upon the world, it cannot be so rigid a thing as I had originally intended. The world would have to have some malleability, with enough looseness for you to write upon its properties. It will make your world more unpredictable than it would have been—I may not be able to warn you about impending dangers like droughts and hurricanes as effectively as I could have—but I can make it such if you want." And with that Adam brought all host of uncertainties to his life, but he gained a world where his deeds and actions mattered.

> This negativeness, however, and this opacity of the chance-thing when thus considered *ab extra*, or from the point of view of previous things or distant things, do not preclude its having any amount of positiveness and luminosity from within, and at its own place and moment. All that its chance-character asserts about it is that there is something in it really of its own, something that is not the unconditional property of the whole. If the whole wants this property, the whole must wait till it can get it, if it be a matter of chance. That the universe may actually be a sort of joint-stock society of this sort, in which the sharers have both limited liabilities and limited powers, is of course a simple and conceivable notion.

And once again [120],

> Why may not the world be a sort of republican banquet of this sort, where all the qualities of being respect one another's personal sacredness, yet sit at the common table of space and time?
>
> To me this view seems deeply probable. Things cohere, but the act of cohesion itself implies but few conditions, and leaves the rest of their qualifications indeterminate. As the first three notes of a tune comport many endings, all melodious, but the tune is not named till a particular ending has actually come,—so the parts actually known of the universe may comport many ideally possible complements. But as the facts are not the complements, so the knowledge of the one is not the knowledge of the other in anything but the few necessary elements of which all must partake in order to be together at all. Why, if one act of knowledge could from one point take in the total perspective, with all mere possibilities abolished, should there ever have been anything more than that act? Why duplicate it by the tedious unrolling, inch by inch, of the foredone reality? No answer seems possible. On the other hand, if we stipulate only a partial community of partially independent powers, we see perfectly why no one part controls the whole view, but each detail must come and be actually given, before, in any special sense, it can be said to be determined at all. This is the moral view, the view that gives to other powers the same freedom it would have itself.

The train of logic back to QBism is this. If James and our analysis of "Wigner's friend" are right, the universe is not *one* in a very rigid sense, but rather more truly a pluriverse([28]). To get some sense of what this can mean, it is useful to start by thinking

([28]) The term "pluriverse" is again a Jamesian one. He used it interchangeably with the word "multiverse," which he also invented [121]. Unfortunately the latter has been coopted by the Everettian movement for their own—in the end monistic—purposes: "The world is one; it *is* the deterministically evolving universal quantum state, the 'multiverse'." Too bad. Multiverse is a tempting word, but we stick with pluriverse to avoid any confusion with the Everettian usage.

about what it is not. A good example can be found by taking a solution to the vacuum Maxwell equations in some extended region of spacetime. Focus on a compact subregion and try to conceptually delete the solution within it, reconstructing it with some new set of values. It can't be done. The fields outside the region (including the boundary) uniquely determine the fields inside it. The interior of the region has no identity but that dictated by the rest of the world—it has no "interiority" of its own. The pluriverse conception says we'll have none of that. And so, for any agent immersed in this world there will always be uncertainty for what will happen upon his encounters with it. To wit, where there is uncertainty there should be Bayesian probabilities, and so on and so on until much of the story we have already told.

What all this hints is that for QBism the proper way to think of our world is as the empiricist or the radical metaphysical pluralist does. Let us launch into making this clearer, for that process more than anything will explain how QBism hopes to interpret Hilbert-space dimension.

The metaphysics of empiricism can be put like this. Everything experienced, everything experienceable, has no less an ontological status than anything else. A child awakens in the middle of the night frightened that there is a monster under her bed, one soon to reach up and steal her arm—that *we-would-call-imaginary* experience has no less a hold on onticity than a Higgs-boson detection event at the LHC, or the minuscule wobbles at LIGO that shook the scientific world. They are of equal status from this point of view—they are equal elements in the filling out and making of reality. There is indeed no doubt that we should call the child's experience *imaginary*. That, however, is a statement about the experience's meaning and interpretation, not its existence. The experience *as it is* exists, period. It is what it is. Like the biblical burning bush, each experience declares, "I am that I am." Most likely in the present example, the experience will be a little piece of the universe isolated, on its own, and of no great consequence. But one never knows until all future plays out. Some lucky dreams have built nations. Maybe the same is true of some lucky Higgs-boson events. Most though, surely, will be of the more minor fabric of existence. All in all, the world of the empiricist is not a sparse world like the world of Democritus (*nothing but* atom and void) or Einstein (*nothing but* unchanging spacetime manifold equipped with this or that field), but a world overflowingly full of variety—a world whose details are beyond anything grammatical (rule-bound) expression can articulate.

Yet this is no statement that physics should give up, or that physics has no real role in coming to grips with the world. It is only a statement that physics should better understand its function. What is being aimed for here finds its crispest, clearest contrast in a statement Richard Feynman once made [122]:

> If, in some cataclysm, all of scientific knowledge were to be destroyed, and only one sentence passed on to the next generation of creatures, what statement would contain the most information in the fewest words? I believe it is the atomic hypothesis (or the atomic fact) that all things are made of atoms—little particles that move around in perpetual motion, attracting each other

> when they are a little distance apart, but repelling upon being squeezed into one another. [...]
>
> Everything is made of atoms. That is the key hypothesis.

The issue for QBism hangs on the imagery that usually lies behind the phrase "everything is made of." William James called it the great original sin of the rationalistic mind [123]:

> Let me give the name of "vicious abstractionism" to a way of using concepts which may be thus described: We conceive a concrete situation by singling out some salient or important feature in it, and classing it under that; then, instead of adding to its previous characters all the positive consequences which the new way of conceiving it may bring, we proceed to use our concept privatively; reducing the originally rich phenomenon to the naked suggestions of that name abstractly taken, treating it as a case of "nothing but" that, concept, and acting as if all the other characters from out of which the concept is abstracted were expunged. Abstraction, functioning in this way, becomes a means of arrest far more than a means of advance in thought. It mutilates things; it creates difficulties and finds impossibilities; and more than half the trouble that metaphysicians and logicians give themselves over the paradoxes and dialectic puzzles of the universe may, I am convinced, be traced to this relatively simple source. *The viciously privative employment of abstract characters and class names* is, I am persuaded, one of the great original sins of the rationalistic mind.

What is being realized through QBism's peculiar way of looking at things is that physics *actually can be done* without any accompanying vicious abstractionism. You do physics as you have always done it, but you throw away the idea "everything is made of [Essence X]" before even starting.

Physics—in the right mindset—is not about identifying the bricks with which nature is made, but about identifying what is *common to* the largest range of phenomena it can get its hands on. The idea is not difficult once one gets used to thinking in these terms. Carbon? The old answer would go that it is *nothing but* a building block that combines with other elements according to the following rules, blah, blah, blah. The new answer is that carbon is a *characteristic* common to diamonds, pencil leads, deoxyribonucleic acid, burnt pancakes, the space between stars, the emissions of Ford pick-up trucks, and so on—the list is as unending as the world is itself. For, carbon is also a characteristic common to this diamond and this diamond and this diamond and this. But a flawless diamond and a purified zirconium crystal, no matter how carefully crafted, have no such characteristic in common: Carbon is not a *universal* characteristic of all phenomena. The aim of physics is to find characteristics that apply to as much of the world in its varied fullness as possible. However, those common characteristics are hardly what the world is made of—the world instead is made of this and this and this. The world is constructed of every particular there is and every way of carving up every particular there is.

An unparalleled example of how physics operates in such a world can be found by looking to Newton's law of universal gravitation. What did Newton really find? Would he be considered a great physicist in this day when every news magazine presents the most cherished goal of physics to be a Theory of Everything? For the law of universal gravitation is hardly that! Instead, it *merely* says that every body in the universe tries to accelerate every other body toward itself at a rate proportional to its own mass and inversely proportional to the squared distance between them. Beyond that, the law says nothing else particular of objects, and it would have been a rare thinker in Newton's time, if any at all, who would have imagined that all the complexities of the world could be derived from that limited law. Yet there is no doubt that Newton was one of the greatest physicists of all time. He did not give a theory of everything, but a Theory of One Aspect of Everything. And only the tiniest fraction of physicists of any variety, much less the TOE-seeking variety, have ever worn a badge of that more modest kind. It is as H. C. von Baeyer wrote in one of his books [124],

> Great revolutionaries don't stop at half measures if they can go all the way. For Newton this meant an almost unimaginable widening of the scope of his new-found law. Not only Earth, Sun, and planets attract objects in their vicinity, he conjectured, but all objects, no matter how large or small, attract all other objects, no matter how far distant. It was a proposition of almost reckless boldness, and it changed the way we perceive the world.

Finding a theory of "merely" one aspect of everything is hardly something to be ashamed of: It is the loftiest achievement physics can have in a living, breathing nonreductionist world.

Which leads us back to Hilbert space. Quantum theory—that user's manual for decision-making agents immersed in a world of *some* yet to be fully identified character—makes a statement about the world to the extent that it identifies a quality common to all the world's pieces. QBism says the quantum state is not one of those qualities. But of Hilbert spaces themselves, particularly their distinguishing characteristic one from the other, *dimension*([29]), QBism carries no such grudge. Dimension is something one posits for a body or a piece of the world, much like one posits a mass for it in the Newtonian theory. Dimension is something a body holds all by itself, regardless of what an agent thinks of it.

That this is so can be seen already from reasons internal to the theory. Just think of all the arguments rounded up for making the case that quantum states should be interpreted as of the character of Bayesian degrees of belief. None of these work for

([29]) Hardy [125, 126] and Dakić and Brukner [127] are examples of foundational efforts that also emphasize this quantum analogue to what Eötvös tested on platinum and copper [128]. Hardy put it this way in one of his axioms, "There exist systems for which $N = 1, 2, \cdots$, and, furthermore, all systems of dimension N, or systems of higher dimension but where the state is constrained to an N dimensional subspace, have the same properties".

Hilbert-space dimension. Take one example, an old favorite—Einstein's argument about conditioning quantum states from afar. In sect. **5** of this paper we repeated the argument verbatim, but it is relevant to note that before Einstein could write down his ψ_{12}, he would have had to associate some Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ with S_1 and S_2 and take their tensor product $\mathcal{H}_1 \otimes \mathcal{H}_2$. Suppose the dimensionalities of these spaces to be d_1 and d_2, respectively. The question is, is there anything similar to Einstein's argument for changing the value of d_2 from a distance? There isn't. ψ_2 may be forced into this or that subspace by choosing the appropriate measurement on S_1, but there is no question of the whole Hilbert space $\mathcal{H}_2$ remaining intact. When it is time to measure S_2 itself, one will still have the full arsenal of quantum measurements appropriate to a Hilbert space of dimension d_2 to choose from—none of those fall by the wayside. In Einstein's terms, d_2 is part of the "real factual situation" of S_2([30]).

The claim here is that quantum mechanics, when it came into existence, implicitly recognized a previously unnoticed capacity inherent in all matter—call it *quantum dimension*. In one manifestation, it is the fuel upon which quantum computation runs [128, 132]. In another it is the raw irritability of a quantum system to being eavesdropped upon [76, 133]. In eqs. (19) and (20) it was a measure of deviation from the Law of Total Probability induced by hypothetical thinking. And in a farther-fetched scenario to which we will come back, its logarithm *might* just manifest itself as the squared gravitational mass of a Schwarzschild black hole [134, 135].

When quantum mechanics was discovered, something was *added* to matter in our conception of it. Think of the apple that inspired Newton to his law. With its discovery the color, taste, and texture of the apple didn't disappear; the law of universal gravitation didn't reduce the apple privatively to *just* gravitational mass. Instead, the apple was at least everything it was before, but afterward even more—for instance, it became known to have something in common with the Moon. A modern-day Cavendish would be able to literally measure the further attraction an apple imparts to a child already hungry to pick it from the tree. So similarly with Hilbert-space dimension. Those diamonds we have already used to illustrate the idea of nonreductionism, in very careful conditions, could be used as components in a quantum computer [136]. Diamonds have among their

([30]) Take a coin, and imagine flipping it. We generally write down a (subjective) probability distribution over two outcomes to capture our degrees of belief of which way the flip will go. But of course it is a judgement call that it can only go two ways. Steven van Enk would say it could always land on its side; so he would always write down a probability distribution over three outcomes. If one takes (p_0, p_1) as a subjective assignment, the number 2 is objective with respect to it: It is something we imagine or hypothesize about the coin. If one takes the (p_0, p_1, p_2) as a subjective assignment, then the number 3 is objective with respect to it: It will fall one of three ways regardless of what we believe about which of the three ways it will fall. So objectivity/subjectivity comes in layers. We call something objective, and then make probability assignments in the subjective layer above it. But of course, the first "calling something objective" has a personal element in itself. Recently, techniques have started to become available to "test" the supposition of a dimension against one's broader mesh of beliefs; see [129-131].

many properties something not envisioned before quantum mechanics—that they could be a source of relatively accessible Hilbert space dimension and as such have this much in common with any number of other proposed implementations of quantum computing. Diamonds not only have something in common with the moon, but now with the ion-trap quantum-computer prototypes around the world.

Diamondness is not something to be derived from quantum mechanics. It is that quantum mechanics is something we *add* to the repertoire of things we already say of diamonds, to the things we do with them and the ways we admire them. This is a very powerful realization: For diamonds already valuable, become ever more so as their qualities compound. And saying more of them, not less of them as is the goal of all reductionism, has the power to suggest all kinds of variations on the theme. For instance, thinking in quantum mechanical terms might suggest a technique for making "purer diamonds"—though to an empiricist this phrase means not at all what it means to a reductionist. It means that these similar things called diamonds can suggest exotic variations of the original objects with various pinpointed properties this way or that. Purer diamond is not *more* of what it already was in nature. It is a new species, with traits of its parents to be sure, but nonetheless stand-alone, like a new breed of dog.

To put it still differently, and now in the metaphor of music, a jazz musician might declare that a tune once heard thereafter plays its most crucial role as a substrate for something new. It is the fleeting solid ground upon which something new can be born. The nine tracks titled *Salt Peanuts* in CAF's mp3 player([31]) are moments of novelty in the universe never to be recreated. So of diamonds, and so of all this quantum world. Or at least that is the path QBism seems to indicate([32]).

To the reductionist, of course, this seems exactly backwards. But then, it is the reductionist who must live with a seemingly infinite supply of conundrums arising from quantum mechanics. It is the reductionist who must live in a state of arrest, rather than moving on to the next stage of physics. Take a problem that has been a large theme of the quantum foundations meetings for the last 30 years. To put it in a commonly heard question, "Why does the world look classical if it actually operates according to quantum mechanics?" The touted mystery is that we never "see" quantum superposition and entanglement in our everyday experience. But have you ever seen a probability distribution sitting in front of you? Probabilities in personalist Bayesianism are not the sorts of things that can be seen; they are the things that are thought. It is *events* that are seen.

The real issue is this. The expectation of the quantum-to-classical transitionists([33]) is that quantum theory is at the bottom of things, and "the classical world of our experience" is something to be derived out of it. QBism says "No. Experience is neither

([31]) Charlie Parker, Dizzy Gillespie, Charlie Parker, Charlie Parker, Charlie Parker, Joshua Redman, Miles Davis Quintet, Arturo Sandoval, "The Quintet" (Massey Hall, 1953).
([32]) A nice *logical* argument for this can be found in [137].
([33]) See [138, 139] for particularly clear discussions of the subject.

classical nor quantum. Experience is experience with a richness that classical physics of any variety could not remotely grasp". Quantum mechanics is something put on top of raw, unreflected experience. It is additive to it, suggesting wholly new types of experience, while never invalidating the old. To the question, "Why has no one ever *seen* superposition or entanglement in diamond before?", the QBist replies: "It is simply because before recent technologies and very controlled conditions, as well as lots of refined analysis and thinking, no one had ever mustered a mesh of beliefs relevant to such a range of interactions (factual and hypothetical) with diamonds". No one had ever been in a position to adopt the extra normative constraints required by the Born rule. For QBism, it is not the emergence of classicality that needs to be explained, but the emergence of our new ways of manipulating, controlling, and interacting with matter that do.

In this sense, QBism declares the quantum-to-classical research program unnecessary (and actually obstructive([34])) in a way not so dissimilar to the way Bohr's 1913 model of the hydrogen atom declared another research program unnecessary (and actually obstructive). Before Bohr, everyone thought that the only thing that could count as an explanation of the hydrogen atom's stable spectrum was a mechanical model. Bohr's great genius in comparison to all the other physicists of his day was in being the first to say, "Enough! I shall not give a mechanistic explanation for these spectra we see. Here is a way to think of them with no mechanism". Researchers had wasted years seeking an unfulfillable vision of the world, and that certainly was an obstruction to science.

All is not lost, however, for the scores of decoherentists this policy would unforgivingly unemploy. For it only suggests that they redirect their work to the opposite task. The thing that needs insight is not the quantum-to-classical transition, but the classical-to-quantum! The burning question for the QBist is how to model in Hilbert-space terms the common sorts of measurements we perform just by opening our eyes, cupping our ears, and extending our fingers.

Take a professional baseball player watching a ball fly toward him: He puts his whole life into when and how he should swing his bat. But what does this mean in terms of the immense Hilbert space a quantum theoretical description would associate with the ball? Surely the player has an intuitive sense of both the instantaneous position and instantaneous momentum of the baseball before he lays his swing into it—that's what "keeping his eye on the ball" means. Indeed it is from this intuition that Newton was able to lay down his laws of classical mechanics. Yet, what can it mean to say this given quantum theory's prohibition of simultaneously measuring complementary observables? It means that whatever the baseball player is measuring, it ain't that—it ain't position *and* momentum as usually written in operator terms. Instead, a quantum model of what he is doing would be some interesting, far-from-extremal *single* POVM—perhaps

([34]) Without an ontic understanding of quantum states, quantum operations, and unitary time evolutions—all of which QBism rejects [2, 5, 37]—how can the project even get off the ground? As one can ask of the Big Bang, "What banged?", the QBist must ask, "In those days of the world before agents using quantum theory, what decohered?".

even one that takes into account some information that does not properly live within the formal structure of quantum theory (the larger arena that Howard Barnum calls "meaty quantum physics" [140]). For instance, that an eigenvector $|i\rangle$ of some Hermitian operator, though identically orthogonal to fellow eigenvectors $|j\rangle$ and $|k\rangle$ in the Hilbert-space sense, might be *closer* in meaning to $|k\rangle$ than to $|j\rangle$ for some issue at hand.

So the question becomes how to take a given common-day measurement procedure and *add to it* a consistent quantum description? The original procedure was stand alone—it can live without a quantum description of it—but if one wants to move it to a new level or new direction, having added a consistent quantum description will be most helpful to those ends. Work along these lines is nascent, but already some excellent examples exist [141]. Of course, unconsciously it is what has been happening since the founding days of quantum mechanics. Here, we find an affinity with a comment of John Stuart Bell, one buried in a letter to Rudolf Peierls and almost lost to physics history [10]:

> I have the impression as I write this, that a moment ago I heard the bell of the tea trolley. But I am not sure because I was concentrating on what I was writing. [...] The ideal instantaneous measurements of the textbooks are not precisely realized anywhere anytime, and more or less realized, more or less all the time, more or less everywhere.

QBism thinks of the textbook "ideal instantaneous measurements" as on the same continuum as listening for the tea trolley. But what POVM elements should one write for the latter? Only time, actively spent in new research, will tell.

The important question is how matter can be coaxed to do new things. It is in the ways the world yields to our desires, and the ways it refuses to, that we learn the depths of its character.

> I give you an object of this much gravitational mass. What can you do with it? What can you not? And when you are not about, what does it cause?
> I give you an object of this much quantum dimension. What can you do with it? What can you not? And when you are not about, what does it cause?

If taken seriously what do these questions imply by their very existence? That they should have meaningful answers! Here is one example. A knee-jerk reaction in many physicists upon hearing these things is to declare that dimension as a capacity collapses to a triviality as soon as it is spoken. "All real-world systems possess infinite-dimensional Hilbert spaces. And it doesn't take quantum field theory to be completely correct to make that true; a simple one-dimensional harmonic oscillator will do. It has an infinite-dimensional Hilbert space." But maybe not. Maybe no real-world quantum system has that much oomph. Just as one can treat the Earth's inertial mass as infinite for many a freshman mechanics problem, or a heat bath as infinite for many a thermodynamical one, maybe this is all that has ever been going on with infinite-dimensional Hilbert spaces. It is a useful artifice when a problem can be economically handled with a differential equation. (Ask Schrödinger.) It is worth noting that when the algebraists set about

making a rigorous statement of what a quantum field theory ought to be, they seem only to be able to make progress by imposing a postulate that says, roughly, "In a QFT, the states that are localized in space and bounded in energy form a finite-dimensional space" [142].

And with this, we come to nearly the farthest edge of QBism. It is the beginning of a place where quantum mechanics must step past itself. To make quantum dimension meaningful in ontic terms, as a quality common to all physical objects, is to say it should be finite—going up, going down from this object to the next, but always finite. Every region of space where electromagnetism can propagate, finite. Every region of space where there is a gravitational "field," finite.

It means that despite its humble roots in nonrelativistic quantum mechanics, there is something already cosmological about QBism. It tinkers with spacetime, saying that in every "hole" (every bounded region) there is an interiority not given by the rest of the universe and a common quality called dimension. It says that there is probably something right about the "holographic principles" arising from other reaches of physics [143]. Recognizing entropy as a personal concept (entropy is a function of probability), QBism would suspect that it is not an entropy bound that arises from these principles, but perhaps a dimension bound [8, 128].

Invocations of a "holographic principle" in quantum gravity research traditionally contain a statement along the lines of, "The information in a volume actually lives on its boundary." These locutions grow more opaque the more closely they are studied. To talk of the "degrees of freedom existing on the boundary" is to trap oneself within obsolete intuitions. Fundamentally, probabilities do not exist without a gambler, and likewise, "information" does not exist without an agent concerned with communication and computation. (The latter statement is the logarithm of the former.) Two orthogonal quantum states for a system are not two distinct physical configurations, in the sense of classical physics. Rather, they are two maximally distinct hypotheses for its possible future behavior consequent upon an agent's action. Alice's quantum state for a system—whether a benzene ring or a black hole—does not live on the system's boundary, nor in its bulk. It lives in Alice's mesh of beliefs, along with all her other fears and aspirations.

A novel perspective requires new images and metaphors, which can in turn stimulate novel technical developments. How do we distance ourselves from the language that Hilbert-space dimension quantifies "the number of distinct states a system can be in"? For this mode of thought is at the root of all the loose talk in trying to interpret those holographic principles.

Suppose that Alice has access to a localized physical phenomenon that she wishes to employ in a quantum communication scheme. Her goal is to detect, as well as possible, whether her communiqués are being eavesdropped upon. A technical result from a few years ago indicates that the maximal achievable sensitivity to eavesdropping is a simple function of the Hilbert-space dimension [76]. (In fact, a SIC furnishes a set of states that saturates this bound.) If we take the holographic principle to say that the maximal Hilbert-space dimension of a phenomenon grows with the area that bounds it, then we have a relation between optimal "sensitivity to the touch" and a boundary area. Indeed,

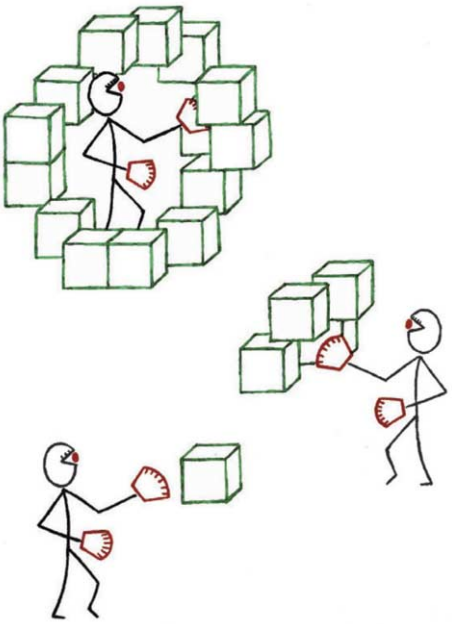


Fig. 8. – *Quantum cosmology from the inside.* The agent in fig. 1 can consider measurements on ever larger systems. There is nothing in quantum mechanics to bar the systems considered from being larger and larger, to the point of eventually surrounding the agent. Pushed far enough, this is quantum cosmology! Why all this insistence on thinking that "an agent must be outside the system he measures" in the cosmological context should mean "outside the physical universe itself"? It means outside the system of interest, and that is the large-scale universe. Nor is there any issue of self-reference at hand. One would be hard pressed to find a cosmologist who wants to include his beliefs about how the beats of his heart correlate with the sidereal cycles in his quantum-state assignment for the external universe. The symbol $|\Psi_{\text{universe}}\rangle$ refers to the green boxes alone.

sensitivity being tied to boundary area is an appealing image: For the boundary is the only thing an agent can touch in the first place!

10. – Quantum cosmology from the inside

> Theodore Roosevelt's decision to build the Panama Canal shows that free will moves mountains, which implies, by general relativity, that even the curvature of space is not determined. The stage is still being built while the show goes on.
>
> — John Conway and Simon Kochen [55]

Let us, however, step back from that farthest edge for a moment and discuss cosmology as it is presently construed before taking a final leap!

Sometimes it is claimed that a point of view about quantum theory like QBism's would make the enquiries of quantum cosmology impossible. For instance, David Deutsch once wrote [144]:

> The best physical reason for adopting the Everett interpretation lies in quantum cosmology. There one tries to apply quantum theory to the universe as a whole, considering the universe as a dynamical object starting with a big bang, evolving to form galaxies and so on. Then when one tries, for example by looking in a textbook, to ask what the symbols in the quantum theory mean, how does one use the wave function of the universe and the other mathematical objects that quantum theory employs to describe reality? One reads there, "The meaning of these mathematical objects is as follows: first consider an observer outside the quantum system under consideration [...]." And immediately one has to stop short. Postulating an outside observer is all very well when we're talking about a laboratory: we can imagine an observer sitting outside the experimental apparatus looking at it, but when the experimental apparatus—the object being described by quantum theory—is the entire universe, it's logically inconsistent to imagine an observer sitting outside it. Therefore the standard interpretation fails. It fails completely to describe quantum cosmology. Even if we knew how to write down the theory of quantum cosmology, which is quite hard incidentally, we literally wouldn't know what the symbols meant under any interpretation other than the Everett interpretation.

But this is nonsense. It is not hard to imagine how to measure the universe as a whole: You simply live in it.

What are the typical observables and predictables of cosmology? The Hubble constant, the cosmological constant, the degree of inhomogeneity of the cosmic microwave background radiation, total baryon number in this or that era of the universe, perhaps others. To do quantum cosmology is to ask how an application of quantum mechanics can be made with regard to these quantities. For the QBist quantum theory would be used *as it always is*: As a normative calculus of consistency for all probability assignments concerned. Quantum theory advises an agent to make all his probability assignments derivable from a single quantum state. Write it like this if you wish—a big, fat wave function (it's the whole universe after all):

$$|\Psi_{\text{universe}}\rangle. \tag{36}$$

Why not? We are swimming in this ocean called the universe, and we have to do physics from inside of it. But then all the rest of the universe is outside each of us. This wave

function represents an agent's catalogue of beliefs for the relevant things outside([35]).

The only point here is that QBism has every bit as much right to do cosmology as any other interpretation of quantum mechanics. The only difference is that QBism does it from the inside.

More exciting is the possibility that once it does all that (its own version of what the other interpretations might have done), its power may not be exhausted. For, noting how the Big Bang itself is a moment of creation with some resemblance to every individual quantum measurement, one starts to wonder whether even it "might be on the inside." Certainly QBism has creation going on all the time and everywhere; quantum measurement is just about an agent hitching a ride and partaking in that ubiquitous process.

At the end of a long article it doesn't hurt to speculate. We let William James and John Archibald Wheeler do the work for us. First more sweepingly [145],

> Our acts, our turning-places, where we seem to ourselves to make ourselves and grow, are the parts of the world to which we are closest, the parts of which our knowledge is the most intimate and complete. Why should we not take them at their facevalue? Why may they not be the actual turning-places and growing-places which they seem to be, of the world—why not the workshop of being, where we catch fact in the making, so that nowhere may the world grow in any other kind of way than this?
>
> Irrational! we are told. How can new being come in local spots and patches which add themselves or stay away at random, independently of the rest? There must be a reason for our acts, and where in the last resort can any reason be looked for save in the material pressure or the logical compulsion of the total nature of the world? There can be but one real agent of growth, or seeming growth, anywhere, and that agent is the integral world itself. It may grow all-over, if growth there be, but that single parts should grow *per se* is irrational.
>
> But if one talks of rationality—and of reasons for things, and insists that they can't just come in spots, what *kind* of a reason can there ultimately be why anything should come at all?

Then, more modernly [39],

> Each elementary quantum phenomenon is an elementary act of "fact creation." That is incontestable. But is that the only mechanism needed to

([35]) There is one issue with assigning a state vector $|\Psi_{\text{universe}}\rangle$. One doesn't even write down a pure quantum state for laser light when its phase is unknown; a mixed state is more appropriate [67]. It is hard to imagine why one would write down a pure state for the large-scale universe. Who would have beliefs that strict of it? Be that as it may, a pure state is certainly allowed in principle. Even people with the most unreasonable of initial beliefs (from an outsider's perspective) want to gamble consistently.

> create all that is? Is what took place at the big bang the consequence of billions upon billions of these elementary processes, these elementary "acts of observer-participancy," these quantum phenomena? Have we had the mechanism of creation before our eyes all this time without recognizing the truth? That is the larger question implicit in your comment ["Is the big bang here?"].

When cosmology hails from the inside, the world stands a chance of being anything it wants to be.

11. – The future

> It is difficult to escape asking a challenging question. Is the entirety of existence, rather than being built on particles or fields of force or multidimensional geometry, built upon billions upon billions of elementary quantum phenomena, those elementary acts of "observer-participancy," those most ethereal of all the entities that have been forced upon us by the progress of science?
>
> — John Archibald Wheeler

Imagine our universe at a time when there were no agents about to use the laws of probability theory as an aid in their gambles—*i.e.*, before the cruel creativity of Darwinian selection had brought forth any such agents. Were there any quantum states in the universe then? We QBists say *No.* It's not a matter of the quantum state of the universe waiting until a qualified PhD student came along before having its first collapse, as John Bell joked, but that there simply weren't any quantum states. Indeed, though we know little about elsewhere under Heaven, here on Earth there weren't any quantum states until 1926 when Erwin Schrödinger wrote the first one down. The reason is simple: The universe is made of something else than ψ-flavored gelatin. But then, what of the Born rule? To this, in contrast, a QBist would say, "Aha, now there's a sensible question." For the Born rule is among the set of relations an agent should strive to attain in his larger mesh of probability assignments. That normative rule indicates the character of the natural world, a character that is present even when there are no agents to make use of it. As Craig Callender once paraphrased it, in QBism, it is the normative rule which is *nature's whisper*, not the specific terms within it.

Any of us can use quantum theory, but you can only use it for yourself. By way of analogy, consider the single-celled organisms called *Euglena.* These are "flagellate protists"—microbes with tails coming off of them. The tail arose from evolutionary pressures, so that a Euglena can move from environments where there are depleted nutrients to environments where there's an abundance of nutrients. It's a tool. Quantum mechanics is like the Euglena's tail. It's something we evolved in the 1920s, and since it's been shown to be such a good tool, we keep using it and we pass it on to our children.

The tail of a Euglena is a single-user tail. But we can look at the tail and ask, "What might we learn about the environment by studying the tail's structure?" We might notice the tail is not completely circular, and that might tell us something about the viscosity of the medium the Euglena travels through. We might look at the ratio of the length of it to the width of it in various places, and that might tell us about features of the microbial environment. Likewise, quantum mechanics is a single-user theory, but by dissecting it, we can learn something about the world that all of us are immersed in.

In this way, QBism is carrying out what Einstein called the program of the real [146]:

> A basic conceptual distinction, which is a necessary prerequisite of scientific and pre-scientific thinking, is the distinction between "sense-impressions" (and the recollection of such) on the one hand and mere ideas on the other. There is no such thing as a conceptual definition of this distinction (aside from circular definitions, *i.e.*, of such as make a hidden use of the object to be defined). Nor can it be maintained that at the base of this distinction there is a type of evidence, such as underlies, for example, the distinction between red and blue. Yet, one needs this distinction in order to be able to overcome solipsism. Solution: we shall make use of this distinction unconcerned with the reproach that, in doing so, we are guilty of the metaphysical "original sin." We regard the distinction as a category which we use in order that we might the better find our way in the world of immediate sensations. The "sense" and the justification of this distinction lies simply in this achievement. But this is only a first step. We represent the sense-impressions as conditioned by an "objective" and by a "subjective" factor. For this conceptual distinction there also is no logical-philosophical justification. But if we reject it, we cannot escape solipsism. It is also the presupposition of every kind of physical thinking. Here too, the only justification lies in its usefulness. We are here concerned with "categories" or schemes of thought, the selection of which is, in principle, entirely open to us and whose qualification can only be judged by the degree to which its use contributes to making the totality of the contents of consciousness "intelligible." The above mentioned "objective factor" is the totality of such concepts and conceptual relations as are thought of as independent of experience, viz., of perceptions. So long as we move within the thus programmatically fixed sphere of thought we are thinking physically. Insofar as physical thinking justifies itself, in the more than once indicated sense, by its ability to grasp experiences intellectually, we regard it as "knowledge of the real."
>
> After what has been said, the "real" in physics is to be taken as a type of program, to which we are, however, not forced to cling *a priori*.

There is so much still to do with the physics of QBism, and this article has just started scratching the surface. Just one example: The technical problems with SICs

are manifest. For instance, there must be a reason a proof of their existence has been so recalcitrant. An optimist would say it is because they reach so deeply into the core of what the quantum is telling us! In any case, we do suspect that when we get the structure of SICs down pat, eq. (20), though already so essential to QBism's distillation of quantum theory's message, will seem like child's play in comparison to the vistas the further knowledge will open up.

But the technical also complements and motivates the conceptual. So far we have only given the faintest hint of how QBism should be mounted onto a larger empiricism. It will be noticed that QBism has been quite generous in treating agents as physical objects when needed. "I contemplate you as an agent when discussing your experience, but I contemplate you as a physical system before me when discussing my own". Our solution to "Wigner's friend" is the great example of this. Precisely because of this, however, QBism knows that its story cannot end as a story of gambling agents—that is only where it starts. Agency, for sure, is not a derivable concept as the reductionists and vicious abstractionists would have it, but QBism, like all of science, should strive for a Copernican principle whenever possible. We have learned so far from quantum theory that before an agent the world is really malleable and ready through their intercourse to give birth. Why would it not be so for every two parts of the world? And this newly defined valence, quantum dimension, might it not be a measure of a system's potential for creation when it comes into relationship with those other parts?

In this article, we have focused on what QBism has to say about small quantum systems. This is, in part, an example of QBism showing its ancestry in quantum information theory. But, as the science journalist James Burke once said, *the revolutionary ideal admits no half-measures*: The lessons of QBism must apply more broadly than qubits and qutrits. What does taking the principles of QBism on board mean for how one thinks about special relativity [7]? What about the practice of renormalization in statistical physics and field theory [147]? Or classical probability, information theory and machine learning [148, 149]? The study of SICs has already changed the shape of the boundary between physics and pure mathematics [95, 104, 106]. How far must the changes go? What, we might even ask, do Jamesian pragmatism, empiricism and radical metaphysical pluralism mean for the nature of mathematical truth?

It is a large research program whose outline is just taking shape. It hints of a world, a pluriverse, that consists of an all-pervading "pure experience," as William James called it([36]). Expanding this notion, making it technical, and letting its insights tinker with spacetime itself is the better part of future work. Quantum states, QBism declares, are not the stuff of the world, but quantum *measurement* might be. Might a one-day future Shakespeare write with honesty,

([36]) Aside from James's originals [114, 117], further reading on this concept and related subjects can be found in refs. [150-156].

Our revels are now ended. These our actors,
As I foretold you, were all spirits and
Are melted into air, into thin air skip ...
We are such stuff as
quantum measurement is made on.

* * *

This research was supported by the Foundational Questions Institute Fund on the Physics of the Observer (grant FQXi-RFP-1612), a donor advised fund at the Silicon Valley Community Foundation. Work reported in sect. **7** was supported in part by the U. S. Office of Naval Research (Grant No. N00014-09-1-0247).

REFERENCES

[1] Caves C. M., Fuchs C. A. and Schack R., *Phys. Rev. A*, **65** (2002) 022305.
[2] Fuchs C. A., *Quantum mechanics as quantum information (and only a little more)*, in proc. of *Quantum Theory: Reconsideration of Foundations*, edited by Khrennikov A. (Växjö University Press) 2002, p. 463.
[3] Fuchs C. A. and Schack R., *Unknown quantum states and operations, a Bayesian view*, in proc. of *Quantum Estimation Theory*, edited by Paris M. G. A. and Řeháček J. (Springer-Verlag) 2004, p. 151.
[4] Caves C. M., Fuchs C. A. and Schack R., *Stud. Hist. Philos. Mod. Phys.*, **38** (2007) 255.
[5] Fuchs C. A. and Schack R., *Rev. Mod. Phys.*, **85** (2013) 1693, arXiv:0906.2187 [quant-ph].
[6] Fuchs C. A., *Coming of Age with Quantum Information* (Cambridge University Press) 2010.
[7] Mermin N. D., *QBism as CBism: Solving the Problem of 'the Now'*, (2013) n.p., arXiv:1312.7825 [quant-ph].
[8] Fuchs C. A., *My Struggles with the Block Universe* (Preprint) 2014, arXiv:1405.2390 [quant-ph].
[9] Fuchs C. A., Mermin N. D. and Schack R., *Am. J. Phys.*, **82** (2014) 749, arXiv:1311.5253 [quant-ph].
[10] Mermin N. D., *Why QBism is not the Copenhagen interpretation and what John Bell might have thought of it*, in *Quantum [Un]Speakables II* (Springer-Verlag) 2016, arXiv:1409.2454 [quant-ph].
[11] Stacey B. C., *Philos. Trans. Roy. Soc. A*, **374** (2016) 20150235, arXiv:1412.2409 [physics.hist-ph].
[12] von Baeyer H. C., *QBism: The Future of Quantum Physics* (Harvard University Press) 2016.
[13] Healey R., *Quantum-Bayesian and pragmatist views of quantum theory* in *The Stanford Encyclopedia of Philosophy*, edited by Zalta E. N., Winter 2016 edition (Metaphysics Research Lab, Stanford University) 2016, p. N/A, `https://plato.stanford.edu/archives/win2016/entries/quantum-bayesian/`.
[14] Feynman R. P., Morinigo F. B. and Wagner W. G., *The Feynman Lectures on Gravitation* (Westview Press, Boulder, CO) 2003.

[15] GISIN N., *Why Bohmian Mechanics? One and two-time position measurements, Bell inequalities, philosophy and physics*, (2015) n.p., arXiv:1509.00767 [quant-ph].
[16] CAVES C. M. and SCHACK R., *Ann. Phys.*, **315** (2005) 123, arXiv:quant-ph/0409144.
[17] KENT A., *Found. Phys.*, **45** (2014) 211.
[18] KASTNER R. E., *Stud. Hist. Philos. Mod. Phys.*, **48** (2014) 56, arXiv:1406.4126 [quant-ph].
[19] ADLAM E., *Stud. Hist. Philos. Mod. Phys.*, **47** (2014) 21.
[20] JANSSON L., *Stud. Hist. Philos. Mod. Phys.*, **53** (2016) 45.
[21] SAUNDERS S., *What is probability?*, in *Quo Vadis Quantum Mechanics?* (Springer) 2005, p. 209, arXiv:quant-ph/0409144.
[22] WALLACE D., *Preprint*, (2009) n.p., arXiv:0906.2718 [quant-ph].
[23] PRICE H., *Decisions, decisions, decisions: Can Savage salvage Everettian probability?*, in *Many Worlds? Everett, Quantum Theory and Reality* (Oxford University Press) 2010, p. 369, arXiv:0802.1390 [quant-ph].
[24] KENT A., *One World versus Many: The inadequacy of Everettian accounts of evolution, probability, and scientific confirmation*, in *Many Worlds? Everett, Quantum Theory and Reality* (Oxford University Press) 2010, p. 307, arXiv:0905.0624 [quant-ph].
[25] HARTLE J. B., *Am. J. Phys.*, **36** (1968) 704.
[26] SPEKKENS R. W., *Phys. Rev. A*, **75** (2007) 032110.
[27] FUCHS C. A., *QBism, The Perimeter of Quantum Bayesianism*, (2010) n.p., arXiv:1003.5209 [quant-ph].
[28] GOOD I. J., *46656 varieties of bayesians*, in *Good Thinking: The Foundations of Probability and Its Applications* (University of Minnesota Press) 1983, p. 20.
[29] MENAND L., *The Metaphysical Club: A Story of Ideas in America* (Farrar, Straus and Giroux) 2001.
[30] BERNARDO J. M. and SMITH A. F. M., *Bayesian Theory* (Wiley) 1994.
[31] LINDLEY D. V., *Understanding Uncertainty* (Wiley) 2006.
[32] DE FINETTI B., *Theory of Probability* (Wiley) 1990.
[33] VON MISES R., *Die Naturwissenschaften*, **10** (1922) n.p., translated in STÖLTZNER M., "Vienna Indeterminism II: From Exner's Synthesis to Frank and von Mises," in *Logical Empiricism. Historical and Contemporary Perspectives*, edited by PARRINI P., SALMON W. and SALMON M. (University of Pittsburgh Press) 2003, p. 194.
[34] WIGNER E. P., *Remarks on the mind-body question*, in *Symmetries and Reflections: Scientific Essays of Eugene P. Wigner* (Ox Bow Press) 1979, p. 171.
[35] ALBERT D. Z., *Quantum Mechanics and Experience* (Harvard University Press) 1994.
[36] NIELSEN M. A. and CHUANG I. L., *Quantum Computation and Quantum Information* (Cambridge University Press) 2000.
[37] LEIFER M. S., *Conditional density operators and the subjectivity of quantum operations*, in proc. of *Foundations of Probability and Physics – 4*, edited by ADENIER G. *et al.* (American Institute of Physics, Melville, NY) 2007, p. 438.
[38] PAULI W., *Writings on Physics and Philosophy* edited by ENZ C. P. and VON MEYENN K. (Springer-Verlag) 1994.
[39] WHEELER J. A., *Bohr, Einstein, and the strange lesson of the quantum*, in proc. of *Mind in Nature: Nobel Conference XVII, Gustavus Adolphus College, St. Peter, Minnesota*, edited by ELVEE R. Q. (Harper & Row, San Francisco, CA) 1982, pp. 1–23, with discussions pp. 23–30, 88–89, 112–113, and 148–149.
[40] JAYNES E. T., *Probability Theory: The Logic of Science* (Cambridge University Press) 2003.
[41] PERES A., *Quantum Theory: Concepts and Methods* (Kluwer, Dordrecht) 1995.
[42] GOLDSTEIN H., *Classical Mechanics* (Pearson Education India) 1965.

[43] Caves C. M. and Fuchs C. A., *Ann. Israel Phys. Soc.*, **12** (1996) 226.
[44] Wigner E. P., *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra* (Elsevier) 2012.
[45] Wigner E. P., *The probability of the existence of a self-reproducing unit*, in *The Logic of Personal Knowledge: Essays Presented to Michael Polanyi on his Seventieth Birthday* (Routledge & Kegan Paul, London) 1961, pp. 231–38.
[46] Gisin N., *Science*, **326** (2009) 1357.
[47] Albert D. Z. and Galchen R., *Sci. Am.*, **300** (2009) 32.
[48] Norsen T., *Found. Phys. Lett.*, **19** (2006) 633.
[49] Maudlin T., *J. Phys. A*, **47** (2014) 424010.
[50] Einstein A., *Dialectica*, **2** (1948) 320, passage translated in D. Howard, "Einstein on Locality and Separability," *Stud. Hist. Philos. Sci. Pt. A*, **16** (1985) 171.
[51] Fine A., *The Shaky Game: Einstein, Realism and the Quantum Theory*, 2nd edition (University of Chicago Press) 1996.
[52] Harrigan N. and Spekkens R. W., *Found. Phys.*, **40** (2010) 125, arXiv:0706.2661 [quant-ph].
[53] Einstein A., *Autobiographical notes*, in *Albert Einstein: Philosopher-Scientist*, edited by Schilpp P. A. (Tudor Publishing Co., New York) 1949.
[54] Stairs A., *Philos. Sci.*, **50** (1983) 578.
[55] Conway J. and Kochen S., *Found. Phys.*, **36** (2006) 1441.
[56] Conway J. and Kochen S., *Not. AMS*, **56** (2009) 226.
[57] Cabello A., Estebaranz J. M. and García-Alcaine G., *Phys. Lett. A*, **212** (1996) 183.
[58] Primas H., *Beyond Baconian quantum physics*, in *Kohti uutta todellisuuskäsitystä. Juhlakirja professori Laurikaisen 75-vuotispäivänä (Towards a New Conception of Reality. Anniversary Publication to Professor Laurikainen's 75th Birthday)*, edited by Ketvel U. (Yliopistopaino, Helsinki) 1990, p. 100.
[59] Caves C. M., Fuchs C. A. and Schack R., *J. Math. Phys.*, **43** (2002) 4537.
[60] Schack R., Brun T. A. and Caves C. M., *Phys. Rev. A*, **64** (2001) 014305.
[61] Fuchs C. A., Schack R. and Scudo P. F., *Phys. Rev. A*, **69** (2004) 062305, arXiv:quant-ph/0307198.
[62] Lo H. K., Chau H. F. and Ardehali M., *J. Crypto.*, **18** (2005) 133.
[63] Renner R., *Nat. Phys.*, **3** (2007) 645.
[64] Renner R., *Int. J. Quant. Inf.*, **6** (2008) 1.
[65] Doherty A. C., Parrilo P. A. and Spedalieri F. M., *Phys. Rev. A*, **71** (2005) 032333.
[66] van Enk S. J., Lütkenhaus N. and Kimble H. J., *Phys. Rev. A*, **75** (2007) 052318.
[67] van Enk S. J. and Fuchs C. A., *Phys. Rev. Lett.*, **88** (2002) 027902//1.
[68] van Enk S. J. and Fuchs C. A., *Quant. Inf. Comp.*, **2** (2002) 151.
[69] Ferrie C. and Emerson J., *New J. Phys.*, **11** (2009) 063040.
[70] Wootters W. K., *Found. Phys.*, **16** (1986) 391.
[71] Appleby D. M., Dang H. B. and Fuchs C. A., *Entropy*, **16** (2007) 1484, arXiv:0707.2071 [quant-ph].
[72] Zauner G., *Quantum designs – foundations of a non-commutative theory of designs*, PhD Thesis, University of Vienna (1999).
[73] Caves C. M., *Symmetric informationally complete POVMs*, posted at `http://info.phys.unm.edu/~caves/reports/infopovm.pdf` (1999).
[74] Renes J. M., Blume-Kohout R., Scott A. J. and Caves C. M., *J. Math. Phys.*, **45** (2004) 2171.
[75] Scott A. J. and Grassl M., *J. Math. Phys.*, **51** (2009) 042203, arXiv:0910.5784 [quant-ph].

[76] FUCHS C. A. and SASAKI M., *Quant. Inf. Comp.*, **3** (2003) 377.
[77] SCOTT A. J., *J. Phys. A*, **39** (2006) 13507.
[78] WOOTTERS W. K. and SUSSMAN D. M., *Discrete phase space and minimum-uncertainty states*, (2007) n.p., arXiv:0704.1277 [quant-ph].
[79] ZHU H., *Phys. Rev. Lett.*, **117** (2016) 120404, arXiv:1604.06974 [quant-ph].
[80] APPLEBY D. M., FUCHS C. A. and ZHU H., *Quant. Inf. Comp.*, **15** (2015) 61, arXiv:1312.0555 [quant-ph].
[81] LING A., PANG S. K., LAMAS-LINARES A. and KURTSIEFER C., *Phys. Rev. A*, **74** (2006) 022309.
[82] MEDENDORP Z. E. D., TORRES-RUIZ F. A., SHALM L. K., FUCHS C. A. and STEINBERG A. M., *Characterizing a Qutrit Directly with Symmetric Informationally Complete (SIC) POVMs* (Optical Society of America) 2010.
[83] BENT N., QASSIM H., TAHIR A. A., SYCH D., LEUCHS G., SÁNCHEZ-SOTO L. L., KARIMI E. and BOYD R. W., *Phys. Rev. X*, **5** (2015) 041006.
[84] BACCIAGALUPPI G. and CRULL E., *Stud. Hist. Philos. Mod. Phys.*, **40** (2009) 374.
[85] BUB J., *Stud. Hist. Philos. Mod. Phys.*, **38** (2007) 232.
[86] JOZSA R., *IBM J. Res. Dev.*, **48** (2004) 79.
[87] APPLEBY D. M., ERICSSON Å. and FUCHS C. A., *Found. Phys.*, **41** (2011) 564, arXiv:0910.2750 [quant-ph].
[88] FUCHS C. A. and SCHACK R., *Found. Phys.*, **41** (2011) 345, arXiv:0912.4252 [quant-ph].
[89] APPLEBY D. M., FLAMMIA S. T. and FUCHS C. A., *J. Math. Phys.*, **52** (2011) 022202, arXiv:1001.0004 [quant-ph].
[90] APPLEBY M., FUCHS C. A., STACEY B. C. and ZHU H., *Introducing the Qplex: A Novel Arena for Quantum Theory*, (2016) n.p., arXiv:1612.03234 [quant-ph].
[91] STUECKELBERG E. C. G., *Helv. Phys. Acta*, **33** (1960) 727.
[92] ADLER S. L., *Quaternionic Quantum Mechanics and Quantum Fields* (Oxford University Press) 1995.
[93] KHATIRINEJAD M., *Regular structures of lines in complex space*, PhD Thesis, Simon Fraser University (2008).
[94] CHIRIBELLA G., CABELLO A. and KLEINMANN M., *The observer observed: a Bayesian route to the reconstruction of quantum theory*, `http://fqxi.org/grants/large/awardees/view/_details/2016/chiribella` (2016).
[95] APPLEBY D. M., FLAMMIA S., MCCONNELL G. and YARD J., *Generating ray class fields of real quadratic fields via complex equiangular lines*, (2016) n.p., arXiv:1604.06098 [math.NT].
[96] STACEY B. C., *Sporadic SICs and the normed division algebras*, (2016) n.p., arXiv:1605.01426 [quant-ph].
[97] BERNSTEIN J., *Am. J. Phys.*, **73** (2005) 999.
[98] FEDAK W. A. and PRENTIS J. J., *Am. J. Phys.*, **77** (2009) 128.
[99] SCHOLZ E., *Weyl entering the "new" quantum mechanics discourse*, in Proceedings of *HQ-1: Conference on the History of Quantum Physics*, edited by JOAS C., LEHNER C. and RENN J. (Max Planck Institute for the History of Science) 2008, pp. 249–67. URL `http://www.mpiwg-berlin.mpg.de/Preprints/P350.PDF\#page=262`.
[100] WEYL H., *The Theory of Groups and Quantum Mechanics* (Dover) 1927.
[101] SCHWINGER J., *Quantum Kinematics and Dynamics* (Addison Wesley Longman) 1970.
[102] ZHU H., *Quantum state estimation and symmetric informationally complete POMs*, PhD Thesis, National University of Singapore (2012).
[103] SZYMUSIAK A. and SŁOMCZYŃSKI W., *Phys. Rev. A*, **94** (2016) 012122, arXiv:1512.01735 [quant-ph].

[104] Stacey B. C., *Geometric and Information-Theoretic Properties of the Hoggar Lines*, (2016) n.p., arXiv:1609.03075 [quant-ph].
[105] Stacey B. C., *Mathematics*, **4** (2016) 36, arXiv:1404.3774 [quant-ph].
[106] Bengtsson I., *The Number Behind the Simplest SIC-POVM*, (2016) n.p., arXiv:1611.09087 [quant-ph].
[107] Schappacher N., *On the history of Hilbert's twelfth problem: A comedy of errors*, in *Matériaux pour lhistoire des mathématiques au XX*e *siècle* (Soc. Math. France, Paris) 1998, p. 243.
[108] Baez J. C. and Joyce H., *Ubiquitous octonions, Plus Magazine*, `https://plus.maths.org/content/ubiquitous-octonions` (2005).
[109] Viazovska M., *The sphere packing problem in dimension 8*, (2016) n.p., arXiv:1603.04246 [math.NT].
[110] Norsen T., *Quantum solipsism and non-locality*, `http://www.ijqf.org/wps/wp-content/uploads/2014/12/Norsen-Bell-paper.pdf` (2014).
[111] Rorty R., *Philosophical Papers, Volume 1: Objectivity, Relativism, and Truth* (Cambridge University Press) 1991.
[112] Dupré J., *The Disorder of Things: Metaphysical Foundations of the Disunity of Science* (Harvard University Press) 1993.
[113] Cartwright N., *The Dappled World: A Study of the Boundaries of Science* (Cambridge University Press) 1999.
[114] James W., *A Pluralistic Universe* (University of Nebraska Press, Lincoln, NB) 1996.
[115] Wahl J., *The Pluralist Philosophies of England and America* (Open Court, London) 1925, translated by F. Rothwell.
[116] James W., *Some Problems of Philosophy* (Longmans, Green, and Co., London) 1940.
[117] James W., *Essays in Radical Empiricism* (University of Nebraska Press, Lincoln, NB) 1996.
[118] James W., *The dilemma of determinism*, in *The Will to Believe and Other Essays in Popular Philosophy; Human Immortality—Both Books Bound as One* (Dover) 1884, p. 145.
[119] Thayer H. S., *Meaning and Action: A Critical History of Pragmatism*, 2nd edition (Hackett Publishing Co., Indianapolis) 1981.
[120] James W., *On some Hegelisms*, in *The Will to Believe and Other Essays in Popular Philosophy; Human Immortality—Both Books Bound as One* (Dover) 1882, p. 263.
[121] James W., *Is life worth living?*, in *The Will to Believe and Other Essays in Popular Philosophy; Human Immortality—Both Books Bound as One* (Dover) 1884, p. 32.
[122] Feynman R. P., *The Character of Physical Law* (MIT Press) 1965.
[123] James W., *Abstractionism and "relativismus"*, in *The Meaning of Truth* (Prometheus Books, Amherst, NY) 1997, p. 246.
[124] von Baeyer H. C., *Petite Leçons de Physique dans les Jardins de Paris* (Dunod, Paris) 2009.
[125] Hardy L., *Quantum Theory from Five Reasonable Axioms*, (2001) n.p., arXiv:quant-ph/0101012.
[126] Hardy L., *Why quantum theory?*, in *Non-locality and Modality*, edited by Placek T. and Butterfield J. (Kluwer, Dordrecht) 2002, p. 61.
[127] Dakić B. and Č. Brukner,, *Quantum theory and beyond: Is entanglement special?*, in *Deep Beauty: Understanding the Quantum World through Mathematical Innovation*, edited by Halvorson H. (Cambridge University Press) 2011, pp. 365–92, arXiv:0911.0695 [quant-ph].
[128] Fuchs C. A., *Quant. Inf. Comp.*, **4** (2004) 467.

[129] BRUNNER N., PIRONIO S., ACIN A., GISIN N., MÉTHOT A. A. and SCARANI V., *Phys. Rev. Lett.*, **100** (2008) 210503.
[130] WEHNER S., CHRISTANDL M. and DOHERTY A. C., *Phys. Rev. A*, **78** (2008) 062112.
[131] WOLF M. M. and PEREZ-GARCIA D., *Phys. Rev. Lett.*, **102** (2009) 190504.
[132] BLUME-KOHOUT R., CAVES C. M. and DEUTSCH I. H., *Found. Phys.*, **32** (2002) 1641.
[133] CERF N. J., BOURENNANE M., KARLSSON A. and GISIN N., *Phys. Rev. Lett.*, **88** (2002) 127902.
[134] HOROWITZ G. T. and MALDACENA J., *J. High Energy Phys.*, **2004** (2004) 008.
[135] GOTTESMAN D. and PRESKILL J., *J. High Energy Phys.*, **2004** (2004) 026.
[136] PRAWER S. and GREENTREE A. D., *Science*, **320** (2008) 1601.
[137] OJIMA I., *Acta Inst. Philos. Aesth.*, **10** (1992) 55.
[138] SCHLOSSHAUER M., *Decoherence and the Quantum-to-Classical Transition* (Springer-Verlag) 2007.
[139] SCHLOSSHAUER M. and CAMILLERI K., *Stud. Hist. Philos. Mod. Phys.*, **49** (2015) 73.
[140] BARNUM H., *Quantum Knowledge, Quantum Belief, Quantum Reality. Notes of a QBist Fellow Traveler*, (2010) n.p., arXiv:1003.4555 [quant-ph].
[141] KOFLER J., *Quantum violation of macroscopic realism and the transition to classical physics*, PhD Thesis, University of Vienna (2008), arXiv:0812.0238 [quant-ph].
[142] HAAG R., *Eur. Phys. J. H*, **35** (2010) 263.
[143] BEKENSTEIN J., *Scholarpedia*, **3** (2008) 7374, `http://www.scholarpedia.org/article/Bekenstein_bound`.
[144] DAVIES P. C. W. and BROWN J. R. (EDITORS), *The Ghost in the Atom: A Discussion of the Mysteries of Quantum Physics* (Cambridge University Press) 1986.
[145] JAMES W., *Pragmatism, a New Name for Some Old Ways of Thinking: Popular Lectures on Philosophy* (Longmans, Green and Co., New York) 1922.
[146] EINSTEIN A., *Remarks concerning the essays brought together in this co-operative volume*, in *Albert Einstein: Philosopher-Scientist*, edited by SCHILPP P. A. (Tudor Publishing Co., New York) 1949, pp. 665–88.
[147] DEBROTA J. B., *A Quantum Information Geometric Approach to Renormalization*, (2016) n.p., arXiv:1609.09440 [quant-ph].
[148] FUCHS C. A. and SCHACK R., *Bayesian conditioning, the reflection principle, and quantum decoherence*, in *Probability in Physics* (Springer) 2012, pp. 233–47, arXiv:1103.5950 [quant-ph].
[149] STACEY B. C., *Multiscale structure in eco-evolutionary dynamics*, PhD Thesis, Brandeis University (2016), arXiv:1509.02958 [q-bio.PE].
[150] LAMBERTH D. C., *William James and the Metaphysics of Experience* (Cambridge University Press) 1999.
[151] TAYLOR E. and WOZNIAK R. H., *Pure Experience: The Response to William James* (Thoemmes Press, Bristol, UK) 1996.
[152] WILD J., *The Radical Empiricism of William James* (Doubleday, Garden City, NY) 1969.
[153] GIESER S., *The Innermost Kernel: Depth Psychology and Quantum Physics. Wolfgang Pauli's Dialogue with C. G. Jung* (Springer) 2005.
[154] RUSSELL B., *The Analysis of Mind* (Arc Manor, Rockville, MD) 2008.
[155] BANKS E. C., *Ernst Mach's World Elements: A Study in Natural Philosophy* (Kluwer, Dordrecht) 2003.
[156] HEIDELBERGER M., *Nature from Within: Gustav Theodor Fechner and His Psychophysical Worldview* (University of Pittsburgh Press) 2004.

Proceedings of the International School of Physics "Enrico Fermi"
Course 197 "Foundations of Quantum Theory", edited by E. M. Rasel, W. P. Schleich and S. Wölk
(IOS, Amsterdam; SIF, Bologna) 2019
DOI 10.3254/978-1-61499-937-9-203

Information-theoretic derivation of free quantum field theory(*)

Giacomo Mauro D'Ariano(**)

QUit Group, Dipartimento di Fisica and INFN gruppo IV
Via Bassi 6, 27100 Pavia, Italy

Summary. — Recently the new information-theoretic paradigm of physics advocated by John Archibald Wheeler and Richard Feynman has concretely shown its full power, with the derivation of quantum theory and of free quantum field theory from informational principles. The paradigm has opened for the first time the possibility of avoiding physical primitives in the axioms of the physical theory, allowing a re-foundation of the whole physics over logically solid grounds. In addition to such methodological value, the new information-theoretic derivation of quantum field theory is particularly interesting for establishing a theoretical framework for quantum gravity, with the idea of obtaining gravity itself as emergent from the quantum information processing, as also suggested by the role played by information in the holographic principle. In this lecture notes I review how free quantum field theory is derived without using mechanical primitives, including space-time, special relativity, Hamiltonians, and quantization rules. The theory is provided by the simplest quantum algorithm encompassing a countable set of quantum systems whose network of interactions satisfies the three following simple principles:

(*) The present notes contain only the second of the two lectures delivered by the author at the Varenna School, Course 197 "Foundations of Quantum Theory" (2016): for the first lecture see the textbook [1]. The present notes are an adapted version of the original review [2] in memoriam of David Finkelstein (D'Ariano G. M., *Int. J. Theor. Phys.*, **56**, Issue 1 (2017) 97, © 2016, Springer Nature).

(**) E-mail: dariano@unipv.it

homogeneity, locality, and isotropy. The inherent discrete nature of the informational derivation leads to an extension of quantum field theory in terms of a quantum cellular automata and quantum walks. A simple heuristic argument sets the scale to the Planck one, and the currently observed regime where discreteness is not visible is the so-called "relativistic regime" of small wave vectors, which holds for all energies ever tested (and even much larger), where the usual free quantum field theory is perfectly recovered. In the present quantum discrete theory Einstein relativity principle can be restated without using space-time in terms of invariance of the eigenvalue equation of the automaton/walk under change of representations. Distortions of the Poincaré group emerge at the Planck scale, whereas special relativity is perfectly recovered in the relativistic regime. Discreteness, on the other hand, has some plus compared to the continuum theory: 1) it contains it as a special regime; 2) it leads to some additional features with GR flavor: the existence of an upper bound for the particle mass (with physical interpretation as the Planck mass), and a global De Sitter invariance; 3) it provides its own physical standards for space, time, and mass within a purely mathematical adimensional context. The lecture ends with future perspectives.

1. – Introduction

The logical clash between General Relativity (GR) and Quantum Field Theory (QFT) is the main open problem in physics. The two theories represent our best theoretical frameworks, and work astonishingly well within the physical domain for which they have been designed. However, their logical clash requires us to admit that they cannot be both correct. One could argue that there must exist a common theoretical substratum from which both theories emerge as approximate effective theories in their pertaining domains—though we know very little about GR in the domain of particle physics.

What we should keep and what we should reject of the two theories? Our experience has thought us that of QFT we should definitely keep the Quantum Theory (QT) of abstract systems, namely the theory of the von Neumann book [10] stripped of its "mechanical" part, *i.e.* the Schrödinger equation and the quantization rules. This leaves us with the description of generic systems in terms of Hilbert spaces, unitary transformations, and observables. In other words, this is what nowadays is also called Quantum Information, a research field indeed very interdisciplinary in physics.

There are two main reasons for keeping QT as valid. First, it has been never falsified in any experiment in the whole physical domain—independently of the scale and the kind of system. This has lead the vast majority of physicists to believe that everything must behave according to QT. The second and more relevant reason is that QT, differently from any other chapter of physics, is well axiomatized, with purely mathematical axioms containing no physical primitive. So, in a sense, QT is as valid as a piece of pure mathematics. This must be contrasted with the mechanical part of the theory, with the bad axiomatic of the so-called "quantization rules", which are extrapolated and

generalized starting from the heuristic argument of the Ehrenfest theorem, which in turn is based on the superseded theory of classical mechanics, and with the additional problem of the ordering of canonical non-commuting observables([1]). No wonder then that the quantization procedure doesn't work well for gravity!

To what we said above we should add that today we know that the QT of von Neumann can be derived from six information-theoretical principles [3, 1], whose epistemological value is not easy to give up([2]). On the contrary, it is the mechanical part of QFT that rises the main inconsistencies, *e.g.* the Malament theorem [13], which makes any reasonable notion of particle untenable [14].

The logical conclusion is that what we need is a field theory that is *quantum ab initio*. But how to avoid quantization rules? The idea is simply to consider a countable set of quantum systems in interaction, and to make the easiest assumptions on the topology of their interactions. These are: locality, homogeneity, and isotropy. Notice that we are not using any mechanics, nor relativity, and not even space and time. And what we get? We get: Weyl, Dirac [4], and Maxwell [7]. Namely: we get free quantum field theory!

The new general methodology suggested to the above experience is then the following: 1) no physical primitives in the axioms; 2) physics only as interpretation of the mathematics (based on experience, previous theories, and heuristics). In this way the logical coherence of the theory is mathematically guaranteed. In this paper we will see how the proposed methodology can be actually carried out, and how the informational paradigm has the potential of solving the conflict between QFT and GR in the case of special relativity, with the latter emergent merely from quantum systems in interaction: Fermionic quantum bits at the very tiny Planck scale. In synthesis the program is an *algorithmization of theoretical physics*, aimed to derive the whole physics from quantum algorithms with finite complexity, upon connecting the algebraic properties of the algorithm with the dynamical features of the physical theory, preparing a logically coherent framework for a theory of quantum gravity [8, 9].

Section **2** is devoted to the derivation from principles of the quantum-walk theory. More precisely, from the requirements of homogeneity and locality of the interactions of countably many quantum systems one gets a theory of quantum cellular automata on the Cayley graph of a group G. Then, upon restricting to the simple case of evolution linear in the discrete fields, the quantum automaton becomes what is called in the literature *quantum walk*. We further restrict to the case with physical interpretation in an Euclidean space, resorting to considering only Abelian G.

In sect. **3** the quantum walks with minimal field dimension that follow from the principles of sect. **2** are reported. These represent the Planck-scale version of the Weyl, Dirac, and Maxwell quantum field dynamics, which are recovered in the relativistic regime of small wave vectors. Indeed, the quantum-walk theory, being purely mathematical—and so adimensional—nevertheless contains its own physical LTM standards written in

([1]) The problem of ordering is avoided miraculously thanks to the fortuitous non-occurrence in nature of Hamiltonians with products of conjugated observables.
([2]) For short reviews, see also refs. [11, 12].

the intrinsic discreteness and non-linearities of the theory. A simple heuristic argument based on the notion of mini black-hole (from a matching of GR-QFT) leads to the Planck scale. It follows that the relativistic regime contains the whole physics observed up to now, including the most energetic events from cosmic rays.

In addition to the exact dynamics in terms of quantum walks, a simple analytical method is also available in terms of a dispersive Schrödinger equation, suitable to the Planck-scale physics for narrow-band wave-packets. As a result of the unitarity constraint for the evolution, the particle mass turns out to be upper bounded (by the Planck mass), and has domain in a circle, corresponding to having also the proper time (which is conjugated to the mass) as discrete. Effects due to discreteness that are in principle visible are also analyzed, in particular a dispersive behavior of the vacuum, that can be detected by deep-space ultra-high energy cosmic rays.

Section **4** is devoted to how special relativity is recovered from the quantum-walk discrete theory, without using space-time and kinematics. It is shown that the transformation group is a non-linear version of the Poincaré group, which recovers the usual linear group in the relativistic limit of small wave vectors. For non-vanishing masses generally also the mass gets involved in the transformations, and the De Sitter group $SO(1,4)$ is obtained.

This lecture note end with a brief section on the future perspectives of the theory.

Most of the results reported in the present review have been originally published in refs. [4-7, 15-20] coauthored with members of the QUit group in Pavia.

2. – Derivation from principles of the quantum-walk theory

The derivation from principles of quantum field theory starts from considering the unitary evolution $\mathscr{A}$ of a countable set G of quantum systems, with the requirements of homogeneity, locality, and isotropy of their mutual interactions. These will be precisely defined and analyzed in the following dedicated subsections. All the three requirements are dictated from the general principle of minimizing the algorithmic complexity of the physical law. The physical law itself is described by a finite quantum algorithm, and homogeneity and isotropy assess the universality of the law.

The quantum system labeled by $g \in G$ can be either associated to an Hilbert space $\mathscr{K}_g$, or to a set of generators of a C*-algebra([3])

$$\psi_g \equiv \{\psi_g^\nu\}, \qquad g \in G, \qquad \nu \in [s_g] := \{1, 2, \dots, s_g\}, \qquad s_g < \infty. \tag{1}$$

The evolution occurs in discrete identical steps([4])

$$\mathscr{A}\psi_g = U\psi_g U^\dagger, \qquad U \text{ unitary}, \tag{2}$$

([3]) The two associations can be connected through the GNS construction.
([4]) More generally the map $\mathscr{A}$ is an automorphism of the algebra.

describing the interactions among systems. When the unitary evolution is also *local*, namely $\mathscr{A}\psi_g$ is spanned by a finite subset $S_g \subset G$, then $\mathscr{A}$ is called *Quantum Cellular Automaton*. We restrict to evolution linear in the generators, namely

$$\mathscr{A}\psi_g = U\psi_g U^\dagger = \sum_{g'} A_{g,g'}\psi_{g'}, \tag{3}$$

where $A_{g,g'}$ is an $s_g \times s_{g'}$ complex matrix called *transition matrix*. Here in all respects the quantum cellular automaton is described by a unitary evolution on a (generally infinite) Hilbert space $\mathscr{H} = \bigoplus_{g\in G}\mathscr{H}_g$, with $\mathscr{H}_g = \mathsf{Span}\{\psi_g^\nu\}_{\nu\in[s_g]}$. In this case the quantum cellular automaton is called *quantum walk*. Here the system simply corresponds to a finite-dimensional block component of the Hilbert space, regardless the bosonic/fermionic nature of the field. In the derivation of free quantum field theory from principles, the quantum walk corresponds to the evolution on the single-particle sector of the Fock space, whereas for the interacting theory a generally non-linear quantum cellular automaton is needed. Simple generalization to Fock-space sectors with fixed number of particles are also possible.

2˙1. *The quantum system: qubit, fermion or boson?* – At the level of quantum walks, corresponding to the Fock space description of cellular quantum automata (leading to free QFT in the non-relativistic limit), it does not make any difference which kind of quantum system is evolving. Indeed one can symmetrize or anti-symmetrize products of wave functions, as it is done in usual quantum mechanics, or else just take products with no symmetrization. Things become different when the vacuum is considered, and particles are created and annihilated by operating with algebra generators on the vacuum state, as in the interacting theory. Therefore, as far as we are concerned with free QFT, which kind of quantum system should be used is a problem that can be safely postponed.

However, there are still motivations for adopting a kind of quantum system instead of another. For example, a reason for discarding qubits as algebra generators is that there is no easy way of expressing the operator U making the evolution in eq. (3) linear, whereas, when ψ_g is bosonic or fermionic this is always possible choosing U exponential of bilinear forms in the fields. On the other hand, a reason to chose fermions instead of bosons is the requirement that the amount of information in a finite number of cells be finite, namely one has finite information density in space([5]). The relation between Fermionic modes and finite-dimensional quantum systems, say *qubits* has been studied in the literature, and the two theories have been proven to be computationally equivalent [22]. However, the quantum theory of qubits and the quantum theory of fermions differ in the notion of what are local transformations [23, 24], with local fermionic operations mapped into non-local qubit transformations and vice versa.

([5]) Richard Feynman is reported to like the idea of finite information density, because he felt that: *"There might be something wrong with the old concept of continuous functions. How could there possibly be an infinite amount of information in any finite volume?"* [21].

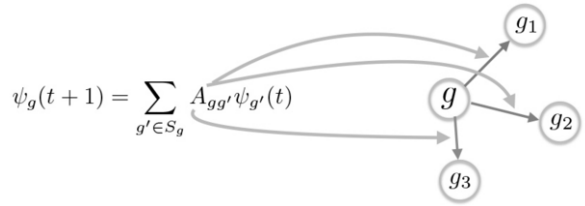


Fig. 1. – The linear eq. (3) endows the set G with a *directed graph structure*. We build a directed graph with an arrow from g to g' wherever the two are connected by a non-null matrix $A_{gg'}$ in eq. (3).

In conclusion, the derivation from informational principles of the fundamental particle statistics still remains an open problem. One could promote the finite information density to the level of a principle, or motivate the Fermionic statistics from other principles of the same nature of those in ref. [3] (see *e.g.* refs. [23, 24]), or derive the Fermionic statistics from properties of the vacuum (*e.g.* having a localized non-entangled vacuum in order to avoid the problem of particle localization), and then recover the *bosonic* statistics as a very good approximation, with the Bosonic mode corresponding to a special entangled state of pairs of Fermionic modes [7], as it will be reviewed in subsect. **3**˙9. This hierarchical construction will also guarantee the validity of the spin-statistic connection in QFT.

2˙2. *Quantum walks on Cayley graphs*([6]). – The linear eq. (3) endows the set G with a *directed graph structure* $\Gamma(G, E)$, with vertex set G and edge set $E = \{(g, g') | A_{g,g'} \neq 0\}$ directed from g to g' (see fig. 1). In the following we will denote by $S_g := \{A_{g,g'} \neq 0\}$ the set of non-null transition matrices with first index g, and by $N_g := \{g' \in G | A_{g,g'} \neq 0\}$ the *neighborhood* of g.

2˙2.1. The homogeneity principle. The assumption of homogeneity is the requirement that every two vertices are indistinguishable, namely for every $g, g' \in G$ there exists a permutation π of G such that $\pi(g) = g'$ which commute with any discrimination procedure consisting of a preparation of local modes followed by a general joint measurement. In ref. [19] it is shown that this is equivalent to the following set of conditions $\forall g \in G$ one has

H1 $s_g = s$;

H2 there exists a bijection $N_g \leftrightarrow N$ with a fixed set N;

H3 S_g contains the same $s \times s$ transition matrices, namely $S_g = S := \{A_{h_1}\}_{i=1}^{|N|}$;

H4 $A_{g,g'} = A_{h_i} \in S \Rightarrow A_{g',g} = A_{h_j} \in S$;

([6]) This subsection is based on results of refs. [4] and [18].

Condition H2 states that $\Gamma(G,E)$ is a *regular graph—i.e.* each vertex has the same degree. Condition H3 makes $\Gamma(G,E)$ a *colored directed graph*, with the arrow directed from g to g' for $A_{g,g'} = A_h \in S$ and the color associated to h([7]). Condition H3 introduces the following formal action of symbols $h_i \in S$ on the elements $g \in G$ as

$$A_{gg'} = A_{h_i} \Rightarrow gh_i = g'. \tag{4}$$

Clearly such action is closed for composition. From condition H4 one has that

$$A_{g'g} = A_{h_j} \Rightarrow g'h_j = g, \tag{5}$$

and composing the two actions we see that $gh_ih_j = g$, and we can write the label h_j as $h_j =: h_i^{-1}$. We thus can build the free group F of words made with the alphabet S. Each word corresponds to a path over $\Gamma(G,E)$, and the words $w \in F$ such that $gw = g$ correspond to closed paths (also called *loops*). Notice that by construction, one has $A_{\pi(g)\pi(f)} = A_{gf} = A_{h_i}$, which implies that $\pi(g)h_i = \pi(f) = \pi(gh_i)$, from which one can prove that $f'w = \pi(f)w = \pi(fw) = \pi(f) = f'$ (see [19]). Thus we have the following:

H5 If a path $w \in F$ is closed starting from $f \in G$, then it is closed also starting from any other $g \in G$.

The subset $R \subset F$ of words w such that $gw = g$ is obviously a group. Moreover R is a normal subgroup of G, since $gwrw^{-1} = (gw)rw^{-1} = (gw)w^{-1} = g$, namely $wrw^{-1} \in R$ $\forall w \in F, \forall r \in R$. Obviously the equivalence classes are just elements of G, which means that $G = F/R$ is a group. Pick up any element of G as the identity $e \in G$. It is clear that the elements of the quotient group F/R are in one-to-one correspondence with the elements of G, since for every $g \in G$ there is only one class in F/R whose elements lead from e to g (write $g = ew$ for every $w \in F$, w representing a path leading from e to g). The graph $\Gamma(G,E)$ is thus what is called in the literature the *Cayley graph of the group* G (see the definition in the following). The Cayley graph is in correspondence with a *presentation* of the group G. This is usually given by arbitrarily dividing the set as $S = S_+ \cup S_-$ with $S_- := S_+^{-1}$([8]), and by considering a set W of generators for the free group of loops R. The group G is then given with the presentation $G = \langle S_+|W\rangle$, in terms of the set of its *generators* S_+ (which along with their inverses S_- generate the group by composition), and in terms of the set of its *relators* W containing group words that are equal to the identity, with the goal of using these words in W to establish if any

([7]) If two transition matrices $A_{h_1} = A_{h_2}$ are equal, we conventionally associate them with two different labels $h_1 \neq h_2$ in such a way that $\sum_{f\in N_{\pi(g)}} A_{\pi(g)f}\psi_{\pi^{-1}(f)} = \sum_{f\in N_g} A_{gf}\psi_f$. If such choice is not unique, we will pick an arbitrary one, since the homogeneity requirement implies that there exists a choice of labeling for which all the construction that will follow is consistent.
([8]) The above arbitrariness is inherent the very notion of group presentation and corresponding Cayley graph, and will be exploited in the following, in particular in the definition of isotropy.

two words of elements of G correspond to te same group element. The relators can also be regarded as a set of generators for R.

The definition of *Cayley graph* is then the following.

Cayley graph of G.. Given a group G and a set S_+ of generators of the group, the Cayley graph $\Gamma(G, S_+)$ is defined as the colored directed graph with vertex set G, edge set $\{(g, gh); g \in G, h \in S_+\}$ with the edge directed from g to gh with color assigned by h (when $h = h^{-1}$ we conventionally draw an undirected edge).

Notice that a Cayley graph in addition to being a *regular* graph, it is also *vertex-transitive*—*i.e.* all sites are equivalent, in the sense that the graph automorphism group acts transitively upon its vertices. The Cayley graph is also called *arc-transitive* when its group of automorphisms acts transitively not only on its vertices but also on its directed edges.

2˙2.2. The locality principle. Locality corresponds to require that the evolution is completely determined by a rule involving a finite number of systems. This means having each system interacting with a finite number of systems (*i.e.*$|N| < \infty$ in H2), and having the set of loops generating F as finite and containing only finite loops. This corresponds to the fact that the group G is *finitely presented*, namely both S_+ and W are finite in $G = \langle S_+|W\rangle$.

The quantum walk then corresponds to a unitary operator over the Hilbert space $\mathscr{H} = \ell^2(G) \otimes \mathbb{C}^s$ of the form

$$A = \sum_{h\in S} T_h \otimes A_h, \tag{6}$$

where T is the *right-regular* representation of G on $\ell^2(G)$, $T_g|g'\rangle = |g'g^{-1}\rangle$.

2˙2.3. The isotropy principle. The requirement of *isotropy* corresponds to the statement that all directions on $\Gamma(G, S_+)$ are equivalent. Technically the principle affirms that there exists a choice of S_+, a group L of graph automorphisms on $\Gamma(G, S_+)$ that is transitive over S_+ and with faithful unitary (generally projective) representation U over $\mathbb{C}^s$, such that the following covariance condition holds:

$$A = \sum_{h\in S} T_h \otimes A_h = \sum_{h\in S} T_{l(h)} \otimes U_l A_h U_l^\dagger, \quad \forall\, l \in L. \tag{7}$$

As a consequence of the linear independence of the generators T_h of the right regular representation of G one has that the above condition (7) implies

$$A_{l(h^{\pm 1})} = U_l A_{h^{\pm 1}} U_l^\dagger. \tag{8}$$

Equation (8) implies that the principle of isotropy requires the Cayley graph $\Gamma(G, S_+)$ to be arc-transitive (see subsect. **2**˙2.1).

We remind that the split $S = S_+ \cup S_-$ is non-unique (and in addition one may add to S the identity element e corresponding to zero-length loops on each element corresponding to self-interactions). Therefore, generally the quantum walk on the Cayley graph $\Gamma(G, S_+)$ satisfies isotropy only for some choices of the set S_+. It happens that for the known cases satisfying all principles along with the restriction to quasi-isometric embeddability of G in Euclidean space (see subsect. **2**˙3) such choice is unique.

2˙2.4. The unitarity principle. The requirement that the evolution be unitary translates into the following set of equations bilinear in the transition matrices as unknown

$$\sum_{h\in S} A_h^\dagger A_h = \sum_{h\in S} A_h A_h^\dagger = I_s, \qquad \sum_{\substack{h,h'\in S \\ h^{-1}h'=h''}} A_h^\dagger A_{h'} = \sum_{\substack{h,h'\in S \\ h'h^{-1}=h''}} A_{h'} A_h^\dagger = 0. \tag{9}$$

Notice that the structure of equations already satisfy the homogeneity and locality principles. The solution of the systems of equations (9) is generally a difficult problem.

2˙3. *Restriction to Euclidean emergent space*. – How a discrete quantum algorithm on a graph can give rise to a continuum quantum field theory on space-time? We remind that the flow of the quantum state occurs on a Cayley graph and the evolution occurs in discrete steps. Therefore the Cayley graph must play the role of a discretized space, whereas the steps play the role of a discretized time, namely the quantum automaton/walk has an inherent Cartesian-product structure of space-time, corresponding to a particular chosen observer. We will then need a procedure for recovering the emergent space-time and a re-interpretation of the notion of inertial frame and of boost in the discrete, in order to recover Poincaré covariance and the Minkowski structure. The route for such procedure is opened by *geometric group theory*, a field in pure mathematics initiated by Mikhail Gromov at the beginning of the nineteen([9]). The founding idea is the notion of *quasi-isometric embedding*, which allows us to compare spaces with very different metrics, as for the cases of continuum and discrete. Clearly an isometric embedding of a space with a discrete metric (as for the word metric of the Cayley graph) within a

([9]) The absence of the appropriate mathematics was the reason of the lack of consideration of a discrete structure of space-time in earlier times. Einstein himself was considering this possibility and lamented such lack of mathematics. Here a passage reported by John Stachel [25]:
"*But you have correctly grasped the drawback that the continuum brings. If the molecular view of matter is the correct (appropriate) one, i.e., if a part of the universe is to be represented by a finite number of moving points, then the continuum of the present theory contains too great a manifold of possibilities. I also believe that this too great is responsible for the fact that our present means of description miscarry with the quantum theory. The problem seems to me how one can formulate statements about a discontinuum without calling upon a continuum (space-time) as an aid; the latter should be banned from the theory as a supplementary construction not justified by the essence of the problem, which corresponds to nothing "real". But we still lack the mathematical structure unfortunately. How much have I already plagued myself in this way!*"

space with a continuum metric (as for a Riemaniann manifold) is not possible. However, what Gromov realized to be geometrically relevant is the feature that the discrepancy between the two different metrics is uniformly bounded over the spaces. More precisely, one introduces the following notion of *quasi-isometry*.

Quasi-isometry. Given two metric spaces (M_1, d_1) and (M_2, d_2), with metric d_1 and d_2, respectively, a map $f : (M_1, d_1) \to (M_2, d_2)$ is a quasi-isometry if there exist constants $A \geq 1$, $B, C \geq 0$, such that $\forall g_1, g_2 \in M_1$ one has

$$\frac{1}{A} d_1(g_1, g_2) - B \leq d_2(f(g_1), \qquad f(g_2)) \leq A d_1(g_1, g_2) + B, \tag{10}$$

and $\forall m \in M_2$ there exists $g \in M_1$ such that

$$d_2(f(g), m) \leq C. \tag{11}$$

The condition in eq. (11) is also called *quasi-onto*.

It is easy to see that quasi-isometry is an equivalence relation. It can also be proved that the quasi-isometric class is an invariant of the group, *i.e.* it does not depend on the presentation, *i.e.* on the Cayley graph. Moreover, it is particularly interesting for us that for finitely generated groups, the quasi-isometry class always contains a smooth Riemaniann manifold [26]. Therefore, for a given Cayley graph there always exists a Riemaniann manifold in which it can be quasi-isometrically embedded, which is unique modulo quasi-isometries, and which depends only on the group G of the Cayley graph. Two examples are graphically represented in fig. 2.

2˙3.1. Geometric group theory. With the idea of quasi-isometric embedding, geometric group theory connects geometric properties of the embedding Riemaniann spaces with algebraic properties of the groups, opening the route to a *geometrization of group theory*, including the generally hard problem of establishing properties of a group that is given by presentation only([10]).

The possible groups G that are selected from our principles are infinitely many, and we need to restrict this set to start the search for solutions of the unitarity conditions (**2**˙3) under the isotropy constraint. Since we are interested in a theory involving infinitely many systems (we take the world as infinite!), we will consider *infinite* groups only. This means that when we consider an Abelian group, we always take it as *free*, namely its only relators are those establishing the Abelianity of the group. This is the case of $G = \mathbb{Z}^d$, with $d \geq 1$.

A paradigmatic result [26] of geometric group theory is that an infinite group G is quasi-isometric to an Euclidean space $\mathbb{R}^d$ if and only if G is *virtually-Abelian*, namely it has an Abelian subgroup $G' \subset G$ isomorphic to $\mathbb{Z}^d$ of finite index (namely with

([10]) One should consider that the Dehn's problem of establishing if two words of generators correspond to the same group element is generally undecidable. The same is true for the problem of establishing if the presentation corresponds to the trivial group!

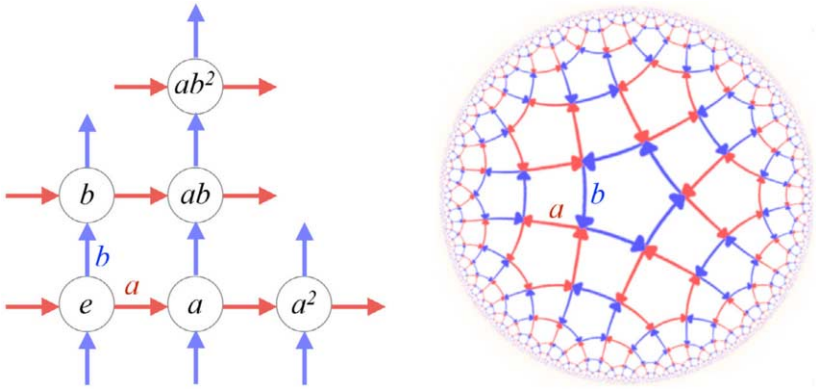


Fig. 2. – From ref. [19] (colors online). Given a group G and a set S_+ of generators, the Cayley graph $\Gamma(G, S_+)$ is defined as the colored directed graph having set of nodes G, set of edges $\{(g, gh); g \in G, h \in S_+\}$, and a color assigned to each generator $h \in S_+$. Left figure: the Cayley graph of the Abelian group $\mathbb{Z}^2$ with presentation $\mathbb{Z}^2 = \langle a, b | aba^{-1}b^{-1} \rangle$, where a and b are two commuting generators. Right figure: the Cayley graph of the non-Abelian group $G = \langle a, b | a^5, b^5, (ab)^2 \rangle$. The Abelian-group graph is embedded into the Euclidean space $\mathbb{R}^2$, the non-Abelian G into the Hyperbolic space $\mathbb{H}_2$ with negative curvature.

a finite number of cosets). Another result is that a group has polinomial growth iff it is *virtually-nihilpotent*, and if it has exponential growth then it is not virtually-nihilpotent, and in particular non-Abelian, and is quasi-isometrically embeddable in a manifold with negative curvature.

In the following we will restrict to groups that are quasi-isometrically embeddable in Euclidean spaces. As we will see soon, such restriction will indeed lead us to free quantum field theory in Euclidean space. It would be very interesting to address also the case of curved spaces, to get hints about quantum field theory in curved space. Unfortunately, the case of negative curvature corresponds to groups, as the Fuchsian group in fig. 2, whose unitary representations (that we need here) are still unknown [27-29]. The virtually-nihilpotent case also would be interesting, since it corresponds to a Riemaniann manifold with variable curvature [29], however, a Cayley graph that can satisfy the isotropy constraint could not be found yet [30].

I close this section with some comments about the remarkable closeness in spirit between the present program and the geometric group theory program. The main general goal of geometric group theory is the *geometrization of group theory*, which is achieved studying finitely-generated groups G as symmetry groups of metric spaces X, with the aim of establish connections between the algebraic structure of G with the geometric properties of X [31]. In a specular way the present program is an *algorithmization of theoretical physics*, with the general goal of deriving QFT (and ultimately the whole physics) from quantum algorithms with finite complexity, upon connecting the algebraic properties of the algorithm with the dynamical features of the physical theory. This will allow a coherent unified axiomatization of physics without physical primitives, preparing a logically coherent framework for a theory of quantum gravity.

3. – Quantum walks on Abelian groups and free QFT as their relativistic regime

As seen in subsect. **2**˙3, from the huge and yet mathematically unexplored set of possibilities for the group G of the quantum walk, we restrict to the case of G virtually-Abelian, which corresponds to G quasi-isometrically embeddable in a Euclidean space. As we will see in the present section, the free QFT that will be derived from such choice exactly corresponds to the known QFT in Euclidean space.

Since we are interested in the physics occurring in $\mathbb{R}^3$, we need to classify all possible Cayley graphs of G having $\mathbb{Z}^3$ as subgroup with finite index, and then select all graphs that allow the quantum walk to satisfy the conditions of isotropy and unitarity. We can proceed by considering increasingly large dimension $s > 0$ (defined in H1), which ultimately corresponds to the dimension of the field—*e.g.* a scalar field for $s = 1$, a spinor field for $s = 2$, etc.

3˙1. *Induced representation, and reduction from virtually-Abelian to Abelian quantum walks*. – An easy way to classify all quantum walks on Cayley graphs with virtually-Abelian groups is provided by a theorem in ref. [20], which establishes the following:

A quantum walk on the Cayley graph of a virtually-Abelian group G with Abelian subgroup $H \subset G$ of finite index i_H and dimension s is also a quantum walk on the Cayley graph of H with dimension $s' = si_H$.

This is just the *induced-representation* theorem [32-34] in group theory, here applied to quantum walks. The multiple dimension $s' = si_h$ corresponds to tiling the Cayley graph of G with a tile made with a particular choice of the cosets of H. The new set of transition matrices of the new walk for H can be straightforwardly evaluated in terms of those for G (generally self-interactions within the same tile can occur, corresponding to zero-length loops in the Cayley graph). In fig. 3 two examples of such tiling procedure are given.

The induced-representation method guarantees that scanning all possible virtually-Abelian quantum walks for increasing s is equivalent to scan all possible Abelian quantum walks, since *e.g.* the set of Abelian walks of dimension $s = nm$ will contain all virtually-Abelian walks with $s = n$ and index m, etc. We therefore resort to consider only Abelian groups.

3˙2. *Isotropy and orthogonal embedding in* $\mathbb{R}^3$. – We will also assume that the representation of the isotropy group L in (7) induced by the embedding in $\mathbb{R}^3$ is orthogonal, which implies that the graph-neighborhood is embedded in a sphere $S^2 \subset \mathbb{R}^3$ (we want homogeneity and isotropy to hold locally also in the embedding space $\mathbb{R}^3$). We are then left with the classification of the Cayley graphs of $\mathbb{Z}^3$ satisfying the isotropic embedding in $\mathbb{R}^3$: these are just the Bravais lattices.

3˙3. *Quantum walks with Abelian G.* – When G is Abelian we can greatly simplify the study of the quantum walk by using the wave vector representation, based on the fact

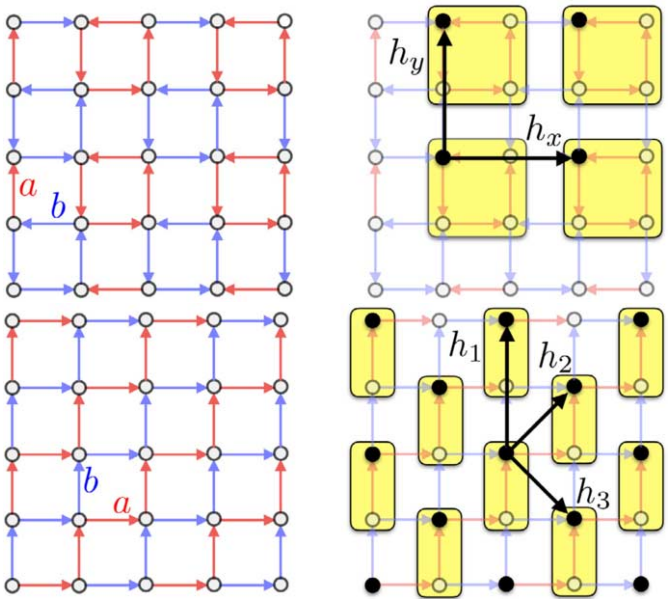


Fig. 3. – From ref. [20] (colors online). Two examples of reduction of a quantum walk on the Cayley graph of a virtually-Abelian group G to that of a quantum walk on the Cayley graph of an Abelian subgroup $H \subset G$ with finite index i_H. The graphs on the left of the figures are the Cayley graph of G (it is easy to see that both groups are non-Abelian). The graphs on the right represents a choice of the Cayley graph of the subgroup $H = \mathbb{Z}^2$, with the tiling corresponding to the induced representation (the elements of H are the black bullets). Top figures: $G = \langle a, b \mid a^4, b^4, (ab)^2 \rangle$. The index is $i_H = 4$. The subgroup generators are $h_x = a^{-1}b$ and $h_y = ba^{-1}$. The tiling is defined by the coset representatives e, a, a^2, a^3. Bottom figures: $G = \langle a, b \mid a^2 b^{-2} \rangle$. The index is $i_H = 2$. The subgroup generators are $h_1 = ba$ and $h_2 = a^2$ (or $h_1 = ba$ and $h_3 = ab^{-1}$), with the tiling the cosets representatives e, a.

that the irreducible representations of G are one-dimensional. The interesting case is for $d = 3$, but what follows holds for any dimension d. We will label the group elements by vectors $\mathbf{g} \in \mathbb{Z}^d$, and use the additive notation for the group composition, whereas the right-regular representation of $\mathbb{Z}^d$ on $\ell_2(\mathbb{Z}^d)$ will be written as $T_{\mathbf{h}}|\mathbf{g}\rangle = |\mathbf{g} - \mathbf{h}\rangle$. This can be diagonalized by Fourier transform, corresponding to write the operator A in block-form in terms of the following direct integral:

$$A = \int_B \mathrm{d}^3\mathbf{k}\, |\mathbf{k}\rangle\langle\mathbf{k}| \otimes A_{\mathbf{k}}, \qquad A_{\mathbf{k}} := \sum_{\mathbf{h}\in S} e^{-i\mathbf{k}\cdot\mathbf{h}} A_{\mathbf{h}}, \qquad |\mathbf{k}\rangle := \frac{1}{\sqrt{|B|}} \sum_{\mathbf{g}\in G} e^{-i\mathbf{k}\cdot\mathbf{g}} |\mathbf{g}\rangle, \tag{12}$$

where B is the *Brillouin zone*, and $|\mathbf{k}\rangle$ is a *plane wave*([11]). Notice that the quantum walk is unitary if and only if $A_{\mathbf{k}}$ is unitary for every $\mathbf{k} \in B$.

([11]) The Brillouin zone is a compact subset of $\mathbb{R}^3$ corresponding to the smallest region containing only inequivalent wave vectors $\mathbf{k}$. (See ref. [4] for the analytical expression.)

3˙4. *Dispersion relation.* – The spectrum $\{e^{-i\omega_{\mathbf{k}}^{(i)}}\}$ of the operator $A_{\mathbf{k}}$ is usually given in terms of the so-called *dispersion relations* $\omega_{\mathbf{k}}^{(i)}$ *versus* $\mathbf{k}$. As in usual wave-mechanics, the speed of the wave-front of a plane wave is given by the *phase velocity* $\omega_{\mathbf{k}}^{(i)}/|\mathbf{k}|$, whereas the speed of a narrow-band packet peaked around the value wave vector $\mathbf{k}_0$ is given by the *group velocity* $\nabla_{\mathbf{k}}\omega_{\mathbf{k}}^{(i)}$ evaluated at $\mathbf{k}_0$.

3˙5. *The relativistic regime.* – As we will see in subsect. **3**˙8.3 an heuristic argument will lead us to set the scale of discreteness of the quantum walk (and similarly the quantum cellular automaton for the interacting theory) at the Planck scale. The domain $|\mathbf{k}| \ll 1$ then corresponds to wave vectors much smaller than the Planck vector, which is much higher than any ever observed wave vector([12]). Such regime includes that of usual particle physics, and is called *relativistic regime*. To be precise, the regime is defined by a set of wave-packets that are peacked around $\mathbf{k} = 0$ with r.m.s. value much smaller than the Planck wave vector, which we will refer shortly to as *narrow-band wave-packets*.

I want to emphasize here that we have never used any mechanical concept in our derivation of the quantum walk, including the notion of Hamiltonian: the dynamics is given in term of a single unitary operator A. A notion of effective Hamiltonian could be considered as the logarithm of A, which would correspond to an Hamiltonian providing the same unitary evolution, and which would even interpolate it between contiguous steps. For this reason we will call such an operator *interpolating Hamiltonian*. In the Fourier direct-integral representation of the operator, the interpolating Hamiltonian will be given by the identity $e^{-iH(\mathbf{k})} := A_{\mathbf{k}}$. It is easy to see that the relativistic limit $H_0(\mathbf{k})$ of $H(\mathbf{k})$, corresponding to consider narrow-band wave-packets centered at $\mathbf{k} = 0$, is achieved by expanding it at the first order in $|\mathbf{k}|$, *i.e.* $H(\mathbf{k}) = H_0(\mathbf{k}) + \mathscr{O}(|\mathbf{k}|^2)$. The interpolated continuum-time evolution in the relativistic regime will be then given by the first-order differential equation in the Schrödinger form

$$i\partial_t\psi(\mathbf{k}, t) = H_0(\mathbf{k})\psi(\mathbf{k}, t). \tag{13}$$

Rigorous quantitative approaches to judge the closeness between free QFT and the relativistic regime of the quantum walk have been provided in ref. [6] in terms of channel discrimination probability, and in ref. [4] in terms of fidelity between the two evolutions. Numerical values will be provided at the end of subsect. **3**˙8.

3˙6. *Schrödinger equation for the ultra-relativistic regime.* – In the ultra-relativistic regime of wave vectors comparable to the Planck vector, an obvious option is that of evaluating the evolution by a numerical evaluation of the exact quantum walk([13]). However, even in such regime we still have an analytical method available for evaluating the

([12]) The highest momentum observed is that of a ultra-high-energy cosmic ray, which is $k \sim 10^{-8}$.

([13]) A fast numerical technique to evaluate the quantum walk evolution numerically exploits the Fourier transform. For an application to the Dirac quantum walk see ref. [35].

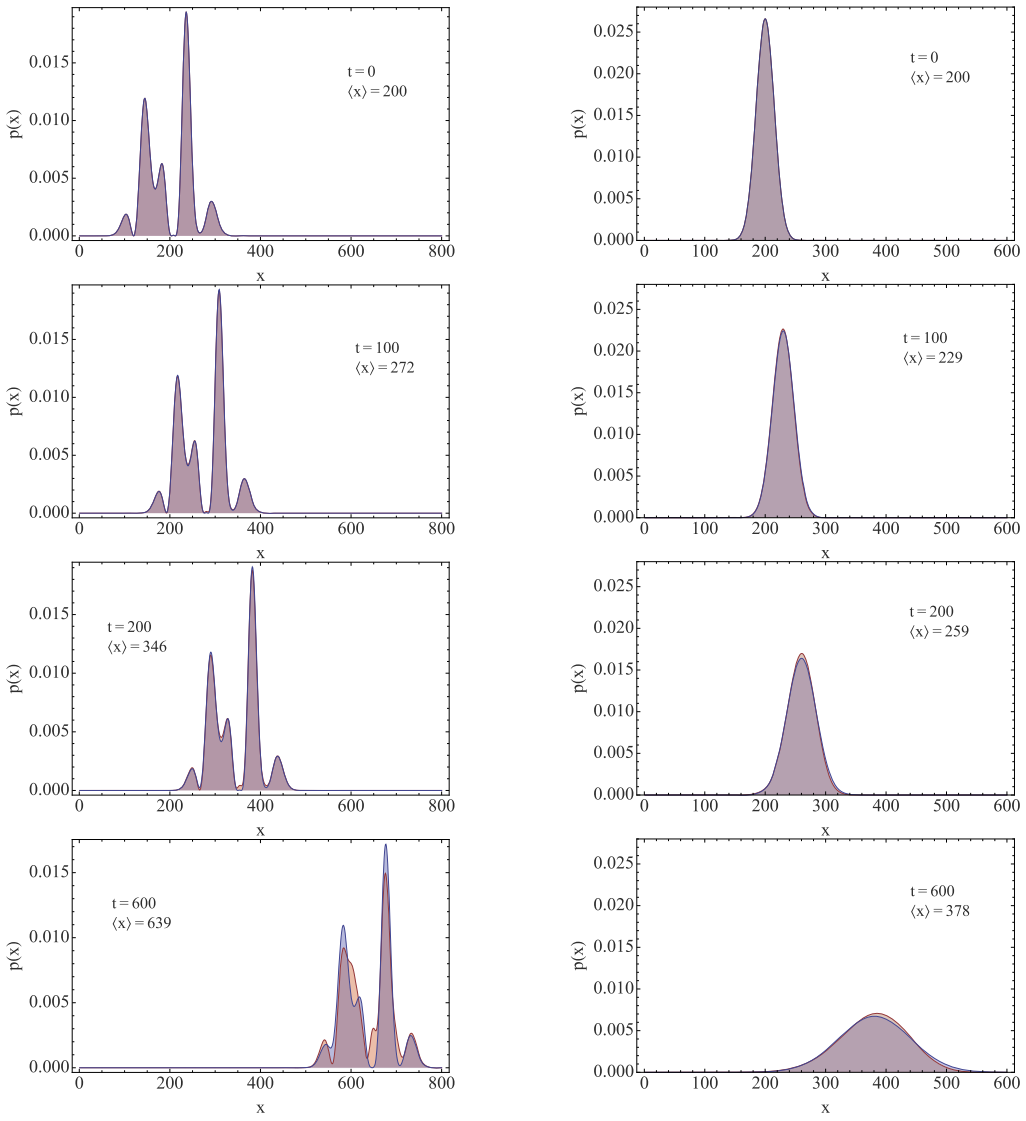


Fig. 4. – From ref. [6] (colors online). Test of the quality of the approximation of the Schrödinger equation (14) at for different time t of the Dirac quantum walk with mass $m = 0.6$ in one space dimension of ref. [6]. Comparison of the probability distribution (in red) and the solution of the Schrödinger equation (in blue). Right figures: the state is a superposition of Hermite functions multiplied by the Gaussian peaked around momentum $k_0 = 3\pi/10$, for drift and diffusion coefficients $v = 0.73$ and $D = 0.31$, respectively. The mean value moves at the group velocity given by the drift coefficient v. The approximation remains accurate even for position spread $\hat{\sigma} = 20$ Planck lengths. Left figures: The same four times comparison for the quantum walk with $m = 0.4$, and an initial Gaussian state peaked around the momentum $k_0 = 0.1$. In this case the drift velocity and the diffusion coefficient are respectively $v = 0.22$ and $D = 2.30$.

evolution of some common physical states. Indeed, for narrow-band wave-packets centered around any value $\mathbf{k}_0$ one can write a dispersive Schrödinger equation by expanding the interpolating Hamiltonian $H(\mathbf{k})$ around $\mathbf{k}_0$ at the second order, thus obtaining

$$i\partial_t\tilde{\psi}(\mathbf{x},t) = \pm\left[\mathbf{v}\cdot\nabla + \tfrac{1}{2}\mathbf{D}\cdot\nabla\nabla\right]\tilde{\psi}(\mathbf{x},t), \tag{14}$$

where $\tilde{\psi}(\mathbf{x},t)$ is the Fourier transform of $\tilde{\psi}(\mathbf{k},t) := e^{-i\mathbf{k}_0\cdot\mathbf{x}+i\omega_0 t}\psi(\mathbf{k},t)$, $\mathbf{v} = (\nabla_{\mathbf{k}}\omega)(\mathbf{k}_0)$ is the drift vector, and $\mathbf{D} = (\nabla_{\mathbf{k}}\nabla_{\mathbf{k}}\omega)(\mathbf{k}_0)$ is the diffusion tensor. This equation approximates very well the evolution, even in the Planck regime and for large numbers of steps, depending on the bandwith (see an example in fig. 4 from ref. [6]).

3˙7. *Recovering the Weyl equation*([14]). – In subsect. **3**˙2 we were left with the classification of the Cayley graphs of $\mathbb{Z}^3$ satisfying the isotropic embedding in $\mathbb{R}^3$, which are just the Bravais lattices. For dimension $s = 1$ it is easy to show that the only solution of the unitarity constraints gives the trivial quantum walk $A = I$([15]). We then consider $s = 2$. Now, the only inequivalent isotropic Cayley graphs are the primitive cubic (PC) lattice, the body centered cubic (BCC), and the rhombohedral. However only in the BCC case, whose presentation of $\mathbb{Z}^3$ involves four vectors $S_+ = \{\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3, \mathbf{h}_4\}$ with relator $\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 + \mathbf{h}_4 = 0$, one finds solutions satisfying all the assumptions of sect. **2**. The isotropy group is given by the group L of binary rotations around the coordinate axes, with the unitary projective representation on $\mathbb{C}^2$ given by $\{I, i\sigma_x, i\sigma_y, i\sigma_z\}$. The group L is transitive on the four BCC generators of S_+. There are only four solutions (modulo unitary conjugation) that can be divided in two pairs $A^\pm$ and $B^\pm$. The two pairs of solutions are connected by transposition in the canonical basis, *i.e.* $A^\pm_{\mathbf{k}} = (B^\pm_{\mathbf{k}})^T$. The solutions $B^\pm_{\mathbf{k}}$ can be also obtained from the solution $A^\pm_{\mathbf{k}}$ by shifting the wave vector $\mathbf{k}$ inside the Brillouin zone([16]) to the vectors [4]

$$\mathbf{k_1} = \frac{\pi}{2}(1,1,1), \qquad \mathbf{k_2} = -\frac{\pi}{2}(1,1,1), \qquad \mathbf{k_3} = -\frac{\pi}{2}(1,0,0). \tag{15}$$

The $A^\pm_{\mathbf{k}}$ solutions in the wave vector representation are

$$A^\pm_{\mathbf{k}} = Iu^\pm_{\mathbf{k}} - i\boldsymbol{\sigma}^\pm\cdot\tilde{\mathbf{n}}^\pm_{\mathbf{k}} \tag{16}$$

with

$$\tilde{\mathbf{n}}^\pm_{\mathbf{k}} := \begin{pmatrix} s_xc_yc_z \mp c_xs_ys_z \\ c_xs_yc_z \pm s_xc_ys_z \\ c_xc_ys_z \mp s_xs_yc_z \end{pmatrix}, \qquad u^\pm_{\mathbf{k}} := c_xc_yc_z \pm s_xs_ys_z, \tag{17}$$

([14]) This section is a synthesis of the results of ref. [4]. It should be noticed that there isotropy is not even assumed in solving eqs. (9). A simplified derivation making use of isotropy and full detailed analysis of all possible Cayley graphs will be available soon [30].

([15]) Also more generally one has $A = T_h$.

([16]) The first Brillouin zone B for the BCC lattice is defined in Cartesian coordinates as $-\sqrt{3}\pi \leq k_i \pm k_j \leq \sqrt{3}\pi$, $i \neq j \in \{x, y, z\}$.

where $c_i := \cos(k_i/\sqrt{3})$, $s_i := \sin(k_i/\sqrt{3})$, and $\boldsymbol{\sigma}^+ = \boldsymbol{\sigma}$, $\boldsymbol{\sigma}^- = \boldsymbol{\sigma}^T$. The spectrum of $A^\pm_{\mathbf{k}}$ is $\{e^{-i\omega^\pm_{\mathbf{k}}}\}$, with dispersion relation given by

$$\omega^\pm_{\mathbf{k}} = \arccos(c_x c_y c_z \mp s_x s_y s_z). \tag{18}$$

It is easy to get the relativistic limit of the quantum walk using the procedure in subsect. **3**˙5. This simply corresponds to substituting $c_i = 1$ and $s_i = k_i/\sqrt{3}$ in eq. (17), thus obtaining

$$i\partial_t \psi(\mathbf{k}, t) = \frac{1}{\sqrt{3}} \boldsymbol{\sigma}^\pm \cdot \mathbf{k}\, \psi(\mathbf{k}, t). \tag{19}$$

Equation (19) are the two Weyl equations for the left and the right chiralities. For $G = \mathbb{Z}^d$ with $d = 1, 2$ one obtains the Weyl equations in dimension $d = 1, 2$, respectively [4]. All the three quantum walks have the same form in eq. (16), namely

$$A_{\mathbf{k}} = u_{\mathbf{k}} I - i\boldsymbol{\sigma} \cdot \tilde{\mathbf{n}}_{\mathbf{k}}, \tag{20}$$

with dispersion relation

$$\omega_{\mathbf{k}} = \arccos u_{\mathbf{k}}, \tag{21}$$

and with the analytic expression of u_k and $\mathbf{n}_k$ depending on d and on the chirality (see ref. [4]). Since the quantum walks in eq. (17) or (20) have the Weyl equations as relativistic limit, we will also call them *Weyl quantum walks*.

The interpolating Hamiltonian is $H(\mathbf{k}) = \boldsymbol{\sigma} \cdot \mathbf{n}_{\mathbf{k}}$, with $\mathbf{n}_{\mathbf{k}} := (\omega_{\mathbf{k}}/\sin\omega_{\mathbf{k}})\tilde{\mathbf{n}}_{\mathbf{k}}$ playing the role of an helicity vector, and with relativistic-limit being given by $H_0(\mathbf{k}) = \frac{1}{\sqrt{d}}\boldsymbol{\sigma} \cdot \mathbf{k}$, which coincides with the usual Weyl Hamiltonian in d dimensions upon interpreting the wave vector $\mathbf{k}$ as the particle momentum.

We conclude the present subsection by emphasizing that one additional advantages of the discrete framework is that the Feynman path-integral is well defined, and it is also exactly calculated analytically in some cases. Indeed, in refs. [36] and [37] the discrete Feynman propagator for the Weyl quantum walk has been analytically evaluated with a closed form for dimensions $d = 1$ and $d = 2$, and the case of dimension $d = 3$ will be published soon [38].

3˙8. *Recovering the Dirac equation.* – From subsect. **3**˙7 we know that all quantum walks derivable from our principles for $s = 2$ give the Weyl equation in the relativistic limit. We now need to increase the dimension s of the field beyond $s = 2$. However, the problem of solving the unitarity equations (9) becomes increasingly difficult, since the unknown are matrices of increasingly larger dimension $s \geq 3$ (we remind that the equations are bilinear non-homogeneus in the unknown transition matrices, and a canonical procedure for the solution is unknown). What we can do for the moment is to provide

only some particular solutions using algebraic techniques. Two ways of obtaining solutions for $s = 4$ is to start from solutions in dimension $s = 2$ and built the direct-sum and tensor product of two copies of the quantum walk in such a way that the obtained quantum walk for dimension $s = 4$ still satisfies the principles. We will see that the quantum walks that we obtain in the relativistic limit give the Dirac equation when using the direct sum, whereas they give the Maxwell equation (plus a static scalar field) when we use the tensor product.

When building a quantum walk in 2×2 block form, all four blocks must be quantum walks themselves. The requirement of locality of the coupling leads to off-diagonal blocks that do not depend on $\mathbf{k}$. A detailed analysis of the restrictions due to the unitarity conditions (9) shows that, modulo unitary change of representation independent on $\mathbf{k}$([17]), we can take the off-diagonal matrix elements as proportional to the identity, whereas the diagonal blocks are just given by the chosen quantum walk and its adjoint, respectively. We then need to weight the diagonal blocks with a constant n and the off-diagonal identities with a constant m, and unitarity requires having $|n|^2+|m|^2 = 1$. Then, starting from the walk $A_{\mathbf{k}}$ that leads us to the Weyl equations for all dimension $d = 1, 2, 3$, the walk, modulo unitary equivalence([18]), can be recast in the form [4]

$$D_{\mathbf{k}} := \begin{pmatrix} nA_{\mathbf{k}} & im \\ im & nA_{\mathbf{k}}^\dagger \end{pmatrix}, \qquad n^2 + m^2 = 1, \qquad n \in \mathbb{R}^+, \qquad m \in \mathbb{R}. \tag{22}$$

Also the sign of m can be changed by a unitary equivalence (a "charge-conjugation"), however, we keep m with changing sign for reasons that will explained in subsect. **3**˙8.2. The walk (22) with $s = 4$ can be conveniently expressed in terms of gamma matrices in the spinorial representation as follows:

$$D_{\mathbf{k}} := nIu_{\mathbf{k}} - in\gamma^0\boldsymbol{\gamma} \cdot \tilde{\mathbf{n}}_{\mathbf{k}} + im\gamma^0, \tag{23}$$

where the functions $u_{\mathbf{k}}$ and $\tilde{\mathbf{n}}_{\mathbf{k}}$ depend on the choice of $A_{\mathbf{k}}$ in eq. (22), *i.e.* on $d = 1, 2, 3$. The dispersion relation of the quantum walk (23) is simply given by

$$\omega_{\mathbf{k}} = \arccos[\sqrt{1 - m^2}u_{\mathbf{k}}]. \tag{24}$$

We will see now that the quantum walks in eq. (22) in the small wave vector limit and for $m \ll 1$ all give the usual Dirac equation in the respective dimension d, with m corresponding to the particle rest mass, whereas n works as the inverse of a refraction index of vacuum. In fact, the interpolating Hamiltonian $H(\mathbf{k})$ is given by

$$H(\mathbf{k}) = \frac{\omega_{\mathbf{k}}}{\sin \omega_{\mathbf{k}}}(n\gamma^0\boldsymbol{\gamma} \cdot \tilde{\mathbf{n}}_{\mathbf{k}} - m\gamma^0), \tag{25}$$

([17]) This can also be *e.g.* the case of an overall phase independent of $\mathbf{k}$.

([18]) Also the solutions with walk $B^\pm = (A_{\mathbf{k}})^T$ are contained in eq. (22), since they can be achieved either by a shift in the Brillouin zone or as $\sigma_y B^\pm \sigma_y = A^{\pm\dagger}$, with the exchange of the upper and lower diagonal blocks that can be done unitarily.

with relativistic limit given by

$$H_0(\mathbf{k}) = \frac{n}{\sqrt{d}}\gamma^0\boldsymbol{\gamma}\cdot\mathbf{k} + m\gamma^0, \tag{26}$$

and to the order $\mathscr{O}(m^2)$ we get the Dirac Hamiltonian

$$H_0(\mathbf{k}) = \frac{1}{\sqrt{d}}\gamma^0\boldsymbol{\gamma}\cdot\mathbf{k} + m\gamma^0. \tag{27}$$

One has the Dirac Hamiltonian, with the wave vector $\mathbf{k}$ interpreted as momentum and the parameter m interpreted as the rest mass of the particle. In the relativistic limit (26) the parameter n plays the role of the inverse of a refraction index of vacuum. In principle this can produce measurable effects from bursts of high-energy particles of different masses at the boundary of the visible universe, and would be complementary to the dispersive nature of vacuum (see subsects. **3**˙8.3 and **3**˙9.2).

In the following we will also call the quantum walk in eq. (22) *Dirac quantum walk*([19]).

In ref. [36] the discrete Feynman propagator for the Dirac quantum walk has been analytically evaluated with a closed formal for dimension $d = 1$, generalizing the solution of ref. [39] for fixed mass value.

3˙8.1. Discriminability between quantum walk and quantum field dynamics. In subsect. **3**˙5 we mentioned that rigorous quantitative approaches to judge the closeness between the two dynamics have been provided in ref. [6], and in ref. [4] in terms of fidelity between the two unitary evolutions. For the Dirac quantum walk for a proton mass one has fidelity close to unit for $N \simeq m^{-3} = 2.2 * 10^{57}$, corresponding to $t = 1.2 * 10^{14}\text{s} = 3.7 * 10^6$ years. The approximation is still good in the ultra-relativistic case $k \gg m$, *e.g.* for $k = 10^{-8}$ (as for an ultra-high-energy cosmic ray), where it holds for $N \simeq k^{-2} = 10^{16}$ steps, corresponding to $5 * 10^{-28}$ s. However, one should notice that practically the discriminability in terms of fidelity corresponds to having unbounded technology, and such a short time very likely corresponds to unfeasible experiments. On the other hand, for a ultra-high-energy proton with wave packet width of 100 fm the time required for discriminating the wave packet of the quantum walk from that of QFT is comparable with the age of the universe.

([19]) For $d = 1$, modulo a permutation of the canonical basis, the quantum walk corresponds to two identical and decoupled $s = 2$ walks. Each of these quantum walks coincide with the one dimensional *Dirac walks* derived in ref. [6]. The last one was derived as the simplest $s = 2$ homogeneous quantum cellular walk covariant with respect to the parity and the time-reversal transformation, which are less restrictive than isotropy that singles out the only Weyl quantum walk in one space dimension.

3˙8.2. Mass and proper-time. The unitarity requirement in eq. (22) restrict the rest mass to belong to the interval

$$m \in [-1, 1]. \tag{28}$$

At the extreme points ± 1 of the interval the corresponding dynamics $D_{\mathbf{k}} = \pm i\gamma^0$ are identical (they differ for an irrelevant global phase factor). This means that the domain of the mass has actually the topology of a circle, namely

$$m \in S^1. \tag{29}$$

From the classical relativistic Hamiltonian [40]

$$H = \vec{p} \cdot \vec{q} + c^2 m\tau - L, \tag{30}$$

with $\vec{p}$ and $\vec{q}$ canonically conjugated position and momentum and L the Lagrangian, we see that the *proper time* τ is canonically conjugated to the rest mass m. This suggests that the Fourier conjugate of the rest mass in the quantum walk can be interpreted as the proper time of a particle evolution, and being the mass a variable in S^1, we conclude that *the proper time is discrete*, in accordance with the discreteness of the dynamical evolution of the quantum walk. This result constitutes a non-trivial logical coherence check of the present quantum walk theory.

3˙8.3. Physical dimensions and scales for mass and discreteness. We want to emphasize that in the above derivation everything is adimensional by construction. Dimensions can be recovered by using as measurement standards for space, time, and mass the discreteness scale for space a_* and time t_* (a_* is half of the BCC cell side, t_* the time-length of the unit step), along with the maximum value of the mass m_* (corresponding to $|m| = 1$ in eq. (22)). From the relativistic limit, the comparison with the usual dimensional Dirac equation leads to the identities

$$c = a_*/t_*, \qquad \hbar = m_* a_* c, \tag{31}$$

which leave only one unknown among the three variables a_*, t_* and m_*. At the maximum value of the mass $|m| = 1$ in eq. (22) we get a flat dispersion relation, corresponding to no flow of information: this is naturally interpreted as a mini black-hole, *i.e.* a particle with Schwarzild radius equal to the localization length, *i.e.* the Compton wavelength. This leads to an heuristic interpretation of m_* as the Planck mass, and from the two identities in eq. (31) we get the Planck scale for discreteness. Notice that the value of m_* can be in principle obtained from the dispersion of vacuum as $m_* \simeq \frac{1}{\sqrt{3}} \frac{\hbar k}{c(k) - c(0)}$ for small k, which can be in principle measured by the Fermi telescope from detection of ultra high energy bursts coming from deep space.

3˙9. *Recovering Maxwell fields*([20]). – In subsects. **3˙7** and **3˙8** we showed how the dynamics of free quantum fields can be derived starting from a countable set of quantum systems with a network of interactions satisfying the principles of locality, homogeneity, and isotropy. Within the present finitistic local-algorithmic perspective one also considers each system as carrying a finite amount of information, thus restricting the quantum field to be Fermionic (see also subsect. **2˙1**). However, one may wonder how the physics of the free electromagnetic field can be recovered in such a way and, generally, how Bosonic fields are recovered from Fermionic ones. In this section we answers to these questions. The basic idea behind is that *the photon emerges as an entangled pair of Fermions evolving according to the Weyl quantum walk* of sect. **3˙7**. Then one shows that in a suitable regime both the free Maxwell equation in 3d and the Bosonic commutation relations are recovered. Since in this subsection we are actually considering operator quantum fields, we will use more properly the quantum automaton nomenclature instead of the quantum walk one.

Consider two Fermionic fields $\psi(\mathbf{k})$ and $\varphi(\mathbf{k})$ in the wave vector representation, with respective evolutions given by

$$\psi(\mathbf{k}, t+1) = W_{\mathbf{k}}\psi(\mathbf{k}, t), \qquad \varphi(\mathbf{k}, t+1) = W^*_{\mathbf{k}}\varphi(\mathbf{k}, t). \tag{32}$$

The matrix $W_{\mathbf{k}}$ can be any of the Weyl quantum walks for $d = 3$ in eq. (16), (the whole derivation is independent on this choice), whereas $W^*_{\mathbf{k}} = \sigma_y W_{\mathbf{k}} \sigma_y$ denotes the complex conjugate matrix. We introduce the bilinear operators

$$G^i(\mathbf{k}, t) := \varphi^T\left(\tfrac{\mathbf{k}}{2}, t\right)\sigma^i\psi\left(\tfrac{\mathbf{k}}{2}, t\right) = \varphi^T(\mathbf{k}, 0)\left(W^\dagger_{\frac{\mathbf{k}}{2}}\sigma^i W_{\frac{\mathbf{k}}{2}}\right)\psi\left(\tfrac{\mathbf{k}}{2}, 0\right) \tag{33}$$

by which we construct the vector field

$$\mathbf{G}(\mathbf{k}, t) := (G^1(\mathbf{k}, t), G^2(\mathbf{k}, t), G^3(\mathbf{k}, t))^T \tag{34}$$

and the transverse field

$$\mathbf{G}_T(\mathbf{k}, t) := \mathbf{G}(\mathbf{k}, t) - \left(\frac{\mathbf{n}_{\frac{\mathbf{k}}{2}}}{|\mathbf{n}_{\frac{\mathbf{k}}{2}}|} \cdot \mathbf{G}(\mathbf{k}, t)\right)\frac{\mathbf{n}_{\frac{\mathbf{k}}{2}}}{|\mathbf{n}_{\frac{\mathbf{k}}{2}}|}, \tag{35}$$

with $\mathbf{n}_{\mathbf{k}} := (\omega_{\mathbf{k}}/\sin\omega_{\mathbf{k}})\tilde{\mathbf{n}}_{\mathbf{k}}$ and $\tilde{\mathbf{n}}_{\mathbf{k}}$ given in eq. (17). By construction the field $\mathbf{G}_T(\mathbf{k}, t)$ satisfies the following relations:

$$\mathbf{n}_{\frac{\mathbf{k}}{2}} \cdot \mathbf{G}_T(\mathbf{k}, t) = 0, \tag{36}$$

$$\mathbf{G}_T(\mathbf{k}, t) = \mathrm{Exp}\left(-i2\mathbf{n}_{\frac{\mathbf{k}}{2}} \cdot \mathbf{J}t\right)\mathbf{G}_T(\mathbf{k}, 0), \tag{37}$$

([20]) The entire subsection is a short review of ref. [7].

where we used the identity

$$\exp\left(-\tfrac{i}{2}\mathbf{v}\cdot\boldsymbol{\sigma}\right)\boldsymbol{\sigma}\exp\left(\tfrac{i}{2}\mathbf{v}\cdot\boldsymbol{\sigma}\right)=\mathrm{Exp}(-i\mathbf{v}\cdot\mathbf{J})\boldsymbol{\sigma}, \tag{38}$$

the matrix $\mathrm{Exp}(-i\mathbf{v}\cdot\mathbf{J})$ acting on $\boldsymbol{\sigma}$ regarded as a vector, and $\mathbf{J}=(J_x,J_y,J_z)$ representing the infinitesimal generators of $\mathbb{SU}(2)$ in the spin 1 representation. Taking the time derivative of eq. (37) we obtain

$$\partial_t\mathbf{G}_T(\mathbf{k},t)=2\mathbf{n}_{\frac{\mathbf{k}}{2}}\times\mathbf{G}_T(\mathbf{k},t). \tag{39}$$

If $\mathbf{E}_G$ and $\mathbf{B}_G$ are two Hermitian operators defined by the relation

$$\mathbf{E}_G:=|\mathbf{n}_{\frac{\mathbf{k}}{2}}|(\mathbf{G}_T+\mathbf{G}_T^\dagger),\qquad \mathbf{B}_G:=i\left|\mathbf{n}_{\frac{\mathbf{k}}{2}}\right|(\mathbf{G}_T^\dagger-\mathbf{G}_T), \tag{40}$$

then eq. (36) and eq. (39) can be rewritten as

$$\begin{aligned}&\partial_t\mathbf{E}_G=i2\mathbf{n}_{\frac{\mathbf{k}}{2}}\times\mathbf{B}_T(\mathbf{k},t),\qquad \partial_t\mathbf{B}_G=-i2\mathbf{n}_{\frac{\mathbf{k}}{2}}\times\mathbf{E}_T(\mathbf{k},t),\\ &2\mathbf{n}_{\frac{\mathbf{k}}{2}}\cdot\mathbf{E}_G=0,\qquad 2\mathbf{n}_{\frac{\mathbf{k}}{2}}\cdot\mathbf{B}_G=0.\end{aligned} \tag{41}$$

Equations (41) have the form of distorted Maxwell equations, with the wave vector $\mathbf{k}$ substituted by $2\mathbf{n}_{\frac{\mathbf{k}}{2}}$, and in the relativistic limit $|\mathbf{k}|\ll 1$ one has $2\mathbf{n}_{\frac{\mathbf{k}}{2}}\sim\mathbf{k}$ and the usual free electrodynamics is recovered.

3˙9.1. Photons made of pairs of fermions. Since in the Weyl equation the field is fermionic, the field defined in eqs. (35) and (40) generally does not satisfy the correct bosonic commutation relations. The solution to this problem is to replace the operator $\mathbf{G}$ defined in eq. (35) with the operator $\mathbf{F}$ defined as

$$\mathbf{F}(\mathbf{k}):=\int\frac{\mathrm{d}\mathbf{q}}{(2\pi)^3}f_{\mathbf{k}}(\mathbf{q})\varphi\left(\tfrac{\mathbf{k}}{2}-\mathbf{q}\right)\boldsymbol{\sigma}\,\psi\left(\tfrac{\mathbf{k}}{2}+\mathbf{q}\right), \tag{42}$$

where $\int\frac{\mathrm{d}\mathbf{q}}{(2\pi)^3}|f_{\mathbf{k}}(\mathbf{q})|^2=1,\forall\mathbf{k}$. In terms of $\mathbf{F}(\mathbf{k})$, we can define the polarization operators $\varepsilon^i(\mathbf{k})$ of the electromagnetic field as follows:

$$\varepsilon^i(\mathbf{k}):=\mathbf{u}^i_{\mathbf{k}}\cdot\mathbf{F}(\mathbf{k},0),\qquad i=1,2, \tag{43}$$

$$\mathbf{u}^i_{\mathbf{k}}\cdot\mathbf{n}_{\mathbf{k}}=\mathbf{u}^1_{\mathbf{k}}\cdot\mathbf{u}^2_{\mathbf{k}}=0,\ |\mathbf{u}^i_{\mathbf{k}}|=1,\qquad(\mathbf{u}^1_{\mathbf{k}}\times\mathbf{u}^2_{\mathbf{k}})\cdot\mathbf{n}_{\mathbf{k}}>0. \tag{44}$$

In order to avoid technicalities from continuum of wave vectors, we restrict to a discrete wave vector space, corresponding to confinement in a cavity. Moreover we assume $|f_{\mathbf{k}}(\mathbf{q})|^2$ to be uniform over a region $\Omega_{\mathbf{k}}$ which contains $N_{\mathbf{k}}$ modes, *i.e.*

$$|f_{\mathbf{k}}(\mathbf{q})|^2=\begin{cases}\dfrac{1}{N_{\mathbf{k}}}, & \text{if } \mathbf{q}\in\Omega_{\mathbf{k}},\\ 0, & \text{otherwise.}\end{cases} \tag{45}$$

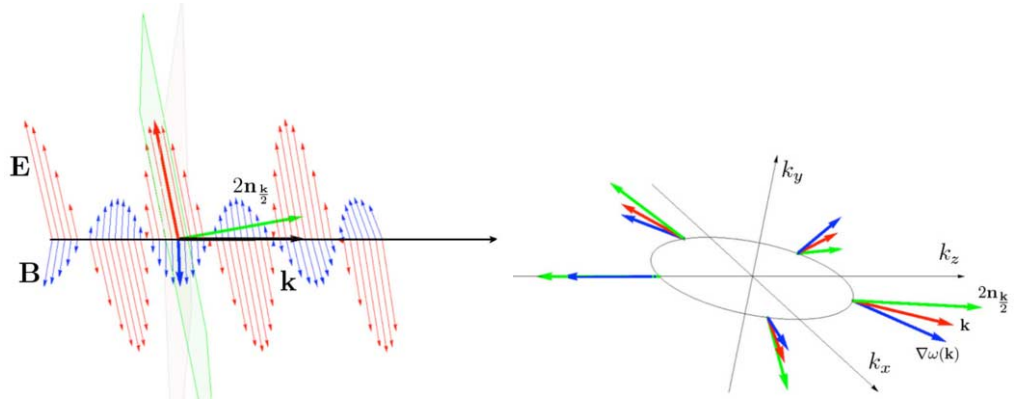


Fig. 5. – From ref. [7] (colors online). Left: In a rectilinear polarized electromagnetic wave, the polarization plane (in green) is slightly tilted with respect the plane orthogonal to $\mathbf{k}$ (in gray). Right: vector $2\mathbf{n}_{\frac{\mathbf{k}}{2}}$ (in green), which is orthogonal to the polarization plane; wave vector $\mathbf{k}$ (in red) and group velocity (in blue) for the value $|\mathbf{k}| = 0.8$ and different directions. Notice that the three vectors are not parallel (the angles between them depend on $\mathbf{k}$).

Then, for a given state ρ of the field we denote by $M_{\varphi,\mathbf{k}}$ (respectively, $M_{\psi,\mathbf{k}}$) the mean number of type φ (respectively, ψ) Fermionic excitations in the region $\Omega_{\mathbf{k}}$. One can then show that, for states such that $M_{\xi,\mathbf{k}}/N_{\mathbf{k}} \leq \epsilon \ll 1$ for both $\xi = \varphi, \psi$ and forall $\mathbf{k}$ we have

$$[\varepsilon^i(\mathbf{k}), \varepsilon^{j\,\dagger}(\mathbf{k}')]_- = \delta_{i,j}\delta_{\mathbf{k},\mathbf{k}'}, \tag{46}$$

i.e. the polarization operators are Bosonic operators.

3˙9.2. Vacuum dispersion. According to eq. (41) the angular frequency of the electromagnetic waves is given by the modified dispersion relation

$$\omega(\mathbf{k}) = 2\left|\mathbf{n}_{\frac{\mathbf{k}}{2}}\right|, \tag{47}$$

which recovers the usual relation $\omega(\mathbf{k}) = |\mathbf{k}|$ in the relativistic regime. In a dispersive medium, the speed of light is the group velocity $\nabla_{\mathbf{k}}\omega(\mathbf{k})$ of the electromagnetic waves, and eq. (47) predict that the vacuum is dispersive, namely the speed of light generally depends on $\mathbf{k}$ (see fig. 5 for directions of vectors). Such dispersion phenomenon has been already analyzed in some literature on quantum gravity, where several authors considered how a hypothetical invariant length (corresponding to the Planck scale) could manifest itself in terms of modified dispersion relations [41-45]. In these models the $\mathbf{k}$-dependent speed of light $c(\mathbf{k})$, at the leading order in $k := |\mathbf{k}|$, is expanded as $c(\mathbf{k}) \approx 1 \pm \xi k^\alpha$, where ξ is a numerical factor of the order 1, while α is an integer. This is exactly what happens in our framework, where the intrinsic discreteness of the quantum cellular automata $A^{\pm}_{\mathbf{k}}$ leads to the dispersion relation of eq. (47) from which one obtains the

following **k**-dependent speed of light:

$$c^{\mp}(\mathbf{k}) \approx 1 \pm 3\frac{k_x k_y k_z}{|\mathbf{k}|^2} \approx 1 \pm \tfrac{1}{\sqrt{3}}k. \tag{48}$$

Equation (48) is obtained by evaluating the modulus of the group velocity and expanding in powers of **k** with the assumption $k_x = k_y = k_z = \frac{1}{\sqrt{3}}k$, $(k = |\mathbf{k}|)$([21]). Notice that the dispersion is not isotropic, and can also be superluminal, though uniformly bounded [4] by a factor $\sqrt{d}$ (which coincides with the uniform bound of the quasi-isometric embedding). The prediction of dispersive behavior, as for the present automata theory of quantum fields, is especially interesting since it is experimentally falsifiable, and, as mentioned in subsect. **3**˙8.3, allows to experimentally set the discreteness scale. In fact, differently to the other mentioned birefringence effects, the disperision effect, although is extremely small in the relativistic regime, it accumulates and become magnified during a huge time of flight. For example, observations of the arrival times of pulses originated at cosmological distances (such as in some γ-ray bursts [46-49]), have sufficient sensitivity to detect corrections to the relativistic dispersion relation of the same order as in eq. (48).

4. – Recovering special relativity in a discrete quantum universe([22])

We have seen how relativistic mechanics, and more precisely free QFT, can be recovered without using any mechanical primitive, and without making any use of special relativity, including the relativity principle itself. However, one may wonder how discreteness can be reconciled with Lorentz transformations, and most importantly, how the relativity principle itself can be restated in purely mathematical terms, without using the notions of space-time and inertial frame. In this section we will see how such goal can be easily accomplished.

The relativity principle is expressed by the statement:

> *Galileo's Relativity Principle:* The physical law is invariant with the inertial frame.

Otherwise stated: the physics that we observe, or, equivalently, its mathematical representation, is independent on the inertial frame that we use.

What is a frame? It is a mathematical representation of physical laws in terms of space and time coordinates. What is special about the *inertial* frame? A convenient way of answering is the following:

> *Inertial frame:* a reference frame where energy and momentum are conserved for an isolated system.

([21]) Notice that, depending on the quantum walk $A^{+}(\mathbf{k})$ of $A^{-}(\mathbf{k})$ in eq. (16) we obtain corrections to the speed of light with opposite sign.

([22]) This entire section is a review of the main results of ref. [18].

When a system is isolated? This is established by the theory. In classical mechanics, a system is isolated if there are no external forces acting on it. In quantum theory a system is isolated when its dynamical evolution is described by a unitary transformation on the system's Hilbert space. At the very bottom of its notion, the inertial frame is the mathematical representation of the physical law that makes its analytical form the simplest. In classical physics, if we include the Maxwell equations among the invariant physical laws, what we get from Galileo's principle is Einstein's special relativity.

The quantum walk/automaton is an isolated system (it evolves unitarily). Mathematically the physical law that brings the information about the constants of the dynamics in terms of their Hilbert eigenspaces is provided by the eigenvalue equation. For the case of virtually-Abelian group G (which ultimately leads to physics in Euclidean space) the eigenvalue equation has the general form corresponding to eqs. (19) and (21)

$$A_{\mathbf{k}}\psi(\omega,\mathbf{k}) = e^{i\omega}\psi(\omega,\mathbf{k}), \tag{49}$$

with the eigenvalues usually collected into s dispersion relations (the two functions $\omega^{\pm}(\mathbf{k})$ for the Weyl quantum walk). This translates into the following re-interpretation of representations of the eigenvalue equation:

> *Quantum-digital inertial frame:* Representation in terms of eigen-spaces of the constants of the dynamics of the eigenvalue equation (49).

Using such notion of inertial frame, the principle of relativity is still the Galileo's principle. The group of transformations that connect different inertial reference frames will be the quantum digital-version of the Poincaré group:

> *Quantum-digital Poincaré group:* group of changes of representations in terms of eigenspaces of the dynamical constants that leave the eigenvalue equation (49) invariant.

It is obvious that the changes of representations make a group. Since the constants of dynamics are $\mathbf{k}$ and $\omega^{\pm}$, a change of representation corresponds to an invertible map $k \to k'(k)$, where with k we denote the four-vector $k := (\omega, \mathbf{k})$.

In the following subsection we will see how the inherent discreteness of the algorithmic description leads to distortions of the Lorentz transformations, visible in principle at huge energies. Nevertheless, Einstein's special relativity is perfectly recovered for $|\mathbf{k}| \ll 1$, namely at energy scales much higher than those ever tested.

On the other hand, as we will see in the following, discreteness has some plus compared to the continuum theory, since it contains the continuum theory as a special regime, and moreover it leads to some additional features with GR flavor: 1) it has a maximal particle mass with physical interpretation in terms of the Planck mass; 2) it leads to a De Sitter invariance (see subsect. **4**˙2). And this, in addition to providing its own physical standards for space, time, and mass within a purely mathematical context (subsect. **3**˙8.3).

4˙1. *Quantum-digital Poincaré group and the notion of particle*([23]). – The eigenvalue equation (49) can now be rewritten in "relativistic notation" as follows:

$$n_\mu(k)\sigma^\mu\psi(k) = 0, \tag{50}$$

upon introducing the four-vectors

$$k = (\omega, \mathbf{k}), \qquad n(k) = (\sin\omega, \mathbf{n}(\mathbf{k})), \qquad \boldsymbol{\sigma} = (I, \sigma), \qquad \sigma = (\sigma_x, \sigma_y, \sigma_z), \tag{51}$$

where the vector $\mathbf{n}(\mathbf{k})$ is defined in eq. (16), namely

$$\mathbf{n}(\mathbf{k})\cdot\sigma := \frac{i}{2}(A_{\mathbf{k}} - A_{\mathbf{k}}^\dagger). \tag{52}$$

As already mentioned, since the constants of dynamics are $\mathbf{k}$ and $\omega^\pm$, a change of representation corresponds to a map $k \mapsto k'(k)$. Now the principle of relativity corresponds to the requirement that the eigenvalue equation (50) is preserved under a change of representation. This means that the following identity must hold:

$$n_\mu(k)\sigma^\mu = \tilde{\Gamma}_k^{-1}\, n_\mu(k')\sigma^\mu\, \Gamma_k, \tag{53}$$

where Γ_k, $\tilde{\Gamma}_k$ are invertible matrices representing the change of representation.

The simplest example of change of observer is the one given by the trivial relabeling $k' = k$ and by the matrices $\Gamma_k = \tilde{\Gamma}_k = e^{i\lambda(\mathbf{k})}$, where $\lambda(\mathbf{k})$ is an arbitrary real function of $\mathbf{k}$. When $\lambda(\mathbf{k})$ is a linear function we recover the usual group of translations. The set of changes of representation $k \mapsto k'(k)$ for which eq. (53) holds are a group, which is the largest group of symmetries of the dynamics. In covariant notation the dispersion relations are rewritten as follows:

$$n_\mu^\pm(k)n^{\mu\pm}(k) = 0, \tag{54}$$

and in the small wave vector regime one has $n(k) \sim k$, recovering the usual relativistic dispersion relation.

In addition to the neighbour of the wave vector $\mathbf{k_0} = (0, 0, 0)$, the Weyl equations can be recovered from the quantum walk (16) also in the neighborhood of the wavevectors in eq. (15). The mapping between the vectors $\mathbf{k}_i$ exchange chirality of the particle and double the particles to four species in total: two left-handed and two right-handed([24]).

([23]) For a simpler analysis in one space dimensions and the connection with doubly-special relativity and relative locality, see ref. [50]. For a connection with Hopf algebras for position and momentum see ref. [51].

([24]) Discreteness has doubled the particles: this corresponds to the well-known phenomenon of fermion doubling [52].

In the following we will therefore more generally refer to the relativistic regime as the neighborhoods of the vectors $\{\mathbf{k}_i\}_{i=0}^3$.

The group of symmetries of the dynamics of the quantum walks (16) contains a non-linear representation of the Poincaré group, which exactly recovers the usual linear one in the relativistic regime. For any arbitrary non-vanishing function $f(k)$ one introduces the four-vector

$$p^{(f)} = \mathscr{D}^{(f)}(k) := f(k)n(k) \tag{55}$$

and rewrite the eigenvalue equation (50) as follows:

$$p_\mu^{(f)}\sigma^\mu\psi(k) = 0. \tag{56}$$

Upon denoting the usual Lorentz transformation by L_β for a suitable f [18] the Brillouin zone splits into four regions B_i, $i = 1, \ldots, 4$ centered around $\mathbf{k}_i$ $i = 0, \ldots 3$, such that the composition

$$\mathscr{L}_\beta^{(f)} := \mathscr{D}^{(f)-1} L_\beta \mathscr{D}^{(f)} \tag{57}$$

is well defined on each region separately. The four invariant regions corresponding to the four different massless Fermionic particles show that the Wigner notion of "particle" as invariant of the Poincaré group survives in a discrete world. For fixed function f the maps $\mathscr{L}_\beta^{(f)}$ provide a non-linear representation of the Lorentz group [53-55]. In figs. 6 the orbits of some wave vectors under subgroups of the non-linear Lorentz group are reported. The distortion effects due to underlying discreteness are evident at large wave vectors and boosts. The relabeling $k \to k'(k) = \mathscr{L}_\beta^{(f)}(k)$ satisfies eq. (53) with $\Gamma_k = \Lambda_\beta$ and $\tilde{\Gamma}_k = \tilde{\Lambda}_\beta$ for the right-handed particles, and $\Gamma_k = \tilde{\Lambda}_\beta$ and $\tilde{\Gamma}_k = \Lambda_\beta$ for the left-handed particles, with Λ_β and $\tilde{\Lambda}_\beta$ being the $(0, \frac{1}{2})$ and $(\frac{1}{2}, 0)$ representation of the Lorentz group, independently of k in each pertaining region.

For varying f, one obtains a much larger group, including infinitely many copies of the non-linear Lorentz one. In the small wave vector regime the whole group collapses to the usual linear Lorentz group for each particle.

4˙2. *De Sitter group for non-vanishing mass.* – Up to now we have analyzed what happens with massless particles. For massive particles described by the Dirac walk (22), the rest-mass m gets involved into the frame transformations, and their group becomes a non-linear realization of the De Sitter group $SO(1,4)$ with infinite cosmological constant, where the rest mass m of the particle plays the role of the additional coordinate. One recovers the previous non-linear Lorentz group at the order $\mathscr{O}(m^2)$.

5. – Conclusions and future perspectives: the interacting theory, . . . , gravity?

The logical connections that have lead us to build up our quantum-walk theory of fields leading to free QFT are summarized in fig. 7. The free relativistic quantum field theory

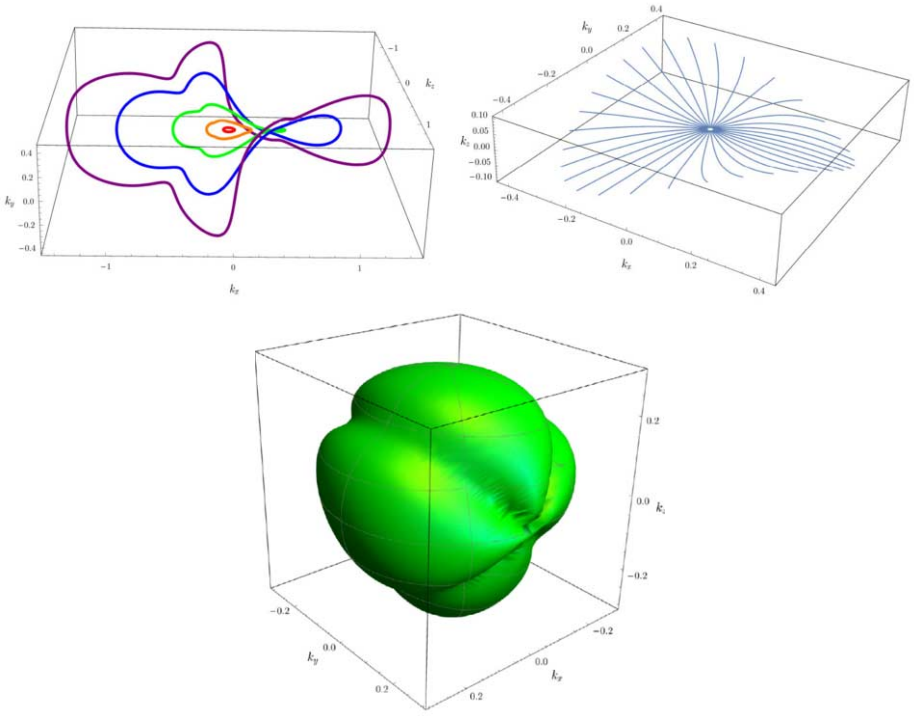


Fig. 6. – From ref. [18] (colors online). The distortion effects of the Lorentz group in the present quantum walk theory leading to the Weyl quantum field in the relativistic limit. *Top left figure*: the orbit of the wave vectors $\mathbf{k} = (k_x, 0, 0)$, with $k_x \in \{.05, .2, .5, 1, 1.7\}$ under the rotation around the z-axis. *Top right figure:* the orbit of wave vectors with $|\mathbf{k}| = 0.01$ for various directions in the (k_x, k_y)-plane under the boosts with β parallel to $\mathbf{k}$ and $|\beta| \in [0, \tanh 4]$. *Bottom figure:* the orbit of the wave vector $\mathbf{k} = (0.3, 0, 0)$ under the full rotation group $SO(3)$.

emerges as a special regime (the relativistic regime) of the evolution of countably many Fermionic quantum bits, provided that their unitary interactions satisfy the principles of homogeneity, locality, and isotropy, and with the restrictions of linearity of the evolution and of quasi-isometric embedding of the graph of interaction in an Euclidean space.

We are left now with the not easy task of recovering also the interacting relativistic quantum field theory, where particles are created and annihilated. We will need to devise which additional principles are missing that will lead to the interacting theory, breaking the linearity assumption. This is likely to be related to the nature of a gauge transformation. How can this be restated in terms of a new principle? From the point of view of a free theory, the interaction can be viewed as a violation of homogeneity, corresponding to the presence of another interacting field—namely the gauge-field. The gauge-field can be regarded as a restoration of homogeneity by a higher-level homogeneous "meta-law". For example, one can exploit the arbitrariness of the local bases of the Hilbert block subspaces $\mathbb{C}^s$ for the Weyl automata, having the bases dependent on

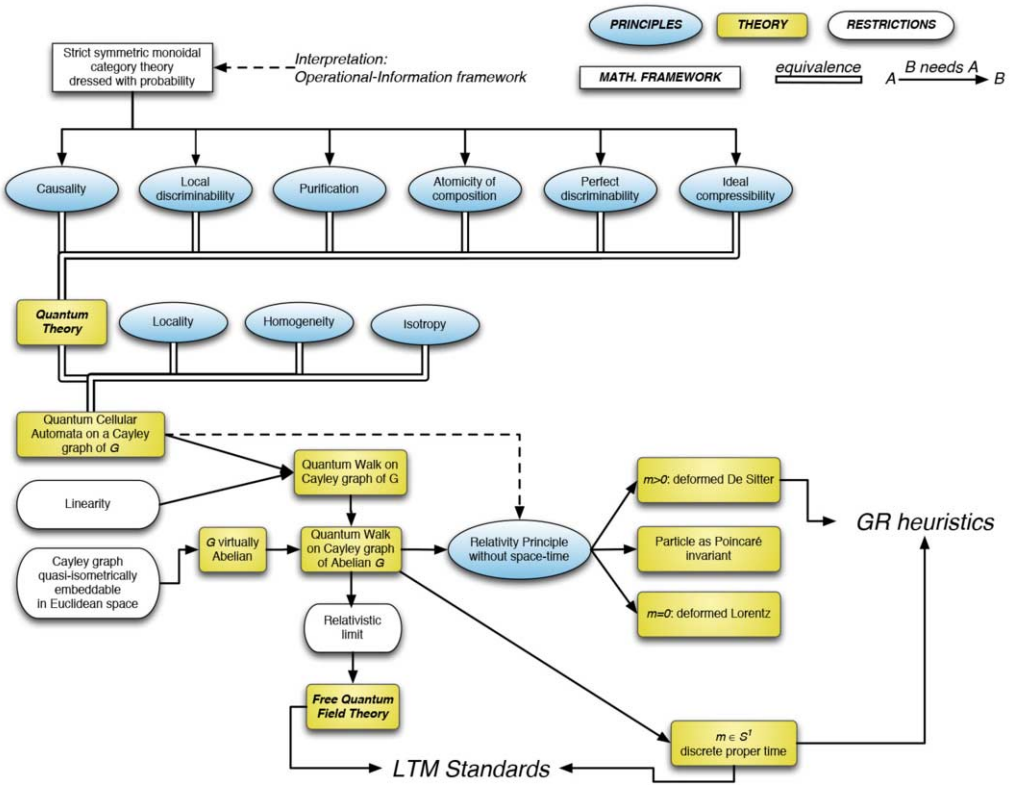


Fig. 7. – Logical scheme of the derivation from principles of the present quantum-walk theory of fields, with the known free quantum field theory as its relativistic limit. The top six principles from which quantum theory of abstract system is derived are not discussed in the present lecture, and can be found in refs. [3, 1].

the local value of the wave function of the gauge automaton made with pairs of entangled Fermions, as for the Maxwell automaton. In order to keep the interaction local, one can consider an *in situ* interaction. In such a way one would have a *quantum ab initio* gauge theory, without the need of artificially quantizing the gauge fields, nor of introducing mechanical Lagrangians. A $d = 1$ interacting theory of the kind of a Fermionic Hubbard quantum cellular automaton, has been very recently analytically investigated by the Bethe ansatz [56], and two-particle bound states have been established. It should be emphasized that for $d = 3$ just the possibility of recovering QED in the relativistic regime would be very interesting, since it will provide a definite procedure for renormalization. Very interesting will be also the analysis of the full dynamical invariance group, leading also to a non-linear version of the Poincaré group, with the possibility that this restricts the choice of the function f in eq. (55). Studying the full symmetry group of the interacting theory will also have the potential of providing additional internal symmetries, *e.g.* the $SU(3)$ symmetry group of QCD, with the Fermion doubling possibly playing

the role in adding physical particles. The mass as a variable quantum observable (as in subsects. **3**˙8.2 and **4**˙2) may provide rules about the lifetimes of different species of particles. The additional quasi-static scalar mode entering in the tensor-product of the two Weyl automata that give the Maxwell field in subsect. **3**˙9 may turn out to play a role in the interacting theory, *e.g.* playing the role of a Higg boson providing the mass value, or even being pivotal for gravitation. But for now we are just in the realm of speculations.

What we can say for sure is that it is not just a coincidence that so much physics comes out from so few general principles. How amazing is the whole resulting theory which, in addition to having a complete logical coherence by construction, it also winks at GR through the two non-trivial features of the maximum mass, and the De Sitter invariance. And with special relativity derived without space-time and kinematics, in a fully *quantum ab initio* theory. So much from so little? This is the power of the new information-theoretical paradigm.

* * *

This work was supported by the Templeton Foundation under the project ID# 43796 *A Quantum-Digital Universe*.

REFERENCES

[1] D'ARIANO G. M., CHIRIBELLA G. and PERINOTTI P., *Quantum Theory from First Principles* (Cambridge University Press, Cambridge) 2017.
[2] D'ARIANO G. M., *Int. J. Theor. Phys.*, **56**, Issue 1 (2017) 97 (Special issue in memoriam of David Finkelstein).
[3] CHIRIBELLA G., D'ARIANO G. M. and PERINOTTI P., *Phys. Rev. A*, **84** (2011) 012311.
[4] D'ARIANO G. M. and PERINOTTI P., *Phys. Rev. A*, **90** (2014) 062106.
[5] BISIO A., D'ARIANO G. M. and TOSINI A., *Phys. Rev. A*, **88** (2013) 032301.
[6] BISIO A., D'ARIANO G. M. and TOSINI A., *Ann. Phys.*, **354** (2015) 244.
[7] BISIO A., D'ARIANO G. M. and PERINOTTI P., *Ann. Phys.*, **368** (2016) 177.
[8] SUSSKIND L., *J. Math. Phys.*, **36** (1995) 6377.
[9] BOUSSO R., *Rev. Mod. Phys.*, **74** (2002) 825.
[10] NEUMANN J. V., *Mathematical foundations of quantum mechanics*, no. 2 (Princeton University Press) 1955.
[11] CHIRIBELLA G., D'ARIANO G. M. and PERINOTTI P., *Quantum from principles*, in *Quantum Theory: Informational Foundations and Foils*, edited by CHIRIBELLA G. and SPEKKENS R. (Springer) 2016, pp. 165–175.
[12] D'ARIANO G. M. and PERINOTTI P., *Found. Phys.*, **46** (2016) 269.
[13] HALVORSON H. and CLIFTON R., *Philos. Sci.*, **69** (2002) 1.
[14] KUHLMANN M., *Sci. Am.*, **24** (2015) 84.
[15] BISIO A., D'ARIANO G. M., PERINOTTI P. and TOSINI A., *Found. Phys.*, **45** (2015) 1137.
[16] BISIO A., D'ARIANO G. M., PERINOTTI P. and TOSINI A., *Found. Phys.*, **45** (2015) 1203.
[17] D'ARIANO G. M., *Phys. Lett. A*, **376** (2012) 697.
[18] BISIO A., D'ARIANO G. M. and PERINOTTI P., *Phys. Rev. A*, **94** (2015) 042120.
[19] D'ARIANO G. M. and PERINOTTI P., *Front. Phys.*, **12** (2017) 120301; Special Topic: Quantum Communication, Measurement, and Computing, in press (arXiv:1608.02004).

[20] D'Ariano G. M., Erba M., Perinotti P. and Tosini A., arXiv preprint arXiv:1511.03992 (2015).
[21] Hey A. J., *Feynman and Computation-Exploring the Limits of Computers* (Perseus Books) 1998.
[22] Bravyi S. B. and Kitaev A. Y., *Ann. Phys.*, **298** (2002) 210.
[23] D'Ariano G. M., Manessi F., Perinotti P. and Tosini A., *Int. J. Mod. Phys. A*, **17** (2014) 1430025.
[24] D'Ariano G. M., Manessi F., Perinotti P. and Tosini A., *EPL*, **107** (2014) 20009.
[25] Colodny R. G., (Editor) *From quarks to quasars: philosophical problems of modern physics* (University of Pittsburgh Press) 1986.
[26] de La Harpe P., *Topics in geometric group theory* (University of Chicago Press) 2000.
[27] Farb B., private communication (2016).
[28] Drutu C., private communication (2016).
[29] Tessera R., private communication (2016).
[30] D'Ariano G. M., Erba M. and Perinotti P., *Quantum field theory from principles and quantum walks on cayley graphs: an invitation*, unpublished (2016).
[31] Drutu C. and Kapovich M., *Geometric Group Theory*, Am. Math. Soc. Colloquium Publications, Vol. **63** (2017).
[32] Mackey G. W., *Am. J. Math.*, **73** (1951) 576.
[33] Mackey G. W., *Ann. Math.*, **55** (1952) 101.
[34] Mackey G. W., *Ann. Math.*, **58** (1953) 193.
[35] D'Ariano G. M., Mosco N., Perinotti P. and Tosini A., *Entropy*, **18** (2016) 228.
[36] D'Ariano G. M., Mosco N., Perinotti P. and Tosini A., *Phys. Lett. A*, **378** (2014) 3165.
[37] D'Ariano G. M., Mosco N., Perinotti P. and Tosini A., *EPL*, **109** (2015) 40012.
[38] D'Ariano G. M., Mosco N., Perinotti P. and Tosini A., *Discrete feynman propagator for the weyl quantum walk in 3+1 dimensions*, unpublished (2015).
[39] Kauffman L. H. and Noyes H. P., *Phys. Lett. A*, **218** (1996) 139.
[40] Greenberger D. M., *J. Math. Phys.*, **11** (1970) 2329.
[41] Ellis J., Mavromatos N. and Nanopoulos D. V., *Phys. Lett. B*, **293** (1992) 37.
[42] Lukierski J., Ruegg H. and Zakrzewski W. J., *Ann. Phys.*, **243** (1995) 90.
[43] 't Hooft G., *Class. Quantum Grav.*, **13** (1996) 1023.
[44] Amelino-Camelia G., *Phys. Lett. B*, **510** (2001) 255.
[45] Magueijo J. and Smolin L., *Phys. Rev. Lett.*, **88** (2002) 190403.
[46] Amelino-Camelia G., Ellis J., Mavromatos N., Nanopoulos D. V. and Sarkar S., *Nature*, **393** (1998) 763.
[47] Abdo A., Ackermann M., Ajello M., Asano K., Atwood W., Axelsson M., Baldini L., Ballet J., Barbiellini G., Baring M. *et al.*, *Nature*, **462** (2009) 331.
[48] Vasileiou V., Jacholkowska A., Piron F., Bolmont J., Couturier C., Granot J., Stecker F. W., Cohen-Tanugi J. and Longo F., *Phys. Rev. D*, **87** (2013) 122001.
[49] Amelino-Camelia G. and Smolin L., *Phys. Rev. D*, **80** (2009) 084017.
[50] Bibeau-Delisle A., Bisio A., D'Ariano G. M., Perinotti P. and Tosini A., *EPL*, **109** (2015) 50003.
[51] Bisio A., D'Ariano G. M. and Perinotti P., *Philos. Trans. R. Soc. A*, **374** (2016) .
[52] Susskind L., *Phys. Rev. D*, **16** (1977) 3031.
[53] Amelino-Camelia G. and Piran T., *Phys. Rev. D*, **64** (2001) 036005.
[54] Amelino-Camelia G., *Int. J. Mod. Phys. D*, **11** (2002) 35.
[55] Magueijo J. and Smolin L., *Phys. Rev. Lett.*, **88** (2002) 190403.
[56] Bisio A., D'Ariano G. M., Perinotti P. and Tosini A., *An interacting fermionic one-dimensional quantum cellular automaton*, draft (2016).

This page intentionally left blank

Proceedings of the International School of Physics "Enrico Fermi"
Course 197 "Foundations of Quantum Theory", edited by E. M. Rasel, W. P. Schleich and S. Wölk
(IOS, Amsterdam; SIF, Bologna) 2019
DOI 10.3254/978-1-61499-937-9-235

Revealing quantum properties with simple measurements

S. Wölk

Department Physik, Naturwissenschaftlich-Technische Fakultät
Universität Siegen - 57068 Siegen, Germany

Summary. — Since the beginning of quantum mechanics, many puzzling phenomena which distinguish the quantum from the classical world, have appeared such as complementarity, entanglement or contextuality. All of these phenomena are based on the existence of non-commuting observables in quantum mechanics. Furthermore, theses effects generate advantages which allow quantum technologies to surpass classical technologies. In this lecture note, we investigate two prominent examples of these phenomena: complementarity and entanglement. We discuss some of their basic properties and introduce general methods for their experimental investigation. In this way, we find many connections between the investigation of complementarity and entanglement. One of these connections is given by the Cauchy-Schwarz inequality which helps to formulate quantitative measurement procedures to observe complementarity as well as entanglement.

1. – Introduction

The quantum world inhabits many features such as complementarity [1], entanglement [2] or contextuality [3, 4] which distinguish it from the well-known classical world. These very puzzling features have been the starting point of many discussions from the very first beginning of quantum mechanics. Furthermore, they are also the key ingredients of the advantages offered by quantum technology nowadays. Therefore, getting

an understanding of these features, as deep as possible, is important to understand the foundations of quantum mechanics as well as developing new applications of quantum technologies.

The most famous of these phenomena is doubtlessly entanglement, which was first discussed by Einstein, Podolski and Rosen [5] and later quantified by Bell [6]. After hearing the first time about entanglement, some people have the impression that entanglement is equal to anticorrelation between the measurement results of two parties. The reason for this misunderstanding is the famous example of the Bell state $|\psi_-\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$ where a spin-measurement in z-direction on particle A reveals exactly the opposite measurement result than a measurement on particle B. However, this behavior has nothing to do with entanglement. First of all, the Bell state $|\phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ is also entangled but leads to perfect correlation of spin measurements in z-direction on particle A and B. Second, perfect anticorrelation also exist for classical measurements. Assume for example a box with a big sock and a small sock. The socks are randomly distributed between Alice and Bob. A measurement of the size (big $\equiv 1$, small $\equiv -1$) also leads to perfect anticorrelation. Therefore, anticorrelation alone is not an indicator for entanglement or quantumness in general.

In general, a measurement in a single measurement-basis alone can never demonstrate quantum properties such as entanglement, discord, coherence or duality, because these measurement can always be simulated by a classical system. Quantum behavior can only be proven by measuring at least two observables which do not commute. In the classical world different observables always commute and are jointly measurable. Consecutively, different correlation between measurements A_j and B_j can influence each other as we illustrate on our example with the socks. Here, we use the measurement of the color (blue $\equiv 1$, red striped $\equiv -1$) as a second measurement-basis. We asume *e.g.* that with a certain probability p_1 the big sock is blue and the small one is red striped and with probability p_2 it is the other way around such that $\langle A_1B_1\rangle = \langle A_2B_2\rangle = \langle A_2B_1\rangle = -1/\sqrt{2}$. By fixing these three correlations, the fourth correlation is bounded by

$$-\frac{1}{\sqrt{2}} \leq \langle A_1B_2\rangle \leq 0 \tag{1}$$

as summarized in table I.

In quantum physics, not all observables are jointly measurable, *e.g.* the observables described by the Pauli-spin matrices $\sigma_x^A \otimes \sigma_x^B$ and $\sigma_z^A \otimes \sigma_z^B$ are jointly measurable but $\sigma_z^A \otimes \sigma_x^A$ is not jointly measurable with the previous two. Therefore, the correlations $\langle \sigma_x^A \sigma_x^B\rangle$ and $\langle \sigma_z^A \sigma_z^B\rangle$ can exist at the same time. However, they destroy the correlation $\langle \sigma_z^A \sigma_x^B\rangle$.

As an example, we choose the observables $A_1 = \sigma_x$, $A_2 = \sigma_z$ and $B_1 = (\sigma_x + \sigma_z)/\sqrt{2}$, $B_2 = (\sigma_z - \sigma_x)/\sqrt{2}$ and the state $|\psi_-\rangle$. In this case, the resulting correlations $\langle A_1B_1\rangle = \langle A_2B_2\rangle = \langle A_2B_1\rangle = -1/\sqrt{2}$ are the same as in the classical example. However, because these three observables are not jointly measurable, they do not bound the correlation $\langle A_1B_2\rangle$ as in the classical way and we get for our example $\langle A_1B_2\rangle = +1/\sqrt{2}$ which is

TABLE I. – *Comparison of classical and quantum correlations. The observable are defined in the following way: i) $A_1 = B_1 =$ size and $A_2 = B_2 =$ color for the classical case, ii) $A_1 = \sigma_x$, $A_2 = \sigma_z$ and $B_1 = (\sigma_x + \sigma_z)/\sqrt{2}$, $B_2 = (\sigma_z - \sigma_x)/\sqrt{2}$ for the quantum state. The correlations $\langle A_1 B_1\rangle, \langle A_2 B_2\rangle$ and $\langle A_2 B_1\rangle$ bound the allowed correlations $\langle A_1 B_2\rangle$ in the classical case but not in the quantum case.*

State	$\langle A_1 B_1\rangle$	$\langle A_2 B_2\rangle$	$\langle A_2 B_1\rangle$	$\langle A_1 B_2\rangle$
p_1  $+p_2$ 	$-1/\sqrt{2}$	$-1/\sqrt{2}$	$-1/\sqrt{2}$	$-1/\sqrt{2} \leq \cdots \leq 0$
$\lvert\psi_-\rangle = (\lvert 01\rangle - \lvert 10\rangle)/\sqrt{2}$	$-1/\sqrt{2}$	$-1/\sqrt{2}$	$-1/\sqrt{2}$	$1/\sqrt{2}$

classically forbidden (see table I).

A more general formulation of the connection between different correlations is given by the Bell inequalities such as the CHSH-inequality [7]

$$|\langle A_1 B_1 + A_2 B_2 + A_2 B_1 - A_1 B_2\rangle| \underset{\text{sep}}{\leq} 2, \tag{2}$$

which limits the correlation between classical observables. However, it can be violated by using non-commuting observables and entangled states.

The phenomenon of non-commuting observables lies at the very heart of quantum mechanics and does not only induce the phenomenon of entanglement but also other quantum phenomena such duality and uncertainty [8-11], contextuality [3] or discord [12].

The aim of this lecture note is to give students with basic knowledge of quantum mechanics a better understanding of these quantum phenomena. Therefore, we will concentrate on a few selected topics, wave-particle duality as an example of complementarity (in sect. **2**) and entanglement (in sect. **3**), rather than trying to give a complete picture.

In detail, we introduce in sect. **2**˙1 the basic observables for wave-like and particle-like behavior and discuss the quantitative formulation of the wave-particle duality. Then, we give an overview of more advanced topics of wave-particle duality such as the quantum eraser (sect. **2**˙2) and higher-order wave-particle duality (sect. **2**˙3). We finish this chapter by pointing out the connection between measurements of wave-particle duality and entanglement in sect. **2**˙4.

We start our investigation of entanglement (sect. **3**) by summarizing the basic definitions and entanglement criteria for bipartite entanglement in sect. **3**˙1. Then, we discuss in detail tripartite entanglement in sect. **3**˙2 and illustrate why tripartite entanglement is more than just the combination of bipartite entanglement. We finish this section by a brief overview on multipartite entanglement in sect. **3**˙3 and a discussion on the spatial distribution of entanglement in sect. **3**˙4.

During the whole article, we will pay special attention on how these quantum features can be quantified and observed. Of special interested is the Cauchy-Schwarz inequality which helps to quantify duality as well as entanglement with a few simple measurements.

2. – Wave-particle duality

Wave-particle duality is closely related to the question of what is light, which is one of the key questions that inspired the development of quantum mechanics. Already the old Greeks discussed the nature of light, if light is a stream of particles (Democritus) or a ray (Plato, Aristotle). In the 17th century, the description of light as a wave (Huygens) made the construction of high quality lenses possible. In the 18th and 19th century, the wave-theory of light was very successful to describe and explain refraction, diffraction and interference (Fresnel, Young).

However, the wave theory of light completely failed to explain effects such as the Compton effect, black body radiation or the photo effect investigated in the late 19th century and the beginning of the 20th century. The contradiction of the wave- and the particle-nature of light might be best expressed by Einstein:

"It seems as though we must use sometimes the one theory and sometimes the other, while at times we may use either. We are faced with a new kind of difficulty. We have two contradictory pictures of reality; separately neither of them fully explains the phenomena of light, but together they do"([1]).

The solution of this dilemma is given by the concept of complementarity best described by:

"If information on one complementary variable is in principle available even if we choose not to know it ... we will lose the possibility of knowing the precise value of the other complementary variable" [13].

Two complementary variable are the which-way information D and the fringe visibility V in a double-slit experiment or an interferometer. The complementarity principle can than be formulated in a quantitative way by

$$D^2 + V^2 \leq 1, \tag{3}$$

with $0 \leq D, V \leq 1$. This inequality was first derived by D.M. Greenberger and A. Yasin for a specific way of defining the which-way information in a neutron interferometer [9] before it was proven in a general way by B.-G. Englert [10].

In the following, we will first discuss in sect. **2**˙1 the two observables of which-way information and fringe visibility and how to measure and interpret them before we derive eq. (3). Then, we discuss the generalization of this inequality to pairs of entangled particles in sect. **2**˙2 leading to the phenomenon of the so-called quantum eraser [14,15]. Finally, we discuss wave-particle duality in the multi-photon case in sect. **2**˙3. Although, the wave-particle inequality eq. (3) is valid in the mulit-photon case, it is very often not very informative. Therefore, we introduce a generalizations of eq. (3) to the multi-photon case in sect. **2**˙3.

([1]) Einstein, source: en.wikipedia.org/wiki/Wave-particle_duality

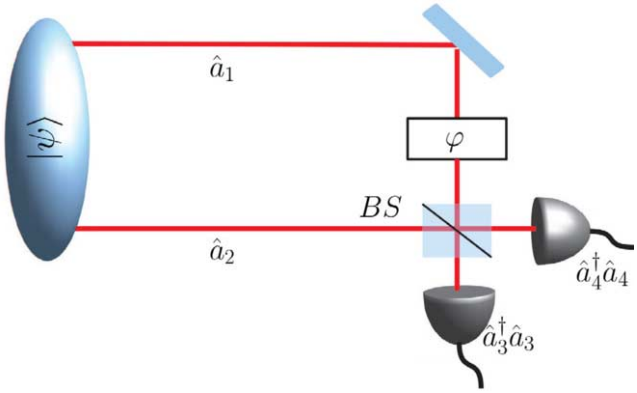


Fig. 1. – General two-mode state $|\psi\rangle$ of an interferometer. Different measurements can be performed by changing the phase φ between the two paths and by inserting or excluding a beam splitter.

2˙1. *Wave-particle duality: an inequality.* – In the following, we will investigate the properties of a two mode quantum state $|\psi\rangle$ with a measurement setup given in fig. 1. $|\psi\rangle$ can be described by

$$|\psi\rangle = f(a_1^\dagger, a_2^\dagger)|0\rangle_1|0\rangle_2 \tag{4}$$

with the creation operator $a_j^\dagger$ acting on mode j. The main properties of the creation and annihilation operator are given by

$$[a_j, a_k^\dagger] = \delta_{j,k}, \tag{5}$$

$$a^\dagger|n-1\rangle = \sqrt{n}|n\rangle, \tag{6}$$

$$a|n\rangle = \sqrt{n}|n-1\rangle, \tag{7}$$

where $|n\rangle$ describes a quantum state consisting of n photons. In the following, we assume that we are able to perform two different kind of measurements i) a simple photon number measurement in both modes described by their mean values

$$n_j = \langle \hat{n}_j \rangle = \langle a_j^\dagger a_j \rangle \tag{8}$$

or ii) an interferometric measurement as shown in fig. 1. Here, first the phase between both modes is changed via a unitary time evolution described by

$$U(\varphi) = \exp[-\mathrm{i}\varphi\, a^\dagger a]. \tag{9}$$

Then, both modes described by a_1 and a_2 interact with each other via a $50:50$ beam splitter leading to the output modes

$$a_3 = \frac{1}{\sqrt{2}}(a_1 + a_2), \quad a_4 = \frac{1}{\sqrt{2}}(a_1 - a_2). \tag{10}$$

Finally, a measurement of the photon number in each output mode is performed. The description of the beam splitter is not unique and depends on the experimental realization of the beam splitter. However, all results obtained in this manuscript are valid independent of the actually used beam splitter transformation.

These two measurements can be used to measure the particle-like behavior, in form of the distinguishability, and the wave-like behavior, in form of the visibility. These two observables can then be used to quantify the wave-particle duality as we will show in this section.

2˙1.1. Distinguishability. If the photon behaves like a particle, it can only be in one of the two modes at the same time. If we always prepare the same quantum state, we find $|\langle a_1^\dagger a_1 - a_2^\dagger a_2\rangle| = 1$. As a consequence, we quantify the particle-like behavior via the distinguishability $D = |\langle \hat{D}\rangle|$ with

$$\hat{D} \equiv \frac{a_1^\dagger a_1 - a_2^\dagger a_2}{\langle a_1^\dagger a_1\rangle + \langle a_2^\dagger a_2\rangle}\,, \tag{11}$$

which corresponds to a measurement setup as shown in fig. 1 without the beam splitter.

In other description of wave-particle duality, D describes the which-way information we gained by a measurement of a state $|\tilde{\psi}\rangle$ before the state reaches the beam splitter [10]. However, in this case our state $|\psi\rangle$ just describes the state $|\tilde{\psi}\rangle$ after it was disturbed by the first measurement. In this case, the first which-way measurement is just part of the preparation of the state and the measurement described in our scenario is just a second measurement which will reveal exactly the same value if nothing happened in between.

In both cases D described the probability of guessing the path of the photon correctly either after the fist measurement [10] or in the next experiment. Here, $D = 1$ corresponds to perfect knowledge and the photon will be always found in the same arm of an interferometer, whereas $D = 0$ corresponds to no knowledge at all and the photon will be detected in both arms with equal probability.

2˙1.2. Visibility. One important property of waves is that they can interfere with each others whereas particles cannot. The quality of an interference signale $S(\varphi)$ as depicted in fig. 2 can be measured by the visibility V which is given be the normalized difference between the maximum and the minimum of the signal

$$V = \frac{S_{\max} - S_{\min}}{S_{\max} + S_{\min}}\,. \tag{12}$$

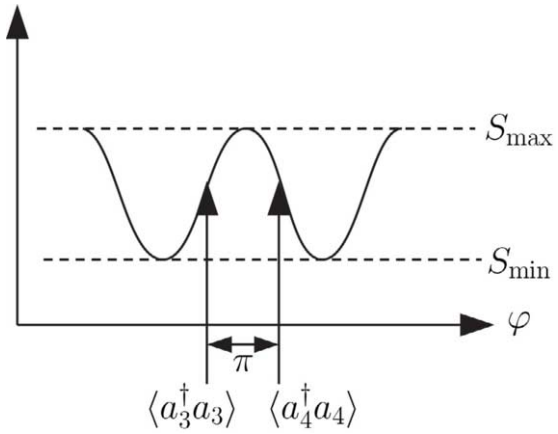


Fig. 2. – Definition of fringe visibility in terms of the observables $\langle a_3^\dagger a_3 \rangle_\varphi$ and $\langle a_4^\dagger a_4 \rangle_\varphi$. The two observables depend on the variable φ and $\langle a_4^\dagger a_4 \rangle_\varphi$ correspond to $\langle a_3^\dagger a_3 \rangle_{\varphi+\pi/2}$ shifted by π.

A value of $V = 1$ corresponds to perfect fringe visibility and perfect wave-like behavior whereas $V = 0$ indicates that no fringes at all are visible. If the phase shift and the beam splitter are present in fig. 1, then the two observables $\langle a_3^\dagger a_3 \rangle$ and $\langle a_4^\dagger a_4 \rangle$ correspond to exactly two measurement points of the interference signal separated by $\delta\varphi = \pi$. Therefore, the fringe visibility is given by $V = \max_\varphi \langle \hat{V} \rangle$ with

$$\hat{V} \equiv \frac{a_3^\dagger a_3 - a_4^\dagger a_4}{\langle a_3^\dagger a_3 \rangle + \langle a_4^\dagger a_4 \rangle}\,. \tag{13}$$

However, we want to describe the fringe visibility as a property of the initial state. As a consequence, we need to rewrite $\hat{V}$ in terms of the input modes of the beam splitter. With the help of the transformations

$$a_3 = \frac{1}{\sqrt{2}}(a_1 \mathrm{e}^{-\mathrm{i}\varphi} + a_2), \tag{14}$$

$$a_4 = \frac{1}{\sqrt{2}}(a_1 \mathrm{e}^{-\mathrm{i}\varphi} - a_2), \tag{15}$$

we arrive finally at

$$\hat{V} = \frac{a_1^\dagger a_2 \mathrm{e}^{\mathrm{i}\varphi} + a_1 a_2^\dagger \mathrm{e}^{-\mathrm{i}\varphi}}{\langle a_1^\dagger a_1 \rangle + \langle a_2^\dagger a_2 \rangle}\,. \tag{16}$$

As a consequence, the fringe visibility is given by

$$V = \max_\varphi \langle \hat{V} \rangle = \frac{2|\langle a_1^\dagger a_2 \rangle|}{\langle a_1^\dagger a_1 \rangle + \langle a_2^\dagger a_2 \rangle}\,. \tag{17}$$

2˙1.3. The wave-particle inequality. After defining a measure of particle-like and wave-like behavior, the main question is if we can observe both at the same time. The complementarity principle forbids $D = 1$ and $V = 1$ at the same time. However, what happens if we have some but not perfect path information, that is $0 < D < 1$. How much fringe visibility is possible in this case? This question can be answered by a simple calculation. The sum of the squares of both observables is given by

$$D^2 + V^2 = 1 - 4\frac{\langle a_1^\dagger a_1\rangle\langle a_2^\dagger a_2\rangle - |\langle a_1^\dagger a_2\rangle|}{\langle a_1^\dagger a_1\rangle + \langle a_2^\dagger a_2\rangle}\,. \tag{18}$$

The Cauchy-Schwarz inequality given by

$$|\langle\phi|\tilde{\phi}\rangle| \leq \langle\phi|\phi\rangle\langle\tilde{\phi}|\tilde{\phi}\rangle \tag{19}$$

leads to

$$\langle a_1^\dagger a_1\rangle\langle a_2^\dagger a_2\rangle - |\langle a_1^\dagger a_2\rangle| \geq 0, \tag{20}$$

if we identify $\langle\phi| = \langle\psi|a_1^\dagger$ and $|\tilde{\phi}\rangle = a_2|\psi\rangle$. This inequality is also true for mixed states, which can be proven either with the convexity of D and V or again with the Cauchy-Schwarz inequality. As a consequence, we finally arrive at

$$D^2 + V^2 \leq 1, \tag{21}$$

quantifying the wave-particle duality. Similar to the Heisenberg uncertainty relation, this inequality needs to be understood as a preparation inequality. That means, independent of how we measure D and V, there exists no state ρ which can be prepared in such a way that $D^2 + V^2 > 1$. This is reflected by the fact that in the derivation of the inequality, we assumed that either the distinguishability or the visibility will be measured in each single run of the experiment, but never both at the same time. However, there exist also setups, with the goal of simultaneous measurements of D and V with the help of *e.g.* entangled photons. For a short overview, see sect. **2**˙2.

In table II we have determined the distinguishability and the visibility for different states. $|\psi_1\rangle$ and $|\psi_2\rangle$ are examples of two pure states with perfect particle-like behavior. The state $\rho = p|\psi_1\rangle\langle\psi_1| + (1-p)|\psi_2\rangle\langle\psi_2|$ is an example of mixed state. The distinguishability $D^2 = (2p-1)^2$ reduces with the mixedness $\gamma = \mathrm{Tr}\,\rho^2 = D^2/2 + 1/2$. However, no fringe visibility is gained by mixing.

$|\psi_3\rangle$ is an example of a state with perfect fringe visibility. In general, every pure one photon state can be written in the form of $|\psi_4\rangle$. By tuning the variable α any amount of D or V can be establish. However, for a given α, the inequality eq. (21) is always saturated. A further investigation of $|\psi_4\rangle$ reveals, that a pure one-photon state leading to fringe visibility is always entangled and in this case V is a measure of entanglement (for further details about entanglement see sect. **3**). However, the relation between fringe

TABLE II. – *Distinguishability and visibility for different two-mode states.*

State	D_1	V_1
$\|\psi_1\rangle = \|01\rangle$	1	0
$\|\psi_2\rangle = \|10\rangle$	1	0
$p\|\psi_1\rangle\langle\psi_1\| + (1-p)\|\psi_2\rangle\langle\psi_2\|$	$(2p-1)^2$	0
$\|\psi_3\rangle = (\|1,0\rangle + \|0,1\rangle)/\sqrt{2}$	0	1
$\|\psi_4\rangle = (\cos\alpha\|1,0\rangle + \sin\alpha\|0,1\rangle)/\sqrt{2}$	$\cos^2(2\alpha)$	$\sin^2(2\alpha)$
$\|\psi_5\rangle = \|1,1\rangle$	0	0
$\|\psi_6\rangle = \|2,0\rangle + \|0,2\rangle$	0	0
$\|\psi_7\rangle = (\|0\rangle + \|1\rangle)(\|0\rangle + \|1\rangle)$	0	1/2
$\frac{1}{4}\|\psi_1\rangle\langle\psi_1\| + \frac{1}{4}\|\psi_2\rangle\langle\psi_2\| + \frac{1}{2}\|\psi_3\rangle\langle\psi_3\|$	0	1/2

visibility and entanglement is only true for single-photon states. If several photons might be present, than also product states such as *e.g.* $|\psi_7\rangle$ are able to produce interferences fringes. In general, eq. (21) is also valid for multi-photon states as exemplified by the lower 4 examples in table II. However, in many cases the inequality does not contain much information because both observables are equal to zero and it is not possible to distinguish entangled states from separable states. For example the product state $|\psi_7\rangle$ and the mixed entangled state $\rho = \frac{1}{4}|\psi_1\rangle\langle\psi_1| + \frac{1}{4}|\psi_2\rangle\langle\psi_2| + \frac{1}{2}|\psi_3\rangle\langle\psi_3|$ inhabit the same values for D and V. However, also for multi-photon states, meaningful duality inequalities can be establish as we will show in sect. **2**˙3.

2˙2. *Simultaneous measurements.* – In the previous section, we discussed the wave-particle duality in a simple scheme, where we either measured the distinguishability D or the visibility V or had at least a defined temporal order of the measurements. Similar to the EPR setup [5] where position and momentum of a particle are measured "simultaneously", we can also measure the distinguishability and the visibility simultaneously by using entanglement and a setup as shown in fig. 3. To perform this task, we entangle our photon A with another photon B with respect to their path information. In this way, we are able to measure the interference of a photon A and later decide if we want to measure also its which-way information with the help of photon B. If we decide to measure the which-way information, the inference pattern must be erased due to the wave-particle duality, whereas the interference survives if we decide to not to measure the which-way information. This phenomenon is known as delayed choice quantum eraser [14, 15]. On the first sight, this phenomenon occurs to be quite puzzling because it seems that the measurement at photon B changes the past of photon A. However, a simple analysis of this setup reveals that nothing is really erased and that eq. (21) is also valid in this case.

To understand this phenomenon, we assume a simple setup as displayed in fig. 3. The photons A and B are prepared in a superposition $|\Psi\rangle_{AB} = (|00\rangle + |11\rangle)/\sqrt{2}$ where either both photons appear in the upper or lower arm of the interferometer. The interferometer

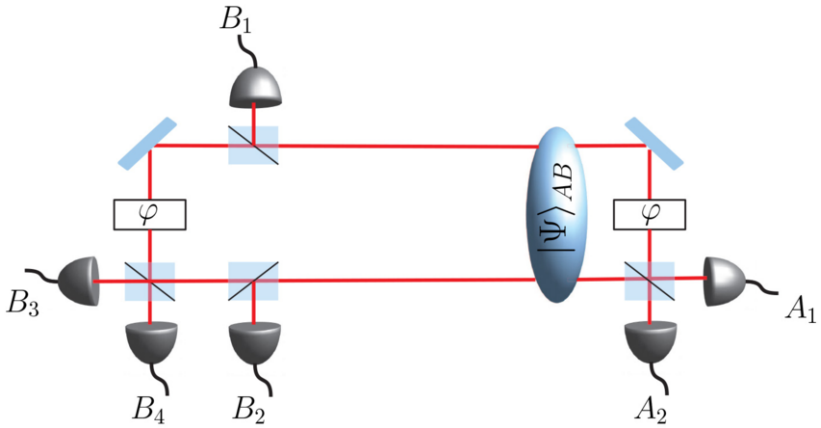


Fig. 3. – Quantum eraser: Two entangled photons are both either prepared in the upper arm or the lower arm of the interferometer. Photon A travels to the right and is detected at detector A_1 or A_2. Photon B travels to the left and is detected at detector B_1, B_2, B_3 or B_4. Coincidence measurement of photon A with detecting photon B at detector 3 or 4 leads to interference whereas clicks at B_1 or B_2 "erase" the interference.

is built in such a way, that photon A reaches the detector A_1 or A_2 before photon B reaches any beamsplitter or detector. Furthermore, the phase φ is chosen in such a way that detector A_1 (B_3) clicks if photon A (B) is in state $(|+\rangle = |0\rangle + |1\rangle)/\sqrt{2}$ and detector A_2 (B_4) clicks for the state $|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$. By inserting the additional beamsplitters and detectors B_1 and B_2 the photon itself randomly chose if the which-way information (detector B_1 or B_2 clicks) or the interference (detector B_3 or B_4 clicks) is measured.

Let us first assume, that Alice (measuring photon A) and Bob (measuring photon B) do not communicate. What does Alice observe? In this case, the reduce state at Alice side is given by

$$\rho_A = \mathrm{Tr}_B(|\Psi\rangle_{AB}\langle\Psi|) = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) \tag{22}$$

which is a completely mixed state. From table II we know that the fringe visibility for this state is given by $V_A = 0$. By rewriting the reduced state $\rho_A = (|+\rangle\langle +| + |-\rangle\langle -|)/2$ we can also interpret this experiment in such a way, that Alice gets randomly the state $|+\rangle$ with fringe visibility $V_+ = 1$ and the state $|-\rangle$ with $V_- = -1$. As long as Alice is not able to distinguish whether she gets the state $|+\rangle$ or $|-\rangle$ via post selection, she is not able to see fringe visibility. The distinguishability D_A would be also equal to zero if Alice would have decided to measure it. As a consequence, Alice can neither see fringes nor does she has any which-path information without communicating with Bob.

If Alice and Bob communicate with each other, we have to rewrite the state $|\Psi\rangle_{AB}$ in the corresponding measurement basis depending on the measurement outcomes. We

rewrite the state

$$|\Psi\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = \frac{1}{2}[(|+\rangle + |-\rangle)|0\rangle_B + (|+\rangle - |-\rangle)|1\rangle_B] \tag{23}$$

for a measurement of distinguishability at Bobs side (detector 1 or 2 clicks) and a fringe visibility measurement at Alice side. As a consequence, a click at detector $B_1 \hat{=} |0\rangle_B$ corresponds to an equal probability that either $A_1 \hat{=} |+\rangle$ or $A_2 \hat{=} |-\rangle$ clicks. The same happens for a click at detector B_2. As a consequence, a click at B_1 or B_2 does not help Alice to postselect between $|+\rangle$ and $|-\rangle$.

However, if B_3 or B_4 clicks, we describe the state by

$$|\Psi\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = \frac{1}{\sqrt{2}}(|++\rangle + |--\rangle). \tag{24}$$

As a consequence, if Alice selects only the measurement outcomes which correspond to a click at $B_3 \hat{=} |+\rangle$ ($B_4 \hat{=} |-\rangle$) at Bob sides, she gets a perfect fringe visibility of $V_{A|B_3} = 1$ ($V_{A|B_4} = -1$).

As a consequence, the wave-particle duality in the form of eq. (21) stays also valid in this scenario. Furthermore, it is not the which-way measurement at Bobs side which erases the fringe visibility at Alice side. However, it is the interference measurement at Bobs side which helps Alice to post select her measurement results to make the fringes appear.

In general, no violation of the wave-particle duality is possible if the experimental data is analyzed correctly. However, sometimes the post selection as described above is not that apparent in a real experiment [16]. In these cases a violation of eq. (21) can occur due to unfair sampling [17, 18].

2˙3. *Higher-order wave-particle duality.* – The wave-particle duality as formulated in eq. (21) is valid for all states independent of the involved number of photons. However, the last four examples of table II also reveal, that eq. (21) might be not very informative if higher photon numbers are involved. In these cases, the generalization of eq. (21) to higher-orders photon numbers as demonstrated in this section come into play.

2˙3.1. Higher-order distinguishability. The idea of a higher-order distinguishability of order k is to be only sensitive to states with at least k photons. Therefore, we define the k-th–order distinguishability

$$\hat{D}_k \equiv \frac{(a_1^\dagger)^k a_1^k - (a_2^\dagger)^k a_2^k}{\langle (a_1^\dagger)^k a_1^k \rangle + \langle (a_2^\dagger)^k a_2^k \rangle}\,, \tag{25}$$

where we have used the k-th–order autocorrelation function. In contrast to the k-th–moment of the number operator given by $(a^\dagger a)^k$, the measurement of $(a^\dagger)^k a^k$ requires a different measurement setup [19]. Therefore, it really takes additional information into

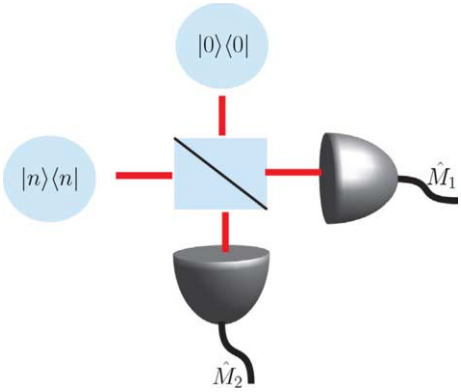


Fig. 4. – The k-th–order autocorrelation function can be interpreted as the probability to get exactly k out of n photons at detector M_1.

account, whereas the higher moments can be estimated by a detailed data analysis of the data already collected for the distinguishability D_1.

Possible measurements can *e.g.* be performed by detectors which need k-times the energy of a single photon to be activated [20] or by photon counting measurements with single-photon resolution.

The action of the k-th–order autocorrelation function on a Fock state $|n\rangle$ is given by

$$\frac{1}{k!}(\hat{a}^\dagger)^k\hat{a}^k|n\rangle = \binom{n}{k}|n\rangle. \tag{26}$$

As a consequence, the k-th–order autocorrelation function is non-zero for states with $n > k$ photons. Contributions with different photon numbers n are weighted with the binomial coefficient $\binom{n}{k}$. This fact can be best understood if we consider a measurement setup as given in fig. 4. Here, a Fock state of exactly n photons and the vacuum state interact on a 50:50 beam splitter. As a consequence, the probability that k photons leave the beam splitter at exit one is given by $\binom{n}{k}/\sqrt{2^n}$ similar to classical probability theory.

The k-th–order distinguishability takes only the diagonal elements of the state ρ in the computational basis into account. We exemplify the different contribution for a state of constant total photon number $\langle a_1^\dagger a_1 + a_2^\dagger a_2\rangle = 4$ for D_3 and D_2 in fig. 5.

2˙3.2. Higher-order visibility. In a similar way, the fringes visibility V can be generalized to

$$\hat{V} \equiv \frac{(a_1^\dagger)^k a_2^k \mathrm{e}^{\mathrm{i}\varphi} + a_1^k (a_2^\dagger)^k \mathrm{e}^{-\mathrm{i}\varphi}}{\langle (a_1^\dagger)^k a_1^k\rangle + \langle (a_2^\dagger)^k a_2^k\rangle}, \tag{27}$$

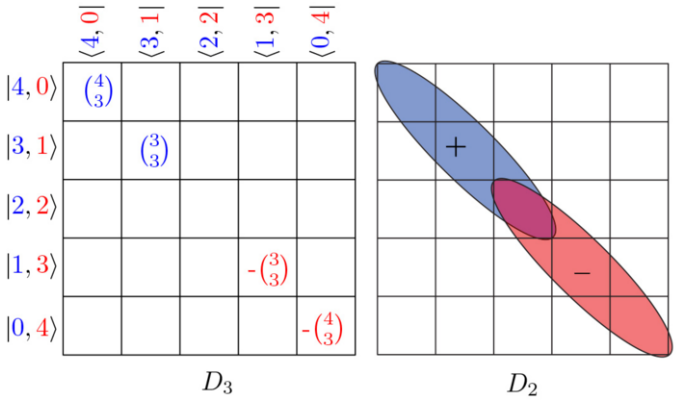


Fig. 5. – The distinguishability D_k is given by the sum of all diagonal terms with at least k excitations in a single mode weighted with different weight factors depending on k and the number of excitations.

where we have used the k-th–order coherence

$$\frac{1}{k!}\langle n+k, j|(\hat{a}_1^\dagger)^k \hat{a}_2^k |n, j+k\rangle = \sqrt{\binom{n}{k}\binom{j}{k}}. \tag{28}$$

As a consequence, the k-th–order visibility is sensitive to states where coherence exist between states where exactly k photons change from one mode to the other. Therefore, V_k is determined by the k-th off-diagonal of ρ in the computational basis as demonstrated in fig. 6.

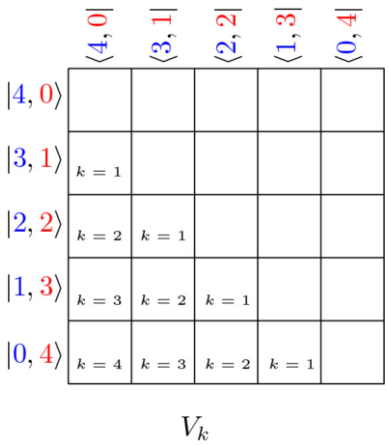


Fig. 6. – The visibility V_k is determined by the off-diagonal elements of the state. Each order k is given by another secondary diagonal.

Higher-order visibilities V_k cannot be measured by a single measurement setting in contrast to the first order visibility V. Nevertheless, the measurement setup depicted in fig. 1 together with detectors described by $(a^\dagger)^k a^k$ can determine V_k if the measurements are performed for several different phases φ_j. We will demonstrate this measurement procedure for $k = 2$. In this case, the sum of the expectation values of both detectors for an arbitrary but fixed phase φ is given by

$$R^+_{2,\varphi} \equiv \langle (a_3^\dagger)^2 (a_3)^2 + (a_4^\dagger)^2 (a_4)^2 \rangle \tag{29}$$

$$= \frac{1}{2} \langle (a_1^\dagger)^2 a_1^2 + (a_2^\dagger)^2 a_2^2 + 4 a_1^\dagger a_1 a_2^\dagger a_2 + \underbrace{(a_1^\dagger)^2 a_2^2 \mathrm{e}^{\mathrm{i}\varphi} + a_1^2 (a_2^\dagger)^2 \mathrm{e}^{-\mathrm{i}\varphi}}_{\sim \hat{V}_2} \rangle. \tag{30}$$

By using the correct phase relations, we are able to isolate the real and the imaginary part of $\langle (a_1^\dagger)^2 a_2^2 \rangle$ and arrive at

$$2|(a_1^\dagger)^k a_2^k|^2 = \underbrace{\left(R^+_{2,\varphi'} - R^+_{2,\varphi'+\pi/2} \right)^2}_{\text{real part}} + \underbrace{\left(R^+_{2,\varphi'-\pi/4} - R^+_{2,\varphi'+\pi/4} \right)^2}_{\text{imaginary part}}. \tag{31}$$

In general, odd orders of V_k are determined by the difference of the detectors M_3 and M_4 whereas even orders by their sum. By defining

$$R^\pm_{k,\varphi} \equiv \frac{2^{k-1}}{k} \langle (a_3^\dagger)^k (a_3)^k \pm (a_4^\dagger)^k (a_4)^k \rangle, \tag{32}$$

we get for odd orders of k

$$2|(a_1^\dagger)^k a_2^k|^2 = \frac{2^{2k-2}}{k^2} \left[\left(\sum_{m=0}^{k-1} R^-_{k,\varphi'+2m\pi/k} \right)^2 + \left(\sum_{m=0}^{k-1} R^-_{k,\varphi'-\pi/(2k)+2m\pi/k} \right)^2 \right] \tag{33}$$

and

$$2|(a_1^\dagger)^k a_2^k|^2 = \frac{2^{2k-2}}{k^2} \left[\left(\sum_{m=0}^{k-1} (-1)^m R^+_{k,\varphi'+m\pi/k} \right)^2 + \left(\sum_{m=0}^{k-1} (-1)^m R^+_{k,\varphi'-\pi/(2k)+m\pi/k} \right)^2 \right] \tag{34}$$

for even orders [21].

2˙3.3. Higher-order wave-particle duality. The higher-order distinguishability and visibility are again connected by the Cauchy Schwarz inequality

$$\left| \langle (a_1^\dagger)^k a_2^k \rangle \right|^2 \leq \langle (a_1^\dagger)^k a_1^k \rangle \langle (a_2^\dagger)^k a_2^k \rangle \tag{35}$$

TABLE III. – *Higher-order distinguishability and visibility for different two-mode states including the examples of table II which involve multi-photon states.*

State	D_1	V_1	D_2	V_2
$\|\psi_5\rangle = \|1,1\rangle$	0	0	0	0
$\|\psi_6\rangle = \|2,0\rangle + \|0,2\rangle$	0	0	0	1
$\|\psi_7\rangle = (\|0\rangle + \|1\rangle)(\|0\rangle + \|1\rangle)$	0	1/2	0	0
$\frac{1}{4}\|\psi_1\rangle\langle\psi_1\| + \frac{1}{4}\|\psi_2\rangle\langle\psi_2\| + \frac{1}{2}\|\psi_3\rangle\langle\psi_3\|$	0	1/2	0	0
$\|\psi_8\rangle = \|20\rangle$	1	0	1	0
$\|\psi_9\rangle = \|4,2\rangle + \|2,4\rangle$	0	0	0	$(6/7)^2$

similiar to the first-order wave-particle duality. As a result, we get in the higher-order case an exactly similar duality relation given by

$$D_k^2 + V_k^2 = 1 + 4\frac{\left|\langle(a_1^\dagger)^k a_2^k\rangle\right|^2 - \langle(a_1^\dagger)^k a_1^k\rangle\langle(a_2^\dagger)^k a_2^k\rangle}{\left(\langle(a_1^\dagger)^k a_1^k\rangle + \langle(a_2^\dagger)^k a_2^k\rangle\right)^2} \leq 1 \tag{36}$$

as for the first order. Determining the higher-order observables for the initial example states given in table II leads to the results shown in table III, which reveals the advantages of using a series of observables and duality relations instead of a single one. The here introduced higher-order duality interprets a collection of k photons as a single larger quasi particle. As a consequence, we get non-zero results for D_2 and V_2 for the states $|\psi_6\rangle = |02\rangle + |20\rangle$, $|\psi_8\rangle = |2,0\rangle$ and $|\psi_9\rangle$. The states $|\psi_5\rangle$ and $|\psi_7\rangle$ including also two photons states lead to $D_k = V_k = 0$ on the other hand because the two photons do not bunch together.

2˙4. *Duality and entanglement.* – The higher-order visibility V_k is similar to the first-order visibility an indicator of entanglement if the total photon number of the state is fixed. However, for states with a varying total photon number additional measurements are necessary for proving entanglement. As we will demonstrate in sect. **3** entanglement can be verified by comparing diagonal and off-diagonal elements with the help of the Cauchy-Schwarz inequality. Entanglement criteria can be developed by applying the Cauchy-Schwarz inequality (CSI) only to subsystems (for more details see sect. **3**) whereas the duality inequality follows from applying the CSI to the total system. As a consequence, the higher-order visibility $\hat{V}_k$ needs to be compared to the higher-order coincidence

$$\hat{C}_k \equiv \frac{(a_1^\dagger)^k a_1^k (a_2^\dagger)^k a_2^k}{\left(\langle(a_1^\dagger)^k a_1^k\rangle + \langle(a_2^\dagger)^k a_2^k\rangle\right)^2}, \tag{37}$$

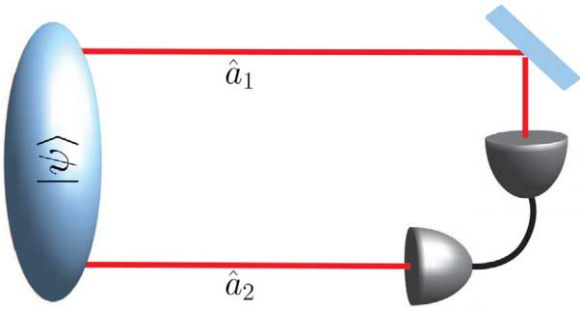


Fig. 7. – Measurement setup of a coincidence measurement C_k necessary to prove entanglement together with V_k.

which can be measured by a setup similar to fig. 7. Separable states obey the inequality

$$V_k^2 \underset{\text{sep}}{\leq} 4C_k, \tag{38}$$

which follows directly from eq. (59), which we will prove in the next section, by using $\hat{A}_1 = (a_1^\dagger)^k$, $\hat{B}_2 = a_2^k$ and $\hat{A}_2 = \hat{B}_1 = \mathbb{1}$. Violation of this inequality demonstrates entanglement.

In general, also the distinguishability D_k can be used to detect entanglement by using the observable $W_k = \max_{\varphi} |\langle \hat{W}_k \rangle|$ with

$$\hat{W}_k \equiv \frac{a_1^k a_2^k \mathrm{e}^{\mathrm{i}\varphi} + (a_1^\dagger)^k (a_2^\dagger)^k \mathrm{e}^{-\mathrm{i}\varphi}}{\langle (a_1^\dagger)^k a_1^k \rangle + \langle (a_2^\dagger)^k a_2^k \rangle}. \tag{39}$$

Entanglement of a state is demonstrated if it violates the inequality

$$D_k^2 + W_k^2 \underset{\text{sep}}{\leq} 1, \tag{40}$$

which can be proven again by eq. (59).

TABLE IV. – *Results of the entanglement test using eq.* (38).

State	D_1	V_1	C_1	W_1	Entangled
$\|\psi_4\rangle = (\cos\alpha\|1,0\rangle + \sin\alpha\|0,1\rangle)/\sqrt{2}$	$\cos^2(2\alpha)$	$\sin^2(2\alpha)$	0	0	yes
$\|\psi_5\rangle = \|1,1\rangle$	0	0	1/4	0	no
$\|\psi_7\rangle = (\|0\rangle + \|1\rangle)(\|0\rangle + \|1\rangle)$	0	1/2	1/16	0	no
$\rho = \frac{1}{4}\|\psi_1\rangle\langle\psi_1\| + \frac{1}{4}\|\psi_2\rangle\langle\psi_2\| + \frac{1}{2}\|\psi_3\rangle\langle\psi_3\|$	0	1/2	0	0	yes

The entangled states of table II can all be detected by the criterion eq. (38) as demonstrated in table IV. On the other side, eq. (40) only detects states with coherence between $|n, j\rangle$ and $|n+k, j+k\rangle$ such as *e.g.* $|\psi\rangle = \sqrt{3}|00\rangle + |11\rangle$ with $D_1 = 0$ but $W_1 = \sqrt{3}$.

3. – Entanglement

Entanglement [2] is the most puzzling and gainful concept of quantum mechanics and a direct consequence of the concept of superposition of quantum states. Originally this effect was formulated 1935 as a contradiction to reality obeying local realism by Einstein, Podolsky and Rosen [5]. It took 29 years until the difference between local realism and quantum mechanics could be quantified [6] and experimentally tested by means of the Bell inequality. Many experiments [22-26] have been performed since then, which strongly support the concept of quantum mechanics. Most importantly entanglement is used in quantum technologies such as quantum information processing [27] and quantum metrology [28] to overcome limits of classical devices.

A deep understanding of entanglement in theory and experiment helps us to develop quantum technologies. Therefore, it is also important to develop entanglement criteria [29] which can be easily applied to experiments to certify and investigate the existing entanglement in a given setup and its behavior under the influence of noise.

The most famous entanglement criterion is given by the Bell inequality in eq. (2). However, not all entangled states violate the Bell inequalities. Moreover, many highly entangled states such as the Greenberger-Horne-Zeilinger (GHZ) state [30] which violate Bell inequalities are very vulnerable with respect to noise and therefore not the best choice for *e.g.* quantum metrology under realistic conditions [31]. Thus, further criteria to characterize entanglement also for weakly entangled states are necessary.

In the following, we will give a small review about entanglement with special attention given to the question of experimental entanglement characterization. We first define bipartite entanglement and give several examples of how to detect it experimentally. Then, we will continue with tripartite and multipartite entanglement and demonstrate that multipartite entanglement is more than the summation of bipartite entanglement.

3˙1. *Bipartite entanglement.* – A pure quantum state is entangled, if it cannot be written as

$$|\Psi\rangle_{A,B} = |\psi\rangle_A \otimes |\phi\rangle_B; \tag{41}$$

a mixed quantum state is entangled, if it cannot be written as

$$\rho_{A,B} = \sum_j p_j |\psi_j\rangle_A\langle\psi_j| \otimes |\phi_j\rangle_B\langle\phi_j|. \tag{42}$$

A problem of this definition is that the decomposition of a mixed quantum state is not unique. Therefore eq. (42) indicates separability if such a decomposition is found. Though, a decomposition including entangled states does not prove that the mixed state

itself is entangled. The definition of bipartite entanglement of pure states, on the other hand, can be used to directly demonstrate entanglement with the help of the Schmidt decomposition [32].

3˙1.1. Schmidt decomposition. The Schmidt decomposition [32] is a method to transform a bipartite pure state $|\psi\rangle_{A,B}$ into the following form:

$$|\Psi\rangle_{A,B} = \sum_{j=1}^{r} p_j |a_j\rangle|b_j\rangle, \tag{43}$$

where $\{|a_j\rangle\}$ and $\{|b_j\rangle\}$ form an orthogonal basis of the two parties A and B, respectively. The maximal summation index r is called Schmidt rank and indicates entanglement if $r > 1$. We will demonstrate the Schmidt decomposition on the example

$$|\Psi\rangle_{A,B} = \frac{1}{\sqrt{6}} \left(|00\rangle + |01\rangle + \sqrt{2}|10\rangle - \sqrt{2}|11\rangle \right). \tag{44}$$

This state is obviously not in the form given by eq. (43) because the basis states $|0\rangle, |1\rangle$ appear twice for each party. To determine the correct basis states for the Schmidt decomposition, we have to determine the eigenstates of the reduced density matrix

$$\rho_B = \mathrm{Tr}_A \left(|\Psi\rangle_{A,B}\langle\Psi| \right), \tag{45}$$

$$= \frac{1}{3}|+\rangle\langle+| + \frac{2}{3}|-\rangle\langle-|, \tag{46}$$

with $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. The Schmidt coefficients are given by $p_j = \sqrt{\lambda_j}$ with the eigenvalues λ_j of ρ_B. The Schmidt decomposition of the state $|\Psi\rangle_{A,B}$ is given by its decomposition into the eigenstates of ρ_B. Thus, the Schmidt decomposition of $|\Psi\rangle_{A,B}$ is given by

$$|\Psi\rangle_{A,B} = \sqrt{\frac{1}{3}}|0\rangle_A|+\rangle_B + \sqrt{\frac{2}{3}}|0\rangle_A|-\rangle_B, \tag{47}$$

and $|\Psi\rangle_{A,B}$ is obviously entangled since $r = 2 > 1$. The Schmidt rank is independent of whether it is determined via the reduced density matrix of system A or system B. In a simplified version, the Schmidt decomposition states that a pure bipartite state is entangled if its reduced density matrices are mixed. However, to apply the Schmidt decomposition the state has not only to be pure, but also we need full knowledge about the state. This is rather difficult and extensive measurements are required. Nevertheless, the Schmidt decomposition can be very helpful to identify the optimal measurement directions for experimental entanglement verification [33] as we will demonstrate with the following example.

One way to detect entanglement of a two-qubit system is to measure various combinations of the Pauli matrices, that is

$$M_{j,k} = \sigma_j \otimes \sigma_k. \tag{48}$$

The sum of these observables is bounded from above by

$$\sum_{j,k=1}^{3} \langle M_{j,k} \rangle_{\text{sep}} \leq 1 \tag{49}$$

for separable states, whereas this bound can be violated by entangled states (for proof see Appendix A). In the worst case, 9 observables are needed to be measured to detect entanglement via this criterion. Though, for pure entangled states, it is sufficient to measure two observables, if the right directions are chosen [33]. To find these directions, Alice and Bob need to determine their reduced density matrices by measuring their local bloch vectors defined by $\vec{v} = (\langle\sigma_x\rangle, \langle\sigma_y\rangle, \langle\sigma_z\rangle)^T$. With the help of the eigenstates $|v\rangle$ and $|v_\perp\rangle$ of the operator $V = \vec{v}\vec{\sigma}$ they define the new measurement directions

$$\sigma_{z'} = |v\rangle\langle v| - |v_\perp\rangle\langle v_\perp|,\ \sigma_{y'} = \mathrm{i}|v_\perp\rangle\langle v| - \mathrm{i}|v\rangle\langle v_\perp|. \tag{50}$$

Every pure entangled state can then be detected by solely measuring $M_{z'z'}$ and $M_{y'y'}$ and using eq. (49). In contrast to the Schmidt decomposition, the states do not necessarily need to be pure. As soon as we find $M_{z'z'} + M_{y'y'} > 1$ we know that the state is entangled independent of whether the state is pure or not. The only difference is, that for strongly mixed states, maybe additional terms have to be measured to exceed the threshold.

3˙1.2. The positive partial transpose criterion. In a real quantum experiment, we can never guarantee pure quantum states. Furthermore, interesting phenomena of entanglement appear when we consider not only pure but mixed states (see sect. **3**˙2). Therefore, it is very convenient to have entanglement criteria, which are also valid for mixed states. The most famous criterion for entanglement of mixed states is the positive partial transpose (PPT) criterion [34,35]. It is based on the fact, that the transposition is a positive but not completely positive map. This means that the transpose of ρ_A is positive semidefinite ($\rho_A^T \geq 0$) if ρ_A itself is positive semidefinite. Still, if the transpose is only applied to a subsystem of a composed quantum state $\rho_{A,B}$, this is not necessarily the case. Assume for example the Bell state $|\psi^+\rangle_{A,B} = (|00\rangle + |11\rangle)/\sqrt{2}$ and its corresponding density matrix

$$\rho_{A,B} \equiv \frac{1}{2}\left(|00\rangle\langle 00| + |11\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 11|\right). \tag{51}$$

Its partial transpose is given by

$$\rho_{A,B}^{PT} = \frac{1}{2}\left(|00\rangle\langle 00| + |01\rangle\langle 10| + |10\rangle\langle 01| + |11\rangle\langle 11|\right) \tag{52}$$

with the resulting eigenvalues $\lambda = \pm 1/2$. In the matrix representation (here in the standard computational basis $|00\rangle, |01\rangle, |10\rangle, |11\rangle$) the partial transposition is calculated by dividing the matrix into 4 submatrices is seen below

$$\left(\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{array}\right) \underset{PPT}{\Rightarrow} \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right). \tag{53}$$

The partial transpose is then calculated by either exchanging the two off-diagonal matrices as a total or by transposing each submatrix depending on whether the partial transpose should be calculated with respect to partition A or B. This example shows that the partial transpose of an entangled state can lead to negative eigenvalues. We call states with negative partial transpose such as $|\psi^+\rangle$ NPT-states.

On the other hand, the partial transpose of a separable state

$$\rho_{\text{sep}}^{PT} = \sum_j p_j |\psi_j\rangle_A\langle\psi_j| \otimes |\phi_j^*\rangle_B\langle\phi_j^*| \tag{54}$$

with $\rho_B^T = \rho_B^*$, stays a valid quantum state with positive eigenvalues. As a consequence, we find that a state is entangled if it possesses a negative partial transpose. Yet, this criterion is only necessary and sufficient for 2×2 and 2×3 systems [35]. For all other systems, we can make no statement about the entanglement if a state is PPT.

Nevertheless, if a state is PPT or not is a very important property. First of all, even if a state with PPT is entangled, it is only bound entangled [36]. This means, even if Alice and Bob possess several copies of this state, they cannot create the singlet state $|\psi^-\rangle$ out of them by using only local operations and classical communication (LOCC). In addition, bipartite bound entangled states cannot be directly used for teleportation or densed coding. Though, they can support these tasks if used as additional resources [37]. Furthermore, PPT can be much easier tested and can be used to detect multipartite entanglement with the help of semidefinite programming [38]. As a consequence, the PPT criterion is a very important criterion for theoretical entanglement investigations. However, the experimental application of the PPT criterion suffers from several disadvantages. First of all, total knowledge of the state is necessary (gained *e.g.* by state tomography) which leads to great experimental effort. Furthermore, experimental state tomography can lead to negative eigenvalues of the state ρ itself [39] if using linear inversion. The use of the maximum-likelihood method to avoid this negativity can lead to an overestimation of the entanglement [40]. Hence, the most convenient way to experimentally verify entanglement is given by inequalities directly based on experimental observables as we will demonstrate in the next section.

3˙1.3. Detecting entanglement with the help of the Cauchy-Schwarz inequality. The most convenient way to experimentally detect entanglement are inequalities of expectation values. There exist many different kinds of entanglement criteria based on inequalities, *e.g.*

for discrete and continuous variables [41,42] in the bipartite and multipartite case [43-45]. Originally, they have been developed with different methods such as using uncertainty relations or using the properties of separable states. Although all the above-cited criteria did not seem to have much in common on the first sight, all of them can be proven with a single general method [46]. This general method is given by the Cauchy-Schwarz inequality as we will demonstrate in this section.

The expectation value of an operator can be upper bounded by the Cauchy-Schwarz inequality by interpreting it as the scalar product of two vectors. In general, an operator O can be part of the bra- or the ket-vector or can be split into $O = O_1 O_2$. Therefore, the most general bound is given by

$$|\underbrace{\langle\psi|O_1}_{\vec{x}^\dagger}\underbrace{O_2|\psi\rangle}_{\vec{y}}|^2 \leq \underbrace{\langle\psi|O_1 O_1^\dagger|\psi\rangle}_{|\vec{x}|^2}\underbrace{\langle\psi|O_2^\dagger O_2|\psi\rangle}_{|\vec{y}|^2}. \tag{55}$$

Hence, we get the general bound

$$|\langle A_1 A_2 B_1 B_2\rangle|^2 \leq \langle A_1 A_1^\dagger B_1 B_1^\dagger\rangle\langle A_2^\dagger A_2 B_2^\dagger B_2\rangle \tag{56}$$

for a bipartite state where the operators A and B acting solely on subsystem A or B, respectively. The expectation value $\langle AB\rangle$ factorizes for product states (ps). As a result, we find the upper bound

$$|\langle A_1 A_2 B_1 B_2\rangle|^2 \underset{\text{ps}}{=} |\langle A_1 A_2\rangle||\langle B_1 B_2\rangle| \tag{57}$$

$$\underset{\text{ps}}{\leq} \langle A_1 A_1^\dagger\rangle\langle A_2 A_2^\dagger\rangle\langle B_1 B_1^\dagger\rangle\langle B_2 B_2^\dagger\rangle \tag{58}$$

for product states. Expectation values of different subsystems can be recombined again, which leads us to the general entanglement criterion

$$|\langle A_1 A_2 B_1 B_2\rangle|^2 \underset{\text{sep}}{\leq} \langle A_1 A_1^\dagger B_2^\dagger B_2\rangle\langle A_2^\dagger A_2 B_1 B_1^\dagger\rangle. \tag{59}$$

This inequality is also valid for mixed separable states [46]. If eq. (59) provides a stricter bound than eq. (56), then it can be used to detect entanglement. For example, the criterion [42]

$$|\langle a^m (b^\dagger)^n\rangle|^2 \underset{\text{sep}}{\leq} \langle (a^\dagger)^m a^m (b^\dagger)^n b^b\rangle, \tag{60}$$

where a and b denote the annihilation operators of system A and B, follows directly by choosing $A_2 = a$, $B_1 = b^\dagger$ and $A_1 = B_2 = \mathbb{1}$. The entanglement criterion

$$|\rho_{01,10}|^2 \underset{\text{sep}}{\leq} \rho_{00,00}\rho_{11,11}, \tag{61}$$

with $\rho_{jk,nm}$ denoting the matrix entries of a two-qubit state, follows from $A_1 = B_2 = |0\rangle\langle 0|$ and $A_2 = B_1^\dagger = |0\rangle\langle 1|$.

The best lower bound for separable states is achieved by using operators of the form $A_1 = |a\rangle\langle\varphi|$ and $A_2 = |\varphi\rangle\langle\alpha|$ with $|a\rangle, |\alpha\rangle$ and $|\varphi\rangle$ being arbitrary states of system A and likewise choosing the operators B_1 and B_2. The criterion is independent of the state φ and linear combinations of such operators lead to weaker bounds [46]. In the bipartite case, the entanglement criterion given in eq. (59) detects only states with NPT and it is necessary and sufficient for two-qubit states if we optimize over all measurement directions. In addition, it detects all NPT states of the form

$$\rho = p|\psi\rangle\langle\psi| + \frac{1-p}{d}\mathbb{1}_d. \tag{62}$$

The choice of the best operator given by $A_1 A_2 = |a\rangle\langle\alpha|$ with two arbitrary but different states $|a\rangle, |\alpha\rangle$ leads to a measurement of mean values of non-Hermitian operators. As a consequence, $|\langle A_1 A_2 B_1 B_2\rangle|$ cannot be measured directly. Nevertheless, the expectation value of a non-Hermitian operator O can be estimated via the two Hermitian operators

$$2\mathrm{Re}\,[\langle O\rangle] = \langle O + O^\dagger\rangle, \tag{63}$$

$$2\mathrm{Im}\,[\langle O\rangle] = \langle \mathrm{i}O^\dagger - \mathrm{i}O\rangle. \tag{64}$$

In photon experiments, the real and imaginary part can be measured *e.g.* with the help of a beam splitter (compare sect. **2**˙1.2). In experiments with trapped ions, the off-diagonal terms can be obtained by measurements in the basis

$$|\leftarrow\rangle = (|0\rangle + \mathrm{e}^{\mathrm{i}\phi}|1\rangle)/\sqrt{2}, \quad |\rightarrow\rangle = (|0\rangle + \mathrm{e}^{\mathrm{i}(\phi+\pi)}|1\rangle)\sqrt{2}. \tag{65}$$

The parity of the measurement given by the probability that the blochvectors of both ions point in the same direction

$$\Pi(\varphi) = P_{\leftarrow,\leftarrow} + P_{\rightarrow,\rightarrow} \tag{66}$$

then determines the off-diagonal elements via $2|\rho_{00,11}| = \Pi_{\max} - \Pi_{\min}$.

3˙2. *Tripartite entanglement.* – One might think first, that multipartite entanglement can be characterized by investigating the entanglement of all possible bipartite splits. However, this is only true for pure states. Mixed entangled states cannot be fully characterized by this attempt.

There exist 4 different groups of tripartite states. If A, B and C denoting the three partition, then the different groups can be characterized in the following way [47]:

1. Fully separable states

$$\rho_{\mathrm{sep}} = \sum_j p_j\ \rho_j^A \otimes \rho_j^B \otimes \rho_j^C. \tag{67}$$

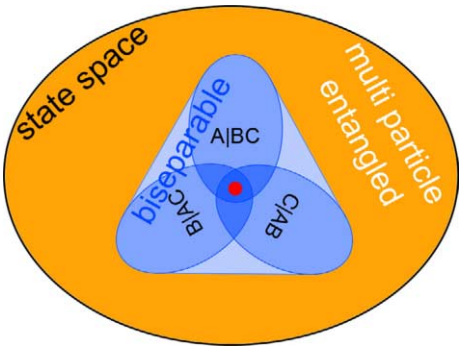


Fig. 8. – Demonstration of the characterization of entanglement in a tripartite systems. The inner red dot represents the fully separable states which is a subset of the biseparable states (inner blue triangle). States which lie outside of the triangle are genuine multipartite entangled. There exist different kinds of bipartite states: i) bipartite states with respect to every bipartition (overlapped of all three ellipses), ii) bipartite states with respect to a given bipartition (single ellipse), iii) convex combination of bipartite states (convex triangular hull) [38].

2. Biseparable states with respect to a given partition, *e.g.* $A|BC$

$$\rho_{A|BC} = \sum_j p_j \, \rho_j^A \otimes \rho_j^{BC}. \tag{68}$$

3. General bipartite states, which are convex combination of bipartite states with respect to different bipartitions

$$\rho_{\mathrm{bs}} = p_1 \, \rho_{A|BC} + p_2 \, \rho_{AB|C} + p_3 \, \rho_{AC|B}. \tag{69}$$

4. Genuine multipartite entangled states given by all states which cannot be represented by eq. (69).

So, on the one hand there exist states which are biseparable under every bipartition but not fully separable. On the other hand, there also exist states which are not separable under any given bipartition, but are still not genuine multipartite entangled. The different types of tripartite states form a nested structure consisting of convex groups as shown in fig. 8.

Entanglement in the tripartite case can be also detected with the help of inequalities of expectation values. In ref. [42], the authors suggested that their entanglement criterion can be generalized to the multipartite case by

$$\left| \left\langle \prod_{k=1}^{N} O_k \right\rangle \right|^2 \underset{\mathrm{sep}}{\leq} \left\langle \prod_{k=1}^{j} O_k^\dagger O_k \prod_{k=j+1}^{N} O_k O_k^\dagger \right\rangle, \tag{70}$$

with O_k being an operator acting on system k. Still, this generalization just divides all parties into the two subgroups $\{1,2,\ldots,j\}$ and $\{j+1,\ldots,N\}$. As a result, it does not investigate real multipartite entanglement but only bipartite entanglement with respect to a given bipartition.

A more advanced generalization is given by

$$\left|\left\langle \prod_{k=1}^{n} O_k \right\rangle\right| \underset{\text{sep}}{\leq} \prod_{k=1}^{n} \langle (O_k^\dagger O_k)^{n/2}\rangle^{1/n}, \tag{71}$$

which is not equivalent to checking a given single bipartition. Nevertheless, it cannot reveal the complex structure of tripartite entanglement because it is still based on combinations of bipartite entanglement [46]. For example, eq. (71) is not only valid for separable states, but also for states were for each pair of subsystems k and j we find a bipartition $M|\bar{M}$ such that $k \in M$ and $j \in \bar{M}$ and the state ρ is biseparable for this bipartition $M|\bar{M}$. In the tripartite case, this means that *e.g.* states which are biseparable under the partition $A|BC$ and $B|AC$ do not violate eq. (71).

A more complete characterization of tripartite entanglement can be achieved with the criteria developed in ref. [44]. Here, the entanglement of three-qubit states are investigated by developing entanglement criteria based on the matrix entries $\rho_{j,k}$ given in the standard product basis $\{|000\rangle, |001\rangle, \ldots, |111\rangle\}$. General biseparable states (bs) satisfy the inequalities

$$|\rho_{1,8}| \underset{\text{bs}}{\leq} \sqrt{\rho_{2,2}\rho_{7,7}} + \sqrt{\rho_{3,3}\rho_{6,6}} + \sqrt{\rho_{4,4}\rho_{5,5}}, \tag{72}$$

$$|\rho_{2,3}| + |\rho_{2,5}| + |\rho_{3,5}| \underset{\text{bs}}{\leq} \sqrt{\rho_{1,1}\rho_{4,4}} + \sqrt{\rho_{1,1}\rho_{6,6}} + \sqrt{\rho_{1,1}\rho_{7,7}} + \frac{1}{2}(\rho_{2,2} + \rho_{3,3} + \rho_{5,5}). \tag{73}$$

Violation of at least one of the two inequalities indicates genuine multipartite entanglement. Fully separable states obey the inequalities

$$|\rho_{1,8}| \underset{\text{sep}}{\leq} (\rho_{2,2}\rho_{3,3}\rho_{4,4}\rho_{5,5}\rho_{6,6}\rho_{7,7})^{1/6}, \tag{74}$$

$$|\rho_{1,8}| \underset{\text{sep}}{\leq} (\rho_{1,1}\rho_{4,4}^3\rho_{5,5}\rho_{6,6}\rho_{7,7})^{1/6}. \tag{75}$$

These inequalities are also able to detect weak entangled states such as states which are bisarable under every bipartition but still not fully separable [44].

Criteria for genuine multipartite entanglement can be generated by convex combinations of biseparable criteria. There exists also a general method to derive entanglement criteria for weak entangled states such as multipartite states which are PPT under every bipartition. This general method is based on the Hölder inequality, which is a generalization of the Cauchy-Schwarz inequality.

For product states, we factorize again the expectation value

$$\langle A_1A_2B_1B_1C_1C_2\rangle \underset{\text{sep}}{\leq} \sum_j p_j |\langle A_1A_2\rangle_j| \cdot |\langle B_1B_2\rangle_j| \cdot |\langle C_1C_2\rangle_j|, \tag{76}$$

where we have used that each separable state ρ can be written as a convex combination $\rho = \sum_j p_j |\psi_j\rangle_{\rm ps}\langle\psi_j|$ of product states. Consecutively, we use the Cauchy-Schwarz inequality for each single subsystem A, B or C and recombine them again. At this point, it is also possible to increase the number of expectation values by using $\langle OO^\dagger\rangle = \sqrt[k]{\langle OO^\dagger\rangle^k}$. In this way, we find *e.g.* the inequality

$$|\langle A_1A_2B_1B_2C_1C_2\rangle| \underset{\rm sep}{\leq} \sum_j p_j \sqrt[4]{\langle A_1A_1^\dagger B_1B_1^\dagger C_2^\dagger C_2\rangle_j \langle A_1A_1^\dagger B_2^\dagger B_2 C_1C_1^\dagger\rangle_j} \times \sqrt[4]{\langle A_2^\dagger A_2 B_1B_1^\dagger C_1C_1^\dagger\rangle_j \langle A_2^\dagger A_2 B_2^\dagger B_2 C_2^\dagger C_2\rangle_j}. \tag{77}$$

This inequality still depends on the representation $\rho = \sum_j p_j |\psi_j\rangle_{\rm ps}\langle\psi_j|$ which is not unique. However, with the help of the generalized Hölder inequality

$$\sum_j p_j x_j y_j \leq \Big(\sum_j p_j x_j^{1/r}\Big)^r \Big(\sum_j p_j y_j^{1/s}\Big)^s \tag{78}$$

and $\langle O\rangle = \sum_j p_j \langle O\rangle$ we finally arrive at

$$|\langle A_1A_2B_1B_2C_1C_2\rangle| \underset{\rm sep}{\leq} \sqrt[4]{\langle A_1A_1^\dagger B_1B_1^\dagger C_2^\dagger C_2\rangle_{\rm sep} \langle A_1A_1^\dagger B_2^\dagger B_2 C_1C_1^\dagger\rangle} \times \sqrt[4]{\langle A_2^\dagger A_2 B_1B_1^\dagger C_1C_1^\dagger\rangle_{\rm sep} \langle A_2^\dagger A_2 B_2^\dagger B_2 C_2^\dagger C_2\rangle}. \tag{79}$$

This inequality cannot be derived by combining bipartite entanglement criteria and can therefore also detect PPT states such as

$$\varrho_\alpha = \frac{1}{8+8\alpha} \begin{pmatrix} 4+\alpha & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & \alpha & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & \alpha & 0 & 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & \alpha & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & \alpha & 0 & 0 & 0 \\ 0 & 0 & -2 & 0 & 0 & \alpha & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & \alpha & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 4+\alpha \end{pmatrix}, \tag{80}$$

which is a PPT entangled state for $2 \leq \alpha \leq 2\sqrt{2}$ [48, 49, 46]. The entanglement of this state can be detected with the help of eq. (79) for $2 \leq \alpha < 2.4$.

The entanglement criteria eq. (71), eq. (74) and eq. (75) can be also derived with the above described scheme. However, eq. (71) and eq. (74) cannot detect the entanglement of the state ϱ_α [46].

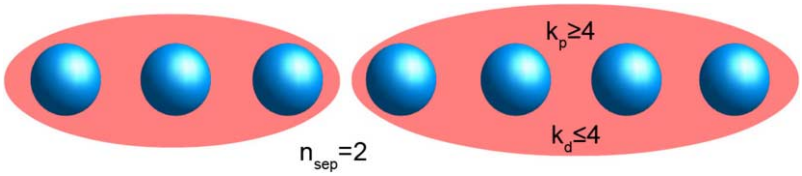


Fig. 9. – Comparison of the entanglement categories: n-separability ($n_{\rm sep}$), k-producibility (k_p) and entanglement depth (k_d).

3˙3. *Multipartite entanglement.* – The exact characterization of tripartite entanglement includes already many different kinds of entanglement as we have seen in the previous section. Therefore, the investigation of mulit-partite entanglement concentrates on a few very characteristic properties instead of pursuing for a complete characterization which would be too complicated. In general, there exist different ways of categorizing multipartite entanglement: i) n-separability and ii) k-producibility/entanglement depth k [50, 51].

The first categorization asks whether a pure state $|\psi\rangle$ can be separated into the product of n groups. As a result, we call a pure state $|\psi\rangle$ n-separable if it can be written as

$$|\psi\rangle = |\phi_1\rangle \otimes |\phi_2\rangle \otimes \cdots \otimes |\phi_n\rangle. \tag{81}$$

A mixed state is n-separable, if it is a convex combination of n-separable pure states. In general, each of the states $|\phi_j\rangle$ can consist of an arbitrary number of parties not necessarily of equal dimension. Moreover, a n-separable state is also $n-1$ separable and so on. Thus, the declaration of a state as n-separable is in general only a lower bound on the separability if not declared differently. To proof that a N-partite state is entangled, we need to proof that this state is not N-separable. For genuine multipartite entanglement, we need to show that is not biseparable.

The second categorization does not ask about the number of groups but how many parties are in each group (see fig. 9). This question can be asked in two different ways: i) the k_p-producibility asks how many entangled particles are needed to create a state and ii) the entanglement depth k defines how much entanglement can be extracted from the state. On the one hand, a k_p-producible state is also k_p+1 producible and so on. Hence, it is an upper bound of the actual entanglement. On the other hand, a state of entanglement depth k_d inhabits also a depth of k_d-1 and so on. Therefore, the entanglement depth is a lower bound. We call a state k-partite entangled if $k_p = k_d = k$.

Multipartite entanglement plays an important role in quantum metrology schemes [28]. Here, the precision $(\triangle\theta)^2$ of determining an unknown parameter θ is lower bounded by

$$(\triangle\theta)^2 \geq \frac{1}{sk^2+r}\,, \tag{82}$$

for a k-partite entangled state of a total number of $N = sk + r$ parties [52].

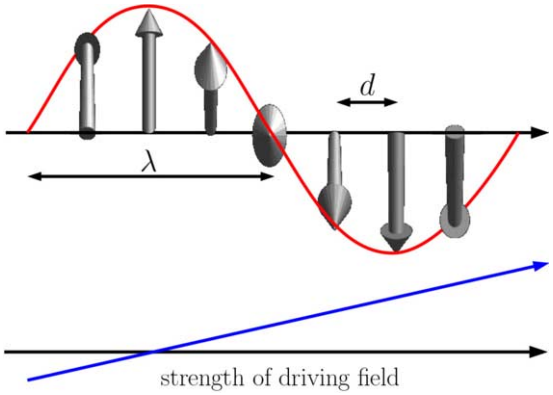


Fig. 10. – The magentic gradient field (lower picture) give rise to a position-dependent rotation of the spin (upper picture). The projection of the spin in one direction leads to a sinusoidal observable.

Multipartite entanglement can be detected *e.g.* by measuring the global spin $\vec{J} = \sum_k \vec{\sigma}_k$ [53, 54]. For k-partite entangled states, the variance $(\triangle J_z)^2$ is limited by

$$(\triangle J_z)^2 \geq J_{\max} \cdot F_{k/2}\left(\frac{\sqrt{\langle J_x^2 + J_y^2\rangle}}{J_{\max}}\right), \tag{83}$$

with the effective spin length $\sqrt{\langle J_x^2 + J_y^2\rangle}$ and the maximal spin $J_{\max} = N/2$ for spin 1/2 particles. The function $F_{k/2}$ can be determined numerically by using that the state minimizing $(\triangle J_z)^2$ is symmetric and given by $|\Psi\rangle = |\psi\rangle^{\otimes N/k}$ [53]. In this way, entanglement of more than 680 ± 35 ^{87}Rb atoms has been demonstrated experimentally [54].

3˙4. *The spatial distribution of entanglement.* – The entanglement depth k is an important variable characterizing a state as long as no spatial dependencies are involved. Though, the size of experimental available quantum systems has grown in recent times. As a consequence, the approximation that our quantum system under investigation couples only to spatially constant fields becomes weaker. Furthermore, the spatial distribution of entanglement helps also to investigate quantum phase transitions [55-59] or to distinguish different ground states of the generalized Heisenberg spin-chain [60, 61].

An example of a position-dependent field coupled to a quantum systems is given *e.g.* by a magnetic gradient field inducing a position dependent rotation of the spin as depicted in fig. 10. The projection of the spin leads to the sinusoidal observable

$$\vec{B} = \sum_j \sin\left(2\pi \frac{x_j}{\lambda}\right)\vec{\sigma}_j \tag{84}$$

with the position x_j of particle j and the wavelength λ determined by the gradient and

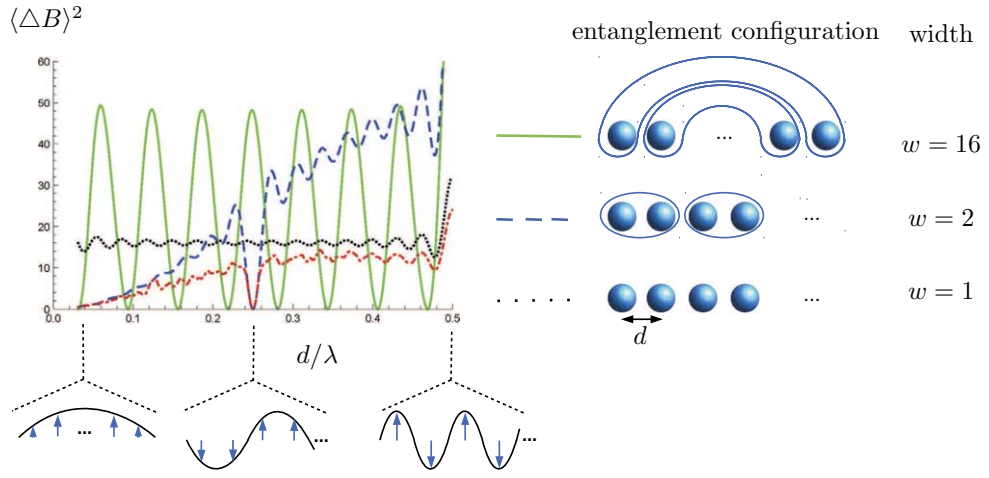


Fig. 11. – Variance of the observable $\vec{B}$ for different parameters λ and entanglement configurations for $N = 16$ particles. Pairs of encircled particles form together the state $|\psi^-\rangle$. The product state (non-encircled particles, $w = 1$) is chosen in such a way, that it minimizes the variance. Although, both entangled states possess the same entanglement depth $k = 2$, they exhibit quite different behavior due to their different spatial distribution of entanglement. The red dashed-dotted line corresponds to the $w = 2$ entanglement configuration (blue dashed line) optimized over all states with the given entanglement configuration.

the interaction time. The variance $(\triangle\vec{B})^2 = (\triangle B_x)^2 + (\triangle B_y)^2 + (\triangle B_z)^2$ can be used to determine the entanglement of the state similar to the variance of the total spin. However, $(\triangle\vec{B})^2$ not only depends on the amount of entanglement but also on its spatial distribution. Assume for example a state $|\Psi\rangle = \bigotimes_{\{j,k\}} |\psi^-\rangle_{j,k}$ consisting of pairs of particles in the state $|\psi^-\rangle_{j,k} = (|01\rangle_{j,k} - |10\rangle_{j,k})/\sqrt{2}$. The variance of this state now depends on the spatial distribution of the $|\psi^-\rangle_{j,k}$ states as shown in fig. 11. Whereas $(\triangle J)^2$ is equal to zero for all configurations, $(\triangle B)^2$ is only equal to zero if the entangled pairs $|\psi^-\rangle_{j,k}$ are situated in such a way, that the coupling strength $a_j = \sin(2\pi x_j/\lambda)$ is equal for entangled particles j and k. The variance becomes larger, the more the coupling strengths a_j and a_k differ from each other. This behavior is not only valid for the singlet state $|\psi^-\rangle$ but stays also valid if we minimize over all states. In general, we find the lower limit of the variance [62]

$$\min_{|\psi\rangle}(\Delta\vec{B})^2_{(j,k)} = \begin{cases} a_j^2\left(2 + 2\varepsilon^2 - \frac{4\varepsilon^2}{(1-\varepsilon)^2}\right), & -1 \leq \varepsilon \leq \varepsilon_0, \\ 3a_j^2(1-\varepsilon)^2, & \varepsilon_0 \leq \varepsilon \leq 1, \end{cases} \tag{85}$$

where we assumed with out loss of generality $|a_j| > |a_k|$ and defined $\varepsilon = a_k/a_j$ and $\varepsilon_0 = 2 - \sqrt{3} \approx 0.27$. Here, we want to note that the state minimizing $(\Delta\vec{B})^2_{(j,k)}$ is given by the singlet state $|\psi^-\rangle$ for $\varepsilon > 2 - \sqrt{3}$. In this way, the minimal variance for states,

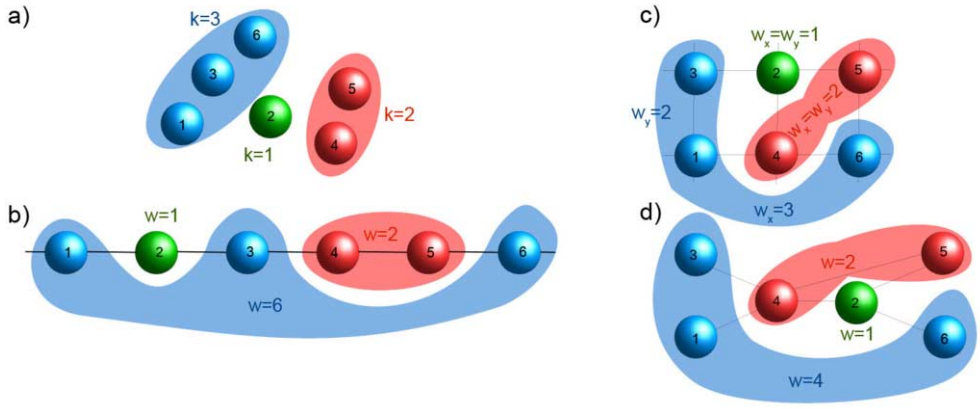


Fig. 12. – Comparison of entanglement depth (a) and entanglement width in 1D (b) and 2D (c) or a graph (d) of the state $|\Psi\rangle = |\psi_{1,3,6}\rangle \otimes |\psi_{4,5}\rangle \otimes |\psi_2\rangle$: whereas the entanglement depth disregards any spatial ordering, the definition of entanglement width requires the particles to be spatially ordered, *e.g.* in a spin chain or a grid. The entanglement depth of the state $|\Psi\rangle$ in (a) is given by $k = 3$ (since maximally three particles are entangled). This is a lower bound on the entanglement width in (b), which equals $w = 6$ (since entanglement occurs over a distance of six particles in the chain). In the case of the 2d grid in (c) we distinguish between the width in the horizontal direction ($w_x = 3$) and vertical direction ($w_y = 2$). In a general graph (d) where edges denote possible interaction between two particles there might exist different ways to connect two particles. In this case, we take always the shortest way to determine the entanglement width [62].

where only nearest-neighbor entanglement between particles j and $j + 1$ for odd j is allowed, can be determined. As can be seen in fig. 11 this bound (red dashed-dotted line) is valid for the product states and the nearest-neighbor configuration of $|\psi^-\rangle$ but is violated by other spatial configurations.

Similar to multipartite entanglement, the exact investigation of the spatial distribution of entanglement is very resource intensive or even impossible for many-particle states. Hence, we define instead the width of entanglement (see fig. 12) as a characteristic measure of the spatial distribution of entanglement similar to the entanglement depth for multipartite entanglement.

To define the width of entanglement w we need to define a spatial ordering given *e.g.* by a linear chain of trapped ions (fig. 12 b)), a lattice of cold atoms (fig. 12 c)) or a graph with given interactions (fig. 12 d)). The distance between two particles is then determined by following the given structure and counting particles in between. As a consequence, the width of entanglement w of a pure state $|\Psi\rangle = \bigotimes_j |\psi_j\rangle$ is defined as the maximal distance w of two entangled particles within the states $|\psi_j\rangle$. A completely separable state exhibits an entanglement width of $w = 1$. The entanglement width of a mixed state is defined by the minimum with w over all decomposition $\varrho = \sum_k p_k |\psi_k\rangle\langle\psi_k|$,

that is

$$w(\varrho) = \min_{\text{decompositions}} \Big[\max_k \{ w(\psi_k) \} \Big]. \tag{86}$$

By definition, the entanglement depth is a lower bound of the entanglement width. However, the entanglement width does not make any statement about the entanglement depth.

To use the variance $(\triangle B)^2$ to detect the width of entanglement, we need to not only optimize over the state but to also find the optimal pairing. For example, for the width $w = 2$ ion j might be entangled with $j-1$ or $j+1$ or not entangled at all. Still, these is a classical optimization problem which for some configurations is easy to solve or can be at least approximated [62]. For example, for N particles situated at $x_j = x_0 + j \cdot d$ with $d = \lambda/(2N)$ and $x_0 = d/2$ the optimal pairing for only nearest-neighbor entanglement is given by entangling all odd particles j with their right neighbor $j+1$. As a consequence, the variance $(\triangle B)^2$ for states with entanglement width $w \geq 2$ are bounded from below by

$$(\Delta \vec{B})^2 \geq \frac{3}{2} N \left[1 - \cos\left(\frac{\pi}{N} \right) \right] \approx \frac{3\pi^2}{4N}. \tag{87}$$

In this way, the width of entanglement can be determined without addressing of single subsystems, solely with global measurements.

4. – Conclusion

In summary, we have discussed in this lecture note different quantum properties such as the wave-particle duality and entanglement. However, both phenomena are based like many other quantum properties on non-commuting observables. As a result, measurement procedures to observe wave-particle duality or entanglement are based on similar principles such as the Cauchy-Schwarz inequality. The only difference is that to observe wave-particle duality or the Heisenberg uncertainty, we apply these principles on the whole system whereas we apply them to single subsystems to observe entanglement.

* * *

I thank M.S. Zubairy, O. Gühne and W.P. Schleich for fruitful discussions on these topics and D. Kiesler, S. Altenburg, T. Kraft, H. Siebeneich, P. Huber and J. Hoffmann for their careful reading of the manuscript.

Appendix A.

Proof of eq. (49)

A separable state can be decomposed into $\rho = \sum_n p_n |\alpha_n\rangle\langle\alpha_n| \otimes |\beta_n\rangle\langle\beta_n|$ with $\sum_n p_n = 1$ and $|\alpha_n\rangle$,$|\beta_n\rangle$ being arbitrary normalized states of system A and B, respectively. The

expectation values of the joined observables $M_{j,k}$ factorize for the product states $|\alpha_n\rangle|\beta_n\rangle$. Therefore, the sum of the observables can be rewritten by

$$\sum_{j,k=1}^{3} \langle M_{j,k}\rangle_{\text{sep}} = \sum_n p_n \langle\alpha_n|(\sigma_x+\sigma_y+\sigma_z)|\alpha_n\rangle\langle\beta_n|(\sigma_x+\sigma_y+\sigma_z)|\beta_n\rangle \tag{A.1}$$

for separable states. Since $\langle\sigma_x+\sigma_y+\sigma_z\rangle \leq 1$ and $\sum_n p_n = 1$ we find

$$\sum_{j,k=1}^{3} \langle M_{j,k}\rangle_{\text{sep}} \leq 1. \tag{A.2}$$

REFERENCES

[1] Bohr N., *Nature*, **19** (1928) 473.
[2] Horodecki R., Horodecki P., Horodecki M. and Horodecki K., *Rev. Mod. Phys.*, **81** (2009) 865.
[3] Kochen S. and Specker E. P., *J. Math. Mech.*, **17** (1967) 59.
[4] Bell J. S., *Rev. Mod. Phys.*, **38** (1966) 447.
[5] Einstein A., Podolsky B. and Rosen N., *Phys. Rev.*, **47** (1935) 777.
[6] Bell J., *Physics*, **1** (1964) 195.
[7] Clauser J. F., Horne M. A., Shimony A. and Holt R. A., *Phys. Rev. Lett.*, **23** (1969) 880.
[8] Wooters W. K. and Zurek W. H., *Phys. Rev. D*, **19** (1979) 473.
[9] Greenberger D. M. and Allaine Y., *Phys. Rev. Lett.*, **8** (1988) 391.
[10] Englert B.-G., *Phys. Rev. Lett.*, **77** (1996) 2154.
[11] Schleich W. P., *Wave-particle dualism in action*, in *Optics in Our Time*, edited by Al-Amri M. D., El Gomati M. M. and Zubairy M. S. (Springer, Cham) 2016.
[12] Ollivier H. and Zurek W. H., *Phys. Rev. Lett.*, **88** (2002) 017901.
[13] Scully R. J., *The Demon and the Quantum* (Wiley-VCH, Weinheim, Germany) 2007.
[14] Scully M. O. and Drühl K., *Phys. Rev. A*, **25** (1981) 2208.
[15] Kim Y.-H., Yu R., Kulik S. P., Shih Y. and Scully M. O., *Phys. Rev. Lett.*, **84** (2000) 1.
[16] Menzel R., Puhlmann D., Heuer A. and Schleich W. P., *Proc. Natl. Acad. Sci. U.S.A.*, **109** (2012) 9314.
[17] Menzel R., Heurer A., Puhlmann D., Dechoum K., Hillery M., Spähn M. J. A. and Schleich W. P., *J. Mod. Opt.*, **60** (2013) 86.
[18] Leach J., Bolduc E., Miatto F. M., Piché K., Leuchs G. and Boyd R. W., *Sci. Rep.*, **6** (2016) 19944.
[19] Allevi A., Olivares S. and Bondani M., *Phys. Rev. A*, **85** (2012) 063835.
[20] Hemmer P. R., Muthukrishnan A., Scully M. O. and Zubairy M. S., *Phys. Rev. Lett.*, **96** (2006) 163603.
[21] Huang J.-H., Wölk S., Zhu S.-Y. and Zubairy M. S., *Phys. Rev. A*, **87** (2013) 022107.
[22] Freedman S. J. and Clauser J. F., *Phys. Rev. Lett.*, **28** (1972) 938.
[23] Clauser J. F., *Phys. Rev. Lett.*, **36** (1976) 1223.
[24] Fry E. S. and Thompson R. C., *Phys. Rev. Lett.*, **37** (1976) 465.

[25] Giustina M. *et al.*, *Phys. Rev. Lett.*, **115** (2015) 250401.
[26] Shalm L. K. *et al.*, *Phys. Rev. Lett.*, **115** (2015) 250402.
[27] Monroe C., *Nature*, **416** (2002) 238.
[28] Tóth G. and Appelaniz I., *J. Phys. A.: Math. Theor.*, **47** (2014) 424006.
[29] Gühne O. and Tóth G., *Phys. Rep.*, **474** (2009) 1.
[30] Greenberger D. M., Horne M. A., Shimony A. and Zeilinger A., *Am. J. Phys.*, **58** (1990) 1131.
[31] Altenburg S., Wölk S., Tóth G. and Gühne O., *Phys. Rev. A*, **94** (2016) 052306.
[32] Ekert A. and Knight P., *Am. J. Phys.*, **63** (1995) 415.
[33] Laskowski W., Schwemmer C., Richart D., Knips L., Paterek T. and Weinfurter H., *Phys. Rev. Lett.*, **108** (2012) 240501.
[34] Peres A., *Phys. Rev. Lett.*, **77** (1996) 1413.
[35] Horodecki P., Horodecki R. and Horodecki M., *Phys. Lett. A*, **223** (1996) 1.
[36] Horodecki M., Horodecki P. and Horodecki R., *Phys. Rev. Lett.*, **80** (1998) 5239.
[37] Horodecki P., Horodecki M. and Horodecki R., *Phys. Rev. Lett.*, **82** (1999) 1056.
[38] Hofmann M., Moroder T. and Gühne O., *J. Phys. A: Math. Theor.*, **47** (2014) 155301.
[39] Moroder T., Kleinmann M., Schindler P., Monz T., Gühne O. and Blatt R., *Phys. Rev. Lett.*, **110** (2013) 180401.
[40] Schwemmer C., Knips L., Richart D., Moroder T., Kleinmann M., Gühne O. and Weinfurter H., *Phys. Rev. Lett.*, **114** (2015) 080403.
[41] Shchukin E. and Vogel W., *Phys. Rev. Lett.*, **95** (2005) 230502.
[42] Hillery M. and Zubairy M. S., *Phys. Rev. Lett.*, **96** (2006) 050503.
[43] Hillery M., Dung H. T. and Zheng H., *Phys. Rev. A*, **81** (2010) 062322.
[44] Gühne O. and Seevinck M., *New J. Phys.*, **12** (2010) 053002.
[45] Huber M., Mintert F., Gabriel A. and Hiesmayr B. C., *Phys. Rev. Lett.*, **104** (2010) 210501.
[46] Wölk S., Huber M. and Gühne O., *Phys. Rev. A*, **90** (2014) 022315.
[47] Dür W. and Cirac J. I., *Phys. Rev. A*, **61** (2000) 042314.
[48] Kay A., *Phys. Rev. A*, **83** (2011) 020303.
[49] Gühne O., *Phys. Lett. A*, **375** (2011) 406.
[50] Sørensen A. S. and Mølmer K., *Phys. Rev. Lett.*, **86** (2001) 4431.
[51] Gühne O., Tóth G. and Briegel H. J., *New J. Phys.*, **7** (2005) 229.
[52] Hyllus P., Laskowski W., Krischeck R., Schwemmer C., Wieczorek W., Weinfurter H., Pezzé L. and Smerzi A., *Phys. Rev. A*, **85** (2012) 022321.
[53] Lücke B., Peise J., Vitagliano G., Arlt J., Santos L., Tóth G. and Klempt C., *Phys. Rev. Lett.*, **112** (2014) 155304.
[54] Hosten O., Engelsen N. J., Krishnakumar R. and Kasevich M. A., *Nature*, **529** (2016) 505.
[55] Osterloh A., Amico L., Falci G. and Fazio R., *Nature*, **416** (2002) 608.
[56] Osborne T. J. and Nielsen M. A., *Phys. Rev. A*, **66** (2002) 032110.
[57] Gu S.-J., Li H., Li Y.-Q. and Lin H.-Q., *Phys. Rev. A*, **70** (2004) 052302.
[58] Biswas A., Prabhu R., Sen(De) A. and Sen U., *Phys. Rev. A*, **90** (2014) 032301.
[59] Richerme P., Gong Z. X., Lee A., Senko C., Smith J., Foss-Feig M., Michalakis S., Gorshkov A. and Monroe C., *Nature*, **511** (2014) 198.
[60] Eckert K., Romero-Isart O., Rodriguez M., Lewenstein M., Polzik E. S. and Sanpera A., *Nat. Phys.*, **4** (2008) 50.
[61] De Chiara G., Romero-Isart O. and Sanpera A., *Phys. Rev. A*, **83** (2011) 021604 (R).
[62] Wölk S. and Gühne O., arXiv: 1507.07226.

Proceedings of the International School of Physics "Enrico Fermi"
Course 197 "Foundations of Quantum Theory", edited by E. M. Rasel, W. P. Schleich and S. Wölk
(IOS, Amsterdam; SIF, Bologna) 2019
DOI 10.3254/978-1-61499-937-9-267

Complementarity and light modes

R. Menzel

University of Potsdam, Institute for Physics and Astronomy
Karl-Liebknecht-Str. 24/25, 14476 Potsdam, Germany

Summary. — In quantum physics we are confronted with new entities which consist indivisible of an energy packet and a coupled wave. The complementarity principle for certain properties of these quantum objects may be their main mystery. Photons are especially useful to investigate these complementary properties. A series of new experiments using spontaneous parametric down-conversion (SPDC) as a tool allowed the detailed analysis of the physical background of this complementarity and offers a new conceptual perspective. Based on these results a straightforward explanation of these sometimes counterintuitive effects is given.

1. – Introduction

Complementarity is one of the most fundamental and important principles of quantum physics. It was mentioned the first time by Niels Bohr in 1927 (at the Fermi conference in Como) as a consequence of the uncertainty of certain pairs of quantum physical parameters and published one year later [1]. The most prominent pair of such parameters are position and momentum. Other pairs, as *e.g.* electric and magnetic field strengths of photons, can be derived from them. But complementarity is also a consequence of the wave-particle-duality and as such very widely investigated in quantum optics. The most prominent example of this duality appears in the double-slit experiment with single photons, where we obtain interference structures in the light pattern behind the double

slit if no which-slit information is available and no interference (but diffraction) if the knowledge about which slit the photon has passed is available [2].

In his famous lectures on physics Richard Feynman wrote in 1964 that the wave-particle dual behavior contains the basic mystery of quantum mechanics [3]. Later he even pronounced it more by saying it may be even the only mystery of quantum mechanics. In 1984 Marlan Scully, Berthold Englert and Herbert Walther stated that: Complementarity is deeper and more general and fundamental to quantum mechanics than uncertainty [4].

Although the present theoretical concepts of quantum physics provide a complete set of recipes to calculate all results which have so far been observed with quantum systems these experimental results are sometimes quite counterintuitive. Therefore it seems to be worthwhile to investigate the physical background of complementarity in further detail to reach a better conceptual understanding. As will turn out at the end of this manuscript it seems possible to identify a more or less hidden physical entity as responsible for these astonishing effects. Its action is completely included in our theoretical concepts in a very optimal way. But it is only indirectly observable: the quantum vacuum [5].

For the investigation of complementarity and for the detailed analysis of the wave-particle-duality photons as quantum objects are especially useful. In our everyday life at room temperature only about 1% of the light modes are occupied by photons, 99% are empty. So far the energy packets of photons can easily be separated with their related modes in contrast to the spectral range of radio waves. On the other side in the high-energy region the wave properties are very difficult to observe because of technical difficulties in this very-short-wavelength range.

As a result in the measuring process the energy packet of the photon is measured as a click. But the photon modes show interference and diffraction. Clicks are not observed at spots of distractive interference. On the other side, even single photons can be detected in spatial regions which would never be illuminated without the diffraction, *e.g.* in the dark region behind the knife edge. Therefore it can be concluded that the single photon (as any other quantum object) consists of an energy packet and an associated wave which are not separable. This duality has far-reaching consequences in the measuring process.

In general light or single photons as the constitutents of light are distributed **spectrally, temporarily, spatially and in their polarization**. But usually the detection system of light is only sensitive in certain ranges of these distributions. As a result in the measurement process the single photons are detected only within the chosen distribution of frequency, time, space and polarization. These distributions define the modes of the detection system which are known from classical optics.

Only photons within the detection mode can be registered as clicks.

As a result in each physical situation only a certain selection of the photons and thus only a certain selection of the reality of that quantum system is registered. As will be shown below this selection determines the properties of the observed reality. In other words:

It is the measuring process which selects the properties of the quantum object in a certain physical situation.

Of course photons cannot be detected if they are never generated. Thus the result of the measurement is the convolution of the modes of generation with the modes of detection. (In many cases the distributions of the generation process are much wider than the distributions of the detection process and therefore in these cases the detection modes play the dominant role).

Regarding complementarity two fundamental aspects can be extracted from the non-separable energy packet and wave nature of single photons (as for all of other quantum objects):

1) The wave nature of the photon results in a number of uncertainty relations as *e.g.* for space and momentum. The spatial modes of single photons as used in quantum field theory are solutions of the classical Maxwell or Helmholtz equations. These modes exist in free space either occupied or not occupied by photon energy packets. The classical light mode with the smallest product of beam waist diameter and divergence is the TEM_{00}-Gauss-mode. The spatial uncertainty within the beam waist and the momentum uncertainty within the divergence fulfill exactly the Heisenberg uncertainty relation between space and momentum [6], which is no surprise because the underlying quantum theory is a wave description.

The uncertainty relations describe the complementarity between the involved parameters of the considered quantum object (as *e.g.* the photon).

2) The photon modes in consideration may show certain coherence properties. If we assume for simplicity the mode function system of Gauss-Laguerre and Gauss-Hermite modes the longitudinal and the transversal coherence lengths of these modes can be defined [7]. Therefore the question arises how coherent are single photons in these modes and how distinguishable are they. The coherence is usually characterized by the visibility V of the fringes of an interference experiment which is calculated from $V = (C_{max} - C_{min})/(C_{max} + C_{min})$. C_{max} and C_{min} are the maximum and the minimum photon count rates of the fringes. The distinguishability can be measured in coincidence with the reference photon as a marker for one of the two paths the photon can take. The distinguishability D follows from $D = (R_{path1} - R_{path2})/(R_{path1} + R_{path2})$. In this formula R_{path1} and R_{path2} are the coincidence count rates measured for the observed photon in path 1 and in path 2 in coincidence with a reference photon marking path 1. As an example think of the double-slit experiment, there path 1 relates to slit 1 and path 2 relates to slit 2. From a fundamental quantum optical calculation there follows that $D^2 + V^2$ equals in maximum 1 [8] and [9]. This relation is also true for physical situations of uncertainties larger than Heisenberg's uncertainty relation would demand, as will be shown below.

The relation between visibility and distinguishability is the second fundamental aspect of the complementarity principle for quantum objects.

Both aspects of complementarity are a consequence of the modes of the photons. Thus it is the wave nature of the photons and not the particle-like aspect which is responsible

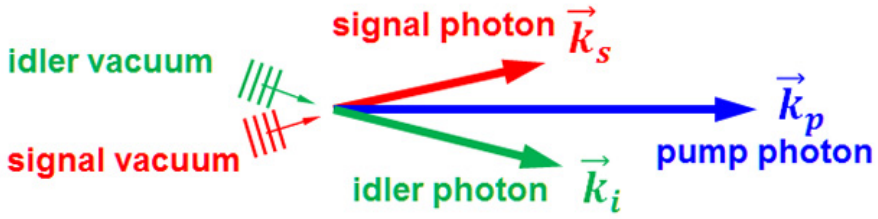


Fig. 1. – Scheme of the emitted light in SPDC: The entangled signal (red) and idler (green) photons occur at opposite sides of the cone.

for complementarity. The not separable energy packets of the single photons are only the reason for occupied (excited) or not occupied (not excited) modes.

To investigate the physical background of these aspects of the complementarity principle a set of new experiments was performed to show in detail the properties of single photons in certain modes.

2. – Spontaneous parametric down-conversion (SPDC) as a tool

In spontaneous parametric down-conversion an optically transparent crystal with a second-order nonlinearity ($\chi^{(2)}$) is used to generate two entangled photons from one pump photon. In this process energy and momentum is conserved. By choosing a suitable orientation of the crystal phase matching between the wave of the pump photon and the two waves of the newly produced photons can be realized and thus the efficiency of the process enlarged significantly [10, 6] and [11]. The two new photons are usually named signal and idler photons. In type-I phase matching which is realized by a special orientation of the crystal the signal and the idler photons belong to the same light cone structure which is emitted cylindrically symmetric around the pump photon direction as illustrated in fig. 1.

Because of momentum conservation the two photons appear on opposite sides of the cone. As a result of the entanglement of the two photons including their temporal synchronization one of the photons (*e.g.* the idler photon) can be used as a reference for the other photon (then the signal photon). This allows in addition to usual single-photon measurements coincidence measurements from the simultaneous observation of two clicks. Because of the entanglement (and correlation), the properties of the signal photon can be determined from measuring the correlated properties of the idler photon. As will be shown below this allows new measurements regarding the investigation of the complementarity principle.

But even more important information is available from using spontaneous parametric down-conversion for the investigation of complementarity. In the quantum description of the generation of the new photons the 3-wave mixing of the pump mode with the vacuum modes is applied.

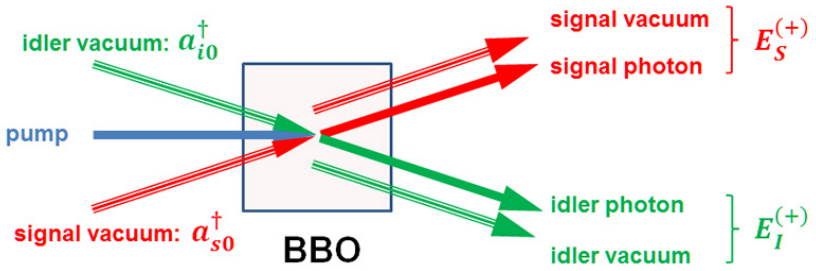


Fig. 2. – Scheme of 3-wave-mixing process with the vacuum field contribution.

The electric field components of the signal and the idler field are given by the following equations:

$$E^{(+)}_{signal}(\mathbf{r},t) = a_{s0}(\mathbf{r},t) + Ca_{pump}(\mathbf{r},t)a^{\dagger}_{i0}(\mathbf{r},t), \tag{1}$$

$$E^{(+)}_{idler}(\mathbf{r},t) = a_{i0}(\mathbf{r},t) + Ca_{pump}(\mathbf{r},t)a^{\dagger}_{s0}(\mathbf{r},t). \tag{2}$$

With this simplified model as used in [12] and [13] an effective Hamiltonian is applied. The couplings are described by $a_{pump}a^{\dagger}_{i0}$ and $a_{pump}a^{\dagger}_{s0}$, where a_{pump}, a_s and a_i ($a^{\dagger}_{pump}$, $a^{\dagger}_s$ and $a^{\dagger}_i$) are photon annihilation (creation) operators for the pump, signal, and idler fields. This way the annihilation (creation) of the pump photon and the simultaneous creation (annihilation) of signal and idler photons in the SPDC 3-wave mixing process with the vacuum field contributions a_{s0} as signal vacuum and a_{i0} as idler vacuum is considered. The electric fields leaving the crystal are the sum of the created photon fields and the undisturbed vacuum fields crossing the crystal. This description of the SPDC process is illustrated in fig. 2.

The important observation here is that in this spontaneous parametric down-conversion the vacuum fields are directly involved in the generation process of the two new photons. This allowed the investigation of the physical background of complementarity with new experiments using SPDC as a tool. By measuring a single signal photon or single idler photon these vacuum fields will not play any role. But if the two photons are measured in coincidence these vacuum fields will mix in the formulas of the measured intensity as Glauber described in his photodetection theory of the measuring process [14] and thus entanglement is provided.

In summary, spontaneous parametric down-conversion (SPDC) allows a detailed investigation of the influence of the vacuum fields in the generation process of single photons and in the entanglement process. This will be, as shown below, a very powerful tool to investigate complementarity.

3. – Induced coherence in the 3-crystal set up

As published in [13] the addition of the third crystal in the known induced coherence setups [15] and with higher visibility described in [16] allows a very direct observation of the randomness of the vacuum fields. The experimental setup is shown in fig. 3.

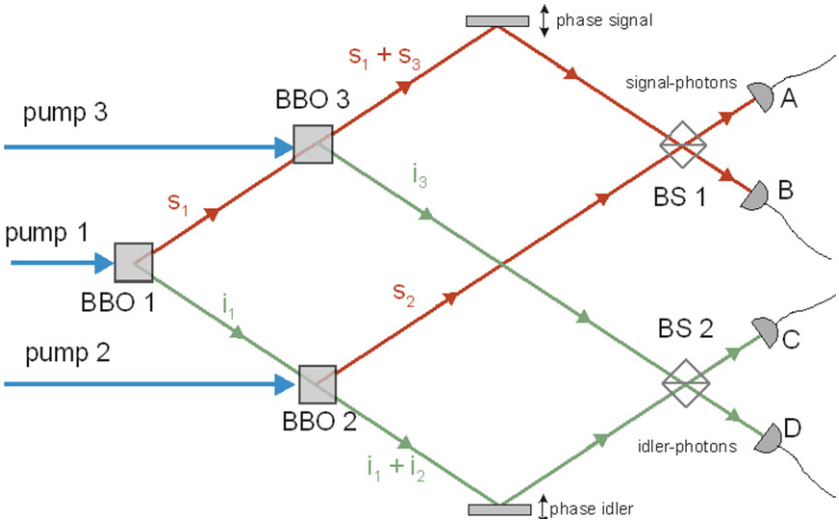


Fig. 3. – Scheme of the induced coherence setup with 3 crystals as published and described in detail in [13].

In this set up the crystals BBO 1 and BBO 2 are aligned for the same idler mode and crystal BBO 1 and BBO 3 are aligned for the same signal mode. Therefore the two signal photons s1 and s2 are perfectly coherent as well as the idler photons i1 and i3 if the whole setup is aligned well. As a result at detector A perfect fringes with visibilities of almost 1 could be measured if crystal BBO 3 was not pumped and in the same way with detector D visibilities of almost 1 could be detected if crystal BBO 2 was not pumped. But if all three crystals are pumped, simultaneously, both interference diagrams show an incoherent background as is shown in fig. 4.

As was shown in ref. [13] this result can be explained using the set of equations as given in eqs. (1) and (2) from above. The result is illustrated in fig. 5. As can be seen from this figure and also shown in eq. (3) the fields of the signal photon 1 and of the signal photon 2 are coherent because in the generation process of these two photons the pump is a coherent state and therefore the fields of pump 1 and pump 2 have a fixed phase relation. The idler vacuum responsible for the 3-wave mixing in the generation process of the two photons is the same. Therefore the 3-wave mixing process in the two crystals BBO1 and BBO2 is realized with the same phase relations and thus the two signal photons are generated with the same phase as can be seen in the following equation (of course the path delays between the two crystals have to be recognized):

$$E^{(+)}_{S-BS}(\mathbf{r},t) = a_{s10} + ia_{s20}e^{i\phi_{s2}} + C_1 a_{pump1} a^{\dagger}_{i0} + iC_2 a_{pump2} a^{\dagger}_{i0} e^{i\phi_{s2}}. \tag{3}$$

The situation for the generation of the two idler photons, idler photon i1 and idler

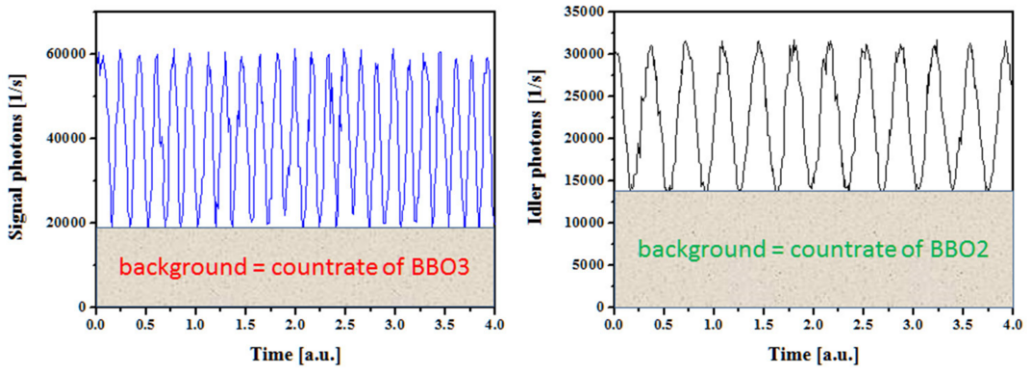


Fig. 4. – Experimental results of the interference fringes in the 3-crystal–induced coherence measurement of fig. 3 as published and described in [13] in detail. On left the count rate of the signal photons measured with detector A and on right the count rate of the idler photons measured with detector D are shown. Although the signal photons of crystal BBO1 and BBO3 are perfectly coherent all photons of crystal BBO2 are not coherent and produce the background signal in this figure on the left. The same argument is valid for the idler photons at detector D.

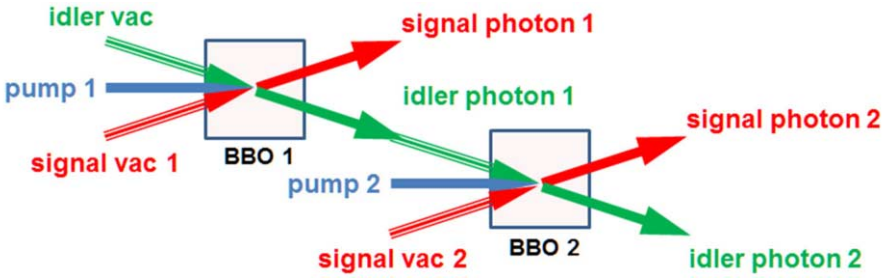


Fig. 5. – Scheme of the induced coherence setup with the vacuum field contribution.

photon i2, is completely different as illustrated in fig. 5 and in the following equation:

$$E^{(+)}_{I-BS}(\mathbf{r},t) = a_{i10} + ia_{i20}e^{i\phi_{i2}} + C_1 a_{pump1}a^{\dagger}_{s10} + iC_2 a_{pump2}a^{\dagger}_{s20}e^{i\phi_{i2}}. \tag{4}$$

Although the two idler photons from the two crystals BBO 1 and BBO 2 are excitations of the same spatial vacuum mode as already shown in fig. 5 the generation process of the two idler photons is a 3-wave mixing of the pump, which is again coherent, but of two clearly distinguishable signal vacuum fields signal vac1 and signal vac2. These two vacuum fields are of course not at all coherent (which is also demonstrated by the result of the experiment, see fig. 4, right). As result the idler photon i1 and the idler photon i2 have a random phase relation. Thus in the induced coherence experiment with three crystals as shown in fig. 3 for the idler photons detected with detector D a coherent signal can be observed for the photons produced in crystal BBO 1 or BBO 3. But all the idler photons generated in crystal BBO 2 are not coherent neither to the photons from BBO

1 nor to the photons from BBO 3. As result all the idler photons generated in BBO 2 appear as an incoherent background in the interference measurement at detector D as observed in fig. 4, right.

The same argumentation is valid for the measurement at detector A. All the signal photons generated in crystal BBO 3 are not coherent to the signal photons generated in crystal BBO 1 or in crystal BBO 2. As result all the signal photons from crystal BBO 2 appear as incoherent background in the interference measurement at detector A.

Regarding the question of complementarity from these experimental results it can be summarized:

Photons generated by the same vacuum field will be coherent if all other participating fields are also coherent. But if the photons are generated by different vacuum fields they are not coherent.

This is also true even if the two photons belong to the same spatial TEM_{00} mode as it was applied in the experiments above for both the idler channel i1 and i2 as well as the signal channel s1 and s3. This means it is possible to generate single photons in the same spatial mode but with different phases. These photons are, as in the experiment above, also distinguishable. In the described measurement they could easily be distinguished by measuring the related signal photon s1 and s2 as reference for the idler photons i1 and i2 or the idler photons i1 and i3 for the signal photons s1 and s3 in coincidence.

As a summary of these experimental results it can be concluded that:

The randomness of the vacuum fields causes temporal complementarity.

4. – Stimulated coherence

The random phase of a vacuum field mode can be overwritten by the coherent mode of a laser [17] as described in detail in [18]. In our experiment the radiation of a HeNe-laser was used as stimulating light field in the 3-wave mixing process inside the SPDC-crystal in the same TEM_{00} mode as was used in the idler channels of the two separated crystals BBO1 and BBO3 in the induced coherence experiment above (see fig. 3). The experimental scheme is depicted in fig. 6.

As can be seen from the scheme of fig. 6 the two crystals BBO1 and BBO3, which are positioned in the same way as in fig. 3, are coupled via the signal channel s1 and s3. As in the experiment of fig. 3 these two signal channels would be not coherent without the HeNe-laser radiation as a consequence of the random phase of the two distinguishable vacuum fields in the idler channels i1 and i3. But by overwriting these two idler vacuum channels with the coherent light field of the HeNe-laser the phases of the two signal channels are fixed and an interference with a high visibility of 0.95 was observed at detector A if one of the pump laser radiation is delayed as shown in fig. 7 [18].

In this case the remaining spontaneous emission triggered by the vacuum fields appears as an incoherent background in this measurement. But their count rate is so low that it is almost not visible in the single-photon interference pattern.

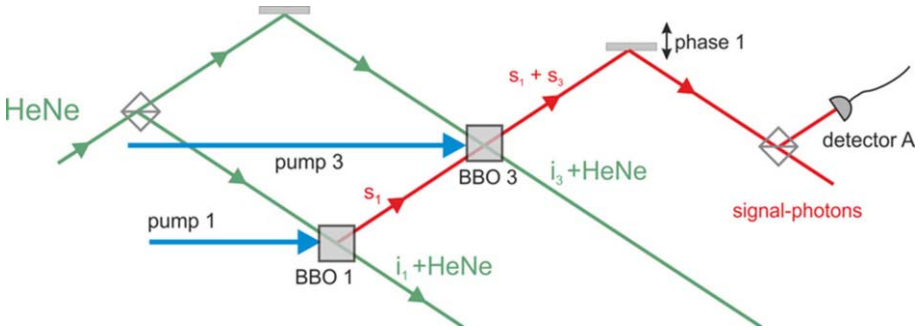


Fig. 6. – Scheme of the setup with two sequential BBO-crystals for measuring stimulated coherence overwriting the random vacuum field as described in detail and published in [18].

This effect can also be demonstrated by using two crystals in a parallel arrangement as is also described in detail in [18]. The two SPDC-crystals BBO2 and BBO3 are pumped coherently from the same pump laser source as in all other experiments before and the HeNe-laser radiation is applied in the idler channels of the two crystals. The experimental scheme is depicted in fig. 8.

If a delay line is inserted in the beam of signal 3 it is possible to observe single-photon interference between the two signal channels s2 and s3 at detector A. The result of this measurement is given in fig. 9.

As can be seen from this fig. 9 the single-photon count rate for the signal photons provided from crystals BBO2 or BBO3 shows clear fringes as a function of the delay of the signal beam with high visibility of 98%. Without the HeNe-laser radiation no

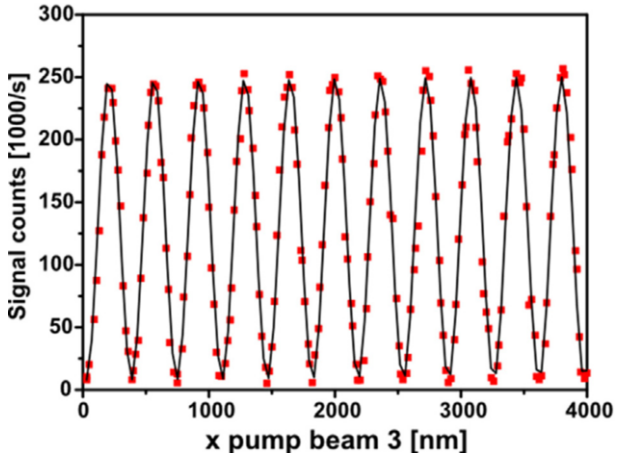


Fig. 7. – Interference of single signal photons emitted from crystals BBO1 or BBO3 of fig. 6 while changing the delay between pump 1 and pump 3 resulting in a visibility of 95% as published in [18].

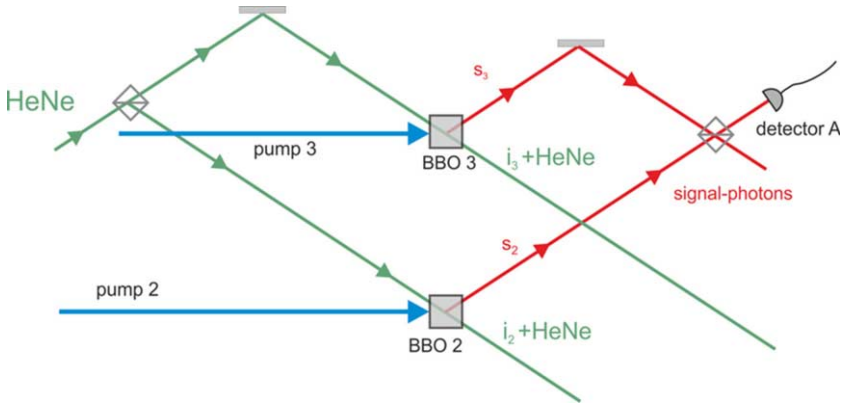


Fig. 8. – Scheme of the setup with two parallel pumped BBO-crystals for measuring stimulated coherence by overwriting the random vacuum field as described in detail and published in [18].

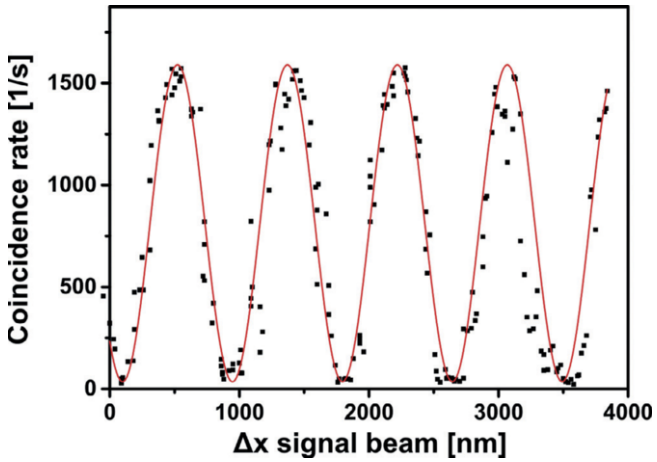


Fig. 9. – Interference measurement between the single signal photons from crystals BBO2 or BBO3 of fig. 8 while changing the delay between signal 2 and signal 3 resulting in a visibility of 98% as published in [18].

single photon interference could be observed at all. Again the remaining spontaneous emission triggered by the random vacuum fields appears as an incoherent background in this measurement, but with very low count rate that it is almost not visible in this measurement.

Summarizing this part of the discussion it can be stated:

The random vacuum fields can be overwritten by coherent light fields, *e.g.* from laser sources.

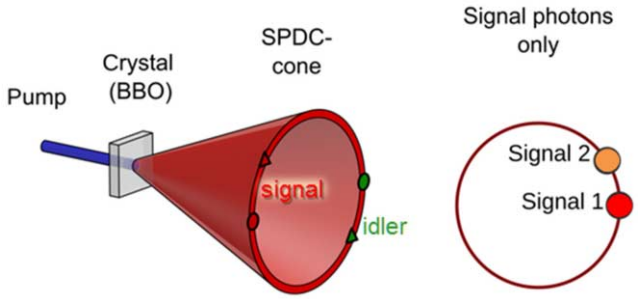


Fig. 10. – Scheme of the spatial emission of spontaneous parametric down-conversion (type I) of a BBO crystal and the position of the two TEM_{00} signal photon modes signal 1 and signal 2 which are selected by the detection system and superimposed for investigating the lateral coherence between them as published in [19].

5. – Complementarity in the spatial dimension

After discussing the complementarity for single photons in the temporal dimension the next step is their investigation in the spatial dimension as published in [19]. While above the temporal coherence of single photons and their distinguishability were analyzed now the coherence of single photons which are potentially belonging to two or more spatial modes will be discussed. Therefore single-photon interference was measured for photons in transversally distributed modes. These modes are superimposed and the resulting visibility and distinguishability of these modes is measured.

For this purpose an experiment was set up in which spontaneous parametric down-conversion (type I) was used as a tool again. The scheme is depicted in fig. 10.

In this case a single light cone is emitted behind the crystal with a cylindrical symmetry around the pump beam direction. The single idler and signal photons appear on

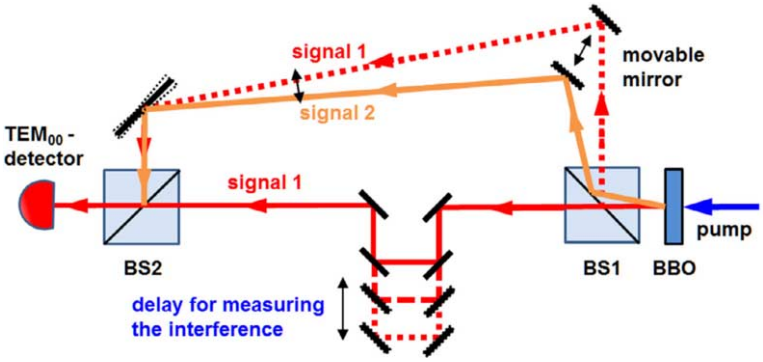


Fig. 11. – Scheme of the Mach-Zehnder-interferometer for measuring the interference of different positions of the selected TEM_{00} modes of signal 1 and signal 2 within the cone.

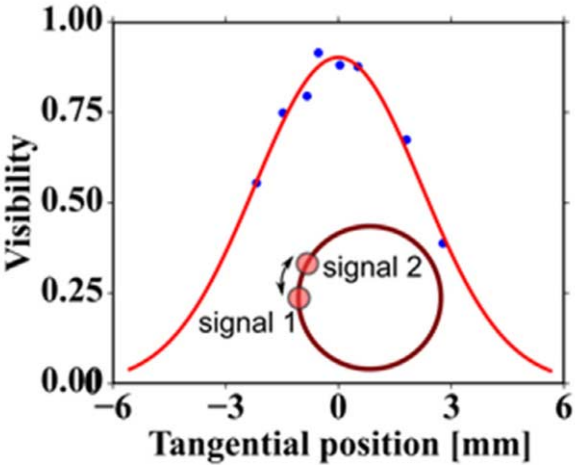


Fig. 12. – Measured visibility of the interference fringes of the two TEM_{00} modes of signal 1 and signal 2 as a function of their tangential distance along the emitted cone structure as published in [19].

opposite sides of the cone. Our spatially selective TEM_{00} mode detector was set up to measure a small emission share from this cone in the areas of signal 1 and signal 2 as depicted in fig. 10. The resulting two TEM_{00} modes of signal 1 and signal 2 were fed into a Mach-Zehnder interferometer as shown in fig. 11.

With this arrangement different TEM_{00} modes of the emitted single photons could be overlaid using a delay line for zero delay and the interference visibility could be measured as a function of the distance of the two TEM_{00} modes within the cone as shown in fig. 10 right. The result of this measurement is shown in fig. 12.

The maximum visibility in this case was determined to about 90% as a consequence of no corrections for any background or other disturbing signals. The experimentally observed width of the visibility curve demonstrated that only photons inside the TEM_{00} detection mode are coherent as described in detail in [19].

The also available idler photon on the opposite side of the light cone (see fig. 10, left) can be used as a reference for the signal photon. From the coincidence measurement of the idler and the signal photons the distinguishability D between the two TEM_{00} measured modes could be determined. The result of this measurement is given in fig. 13.

From these two measurements at each distance between the two TEM_{00} modes signal 1 or signal 2 the complementarity value of $D^2 + V^2$ was determined to values between 0.8 and 0.9 [19]. These values confirm that in this measurement the complementarity principle in the spatial dimension is nicely fulfilled. The reason why the sum is slightly below 100% is a consequence of the not corrected visibility measurement of fig. 12.

These experimental results can be summarized in the following way:

– **Photons in a TEM_{00} mode are only coherent from the same mode.**

– **These photons are not distinguishable.**

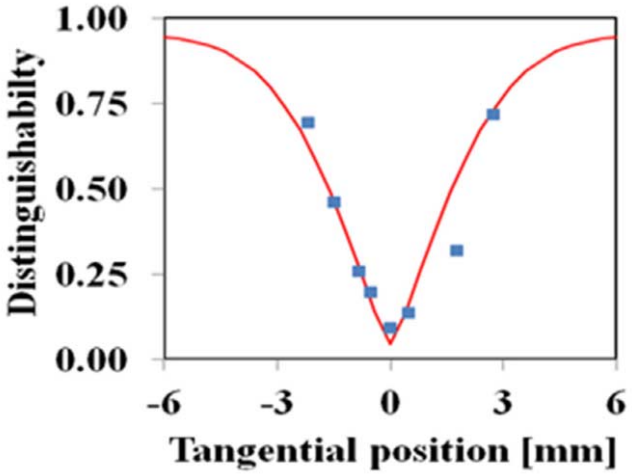


Fig. 13. – Distinguishability between the single photon in the TEM_{00}-modes signal 1 or signal 2 as a function of the tangential distance of signal 2 in relation to the signal 1 as published in [19].

Regarding complementarity in the spatial dimension it can be concluded:

Distinguishable spatial modes are generated from different vacuum fields. Because of the randomness of the vacuum fields these modes are not coherent. Thus also in the spatial dimension complementarity is a consequence of the randomness of the vacuum fields.

6. – Complementarity for single photons in higher-order spatial modes

After having the mode concept applied for single photons in the fundamental mode, successfully, the question about complementarity for photons in higher-order spatial modes will be asked. These modes show intensity structures with at least two or more humps as *e.g.* the TEM_{01} Gauss-Laguerre or Gauss-Hermite modes. The electric field has opposite phases in the neighboring humps and between the humps both the electric field and the intensity is 0 (see fig. 14).

This suggests the questions: Does a single photon in such higher-order mode show interference as the classical mode would do and can the single photon be localized in one of the humps?

For investigating the complementarity principle for single photons in such higher-order spatial modes reference photons are needed to determine the distinguishability as in all previous measurements described above. The known way to realize such a photon correlation is spontaneous parametric down-conversion as described above.

Therefore we applied photons in a TEM_{01} mode to pump at type II SPDC crystal [20]. Indeed the out coming signal and idler photons showed a double-hump structure in the near field of the crystal similar as the pump photon. Therefore the single signal photons in

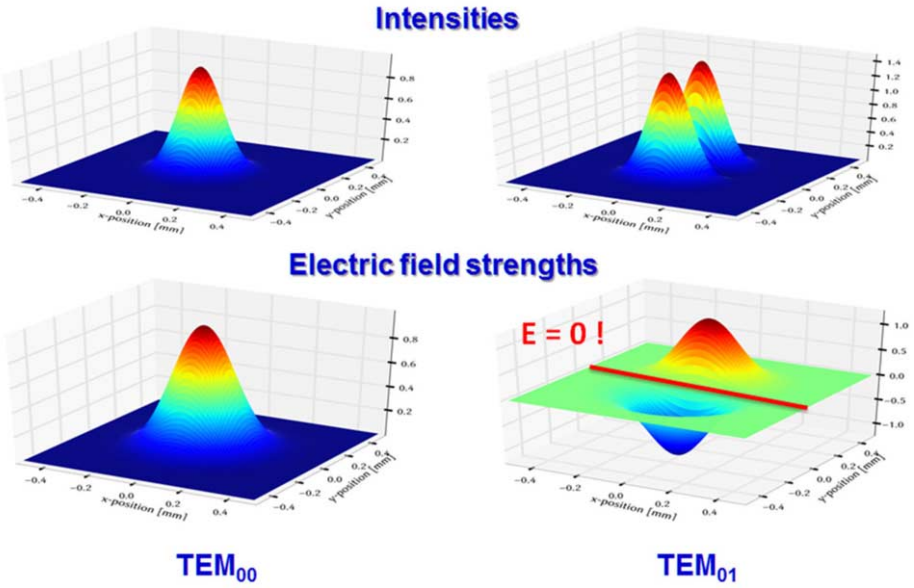


Fig. 14. – Calculated intensity (top) and electric field patterns (bottom) for a TEM_{00} (left) and a TEM_{01} (right) mode. Both the electric field and the intensity of the TEM_{01}-mode are zero between the two humps.

this TEM_{01}-like mode near field structure were used in the Youngs double-slit experiment as shown fig. 15 and described in detail in [20].

The two humps of the signal photon light were imaged onto the two slits. With the idler photon detector as reference "which-slit" information based on the near field correlation could be measured in coincidence if the signal photon detector was placed right behind the double slit. A very strong correlation of more than 95% was found for

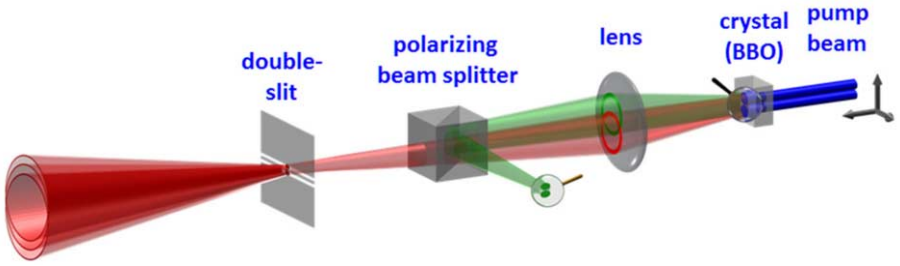


Fig. 15. – Scheme of the double-slit experiment using a TEM_{01} mode to pump at type-II SPDC crystal as described in detail and published in [20]. The two humps of the signal photons were imaged onto the double slit and the idler photons coupled out with a polarizing beam splitter and imaged on the idler detector for reference.

the two photons emitted both from the same hump of the pump mode. We observed this result also for single-photon pairs in higher-order modes as *e.g.* TEM_{02} or TEM_{11} modes.

While imaging the single-signal SPDC-photons onto the double slit as shown in fig. 15 and described in detail in [20] interference fringes with good visibility in the range of 0.6 could be observed in the far field behind the slits. This result seems to be conflict with the complementarity principle. But this simple arrangement of fig. 15 includes an unfair sampling for the measured photons. Much more photons are detected for the "which-slit" information compared to the photons observed in the interference measurement as it was discussed in ref. [21].

A detailed analysis showed that this simple experiment shows a reach variety of transversal light structures behind the SPDC crystal as a consequence of the TEM_{01} pump mode [22] and the resulting complex phase matching conditions. Unfortunately the higher-order pump mode does not directly transform to the same down-converted mode of the signal and idler photons.

But based on the results of this complicated and full theoretical description of the SPDC process without any fitting parameters [22] and the realized almost perfect match of the experimental results it was possible to modify the experimental setup with apertures in the far field between the crystal and the polarizing beam splitter (near the lens in fig. 15) to avoid the unfair sampling. Both the idler and the signal photon detector are measuring the same photon pairs in the near and in the far field of the signal photons, then. Thus which-slit information and interference visibility can directly be combined in this modified setup.

This way it was possible to obtain the double-slit interference with signal photons in a TEM_{01}-like mode to our knowledge for the first time. The result as described in detail in [23] is shown in fig. 16.

As can be seen from the left side of fig. 16 the far field interference pattern of the signal photons shows on this side of the emitted light cone a dip in the middle of the interference structure. This indicates that the electric field of the two humps imaged onto the two slits has a phase shift of 180 because otherwise the maximum of the interference would be obtained in the middle of that structure. So far the TEM_{01}-like mode structure is confirmed from this observation.

On the right side of fig. 16 the distinguishability D, the visibility V and the $D^2 + V^2$ values of the signal photons are given as a function of the vertical position of the idler detector in this coincidence measurement. If the idler detector is positioned in the middle of the TEM_{01} mode the visibility is as high as possible ($V = 1$) and the distinguishability D of the signal photon passing slit 1 or slit 2 is zero. If the idler detector is moved outside of the middle of the mode the visibility drops and the distinguishability increases. The $D^2 + V^2$ value is 1 in maximum which is expected and is a little bit smaller in between.

Therefore also for single photons in a higher-order transversal mode (in our case TEM_{01}) the complementarity principle is fulfilled. But in the context of the previous discussions the physical background of this behavior can be analysed, too. First it can be concluded that even single photons in a higher-order transversal mode show the clas-

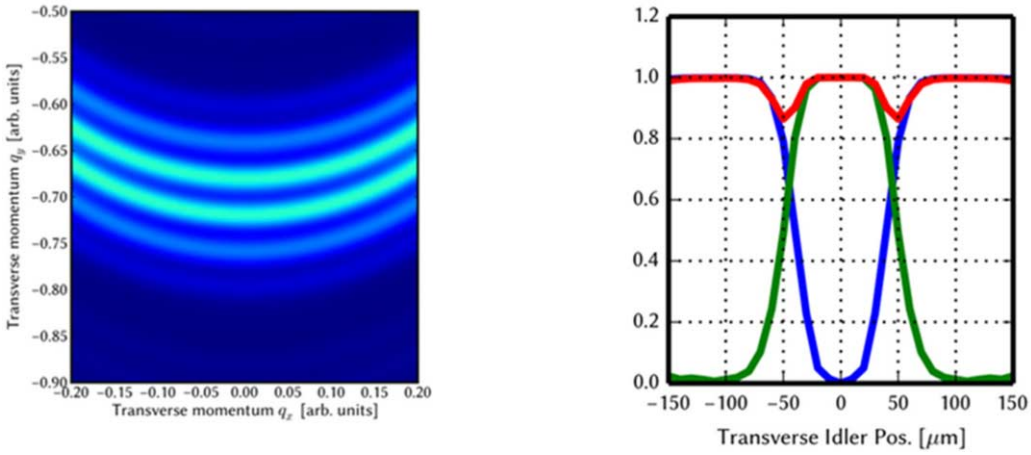


Fig. 16. – Result of the double-slit interference with signal photons in a TEM_{01}-like mode (left) and the associated distinguishability and visibility measurements (right) as described in detail in [23] with the setup of fig. 15 but with apertures for fair sampling.

sically expected interference probability structure. Although the near field coincidence measurement of the signal and idler photon pair shows a strong localization of the emission spot of the bi-photon in just one of the humps the interference pattern is given by the full mode.

It can be concluded from this experiment that also the higher-order transversal modes exist as non-occupied vacuum field modes. With this physical background there is a very intuitive explanation of the result of fig. 16:

As long as our detectors see the vacuum field contributions in the generation process of the bi-photons only in the desired TEM_{01} mode the obtained visibility V is 1 as a consequence of the coherence of this mode. But as soon as the idler detector is moved towards one of the humps photons not belonging to the desired TEM_{01} mode are measured. These photons are generated by different vacuum fields. These vacuum fields are not coherent to the vacuum fields of the first mode. Therefore the visibility drops. But then the two modes become distinguishable and the distinguishability D in this measurement increases. In any case the $D^2 + V^2$ value is maximum one.

Finally these measurements can be summarized in the following way:

Photons in a higher-order transversal modes are coherent and not distinguishable and vice versa ($D^2 + V^2 = 1$).

7. – Conclusion

In summary of the here discussed experiments the complementarity principle was demonstrated as a consequence of the measuring process, which makes a selection of the

mode function and the involved vacuum field contributions. Finally this may give a new conceptual perspective towards quantum optics.

The physics behind the complementarity principle may be summarized as follows:

– **uncertainty is a consequence of the wave nature of quantum objects**

– **duality is a consequence of the energy packet in this wave (mode)**

– **the detection system selects coherent or distinguishable modes**

– **vacuum field modes are random: intrinsic randomness in QM**

– **coherent (laser) modes can overwrite the random vacuum fields**

The most important consequence of this discussion is:

Only modes based on the same vacuum are coherent.

Therefore each experimental situation can be analyzed for the measured modes and their vacuum contributions. Complementarity means that photon modes based on the same vacuum will be coherent because the excitation of these modes does not change their phase. But as in the example of the 3-crystal experiment the same spatial mode can be excited by different vacuum field contributions and then the single-photon modes can be incoherent. Therefore the final consequence of this discussion is:

The physics of the vacuum field modes causes complementarity in quantum optics.

* * *

Detailed and long standing discourses as well as fruitful scientific collaborations over the years with Axel Heuer, Peter Milonni, Martin Ostermeyer, Wolfgang Schleich and Martin Wilkens and with the members of my group are greatfully acknowledged. In addition very helpful have been the discussions with Bob Boyd, Claude Fabre, Berthold Englert, Gerd Leuchs, Pierre Meystre, Harry Paul, Marlan Scully and Ewan Wright and many others.

REFERENCES

[1] BOHR N., *Nature*, **121** (1928) 580.
[2] WOOTERS W. K. and ZUREK W. H., *Phys. Rev. D*, **19** (1979) 473.
[3] FEYNMAN R., *The Feynman lectures on physics (CalTech)*, **III** (1964-2010).
[4] ENGLERT B. G., SCULLY M. O. and WALTHER H., *Sci. Am.*, **12** (1994) 86.
[5] MILONNI P. W., *The Quantum Vacuum* (Academic Press) 1994.
[6] MENZEL R., *Photonics*, 2nd edition (Springer) 2007.
[7] BORN M. and WOLF E., *Principles of Optics*, 7th edition (Cambridge University Press) 2013.
[8] ENGLERT B. G., *Phys. Rev. Lett.*, **77** (1996) 2154.

[9] SCULLY M. O., ENGLERT B. G. and WALTHER H., *Nature*, **351** (1991) 111.
[10] BOYD R., *Nonlinear Optics* (Academic Press) 1992.
[11] WALBORN S. P., MONKEN C. H., PDUA S. and SOUTO RIBEIRO P. H., *Phys. Rep.*, **495** (2010) 87.
[12] MILONNI P. W., FEARN H. and ZEILINGER A., *Phys. Rev. A*, **53** (1996) 4556.
[13] HEUER A., MENZEL R. and MILONNI P. W., *Phys. Rev. Lett.*, **114** (2015) 053601.
[14] GLAUBER R. J., *Phys. Rev.*, **130** (1963) 2529.
[15] OU Z. Y., WANG L. J., ZOU X. Y. and MANDEL L., *Phys. Rev. Lett.*, **67** (1991) 318.
[16] HEUER A., RAABE S. and MENZEL R., *Phys. Rev. A*, **90** (2014) 045803.
[17] OU Z. Y., WANG L. J., ZOU X. Y. and MANDEL L., *Phys. Rev. A*, **41** (1990) 566.
[18] HEUER A., MENZEL R. and MILONNI P. W., *Phys. Rev. A*, **92** (2015) 033834.
[19] MENZEL R., PUHLMANN D. and HEUER A., *JEOS*, **13:8** (2017) 1.
[20] MENZEL R., PUHLMANN D., HEUER A. and SCHLEICH W. P., *Proc. Natl. Acad. Sci. U.S.A.*, **109** (2012) 9314.
[21] BOLDUC E., LEACH J., MIATTO F. M., LEUCHS G. and BOYD R. W., *Proc. Natl. Acad. Sci. U.S.A.*, **111** (2014) 12337.
[22] ELSNER R., PUHLMANN D., PIEPLOW G., HEUER A., MENZEL R. and MENZEL R., *Opt. Soc. Am. B*, **32** (2015) 1910.
[23] MENZEL R., PUHLMANN D., MARX R., HEUER A. and SCHLEICH W. P., to be published (2017).

Proceedings of the International School of Physics "Enrico Fermi"
Course 197 "Foundations of Quantum Theory", edited by E. M. Rasel, W. P. Schleich and S. Wölk
(IOS, Amsterdam; SIF, Bologna) 2019
DOI 10.3254/978-1-61499-937-9-285

Quantum imaging

Robert W. Boyd

Department of Physics and School of Electrical Engineering and Computer Science
University of Ottawa - Ottawa, Ontario, K1N 6N5 Canada

The Institute of Optics and Department of Physics and Astronomy
University of Rochester - Rochester, New York, 14627 USA

School of Physics and Astronomy, University of Glasgow - Glasgow G12 8QQ, UK

Summary. — We present a brief overview of the field of quantum imaging, concentrating on some recent results. Quantum imaging is a specific example of quantum metrology, and we thus start out with a discussion of quantum metrology including the generation of squeezed light and the generation of entangled photon pairs through the process of spontaneous parametric downconversion (SPDC). We then proceed to review three different examples of quantum imaging, namely ghost imaging, imaging based on interaction-free measurements, and imaging based on Mandel's induced coherence.

1. – What is quantum imaging?

The goal of quantum imaging is to produce "better" images using quantum methods. These images can be better in that they are created through use of a very small number of photons, that they possess better spatial resolution, or that the possess a better signal-to-noise ratio. From a more abstract point of view, one can say that quantum imaging is image formation that exploits the quantum properties of the transverse structure of light fields. In this paper, we present a review of some recent work on this topic. A good summary of earlier work is presented in the book Quantum Imaging [1].

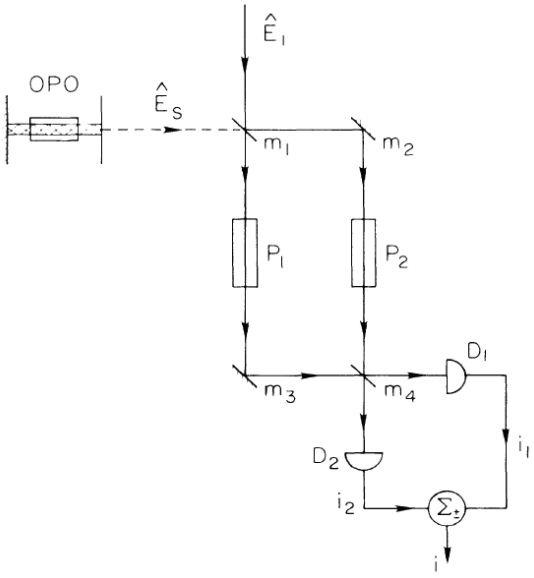


Fig. 1. – Laboratory setup for the demonstration of sub-shot-noise sensitivity through use of squeezed light [4].

2. – Brief history of quantum methods in metrology

Before turning our attention to quantum imaging, it is instructive to first review the usefulness of quantum methods more generally for their use in metrology. Under proper conditions, quantum methods allow one to perform optical measurements with an accuracy that exceeds the "standard quantum limit", the limit imposed by shot noise in a measurement apparatus.

One example of a quantum method in metrology is the use of squeezed light [2]. Squeezed light refers to a light field in which the fluctuations in one conjugate variable are suppressed at the expense of having increased fluctuations in the other conjugate variable. One specific example of squeezed light is quadrature-squeezed light, in which the fluctuations of one quadrature of the field (the part oscillating as $\cos\omega t$, for example) are suppressed and the fluctuations in the other quadrature (the part oscillating as $\sin\omega t$) are increased. Certain nonlinear optical interactions can create light fields with this squeezing property. Quadrature-squeezed light was first demonstrated by the group of Slusher in 1985 [3]. Its application to precision metrology was demonstrated by the group of Kimble [4]. The experimental setup of this work is shown in fig. 1. P_1 and P_2 represent phase modulators placed inside a Mach-Zehnder interferometer. These devices represent the phase objects to be measured by the interferometer. The input to the interferometer is provided by a coherent laser beam E_1 and from the squeezed-vacuum output of an optical parametric oscillator (OPO). Some of the measured results are shown in fig. 2.

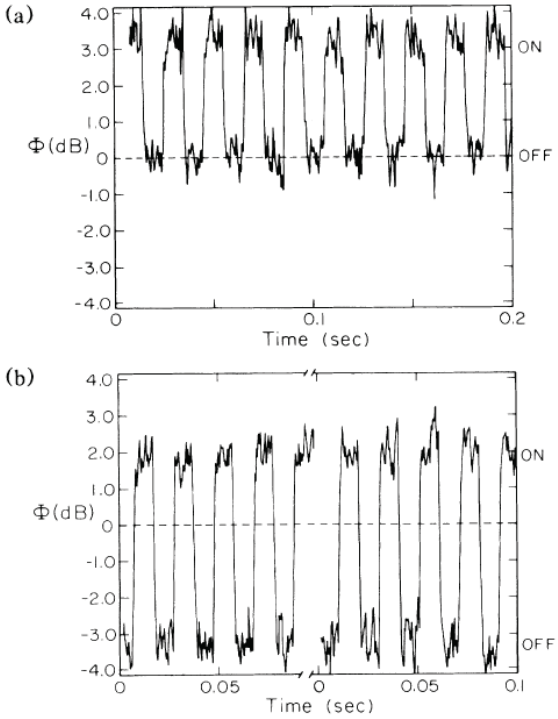


Fig. 2. – Example of data collected with the setup of fig. 1 [4].

The top right figure shows the noise in the output when the OPO is blocked so that the (fluctuating) electromagnetic vacuum enters through the left port. In obtaining this trace, phase modulator P_1 is modulated at 50 Hz to vary the output power. The dashed line at $\Phi = 0$ shows the standard quantum limit (shot noise level). The bottom right figure shows the noise level when the OPO is unblocked so that squeezed vacuum is injected into the left port. One sees that the noise level is decreased by approximately 3 dB. This decreased noise level allows one to make more accurate measurements of any phase shift between the two arms of the interferometer.

Another example of quantum methods in metrology is afforded through use of twin beams. Twin beams are beams of light that contain identical fluctuations. Therefore, although each beam is "noisy", the difference in the intensities of the beams has greatly reduced noise and in principle is entirely noise free. Quantum metrology based on twin beams has been studied extensively by the group of Fabre and coworkers [5]. One example is shown in fig. 3. Here one of the beams falls passes through a potassium vapor cell and is detected by DET2. Its twin falls directly onto detector DET1. The difference in the two photocurrents is measured and is plotted in fig. 4 as a function of the frequency of the light. The trace on the left is seen to be much less noisy than the trace on the right obtained without the use of twin beams.

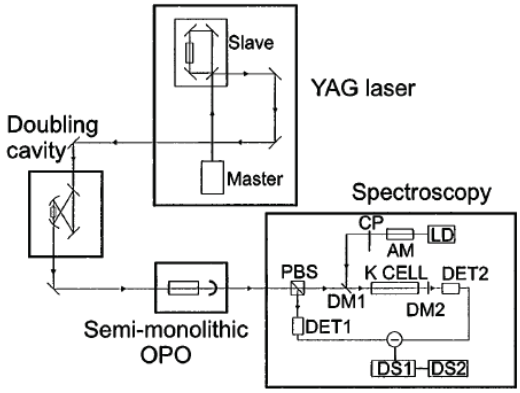


Fig. 3. – Laboratory setup for the demonstration of sub-shot-noise sensitivity through the use of twin beams [5].

Another quantum resource of considerable interest in metrology and in quantum technologies is afforded by entangled light fields. Applications of entangled light fields include quantum teleportation and quantum cryptography. Entangled light sources are also used for performing fundamental tests of quantum mechanics, such as its inherent nonlocality as illustrated in Einstein-Podolsky-Rosen correlations. In the next section, we review one of the standard means of generating entangled photon pairs.

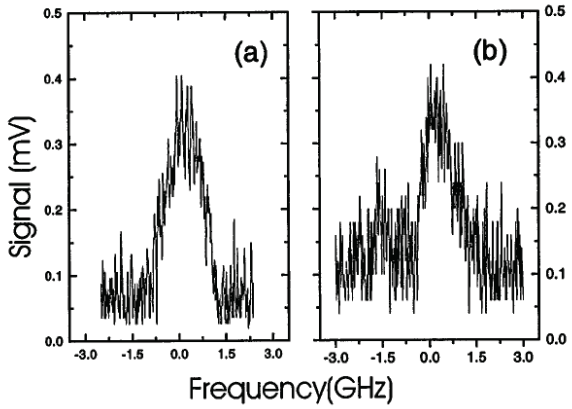


Fig. 4. – Demonstration of sub-shot-noise sensitivity using the setup of fig. 3 [5].

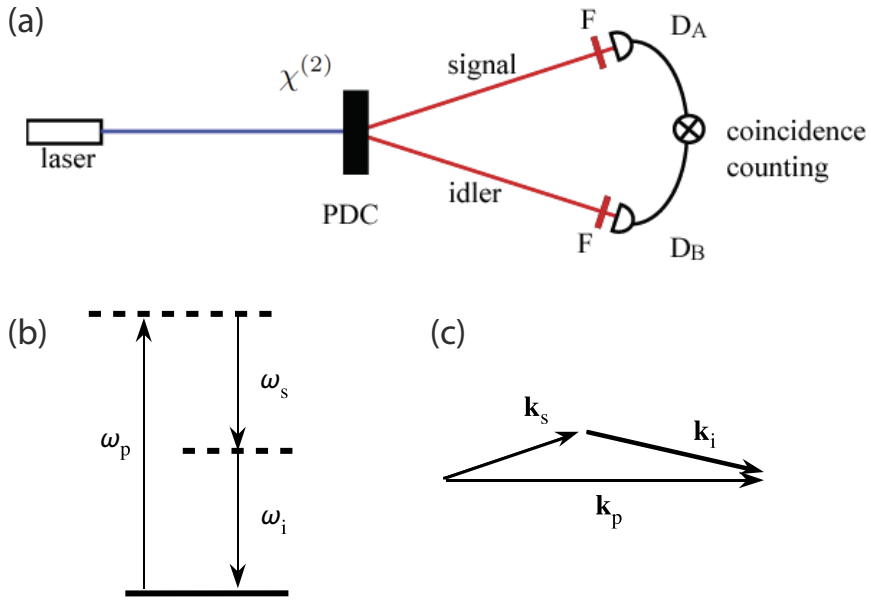


Fig. 5. – (a) Process of parametric downconversion (PDC). (b) Energy conservation in PDC. (c) Wave vector conservation in PDC.

3. – Parametric downconversion and the generation of entangled photons

A primary method for the creation of entangled photons is through the process of parametric downconversion (PDC). This process is illustrated in fig. 5. Part (a) of the figure shows an intense laser beam illuminating a crystal characterized by a second-order nonlinear optical response. On occasion a pump photon splits into two new photons as a consequence of the nonlinear response of the system. For historical reasons these two photons are known as the signal and idler photon, with considerable arbitrariness as to which photon is the signal and which is the idler. Part (b) of the figure illustrates this process in terms of an energy-level diagram, and part (c) shows how photon momentum is conserved in this process.

The photon pairs created by this process are said to be entangled, and they can show entanglement by means of any of their degrees of freedom, such as (a) polarization, (b) time and energy, (c) position and transverse momentum, or (d) angular position and orbital angular momentum. As an example of what is meant by entanglement, we consider the specific case of time-energy entanglement. The photons created by the PDC process of fig. 5 have the property that if the signal frequency ω_s is measured, then one can immediately predict that the idler frequency is given by $\omega_i = \omega_p - \omega_s$. However, if instead of measuring the frequency of the signal photon, one measures the moment of time when it was created, one always finds that the idler photon was created at the same moment of time. However, by measuring the moment of time when the signal photon

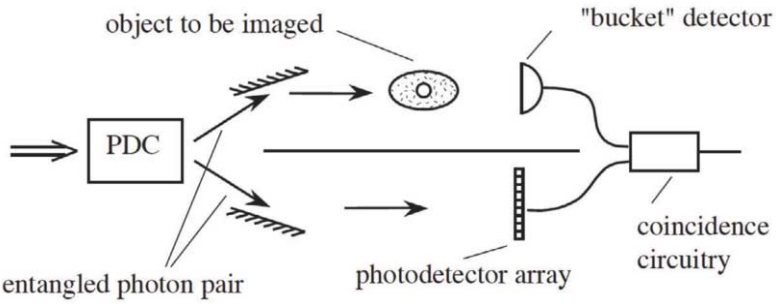


Fig. 6. – Illustration of the process of ghost imaging.

was created, one loses all knowledge of its energy, and in fact one even loses all knowledge of the energy of the idler photon. Likewise, when one measures the energy of the signal photon, one loses all knowledge of the emission time of both signal and idler photons. In fact, one can wait until both photons have long left the crystal and become well separated from one another before deciding which property (energy or emission time) of the signal photon to measure. Nonetheless, the idler photon is always found to have the same property perfectly correlated with that of the signal photon. This property is the key experimental signature of entanglement: the two photons have properties that are completely correlated even in two mutually unbiased bases.

4. – What is ghost imaging and what are its properties?

Ghost imaging, also known as coincidence imaging, is a special sort of imaging technique that can offer significant advantages under certain circumstances. Ghost imaging was originally reported by Strekalov *et al.* [6] and by Pittman *et al.* [7] and has subsequently been studied by many groups [8-20].

The process of ghost imaging is shown schematically in fig. 6. A laser beam incident from the left excites a second-order nonlinear crystal where parametric downconversion (PDC) occurs, leading to the generation of a pair of spatially entangled photons. One of these photons falls onto an object to be imaged. If it falls onto a low-loss region of the object, it will be transmitted and will be detected by the bucket detector shown in the figure. This detector provides no spatial information about the object. The other photon falls onto a photodetector array. This detector records the position of this photon and thus the position of the other photon in the plane of the object. By performing a coincidence measurement between these two measurements, one is able to determine the intensity structure of the object based on measurements of the properties of photons that have never physically interacted with the object. For this reason, this imaging method has been referred to as "ghost imaging".

There has been an ongoing discussion as to whether ghost imaging is a "quantum" phenomenon. The first demonstration of ghost imaging by Strekalov *et al.* [6] made

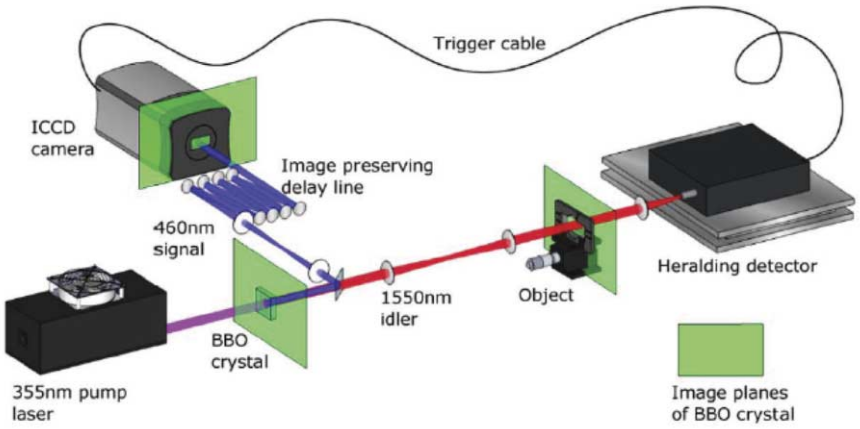


Fig. 7. – Setup for two-color ghost imaging. The BBO nonlinear crystal is designed to split pump photons at 350 nm into signal photons at 460 nm and idler photons at 1550 nm.

use of the correlations of entangled photons and certainly was quantum in this sense. However, the question was still open as to whether other types of correlations of a purely classical nature could be used to perform ghost imaging. One group [8] held that ghost imaging was a purely quantum effect. This claim was refuted by Bennink *et al.* (2002) [9] who reported the observation of ghost imaging through use of light beams that showed only classical correlations. The situation was clarified by the work of Gatti *et al.* [10], who developed a criterion for demonstrating quantum features of ghost imaging, namely the presence of correlations in both the near and far fields of the source of light source. These features were subsequently verified experimentally by Bennink *et al.* (2004) [12]. Specifically, they demonstrated that good ghost images were observed using the correlations of both the near and far fields of a parametric downconversion source, but that a classically correlated source could produce good ghost images in only one conjugate plane.

Aside from questions associated with the quantum or classical origin of ghost imaging, the fact remains that ghost imaging can provide new possibilities for image formation that are not available using traditional techniques. One example of such a modality is that of two-color ghost imaging. In this process, the light that illuminates the object can be of a significantly different wavelength of the light that falls onto the detector array. One achieves entanglement between two beams of very different wavelength by adjusting the orientation of the nonlinear crystal used to perform parametric downconversion to achieve phasematching for nondegenerate (different wavelength) conditions. An example of such a two-color ghost imaging measurement setup [19] is shown in fig. 7. A beta barium borate (BBO) nonlinear crystal is designed to split pump photons at 350 nm into signal photons at 460 nm and idler photons at 1550 nm. The 1550 nm photons fall onto the object, and the transmitted photons are registered by a sensitive "bucket" detector.

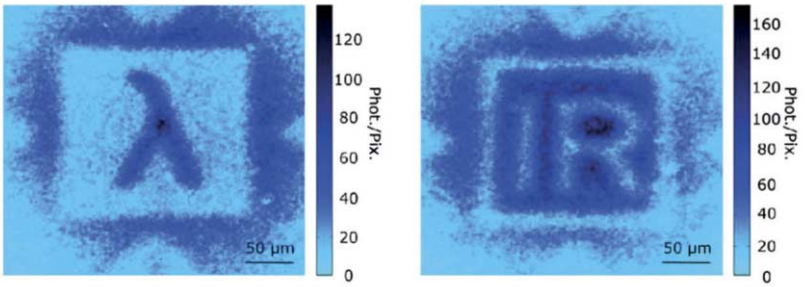


Fig. 8. – Ghost images recorded in the laboratory using the setup of fig. 7. The images show the properties of two stencils at a wavelength of 1550 nm, as recorded by a camera sensitive to 460 nm light. The objects are gold stencils on a silicon substrate.

This detector acts as a trigger for an imaging detector (labeled ICCD) that is sensitive to the 460 nm light. This trigger pulse must arrive approximately 50 ns before the image-bearing photon. These workers thus make use of an image-preserving delay line to ensure that the image-bearing photon arrives at the correct time. Some images obtained with this system are shown in fig. 8.

To summarize this section, we have seen that traditional ghost imaging is not an intrinsically quantum phenomenon, although some methods of ghost imaging can display quantum features. We will next turn to other sorts of quantum imaging that are intrinsically quantum in nature.

5. – Interaction-free imaging

In this section, we describe a quantum imaging procedure known as interaction-free imaging. Before we do so, let us first ask the question of what constitutes a quantum measurement. As a specific example, we consider the process shown in fig. 9. Here a single photon falls onto a beam splitter, and we wish to determine through which output port the photon leaves. In Situation 1, the detector to the right of the beam splitter registers the photon (the detector "clicks"). We thus know with certainty that the photon exited through the right-side output port of the beam splitter. Let us now consider the circumstance of Situation 2. In this case, the detector does not click. If we assume that the detector is ideal in that it registers every photon that falls onto it, we thereby conclude that the photon must have exited through the upper output port of the beam splitter. We thus reach the provocative conclusion that the lack of a detection event can constitute a quantum measurement. Similar situations have been described by Renninger [21] and by Dicke [22].

We next describe what is meant by an interaction-free measurement. The concept of an interaction-free measurement was introduced theoretically by Elitzur and Vaidman [23]. Interaction-free measurements were described experimentally by Kwiat *et al.* [24]. For conceptual clarity, we consider the situation described by White *et al.* [25]

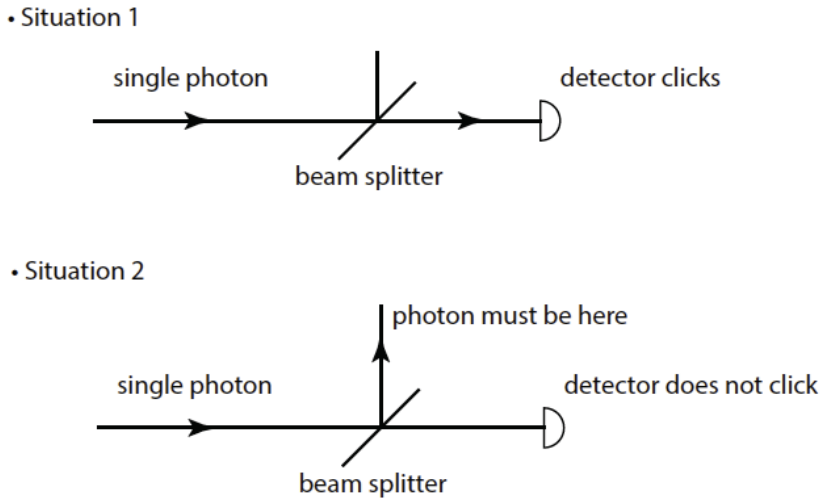


Fig. 9. – Even the lack of a detection event can constitute a quantum measurement.

as shown in fig. 10. Part a shows the situation in which a single photon falls onto a Mach-Zehnder interferometer in which the path lengths have been adjusted so that all light exits through the horizontal output port and falls onto detector D_1. Part b shows what happens if an opaque object is placed into the upper arm of the interferometer. The presence of this object blocks the upper path and thereby frustrates the destructive interference that prevented light from exiting through the port leading to detector D_2. Thus, 25% of the time the incident photon will fall onto D_2 and produce a click. This result is quite perplexing. It shows that one can deduce that an opaque object is located within the interferometer. However, we know that the photon did not physically strike the object, because the object is assumed to be opaque and we also know that the photon was detected by D_2. White *et al.* [25] developed this concept into a form of quantum imaging. They placed a focusing system into the upper arm of the interferometer and translated various objects through the focal region. In this way, they were able to map out the transmission profiles of these objects, as measured by photons that never directly interacted with the object.

6. – Imaging by Mandel's induced coherence

We next turn to another imaging modality, known as imaging by induced coherence. It is also known as imaging with undetected photons, as it was called in the original publication [26] demonstrating this effect. This procedure is fully quantum in nature.

As a first step, let us review the concept of induced coherence as described initially by the group of Mandel [27]. Their experimental setup is shown in fig. 11. Two parametric downconversion crystals NL1 and NL2 are pumped by a UV line of an argon ion laser. The signal beams from each crystal are combined at beam splitter BS_O, and the power

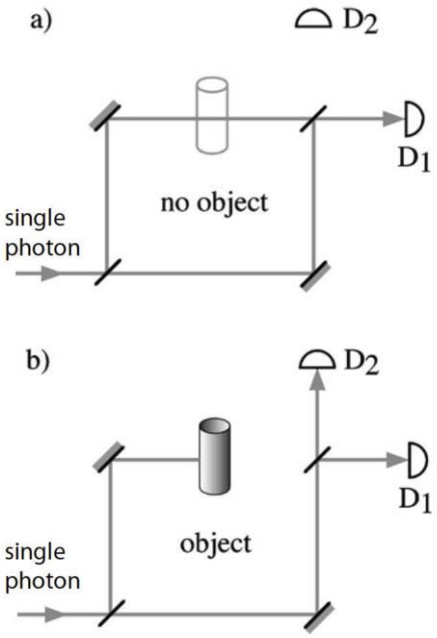


Fig. 10. – The concept of an interaction-free measurement. Adapted from White *et al.* [25] with permission.

hitting detector D_s is measured as a function of the position of BS_O as it is translated vertically. The results are shown in fig. 12. Interference fringes are observed when the idler beams from the two crystals are aligned (curve A). However, these fringes disappear when the idler beam path between the two crystals is blocked (curve B).

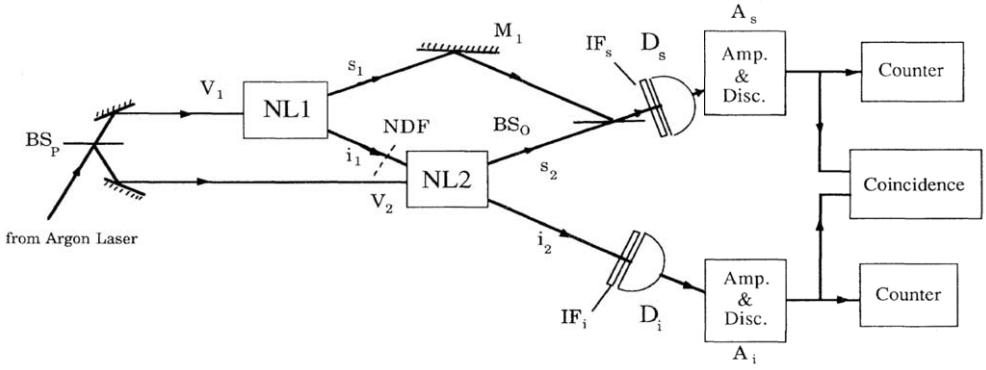


Fig. 11. – Setup for studying induced coherence [27], reproduced with permission.

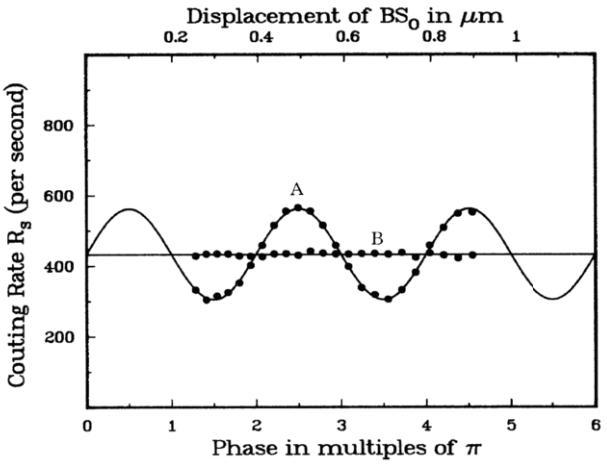


Fig. 12. – Results from the induced-coherence experiment of fig. 11. Curve A corresponds to the neutral density filter (NDF of fig. 11) having a transmission of 91%; curve B corresponds to the NDF having zero transmission.

These results are perhaps unexpected for the following reasons. In performing this measurement, the pump intensity was kept sufficiently low that there was essentially no *induced emission* from NL2. By this, one means, for example, that when the idler beam between the two crystals was blocked, the emission rate from NL2 did not change (decrease) by a measurable amount. Moreover, the emission rates from NL1 and NL2 were sufficiently low that there were essentially never photons from both NL1 and NL2 present simultaneously within the measure-ment device. Nonetheless, interference fringes were observed. The explanation of this effect is that interference occurs in quantum mechanics when two pathways are indistinguishable. Specifically, when the idler paths are unblocked and aligned, there is no way to tell if a photon arriving at D_s came from NL1 and NL2. Conversely, if a beam block is placed between the two crystals, then a "click" at D_i demonstrates that the photon pair was created in NL2. Thus, the pathways to Ds from NL1 and NL2 become distinguishable, and the interference no longer occurs.

These ideas were implemented in an imaging context by the group of Zeilinger in work published in 2014 [26]. Their experimental setup is shown in fig. 13. It is similar to that of fig. 11, except that an object O (the stencil of a cat) is placed in the pathway between NL1 and NL2. Also, the pump laser wavelength is 532 nm. The nonlinear crystals are cut for nondegenerate SPDC producing a signal photon at 810 nm and an idler photon at 1550 nm. D1 is a dichroic beamsplitter. The idler photon is directed along path d and the signal photon along path c. The image of the object O is thus impressed onto the idler photon, which is combined with the pump beam at D2 and both beams enter NL2 where another signal beam is created. The idler beam is then expelled from the setup at D3. The two signal beams from paths c and e are now combined at BS2 where

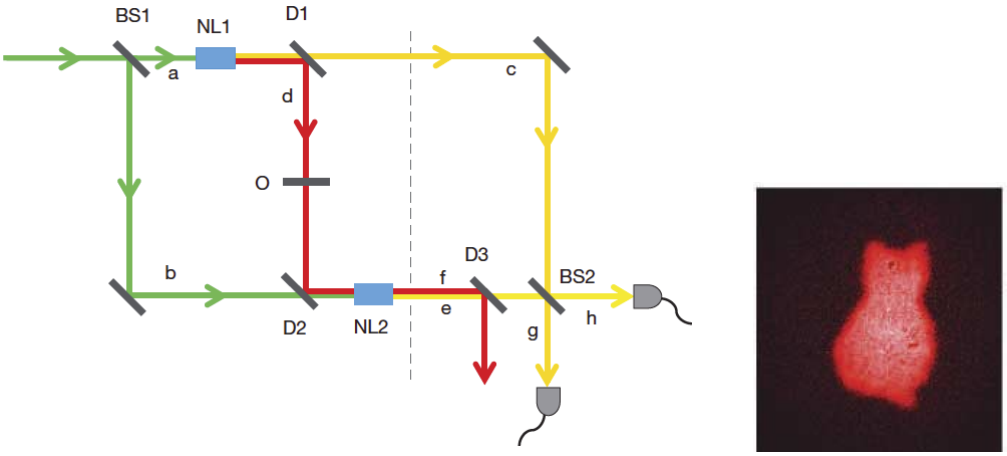


Fig. 13. – Left: Experimental setup to perform imaging based on induced coherence. Right: Image of a cat obtained with this experimental arrangement.

they interfere. The image of the cat, shown to the right of the figure, is created by the interference. As in the experiment of Mandel [27], the intensity of the idler beam in path d is too small to induce emission. Only the coherence of this beam is transferred to the signal beam in the process of PDC in NL2.

7. – Technology for quantum imaging

Significant technological progress has been made in recent years in the development of sensitive low-noise cameras. These cameras have properties that approach the ideal situation of a 100% detection quantum efficiency and a vanishing dark-count rate. Two of these modern cameras are as follows.

- Electron multiplied CCD (EMCCD) cameras have a detection quantum efficiency of about 80%, but have a background dark count rate of about 0.02 counts per pixel per readout. These specifications render these cameras suitable for many applications in quantum information.

- Intensified CCD (ICCD) cameras have a detection quantum efficiency of only about 20%, but can be gated in such a way that there are essentially no dark counts in an integration time. The ICCD camera was mentioned earlier in this chapter. It is the camera used in the work presented earlier in relation to fig. 7.

In the remainder of this section we describe the results of one particular study, that of Edgar *et al.* [28], which made use of an EMCCD camera. We note also the work of the group of Walmsley on similar topics [29]. To establish the context of the study of Edgar *et al.* [28], we present fig. 14, which shows the distribution of light produced by spontaneous parametric downconversion. Clearly the light is emitted into a very large

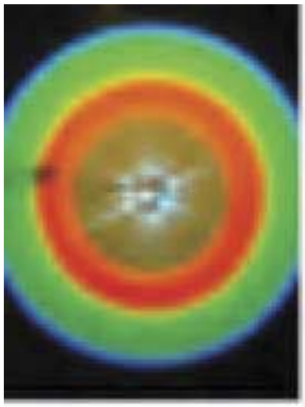


Fig. 14. – Spatial and spectral distribution of light generated by the process of spontaneous parametric downconversion (SPDC).

number of spatial and frequency modes of the field. Light emitted at opposite sides of the distribution are spatially entangled, for the reasons described in the description of fig. 5. Historically, one usually examined the nature of this entanglement through the use of point detectors that are raster-scanned through the intensity pattern. However, in the study of Edgar *et al.*, the entanglement of the entire distribution was measured simultaneously through use of an EMCCD camera.

The experimental setup of Edgar *et al.* [28] is shown in fig. 15. The pump source is a continuously running mode-locked Nd:YAG laser that is frequency tripled to produce an output at 350 nm. A BBO nonlinear crystal cut for type-I degenerate phase matching produces entangled photon pairs at 700 nm through the process of SPDC. In part (a) of the figure the plane of the BBO crystal is imaged onto the EMCCD to allow the measurement of correlations in position space. In part (b) the Fourier plane of the crystal is imaged onto the EMCCD to allow the measurement of correlations in transverse momentum. As spectral filter (not shown in the figure) centered at 700 nm with a bandwidth of 10 nm is placed immediately in front of the camera so that only photons of nearly the same wavelength were detected.

Some of the results of this study are shown in fig. 16. The panel on the left shows that there is a strong correlation in the spatial positions of the signal and idler photons. The panel on the right shows that there is a strong anticorrelation between the momenta of the signal and idler photons. Strong correlations in either position or momentum (whichever one chooses to measure) is the key signature of quantum entanglement. This thought can be rendered quantitative in terms of the Reid criterion which states that

$$\Delta^2_{\min}(x_1|x_2)\ \Delta^2_{\min}(p_{x1}|p_{x2}) > \frac{\hbar^2}{4}\,,$$

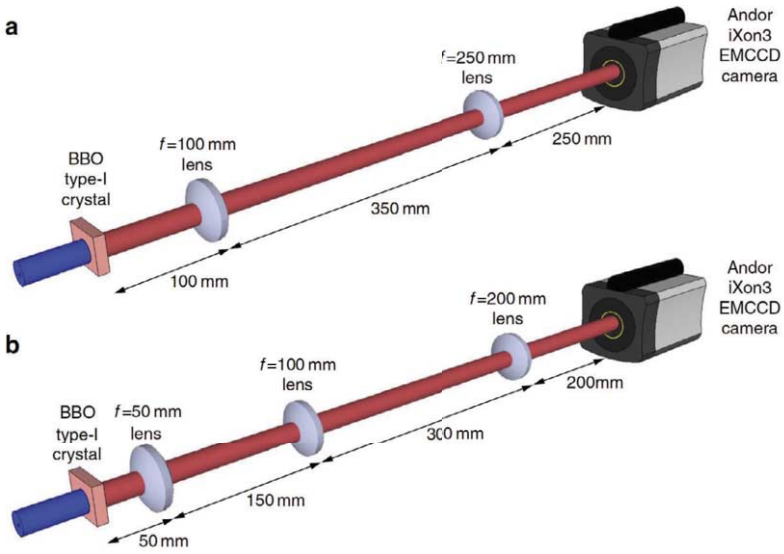


Fig. 15. – Experimental scheme used to measure (a) position and (b) momentum correlations. In (a) the camera is in an image plane of the PDC crystal; in (b) it is in the Fourier plane of the crystal.

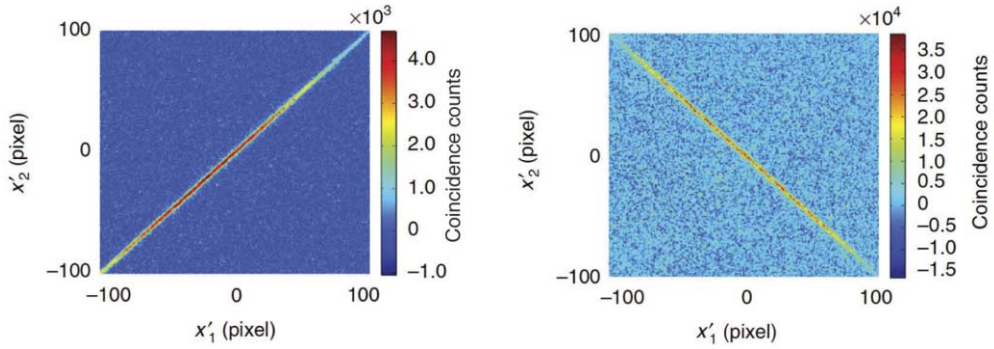


Fig. 16. – Probability distributions for joint detections in the image plane (left) and far-field (right).

where $\Delta^2_{\min}(r_1|r_2)$ is the minimum inferred variance, describing the minimum uncertainty in measuring the variable r_1 conditional on the measurement of variable r_2. The violation of this inequality is a signature of entanglement. Edgar *et al.* [28] report an uncertainty product of

$$\Delta^2_{\min}(x_1|x_2)\ \Delta^2_{\min}(p_{x1}|p_{x2}) > 6 \times 10^{-4}\hbar^2,$$

which is an indication of strong entanglement. Edgar *et al.* also estimate that there are 2500 spatial modes of the light field that are entangled.

8. – Summary and discussion

Quantum imaging is a still-developing field with important implications. Quantum methods can be used to form images that are better than classical images in terms of sensitivity and spatial resolution. From a different perspective, imaging methods can be used the enhance the protocols of quantum information. Image science is capable of exploiting the parallelism that is intrinsic to many of the procedures of quantum information science. One example is the simultaneous entanglement involving a very large number of modes of the optical field [28, 30].

In this paper, we have presented a broad overview of quantum imaging, while concentrating on several imaging protocols of current research interest. Three different examples of quantum imaging are described, namely ghost imaging, imaging based on interaction-free measurements, and imaging based on Mandel's induced coherence.

REFERENCES

[1] KOLOBOV M., (Editor) *Quantum Imaging* (Springer-Verlag New York) 2007.
[2] ANDERSON U. L., GEHRING T., MARQUARDT C. and LEUCHS G., *Phys. Scripta*, **91** (2016) 053001.
[3] SLUSHER R., HOLLBERG L., YURKE B. and MERTZ J., in *Optical Instabilities, Proceedings of the International Meeting on Instabilities and Dynamics of Lasers and Nonlinear Optical Systems, University of Rochester, June 18-21, 1985*, edited by BOYD R. W., RAYMER M. G. and NARDUCCI L. M. (Cambridge University Press); see also SLUSHER R., HOLLBERG L., YURKE B., MERTZ J. and VALLEY J., *Phys. Rev. Lett*, **55** (1985) 2409.
[4] XIAO M., WU L. A. and KIMBLE H. J., *Phys. Rev. Lett.*, **59** (1987) 278.
[5] SOUTO RIBEIRO P. H., SCHWOB C., MAÎTRE A. and FABRE C., *Opt. Lett.*, **24** (1997) 1893.
[6] STREKALOV D. V., SERGIENKO A. V., KLYSHKO D. N. and SHIH Y. H., *Phys. Rev. Lett.*, **74** (1995) 3600.
[7] PITTMAN T. B., SHIH Y. H., STREKALOV D. V. and SERGIENKO A. V., *Phys. Rev. A*, **52** (1995) R3429.
[8] ABOURADDY A. F., SALEH B. E. A., SERGIENKO A. V. and TEICH M. C., *Phys. Rev. Lett.*, **87** (2001) 123602.
[9] BENNINK R. S., BENTLEY S. J. and BOYD R. W., *Phys. Rev. Lett.*, **89** (2002) 113601.
[10] GATTI A., BRAMBILLA E. and LUGIATO L. A., *Phys. Rev. Lett.*, **90** (2003) 133603.
[11] GATTI A., BRAMBILLA E., BACHE M. and LUGIATO L. A., *Phys. Rev. Lett.*, **93** (2003) 093602.
[12] BENNINK R., BENTLEY S. J., BOYD R. W. and HOWELL J. C., *Phys. Rev. Lett.*, **92** (2004) 033601.
[13] O'SULLIVAN M. N., CHAN K. W. C. and BOYD R. W., *Phys. Rev. A B*, (2010) 053803.
[14] CHAN K. W. C., O'SULLIVAN M. N. and BOYD R. W., *Opt. Express*, **18** (2010) 5562.
[15] ZEROM P., SHI Z., O'SULLIVAN M. N., CHAN K. W. C., KROGSTAD M., SHAPIRO J. H. and BOYD R. W., *Phys. Rev. A*, **86** (2012) 063817.

[16] TASCA D. S., ASPDEN R. S., MORRIS P. A., ANDERSON G., BOYD R. W. and PADGETT M. J., *Optics Express*, **21** (2013) 30460.
[17] ASPDEN R. S., TASCA D. S., BOYD R. W. and PADGETT M. J., *New J. Phys.*, **15** (2013) 073032.
[18] ASPDEN R. S., TASCA D. S., FORBES A., BOYD R. and PADGETT M. J., *J. Mod. Optic.*, (2014) http://dx.doi.org/10.1080/09500340.2014.899645.
[19] ASPDEN R. S., GEMMELL N. R., MORRIS P. A., TASCA D. S., MERTENS L., TANNER M. G., KIRKWOOD R. A., RUGGERI A., TOSI A., BOYD R. W., BULLER G. S., HADFIELD R. H. and PADGETT M. J., *Optica*, **2** (2015) 1049.
[20] MORRIS P. A., ASPDEN R. S., BELL J. E. C., BOYD R. W. and PADGETT M. J., *Nat. Commun.*, **6** (2015) 5913.
[21] RENNINGER M., *Z. Phys.*, **15S** (1960) 417.
[22] DICKE R. H., *Am. J. Phys.*, **49** (1981) 925.
[23] ELITZUR A. and VAIDMAN L., *Found. Phys.*, **23** (1993) 987; see also VAIDMAN L., *Quant. Opt.*, **6** (1994) 119.
[24] KWIAT P., WEINFURTER H., HERZOG T., ZEILINGER A. and KASEVICH M. A., *Phys. Rev. Lett.*, **74** (1995) 4763.
[25] WHITE A. G., MITCHELL J. R., NAIRZ O. and KWIAT P. G., *Phys. Rev. A*, **58** (1998) 605.
[26] BARRETO LEMOS G., BORISH V., COLE G. D., RAMELOW S., LAPKIEWICZ R. and ZEILINGER A., *Nature*, **512** (2014) 409.
[27] ZOU X. Y., WANG L. J. and MANDEL L., *Phys. Rev. Lett.*, **6** (1991) 318.
[28] EDGAR M. P., TASCA D. S., IZDEBSKI F., WARBURTON R. E., LEACH J., AGNEW M., BULLER G. S., BOYD R. W. and PADGETT M. J., *Nat. Commun.*, (2012) DOI: 10.1038/ncomms1988.
[29] ZHANG L., NEVES L., LUNDEEN J. S. and WALMSLEY I. A., *J. Phys. B*, **42** (2009) 114011.
[30] O'SULLIVAN-HALE M. N., ALI KHAN I., BOYD R. W. and HOWELL J. C., *Phys. Rev. Lett.*, **94** (2005) 220501.

Proceedings of the International School of Physics "Enrico Fermi"
Course 197 "Foundations of Quantum Theory", edited by E. M. Rasel, W. P. Schleich and S. Wölk
(IOS, Amsterdam; SIF, Bologna) 2019
DOI 10.3254/978-1-61499-937-9-301

Spekkens' toy model and contextuality as a resource in quantum computation

L. Catani and D. E. Browne

University College London, Physics and Astronomy department
Gower St, London WC1E 6BT, UK

N. de Silva

University College London, Computer Science department
Gower St, London WC1E 6BT, UK

Summary. — Spekkens' toy model (SM) is a non-contextual hidden-variable model made to support the epistemic view of quantum theory, where quantum states are states of partial knowledge about a deeper underlying reality. Despite being a classical model, it has reproduced many features of quantum theory (entanglement, teleportation, ...): (almost) everything but contextuality, which therefore seems to be the inherent quantum feature. In addition to the importance in foundation of quantum theory, the notion of contextuality seems to be a crucial resource for quantum computation. In particular it has been proven that, in the case of odd prime discrete dimensional systems, contextuality is necessary for universal quantum computation in state-injection schemes of computation based on stabilizer quantum mechanics (SQM). The latter is a subtheory of quantum mechanics which is very popular in the field of quantum computation and quantum error correction. State-injection schemes consist of a classically-simulable part (like SQM) and a resource state that boosts the computation to a quantum improvement. In the odd-dimensional case, SM is operationally equivalent to SQM. In the even-dimensional case, the equivalence only holds in terms of structure, not in terms of statistical predictions. This because qubit-SQM shows contextuality, while qudit(odd dimensions)-SQM does not. We believe that SM can be a valid tool to study contextuality as a resource in the field of quantum computation. Restricted versions of SM compatible with quantum mechanics (QM) can be used as the non-contextual classically-simulable part of state-injection schemes thus opening other scenarios where studying if contextuality is necessary for quantum computational speed-up.

1. – Spekkens' toy model

In the last decades many attempts to better understand quantum theory through hidden-variable models have been developed [1-3]. Nowadays a big question in quantum foundations is whether to interpret the quantum state according to the ontic view, *i.e.* where it completely describes reality, or to the epistemic view, where it is a state of incomplete knowledge of a deeper reality which can be described by the hidden variables. In 2005 R. Spekkens [4] constructed a non-contextual hidden-variable model to support the epistemic view of quantum mechanics. The aim of the model was to replace quantum mechanics by a hidden-variable theory with the addition of an epistemic restriction (*i.e.* a restriction on what an observer can know about reality). The first version of the model [4] refers only to two-dimensional systems (inspired by the quantum bits) and, despite its simplicity, it has obtained many results that were thought to belong only to quantum mechanics (*e.g.* the no-cloning theorem and teleportation). A later version of the model [5], with a more rigorous mathematical formulation, has extended the theory to all discrete prime and continuous degrees of freedom. The latter has been shown to be operationally equivalent, except for the two-dimensional case, to sub-theories of quantum mechanics, so-called *quadrature quantum mechanics*, which in the discrete case correspond to SQM. Almost all the features of quantum mechanics are reproduced there, approximately everything but contextuality (and the related Bell non-locality), which therefore arises as the signature of quantumness. Spekkens' theory has influenced much research over the years (*e.g.*, [16-20]) and it also addresses many key issues in quantum foundations: whether the quantum state describes reality or not, finding a derivation of quantum theory from intuitive physical principles and classifying the inherent non-classical features.

In this section we describe Spekkens' theory for discrete dimensional systems by defining, in a physically motivated way, what are the states, the measurement observables and the outcomes. The updating rules for the state of a system after a measurement can be found in [6].

We denote with $\Omega \equiv (\mathbb{Z}_d)^{2n}$ the discrete phase space of n d-dimensional systems([1]). Let us consider a fiducial set of quadrature variables in the phase space (capital letters), X_j and P_j (like position and momentum in the standard classical mechanics), on each system j, where $j \in 0, \ldots, n-1$, taking values in $0, \ldots, d-1$. These variables allow us to define the *ontic states* (the reality) associated to the systems. Each ontic state represents a set of values for the fiducial observables X_j and P_j, and so an ontic state is denoted by a point in the phase space $\lambda \in \Omega$. We call X_j and P_j observables because they correspond to proper measurable quantities that uniquely define the ontic state. We can refer to Ω as a vector space where the ontic states are vectors (bold characters) whose components (small letters) are the values of the fiducial variables:

$$\boldsymbol{\lambda} = (x_0, p_0, x_1, p_1, \ldots, x_{n-1}, p_{n-1}). \tag{1}$$

([1]) The dimension d is any positive number, odd or even, prime or non-prime, unless differently specified.

Spekkens' theory imposes a restriction on what an observer can know about the ontic state of a system. This means that what an observer can know about the system is described by an *epistemic state* which is a probability distribution $p(\lambda)$ over Ω.

A generic observable, denoted by O, is defined by any linear combination of fiducial variables:

$$O = \sum_m (a_m X_m + b_m P_m), \tag{2}$$

where $a_m, b_m \in \mathbb{Z}_d$ and $m \in 0, \ldots, n-1$. The observables live in the dual space Ω^*, which is isomorphic to Ω itself. Therefore we can define them as vectors, in analogy with ontic states, $\mathbf{O} = (a_0, b_0, a_1, b_1, \ldots, a_{n-1}, b_{n-1})$. The formalism provides a simple way of *evaluating* the outcome σ of any observable measurement O given the ontic state λ, *i.e.* by computing their *inner product*:

$$\sigma = O^T \lambda = \sum_j (a_j x_j + b_j p_j), \tag{3}$$

where all the arithmetic is over $\mathbb{Z}_d$.

The epistemic restriction of ST is called *classical complementarity principle* and it states that two observables can be simultaneously measured when their Poisson bracket is zero, and in this case we will say that they *commute*. This can be simply rephrased in terms of the *symplectic inner product*:

$$\langle \mathbf{O_1}, \mathbf{O_2} \rangle \equiv {\mathbf{O_1}}^T J \mathbf{O_2} = 0, \tag{4}$$

where $J = \bigoplus_{j=1}^{n} \left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]_j$ is the symplectic matrix.

A subspace of commuting observables, thus satisfying the classical complementarity principle, is called *isotropic*. This means that the subspace of the variables jointly known by the observer is an isotropic subspace. We denote the subspace of known variables as $V = span\{O_1, \ldots, O_n\} \subseteq \Omega$, where O_i denotes one of the generators (commuting observables) of V. Taking into account this definition we can define an epistemic state by the set of known variables V, and also the values $\sigma_1, \ldots, \sigma_n$ that these variables take. This means that $\mathbf{O}_j^T \cdot \mathbf{w} = \sigma_j$, where $w \in V$ is the ontic state that evaluates the known observables and it is called the *shift vector*. The set of ontic states consistent with the epistemic state described by $(V, \mathbf{w})$ is $V^\perp + \mathbf{w}$, where the perpendicular complement of V is, by definition, $V^\perp = \{a \in \Omega \mid \mathbf{a}^T \mathbf{b} = 0 \ \forall \ b \in V\}$. The proof of this result can be found in [6]. By assumption the probability distribution associated to the epistemic state $(V, \mathbf{w})$ is uniform (indeed we expect all possible ontic states to be equiprobable), so the probability distribution of one of the possible ontic states in the epistemic state $(V, \mathbf{w})$ is

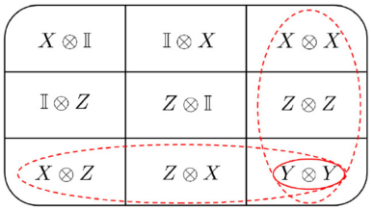
$$P_{(V,\mathbf{w})}(\boldsymbol{\lambda}) = \frac{1}{d^n} \delta_{V^\perp + \mathbf{w}}(\lambda), \tag{5}$$

where the delta is equal to one only if $\boldsymbol{\lambda} \in V^{\perp} + \mathbf{w}$ (note this means that the theory is a *possibilistic* theory).

The aim of Spekkens is to show that epistemic states in his theory are the analogue of quantum states in quantum theory (fig. 1). The analogue of unitary evolutions in quantum theory here corresponds to a subset of all the possible permutations (symplectic affine transformations). The elements of a sharp measurement are here represented as an epistemic state, by virtue of the duality between states and measurements. In the odd-dimensional case the operational equivalence between SM and SQM is proven by using Gross' Wigner functions [6, 15] Figures 1(a) and 1(b) picture the notions defined so far in the two-dimensional case. The notion of entanglement is depicted in fig. 1(c).

2. – Contextuality

In 1967 Kochen-Specker showed that no non-contextual hidden-variable models can ever reproduce all the results of quantum mechanics [3]. The theorem, that can be seen as a complement to Bell's theorem [2], highlights the new concept of contextuality. Probably the most intuitive and popular way to express this concept is through the so-called Mermin square [9]:



The square is composed by nine Pauli observables on a two-qubit system. Each row and each column is composed by commuting (simultaneously measurable) observables. With the assumption that the functional relation between observables is preserved in terms of their outcomes (*e.g.* if an observable C is the product of two observables A, B, also its outcome c is the product of the outcomes a, b of A, B), and that the outcome of each observable does not depend on which other commuting observables are performed with it (non contextuality), the square shows that it is impossible to predict the outcome of each observable among all the rows and columns without falling into contradiction. For example, if we start by assigning values, say ± 1, to the observables starting from the first (top left) row on, the contradiction can be easily seen when we arrive at the last column and last row (red circles), that bring different results to the same observable $Y \otimes Y$, as witnessed by the following simple calculation, $(X \otimes Z) \cdot (Z \otimes X) = Y \otimes Y$, and $(X \otimes X) \cdot (Z \otimes X) = -(Y \otimes Y)$. Measurement contextuality refers to the fact that the outcome of a measurement does depend on the other compatible measurements that we perform with it (*i.e.* on the contexts). More recent versions of contextuality do not only consider sharp measurements, but also preparations and transformations [7]. Non-contextual inequalities have been developed to quantify contextuality [11-13] and in 2014 Howard *et al.* [14] used Cabello-Severini-Winter inequality to prove that contextuality is

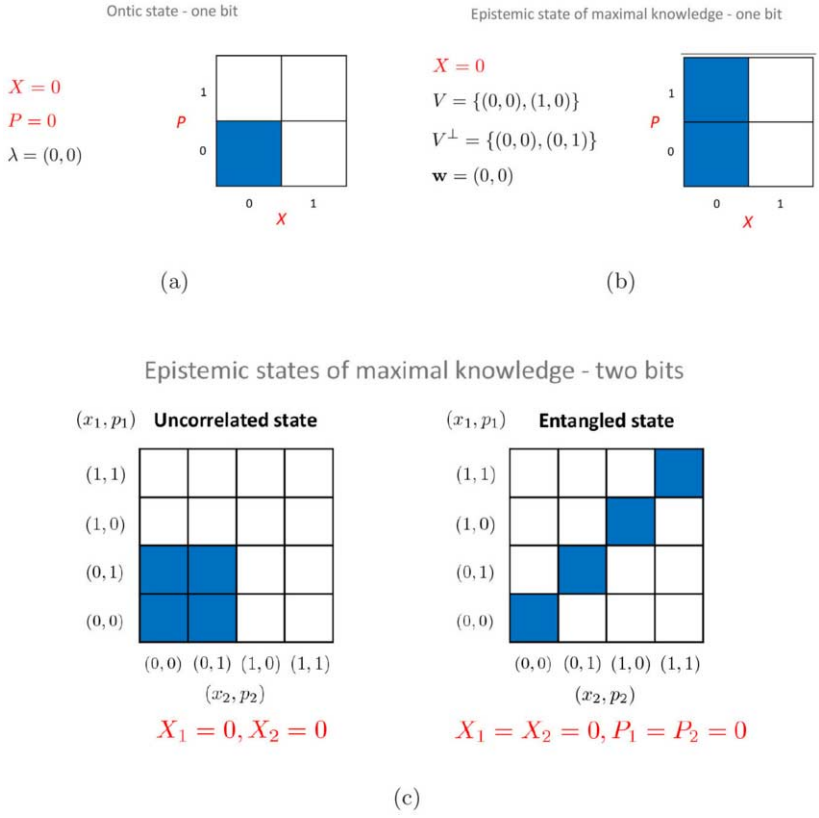


Fig. 1. – Spekkens' toy states of one and two bits. Panels (a) and (b) show the elementary system of Spekkens' theory in two dimension: the bit. One possible ontic state of one bit is shown in (a), where the observer both knows $X = 0$ and $P = 0$, so $\boldsymbol{\lambda} = (0,0)$. The epistemic restriction — classical complementarity principle — in this case corresponds to saying that at maximum the observer has "half" of the knowledge about the ontic state. For example a possible epistemic state is shown in (b), where the observer only knows the variable $X = 0$, so $V = span\{(1,0)\}$ and $\mathbf{w} = (0,0)$. In this case the epistemic state $X = 0$ of one bit can be seen as the analogue of the quantum state $|0\rangle$ of one qubit. Panel (c) shows two kinds of two-bits epistemic states of maximal knowledge. The state on the left is a non-correlated state ($X_1 = 0 = X_2$), indeed we have the knowledge of the states of the individual subsystems, while the state on the right ($X_1 = X_2$ *and* $P_1 = P_2$) is perfectly correlated (*i.e.* entangled), indeed it would be impossible to know the states of the individual subsysytems, but we know exactly the correlation between them (in the case above we know that they have the same ontic states). This trade-off in choosing if knowing the correlation or the states of the individual subsystems is something which is not present in any classical theory.

necessary for universal quantum computation in a state-injection scheme of computation based on stabilizer quantum mechanics. The latter is a subtheory of quantum mechanics where only common eigenstates of tensors of Pauli operators are considered and only unitaries belonging to the Clifford group and Pauli measurements are allowed. The previous state-injection scheme consists of two parts (fig. 2): the "classical" part of the

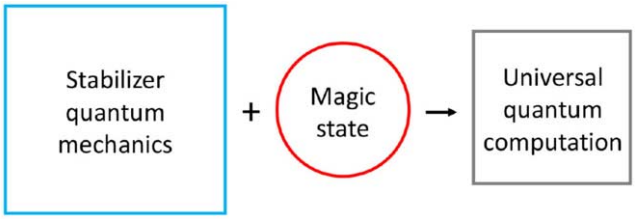


Fig. 2. – State-injection scheme of computation. Quantum computation by state injection consists of a non-contextual and efficiently classically simulable part (like a stabilizer circuit) and a "magic" resource state (given by a non-Clifford gate) that boosts it to universal quantum computation. Howard *et al.* [14] have found that if the resource state can be distilled into a magic one then it must show contextuality. We want to study a similar state-injection scheme for qubits based on other sub-theories of SM compatible with QM instead of SQM.

computation which is composed by stabilizer circuits, which are known to be efficiently classically simulable [8], and the resource state. The classical part of the computation is boosted to quantum universality by injecting a particular resource state, called *magic* state. The result states that if the resource state can be distilled into a magic one then it shows a violation of CSW-inequality, *i.e.* contextuality. The result holds only for qudits (odd dimensions). Note that the contextuality considered in the above scenario is *state-dependent* (indeed it is injected by the magic state), while the one presented in the Mermin square scenario is state-independent. We conclude this section on contextuality by highlighting which are the philosophical assumption behind the notion of contextuality: counterfactual realism and counterfactual compatibility. The former says that results of unperformed tests have the same degree of reality of the results of performed tests. In this sense contextuality consists of a logical contradiction between an *actual* outcome that happens and a *potential* outcome that does not. Counterfactual compatibility roughly says that the common observable (*e.g.* $Y \otimes Y$) considered in the two incompatible contexts (*e.g.* last row and last column of the Mermin square) is exactly the same in the two scenarios. In other words the result of an unperformed test does not depend on the choice of compatible observables that can be performed with it [10].

3. – Restricting SM as a subtheory of QM

Spekkens' model is a hidden-variable model that is not operationally equivalent to quantum theory, indeed it does not show contextuality. Nevertheless we could consider sub-theories of SM that are compatible (thus showing the same statistics of measurements) with QM. A simple example of a sub-theory of SM which is compatible with QM is the following. Let us consider one qubit system where the allowed measurements([2]) are the Pauli X, Z and the allowed gates are H, X, Y, Z, where H is the Hadamard gate

([2]) Note that setting the allowed measurements also sets the allowed states. If the allowed measurements are the Pauli X, Z, the allowed states are the eigenstates of $\pm X, Z$.

and X, Y, Z are the Pauli unitary transformations. All the measurements/states have a faithful representation in SM and the unitaries are symplectic affine transformations. However if we add the Pauli Y measurement and the S gate, where $S = \left[\begin{smallmatrix} 1 & 0 \\ 0 & i \end{smallmatrix}\right]$ in the computational basis, we obtain a theory which is not faithfully represented in SM. The gate S is not symplectic and it would imply a different action in SM and QM. Our aim is to understand the mathematical prerequisites for sub-theories of SM to coincide with QM (even in the case of many qubits with entangling gates) and develop frameworks for that. An interesting idea is to treat these restricted versions of SM compatible with QM as the non-contextual classically simulable part of state-injection schemes. Inspired by Howard's result [14], we can then analyse the role of contextuality in qubit quantum computation by injecting resources that boost valid Spekkens' sub-theories of QM to universal quantum computation. We think that the above construction is just one possible application of Spekkens' toy model in quantum computation. Its correspondence with SQM could suggest other possible applications, like the ones related to SQM in non-prime dimensions [6].

* * *

This work was supported by EPSRC Centre for Doctoral Training in Delivering Quantum Technologies [EP/L015242/1].

REFERENCES

[1] EINSTEIN A., PODOLSKY B. and ROSEN N., *Can Quantum-Mechanical Description of Physical Reality be Considered Complete?*, *Phys. Rev.*, **47** (1935) 777.
[2] BELL J. S., *On the problem of hidden variables in quantum mechanics*, *Rev. Mod. Phys.*, **38** (1966) 447.
[3] KOCHEN S. and SPECKER E. P., *The problem of hidden variables in quantum mechanics*, *J. Math. Mech.*, **17** (1967) 59.
[4] SPEKKENS R. W., *Evidence for the epistemic view of quantum states: A toy theory*, *Phys. Rev. A*, **75** (2007) 032110.
[5] SPEKKENS R. W., *Quasi-quantization: classical statistical theories with an epistemic restriction*, arXiv:1409.5041 (2014).
[6] CATANI L. and BROWNE D. E., *Spekkens' toy model in all dimensions and its relationship with stabiliser quantum mechanics*, *New J. Phys.*, **19** (2017) 073035.
[7] SPEKKENS R. W., *Contextuality for preparations, transformations, and unsharp measurements*, *Phys. Rev. A*, **71** (2005) 052108.
[8] GOTTESMAN D., *The Heisenberg representation of quantum computers*, arXiv:quant-ph/9807006 (1998).
[9] MERMIN N. D., *Simple unified form for the major no-hidden-variables theorems*, *Phys. Rev. Lett.*, **65** (1990) 3373.
[10] PERES A., *Quantum theory: concepts and methods* (Kluwer Academic Publishers) 1995.
[11] CABELLO A., SEVERINI S. and WINTER A., *Graph-Theoretic Approach to Quantum Correlations*, *Phys. Rev. Lett.*, **112** (2014) 040401.
[12] ABRAMSKY S. and BRANDENBURGER A., *The sheaf-theoretic structure of non-locality and contextuality*, *New. J. Phys.*, **13** (2011) 113036.
[13] DE SILVA N., *Graph-theoretic strengths of contextuality*, *Phys. Rev. A*, **95** (2017) 032108.

[14] Howard M., Wallman J., Veitch V. and Emerson J., *Contextuality supplies the magic for quantum computation*, *Nature*, **510** (2014) 351.
[15] Gross D., *Hudson's theorem for finite-dimensional quantum systems*, *J. Math. Phys.*, **47** (2006) 122107.
[16] Pusey M. F., *Stabilizer notation for Spekkens' toy theory*, *Found. Phys.*, **42** (2012) 688.
[17] Coecke B., Edwards B. and Spekkens R. W., *Phase Groups and the Origin of Non-locality for Qubits*, *Electron. Notes Theor. Comput.*, **270** (2011) 29.
[18] Disilvestro L. and Markham D., *Quantum Protocols within Spekkens' Toy Model*, *Phys. Rev. A*, **95** (2017) 052324.
[19] Blasiak P., *Quantum cube: A toy model of a qubit*, *Phys. Lett. A*, **377** (2013) 847.
[20] Larsson J., *A contextual extension of Spekkens' toy model*, *AIP Conf. Proc.*, **1424** (2012) 211.

Proceedings of the International School of Physics "Enrico Fermi"
Course 197 "Foundations of Quantum Theory", edited by E. M. Rasel, W. P. Schleich and S. Wölk
(IOS, Amsterdam; SIF, Bologna) 2019
DOI 10.3254/978-1-61499-937-9-309

Casimir forces in spherically symmetric dielectric media

Yael Avni and Ulf Leonhardt

Department of Physics of Complex Systems
Weizmann Institute of Science - Rehovot 76100, Israel

Summary. — The most well-known manifestation of the Casimir effect is the attraction of two uncharged conducting plates. However, it turns out that Casimir forces are all around us: they originate from vacuum fluctuations of the electromagnetic field that excite dipoles in dielectric and conducting materials. These dipoles then interact with each other, generating measurable forces between macroscopic bodies: the Casimir force. A naive calculation of the Casimir force produces infinities, and though extensive work has been done in the field, there is still no universal prescription to renormalize the force. In this paper, we introduce the subject of Casimir forces and focus on the Casimir self-stress inside a homogeneous sphere. We discuss previous calculations and suggest an additional renormalization scheme that could solve the problem.

1. – Introduction

In 1948 Casimir predicted the attraction of two uncharged conducting plates [1] in a simple calculation, relying only on the zero point energy of the electromagnetic field,

$$E = \sum_{n=0}^{\infty} \frac{1}{2} \hbar \omega_n(a), \tag{1}$$

where a is the distance between the plates, and the summation is over the eigenmodes of the system. The fact that the energy depends on the distance a, indicates the existence of a force, as $\mathbf{F} = -\nabla E$. Casimir calculated this force (per unit area) to be,

$$\frac{F_c}{A} = -\frac{\hbar c \pi^2}{240 a^4} \,. \tag{2}$$

Almost half a century later, his result was verified to good accuracy in experiment [2].

Casimir's calculation, though elegant, does not provide full insight as to the nature of these forces. The picture becomes clearer once one realizes that Casimir's result is a limiting case of a much more general theory. Consider first a system of two neutral atoms. Thanks to the quantum vacuum, even at zero temperature electric and magnetic fields constantly fluctuate in space, excite the atoms and induce dipoles in them. The dipoles interact with one another and become correlated, giving rise to an attractive force between the atoms [3].

Later the theory was extended by Lifshitz *et al.* [4, 5] to apply to macroscopic bodies as well: fluctuations of the quantum vacuum induce fluctuating dipoles in dielectric materials, these dipoles interact with one another and generate measurable forces between macroscopic bodies. Casimir's result (eq. (2)) is recovered in the limit where the dielectrics become perfect conductors.

The spherical problem. – In 1953, Casimir suggested that vacuum fluctuations might cause a conducting spherical shell to contract [6], in a way analogous to the case of two conducting plates. The idea was originally directed to resolve the Lorentz model: a classical model for the structure of an electron, suggesting that it is a conducting, uniformly charged spherical shell. The problem with the model is that enormous electric repulsion will "break" the shell. To settle this, in 1905 Henri Poincaré postulated the existence of some "Poincaré stress", pointing towards the inside of the shell and stabilizing it.

Casimir pointed out that based on similarities to the two-plate case, the zero point energy of a conducting spherical shell should be of the form

$$E = \alpha \frac{\hbar c}{a} \,, \tag{3}$$

where a is the radius of the sphere and α is a pre-factor determined by the spherical geometry. Therefore, he thought, it could be that the Casimir stress is the mysterious Poincaré stress. One can verify that the condition is

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

which is equivalent to α being the fine-structure constant. This means that if the conjecture is true, the fine-structure constant can be derived from geometry, without any reference to physical constants of nature.

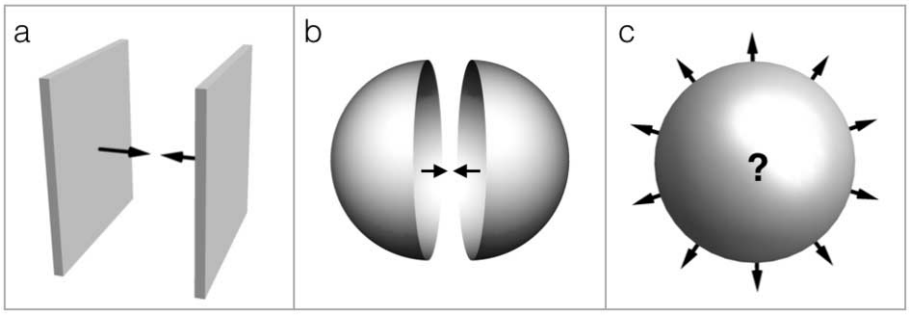


Fig. 1. – Due to Casimir forces (a) two conducting plates attract themselves, (b) two hemispheres attract each other. (c) Can the Casimir force inside a conducting spherical shell be repulsive?

Finally, in 1968 Boyer calculated the Casimir stress on a spherical conducting shell [7]. Although the magnitude of α was relatively close to the desired one, the sign was opposite: the Casimir stress was found to be repulsive.

Boyer's calculation, though impressive, lacked physical justification of its regularization techniques. Ten years later Julian Schwinger *et al.* obtained the same result by a seemingly more physical calculation [8]. Nevertheless, this result still raises questions. For example, in [9] it was shown that the Casimir force between two conducting hemispheres is attractive, regardless of the distance between them. Repulsion in a closed sphere suggests a counterintuitive discontinuity of the force (fig. 1). However, it should be kept in mind that in [9] the force is derived only from the interaction between the two hemispheres. It might be that due to Casimir stress each hemisphere by itself would expand, possibly solving the contradiction.

2. – Renormalization and Lifshitz theory

Naive calculation of the Casimir force usually produces infinity. That is due to the divergence of elements of the form: $\langle 0|\hat{\mathbf{E}}(\mathbf{r})^2|0\rangle, \langle 0|\hat{\mathbf{B}}(\mathbf{r})^2|0\rangle$ which appear in the expression for the energy and for the stress tensor. The Casimir force cannot actually be infinite, which means that we are including unphysical contributions and renormalization is required. So far there is no universal mathematically rigorous prescription to renormalize, which constitutes the main difficulty in Casimir force calculations (apart from numerical complexity).

The best candidate so far to renormalize the Casimir force in realistic materials is Lifshitz theory, where the force density is described by the electromagnetic stress tensor:

$$\mathbf{f} = \nabla \cdot \sigma, \tag{5}$$

$$\sigma = \langle 0|\hat{\mathbf{E}} \otimes \hat{\mathbf{D}}|0\rangle + \langle 0|\hat{\mathbf{B}} \otimes \hat{\mathbf{H}}|0\rangle - \frac{1}{2}\Big(\langle 0|\hat{\mathbf{E}} \cdot \hat{\mathbf{D}}|0\rangle + \langle 0|\hat{\mathbf{B}} \cdot \hat{\mathbf{H}}|0\rangle\Big) I_3, \tag{6}$$

I_3 being the identity 3×3 tensor. A "point split" is made in order to regularize the

Fig. 2. – The renormalization: Subtract the direct interaction between adjacent points (right) from the full interaction (left). Only the scattered waves that probe the geometry remain.

divergence:

$$\langle 0|\hat{\mathbf{E}}(\mathbf{x})\otimes\hat{\mathbf{E}}(\mathbf{x})|0\rangle \to \langle 0|\hat{\mathbf{E}}(\mathbf{x})\otimes\hat{\mathbf{E}}(\mathbf{x}')|0\rangle, \tag{7}$$

where $\mathbf{x}$ is a 4-vector, and the limit $\mathbf{x}\to\mathbf{x}'$ it taken only at the very end. The problem then translates into the problem of finding the fields at point $\mathbf{x}$ that result from a dipole at point $\mathbf{x}'$ (for a derivation see [10,11]). The interaction between the two points can be divided into direct interaction mediated by "outgoing waves", and indirect interaction which consists of waves that scatter from their surroundings [12]. In Lifshitz theory one subtracts the contribution of the direct self-interaction. It corresponds to a homogeneous space with no scattering, which from symmetry considerations, should not generate any Casimir force (fig. 2).

Lifshitz's approach and renormalization works well for piecewise homogeneous media. The general formula it produces for a 3-layer system described in fig. 3a at zero temperature is

$$\frac{F}{A} = -\frac{\hbar}{2\pi^2}\int_0^\infty \mathrm{Re}\int_0^\infty \mathrm{d}qqk_m^z\left(\frac{A_r^M A_l^M e^{2ik_m^z d}}{1-A_r^M A_l^M e^{2ik_m^z d}}+\frac{A_r^E A_l^E e^{2ik_m^z d}}{1-A_r^E A_l^E e^{2ik_m^z d}}\right)\mathrm{d}\omega, \tag{8}$$

where A^E and A^M are the reflection coefficients for the two polarizations (in the direction of the electric field and the magnetic field), defined by

$$A_{r,l}^E = \frac{\mu_{r,l}k_m^z - \mu_m k_{r,l}^z}{\mu_{r,l}k_m^z + \mu_m k_{r,l}^z}, \qquad A_{r,l}^M = \frac{\epsilon_{r,l}k_m^z - \epsilon_m k_{r,l}^z}{\epsilon_{r,l}k_m^z + \epsilon_m k_{r,l}^z} \tag{9}$$

and $k_i^z = \sqrt{\epsilon_i\mu_i\frac{\omega^2}{c^2}-q^2}$, $i=r,l,m$ (see fig. 3a).

Casimir's result (eq. (2)) is recovered by substituting the properties of a perfect mirror, which are $\epsilon_r=\epsilon_l\to-\infty, \epsilon_m=\mu_m=\mu_r=\mu_l=1$.

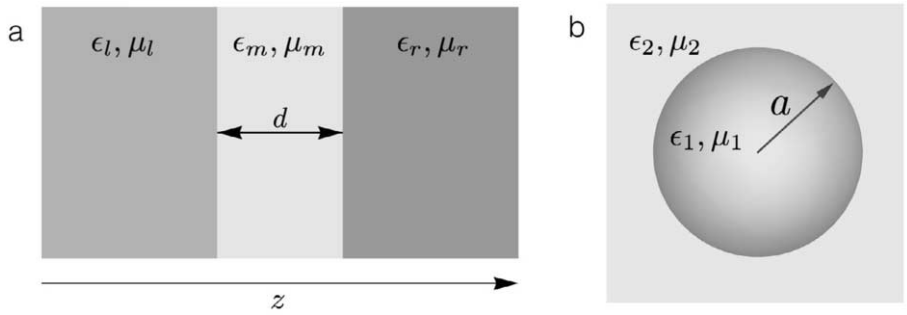


Fig. 3. – (a) Planar dielectric consisting of three homogeneous layers with different permittivities and permeabilities (ϵ and μ). l, r and m stand for left, right and middle. The outer layers are extended to infinity. (b) Dielectric sphere surrounded by dielectric background.

3. – Results

We considered a simple generalization of the spherical geometry: a homogeneous sphere surrounded by a homogeneous background, with different permittivities and permeabilities $\epsilon_{1,2}, \mu_{1,2}$, as shown in fig. 3b. Applying the machinery of Lifshitz theory we calculated the components of the stress tensor as a function of the radius, and found that they diverge as r and r' approach a, the radius of the sphere. However, the force density,

$$f_r = (\nabla \cdot \sigma)_r = \frac{1}{r^2}\partial_r \left(r^2 \sigma_r^r\right) - \frac{1}{r}\sigma_\theta^\theta - \frac{1}{r}\sigma_\phi^\phi, \tag{10}$$

was found to be zero everywhere except on the surface of the sphere. To calculate the total force on the surface of the sphere we integrated f_r over an infinitesimally thin shell:

$$F_r = 4\pi a^2 \left(\sigma_r^r|_{r=a^+} - \sigma_r^r|_{r=a^-}\right) - 8\pi \int_{a-\varepsilon}^{a+\varepsilon} \left(r\sigma_\theta^\theta(r)\mathrm{d}r\right), \tag{11}$$

where we used the relation $\sigma_\theta^\theta = \sigma_\phi^\phi$. We found that the force diverges for any choice of $\epsilon_{1,2}, \mu_{1,2}$. In previous calculations [8,13] the integral in eq. (11) was ignored, but that is not justified since σ_θ^θ diverges like ε^{-4} in the general case, and like ε^{-3} in the limit of a perfectly conducting spherical shell (the ε^{-3} divergence can disappear by taking the Cauchy principal value of the integral, but then a divergence of ε^{-2} remains). Therefore, although the known result of the self-stress in a conducting spherical shell might be real, it cannot be determined yet: an additional renormalization is needed.

Inspired by Lifshitz's renormalization in piecewise homogeneous media, we propose the following additional renormalization. Consider the interaction between the points $\mathbf{r}$ and $\mathbf{r}'$ surrounded by two mirror walls. It can be decomposed into direct interaction and multiple reflections (fig. 4a). The strength of the interaction diverges at two limits. One is the limit $\mathbf{r} \to \mathbf{r}'$, due to the direct interaction term which we already subtract (fig. 2). The second divergence appears when we take the limit of $\mathbf{r}$ and $\mathbf{r}'$ being infinitesimally

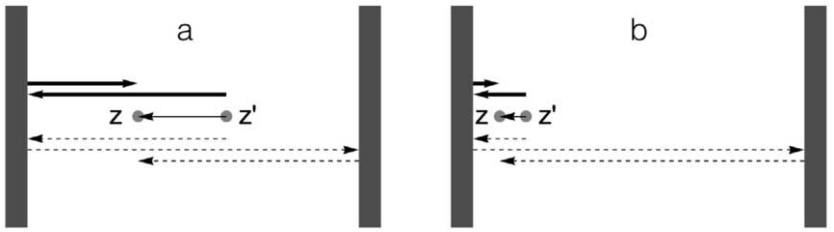


Fig. 4. – (a) First three orders of multiple reflections decomposition. Direct interaction (thin line), first reflection (thick line) and second reflection (dashed line). (b) z and z' are very close to the left wall. The direct interaction and first reflection give infinite, unphysical contributaions.

close to one of the walls. This divergence is caused by the first reflection term, describing a wave emitted from $\mathbf{r}$, hitting the wall one time and bouncing back to $\mathbf{r}'$. It essentially goes through zero distance and hence its strength diverges (fig. 4b). Since we are only interested in the stress on the walls, the first reflection, similar to the direct interaction term, is merely an artifact of the point splitting procedure which we need to exclude.

This source of divergence does not present a problem in the planar case, where the walls are flat: although σ_x^x and σ_y^y diverge, σ_z^z does not, and therefore the force in the z direction is finite. However, in a spherical geometry, the walls are curved. Therefore all the components of the stress tensor are affected by the first reflection, which causes the force to diverge. We expect that identification and subtraction of the first reflection in a sphere will make the renormalization complete, and reveal the true Casimir self-stress in a sphere.

REFERENCES

[1] CASIMIR H. B. G., *Proceedings of the KNAW*, **51** (1948) 793.
[2] LAMOREAUX S. K., *Phys. Rev. Lett.*, **78** (1997) 5.
[3] CASIMIR H. B. G. and POLDER D., *Phys. Rev.*, **73** (1948) 360.
[4] LIFSHITZ E. M., *Sov. Phys.*, **2** (1956) 73.
[5] DZYALOSHINSKII I. E., LIFSHITZ E. M. and PITAEVSKII L. P., *Adv. Phys.*, **10** (1961) 165.
[6] CASIMIR H. B. G., *Physica*, **19** (1953) 846.
[7] BOYER T. H., *Phys. Rev.*, **174** (1968) 1764.
[8] MILTON K. A., DERAAD L. L. and SCHWINGER J., *Ann. Phys. (N.Y.)*, **115** (1978) 388.
[9] KENNETH O. and KLICH I., *Phys. Rev. Lett.*, **97** (2006) 160401.
[10] SCHEEL S., *The Casimir stress in real materials*, in *Forces of the Quantum Vacuum*, edited by SIMPSON W. M. A. and LEONHARDT U. (Weizmann Institute of Science, Israel) 2015, pp. 107-138.
[11] LEONHARDT U., in *Essential Quantum Optics: From Quantum Measurements to Black Holes* (Cambridge University Press) 2010, pp. 231-249.
[12] GRINIASTY I. and LEONHARDT U., arXiv:1703.02211.
[13] MILTON K. A. and NG Y. J., *Phys. Rev. E*, **55** (1997) 4207.

Proceedings of the International School of Physics "Enrico Fermi"
Course 197 "Foundations of Quantum Theory", edited by E. M. Rasel, W. P. Schleich and S. Wölk
(IOS, Amsterdam; SIF, Bologna) 2019
DOI 10.3254/978-1-61499-937-9-315

The strange roles of proper time and mass in physics

Daniel M. Greenberger

City College (CCNY) of the City University (CUNY) of New York
New York, NY 10031, USA

Summary. — Mass is conventionally introduced into physical theories as a passive parameter, m_0. As such, it plays no dynamical role in the theory, nor can it change. But in practice, particles decay and recombine, changing their mass. They also acquire binding energies, changing their mass, and may also have an energy uncertainty, and so also a mass uncertainty. Similarly, the proper time of a particle is described along its trajectory. But quantum mechanically, trajectories can be split and recombined, or they may not be well-defined at all. So the proper time also has a dynamical role to play. We also show that there is a natural extension to the equivalence principle that is needed to include unstable particles. Both proper time and mass should be treated as quantum-mechanical operators, whose values are determined by measurement. The Hamiltonian formalism has a natural extension to include them as an extra coordinate and conjugate momentum, allowing one to construct both a classical and quantum theory of particles that can decay, have binding energies and obey the uncertainty principle.

1. – Mass: The role it plays, and the role it ought to play

Newton weirdly defined the mass of a body as the density times the volume, which set off a controversy that continues today. Mach worried about how an inertial system was defined, and how one could separate "fictitious" forces due to acceleration from real forces in an inertial system. His solution was that reference frames are defined relative

to each other, and settled on the "fixed stars", far away, as defining an inertial system. This question was settled by Einstein's Equivalence Principle, which we shall discuss in sect. **3**.

The property of the mass that we find upsetting is that it enters into both classical physics and relativity in a very passive way. It is a parameter that enters as a property of a body, but that plays no dynamical role in the interaction of bodies. Since there is no mechanism for affecting the mass, it is always conserved. There are no changing masses, as there is no way to change them. Even in "varying mass" problems, like accelerating rockets, or falling chains, the total mass is conserved. It merely moves from one body to another.

In relativity, interactions cause the energy to change, and since energy and mass are sort of equivalent, one would expect the mass to change also. But it does not. If two bodies interact, and form a bound state, they are held together by the binding energy. This should automatically change the mass of the state, but it does not. For example, if an electron and proton come near each other, they can combine into a hydrogen atom, whose total energy is less than that of the rest masses of the two free particles, and maybe give off a photon as they combine, to conserve energy. But the bound system has not changed its mass by virtue of binding. Of course the change can be put in by hand, and it is important, since if the H atom meets another particle, both its inertial mass and its gravitational mass will be that of the atom, not of the original particles. What bothers us here is that the same interaction that forms the bound state should automatically change the mass of the system for us, through the dynamics, via an equation of motion. But the mass is a purely passive actor, and nothing affects it.

Another problem that bothers us is when particles decay. If a particle can decay into two other particles, each of the particles is represented by a field, each of which has its own rest mass. The decay is represented by an interaction, whose role is to swallow the original particle, and replace it by the two daughters, each of which is created with its own mass. One can verify the mass of each by, for example, passing it through a slit and measuring the diffraction pattern. Each particle is created with its own mass, which we verify with the slits. But this description is inconsistent with quantum mechanics.

Actually, these particles are created in an entangled state. We do not know the mass of either one until one of them meets its slit and acquires a definite mass. To think of a particle as having a definite mass before meeting a slit is EPR thinking. It violates Bell's theorem. We say that two entangled particles do not each have a definite spin state until we measure it, but we think of them as having a definite mass state from the time they are created. There must be a nice Bell experiment that proves this to be wrong.

If the original particle had several decay modes instead of one, and one of several pairs of particles is created, and they travel through the Earth's gravitational field before meeting a slit, or something else to determine the mass, the mass of these decay modes is undetermined until the mass of a daughter particle is measured.

In fact, a free particle does not have a mass at all until it is measured. If the decay process uniquely identifies it, that is one thing. But if it does not, the mass is unknown until it is detected. So the mass is a quantum-mechanical operator, just like any observable.

And it is really inconsistent not to treat it as one. Even in the non-relativistic limit, this has consequences. One of the main purposes of these lectures is to point this out.

2. – Proper time: The role it plays, and the role it ought to play

In classical physics the time is a universal parameter, totally unaffected by the physics that takes place in a given reference frame. There is no use for the concept of proper time. In classical special relativity the proper time is kinematically determined, $\dot{\tau} = \sqrt{1 - v^2}$, (we will usually use the convention $c = 1$). One needs only a trajectory, and the speed, to determine it. There is no way to otherwise affect it by whatever physics is taking place. In general relativity, things are different. The proper time is also still determined by the geometry of space, but now there is extra flexibility.

In general relativity $\mathrm{d}\tau = \sqrt{g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu} \underset{NR}{\longrightarrow} \sqrt{(1 + 2\varphi - v^2)}\mathrm{d}t$, where NR means in the non-relativistic region, meaning here the weak field limit. So the passage of proper time is affected by the gravitational potential. One could immerse a local system in a constant potential, and its clock would run at a different rate from one far away. This means that one can not only re-set the clock, but also re-set the rate at which it runs.

In classical mechanics, one feature of an independent dynamical variable (q) is that one can arbitrarily set both the initial value q_0, and its initial rate of change $\dot{q}_0$, and the system will subsequently evolve via an equation of motion, independently of how the initial values came to be. So within general relativity, the flexibility exists to regard the proper time as an independent variable, whose history will be determined by a dynamical equation of motion, even though this freedom is not normally exercised. Here one has the capacity to turn the proper time into an independent degree of freedom.

But why should one want to do this? The best answer comes from quantum-mechanical considerations. Classical particles tend to have unique trajectories. But quantum-mechanically, particles can behave like waves, and if a particle approaches a beam splitter, it can be coherently split into two amplitudes. The two paths may have different histories, involving different proper times. For example look at the situation in fig. 1. Here a beam is split, and one of them goes through an E field to slow it up, and another to speed it up again to its original speed. The other beam remains at its same speed, but its path is longer, so the two beams can meet again at some time t and interfere at a second beam splitter. The two beams have accumulated two different proper times. So, what is the proper time of the recombined beam? Do their proper times interfere? What if the original particle was unstable. When will the recombined particle decay? Will it show proper time interference fringes in the decay pattern? So far as we know, issues like this, which one can theoretically guess the answer to, have never been decided experimentally.

The preceding example seems to show that the proper time is an observable, and should be treated like any other observable, and be represented by an operator, which can be measured along each path. The dynamical variable conjugate to the proper time is the mass, and promoting the mass and proper time to the status of independent variables has consequences, such as an uncertainty principle between them.

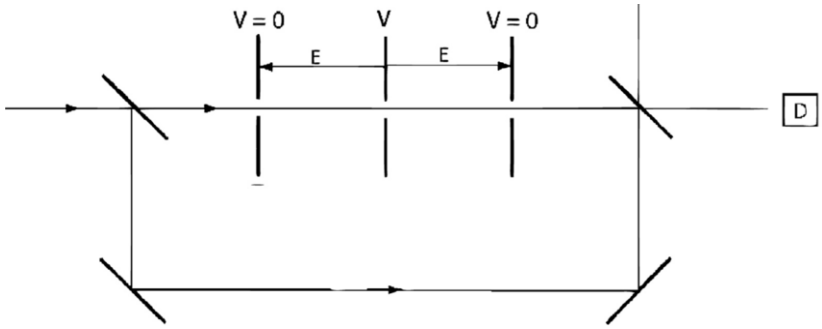


Fig. 1. – The Proper Time of a Superposition. Here a particle is split by an interferometer, and the two amplitudes take different proper times before recombining, which leads to questions about the subsequent behavior of the particle.

3. – The equivalence principle and the extended equivalence principle

3˙1. *Stable particles: The equivalence principle.* – Einstein, when he was looking for a way to extend special relativity, said he suddenly realized that a body in free fall would not feel a gravitational field. He later referred to that as "the happiest moment in my life". It implies that a body in free fall locally behaves like it is in an inertial reference frame. This solves a problem that has always plagued the Newtonian formulation, to which we alluded earlier. Namely, how can one experimentally tell whether one is really in an inertial system. If one is accelerating, one will feel "fictitious forces", like the "centripetal force", and the question arises as to how one tells this apart from "real forces", caused by pushes and pulls.

Einstein's answer was that being in free fall, when these fictitious forces vanish, is equivalent to being in an inertial system locally, and defines experimentally, as best as it can be defined, what an inertial system is. In such a system locally, all particles will move inertially, with no acceleration. If you drop a heavy and a light body, they will float away with no forces on them, and their masses will be irrelevant. Instead, if one is on a body accelerating with respect to this inertial system, all the objects floating along with the original inertial system will appear to be accelerating at the same rate, independently of their masses.

So to a body accelerating with respect to an inertial system, it looks just like there is a gravitational field present, that accelerates all bodies at the same rate, since that is the hallmark of a gravitational field. These observations then make up the Equivalence Principle, which states that being at rest in an accelerating system is equivalent to being at rest in a gravitational field. (This is usually called the "strong equivalence principle".) Of course this can only be true locally, meaning over a very small distance.

There is also what is usually called the "weak equivalence principle", which only concludes that all particles, regardless of their mass, fall with the same acceleration in an

external gravitational field. We like to introduce this principle by saying that in a system with non-gravitational external forces, one only needs one parameter to determine the motion of a body, its mass. That is a minor miracle, as nature could have made things much more complicated. But in the special case of a system with only gravitational external forces, a truly major miracle occurs. One does not need any parameters at all to describe the body. Locally, all bodies behave identically. The motion of the particle is completely determined by its environment. So its motion depends on the geometry of the space surrounding it. The particle itself gives no contribution locally to its motion, and the particle moves along a geodesic in this environment, along as straight a line as is possible, taking the shortest path.

As a special case, as we have pointed out, a free particle has no specific mass attached to it. All free particles act identically. This is an important consequence of the weak equivalence principle. This is why the mass must be a quantum-mechanical operator. It is not until it is measured (by a non-gravitational force) that it collapses into a definite quantum state, with a definite mass eigenvalue. This is the meaning of the statement that it is only when a non-gravitational force acts on the particle, that its mass becomes defined. This is true classically, for a point particle, and it defines the transition to quantum mechanics, but we shall see that quantum-mechanically, other considerations enter to complicate the matter.

To illustrate the equivalence principle, Einstein's most famous example was of an elevator suspended somewhere in space, with no masses around it. A person standing in the elevator, holding two masses in his outstretched arms, feels no forces. But if the elevator is suddenly accelerated upward, the person will feel a "fictitious" force, due to the acceleration, pulling him downward. Furthermore, if he simultaneously drops the two masses in his hands, they will both fall at the same rate, regardless of their masses. Of course, they are standing still, while he is being accelerated upward. Now if instead, the elevator was kept standing still, while a gravitational field was suddenly turned on, the person would feel exactly the same force as before, pulling him downward. Again, if he let go of the two masses, they would fall at the same rate. In this case it is due to a "real" gravitational force. According to the equivalence principle, he cannot tell the two situations apart locally.

One can analyze the two cases to see what is going on. In one case (fig. 2), the elevator is at rest, and there is a gravitational field present [1]. In the elevator, there are two observers, (1), who is at rest on the floor and who feels the gravitational field, and (ff), who at $t = 0$, lets go and enters into free fall. A third observer, (2), is at rest at the top of the elevator, and emits a photon (γ) at time $t = 0$, downward toward (1). If the photon has frequency ω_0, to observer (2) as he emits it, it will also have frequency ω_0 to observer (ff), who is just in free fall, and momentarily at rest with respect to (2). But because (ff) is in free fall, the frequency of the photon will remain ω_0 to him, throughout its flight. If L is the height of the elevator, it will take a time approximately $T \approx L/c$ to reach (ff), who is still close to the bottom of the elevator. But during this short time (ff) will have reached a velocity $v = gT$, with respect to (1), and to him as (1) is approaching

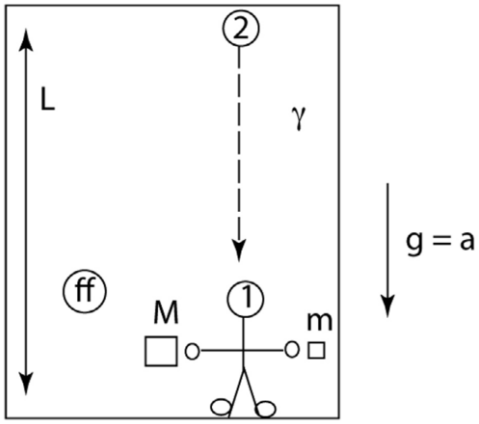


Fig. 2. – According to equivalence, in a gravitational field, observer (1) will see a higher frequency photon than observer (2), who emitted it.

the light beam, the frequency that (1) will receive will be Doppler-shifted by

$$\omega_1 = (1 + v/c)\omega_0 = (1 + gL/c^2)\omega_0 = [1 + (\varphi_2 - \varphi_1)/c^2]\omega_0, \tag{1}$$

with respect to (ff). To (1), who feels gravity, $(\varphi_2 - \varphi_1)$ is the potential difference between observers (2) and (1). Thus the photon at (1), who is at a lower potential, will appear to (ff) to be blue-shifted. Alternatively, one might say that his clocks are running slower, since more waves pass per second,

$$\dot{\tau}_1 = [1 + (\varphi_1 - \varphi_2)/c^2]\dot{\tau}_2. \tag{2}$$

The clock at (1) is running slower than that at (2), and so more ticks have taken place, during the passage of a given number of waves, so their frequency is higher. Clocks are affected by the gravitational potential difference between two points, even if they are at rest with respect to each other.

By the equivalence principle, this scenario is equivalent to the elevator being accelerated upward with acceleration g, in which case the red shift is caused by the Doppler Effect, rather than a gravitational potential difference, as a consequence of looking at the Equivalence Principle in the accelerated frame. If instead of the photon having been sent vertically, it had been sent horizontally, it would have landed at a lower height than it was emitted from. This shows that a photon's trajectory is bent in a gravitational field.

But the photon has no rest mass, only energy. So an even further result of the equivalence principle is that energy concentrations, rather than mass concentrations, are attracted in a gravitational field. Now energy is the fourth component of a 4-vector, and not a scalar like the mass. (The source also attracts with its energy, not its mass.) So in the non-relativistic limit, Newton's law of gravity behaves like the 00 component of a tensor, the energy-momentum tensor, rather than like a scalar.

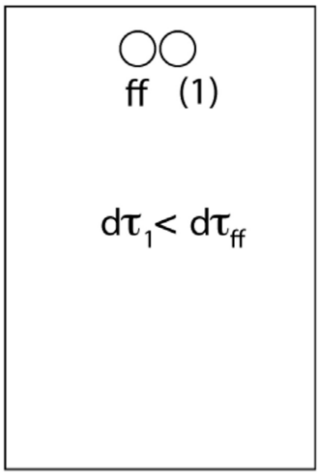


Fig. 3. – An unstable particle is not an inertial system. Compared with the free-fall observer, there is a τ-dependent potential present, even for an unstable particle at rest.

3˙2. *Unstable particles: The extended equivalence principle.* – So far, everything we have said assumes that we are talking about stable particles, and we have been comparing two points, (1) and (2), at different locations. But the equivalence principle can be extended. When a particle decays its mass changes, and $m = m(\tau)$. The proper time is the appropriate variable here, because we know experimentally that the lifetime, t_0, is different in coordinate systems where the particle is moving, and so we need the invariant lifetime, τ_0. We can construct the "Compton frequency" ω_c, $(\omega_c(\tau) = m(\tau)c^2/\hbar)$, to represent the mass of the particle, which in the case of an unstable particle will change with time. The Compton frequency is a scalar, and so is not a true frequency, which would be the fourth component of a 4-vector. But as the particle decays, the change in Compton frequency acts like a Doppler effect, and there will be an "effective" velocity. If we think of the "free fall" system as an inertial system, it will have a constant mass. Now a decaying particle will no longer represent an inertial system, but an accelerating system. The nature of this acceleration will immediately become apparent. Instead of a Doppler effect, if we think of the equivalent gravitational potential, we can write

$$\omega_c = (1 - \varphi(\tau)/c^2)\omega_{c0} \quad \text{or} \quad m = (1 - \varphi(\tau)/c^2)m_{c0}. \tag{3}$$

Here the subscript 0 refers to an inertial system, *i.e.*, one with constant mass.

The meaning of this equation is that we can think of an accelerated system not only in terms of a spatial acceleration, where the equivalent gravitational potential represents the potential difference between two different spatial points (pts. (1) and (2) in fig. 2), but also can represent the potential difference at a single point, at two different proper times (pt. (1) in fig. 3). In fig. 3, the point marked (ff) represents a point in "free-fall",

which in this case means the proper time at the same point (1) in the elevator, which is at rest in free space, but which at (ff) represents the proper time relative to a stable particle that might be sitting there. A stable particle represents an inertial system. A decaying particle at point (1), even though also at rest, is no longer an inertial system, but represents a system accelerating in proper time, rather than in space. The relation between the proper times of the decaying and stable particles is given by eq. (2), or equivalently,

$$\dot{\tau} \approx (1 + \varphi(\tau)/c^2)\dot{\tau}_0. \tag{4}$$

Equation (4) represents the proper time of the decaying system relative to that of the inertial, or free-fall system, to first order. In the free-fall system, $\dot{\tau}_0 = 1$. But the decaying system, even if it is located at the same point in space as the inertial system, will obey eq. (4).

In special relativity, one can construct a "light-clock" to determine the rate of a moving clock relative to a stationary one, but in that case, both clocks are on stable particles. In the case where the particle is unstable, we must resort to the phase of the wave function, which will be

$$e^{i\Phi} = e^{i(p\cdot r - Et)/\hbar} \to e^{im(\gamma v\cdot vt - \gamma t)/\hbar} = e^{-imt/\gamma\hbar} \to e^{-i\int m\mathrm{d}\tau/\hbar} = e^{-i\int L\mathrm{d}t/\hbar}. \tag{5}$$

Here, Φ represents the phase, not the gravitational potential. The phase is the action, controlled by the Lagrangian. The phase is an invariant, and as the mass changes, so will the rate of passage of proper time. Thus at the same point in 3-space, this rate of passage of time will be different for a stable particle and an unstable one. Specifically, we will have

$$\int L\mathrm{d}t = \int m\mathrm{d}\tau = m_0 \int \mathrm{d}t = m_0 \int \frac{\mathrm{d}\tau}{(1 + \varphi/c^2)}\,. \tag{6}$$

(from eq. (3)). While in the free-fall system $\dot{\tau}_0 = 1$, in a decaying system at rest with respect to the free-fall system, eq. (3) gives us

$$\mathrm{d}\tau = (1 + \varphi/c^2)\mathrm{d}t, \tag{7}$$

and in general, in a decaying system moving with speed v, we would have

$$\begin{aligned} \mathrm{d}\tau^2 &= (1 + \varphi(\tau))^2\mathrm{d}t^2 - \mathrm{d}x^2/c^2, \\ \mathrm{d}\tau &= \sqrt{(1 + \varphi(\tau))^2 - v^2/c^2}\mathrm{d}t. \end{aligned} \tag{8}$$

Equivalently,

$$m_0 = m(1 + \varphi/c^2). \tag{9}$$

Equation (8) represents an extension of the geometry of special relativity (and also of general relativity, since now $g_{\mu\nu} = f(\tau)$), however when we place the theory in the context of a Hamiltonian formulation, it will follow as an equation of motion.

Equation (8), which says that $\dot{\tau}$ can be a $f(\tau)$, implies that τ can be a time-varying $f(t)$, which makes the situation for a decaying particle very different from that for a stable one. So one of the unexpected predictions of equivalence, just like the one that a gravitational field will bend a photon, is that a clock sitting on an unstable particle will run at a different rate from that of a stable particle sitting at the very same location.

$$\begin{aligned} &\dot{\tau} = f(x_\mu), && \text{for a stable particle,} \\ &\dot{\tau} = f(x_\mu;\tau), && \text{for a decaying particle.} \end{aligned} \tag{10}$$

(We note that according to eq. (10), for a decaying particle $\dot{\tau}$ changes as a $f(\tau)$. This is not to be confused with the fact that $\dot{\tau}$ is not the same for all systems in free-fall. Different inertial systems have their own rate of $\dot{\tau}$. For example if two inertial systems are moving with respect to each other, then if system (1) is at rest with respect to coordinate time t, then $\dot{\tau} = 1$. But then if one is at rest with respect to a system (2), moving at speed v with respect to system (1), then $\dot{\tau}_2 = \sqrt{1 - v^2}$. Similarly, if a system (3) is at rest with respect to system (1), but it has a constant gravitational potential φ_0, then $\dot{\tau}_3 = 1 + \varphi_0$.)

The quantity $\mathrm{d}\tau = \sqrt{g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu}$ plays a special role in determining the geometry of space-time. In general relativity the $g_{\mu\nu}$ determine the curvature of space and the motion of bodies embedded in it and they are functions of the coordinates x^μ. So although the interaction of particles is determined by the non-gravitational forces between them, the very presence of gravitational forces affects their motion through the geometrical effect they generate on space-time itself, through the $g_{\mu\nu}$.

But by now allowing the $g_{\mu\nu}$ to also depend directly on τ, the $g_{\mu\nu}$ are also affected by the stability of the masses of particles imbedded in space-time, and through this dependence they directly control the rates of decay of these particles. The interactions between particles that make them decay are non-gravitational, but the execution of the decay is affected by the τ-dependent potentials acting on them. This is a completely new role for gravity. It is a dynamical property of geometry and is necessarily quantum-mechanical, as m and τ become conjugate variables, which implies, among other things, an uncertainty relation between them.

Whereas gravitation has traditionally been thought of as the very weak sister of the other forces, so that it can be ignored in the formulation of most of the laws of nature, it turns out that to the contrary, it directly affects how the other forces carry out their roles, and change masses, and it can never be ignored as soon as one concedes that particles are not always stable, and can decay through interactions. This is a completely different role for gravity than merely the conventional one of causing spacetime to curve.

Our belief is that the hope that one could ever correctly and smoothly combine gravity and quantum physics together into a general theory governing all of physics, will remain

unfulfilled so long as one ignores the dynamical interdependence of mass and proper time (which includes the τ-dependence of the $g_{\mu\nu}$).

4. – Incorporating mass and proper time as dynamical variables

We have stated that mass and proper time should be considered as conjugate dynamical variables. The next question is can they be incorporated into the formalism of mechanics? It turns out that the standard Hamiltonian formalism is almost begging to be generalized to include this modification [2]. First let us consider a free particle. The standard variables are x and p as the coordinate and its conjugate momentum. Now they are to be supplemented by τ and m (really mc^2, for dimensional reasons). Before we begin, we should again point out that if mass is to be considered a dynamical variable, then the mass must be considered to be the energy in the rest frame of the system. So if the energy is uncertain, then the mass will also be. Thus the mass, m_0 that appears in, say the Klein-Gordon equation for the particle, or in the Dirac equation, is the "nominal" mass. The actual mass may be quite different. If the mass is experimentally determined, it may well have a Δm associated with it.

Similarly, the proper time is the time as measured in the particle's free fall system. It can also be considered in its rest system, but if there is a gravitational potential present there, it will include that. If the particle is unstable, that includes its τ-dependent gravitational potential.

So the Hamiltonian should be $H(x, p; \tau, m)$. The standard equations of motion are

$$\dot{x} = \frac{\partial H}{\partial p}\,, \quad \dot{p} = -\frac{\partial H}{\partial x}\,. \tag{11}$$

These are to be supplemented by

$$\dot{\tau} = \frac{\partial H}{\partial m}\,, \quad \dot{m} = -\frac{\partial H}{\partial \tau}\,. \tag{12}$$

For a free particle, we use the standard Hamiltonian,

$$H = \sqrt{p^2 + m^2}. \tag{13}$$

There is no x-dependence here, and so there are no forces, and p will be constant. The equation for $\dot{x} = v$ will be, from eq. (11),

$$\dot{x} = v = \frac{p}{\sqrt{p^2 + m^2}}\,, \quad p = \frac{mv}{\sqrt{1 - v^2}} = mv\gamma. \tag{14}$$

where $\gamma = 1/\sqrt{1 - v^2}$. The second equation here comes from solving the first equation for p. The equation for $\dot{\tau}$ is, from eq. (12),

$$\dot{\tau} = \frac{m}{\sqrt{p^2 + m^2}} = \frac{m}{m\gamma} = \sqrt{1 - v^2}. \tag{15}$$

And so, instead of being a geometrical necessity, the equation for $\dot{\tau}$ follows dynamically, as an equation of motion. Since there is no τ dependence, m will be a constant, from eq. (12). But, just as a potential that depends on x destroys the homogeneity of space, and produces a force that changes p, so too would a potential dependent upon τ produce a force that would change m, from eq. (12). Thus we see that we have here a dynamics where m does not have to be a constant, and if an atom, say, were to be excited above its ground state, then the mass could change to accomodate this. Also, one could have a classical theory of decaying particles, something that could not happen in ordinary classical dynamics.

The next simplest case would be an external potential acting on the particle. To be consistent with our discussion of the equivalence principle, as a gravitational potential it would have to act on the energy, rather than the mass, so we take as our Hamiltonian,

$$H = \sqrt{p^2 + m^2}(1 + \varphi(\tau)). \tag{16}$$

This example shows a particle at rest, or moving at constant speed (since there is no x-dependence), decaying into the void. Of course, in a more realistic example, it would be decaying into something, and the problem would need more structure. but the problem is instructive in giving insight into the nature of the theory. Of course, φ could be a function of x, instead of τ, but then it would just be a standard dynamics problem, where m is constant, and the force is a function of x.

Actually, we shall take the simplest case, namely, a particle at rest decaying. For the momentum to remain zero, the decay would have to be uniform in all directions, so one can think of the particle spherically decaying. Then the Hamiltonian simplifies to

$$H = m(1 + \varphi(\tau)). \tag{17}$$

Since m and τ are dynamical variables, and H does not explicitly depend on t, the coordinate time, then H will be a constant in time. So we write

$$H = m(1 + \varphi(\tau)) = m_0, \quad m = \frac{m_0}{1 + \varphi(\tau)}\,. \tag{18}$$

One assumes that one knows $m(\tau)$, and one solves for $\varphi(\tau)$, and $\tau(t)$. So one discovers that even though the particle is at rest, $\tau \neq t$. There is a "decay red shift", as predicted by the extended equivalence principle.

As a simple example, assume that

$$m = m_0 + \Delta e^{-\alpha\tau}. \tag{19}$$

For example this could refer to an atom in an excited state decaying down to the ground state. It could also refer to a particle giving off massive particles, like a nucleus going

into a lighter one. In this case

$$m_0 + \Delta e^{-\alpha\tau} = \frac{m_0}{1+\varphi(\tau)}\,, \tag{20}$$

$$\varphi(\tau) = \frac{m_0}{m_0 + \Delta e^{-\alpha\tau}} - 1 = -\frac{(\Delta/m_0)e^{-\alpha\tau}}{1+(\Delta/m_0)e^{-\alpha\tau}}\,.$$

From eq. (12) we have

$$\dot{\tau} = 1 + \varphi(\tau), \quad t = \tau + \frac{\Delta}{m_0\alpha}(1 - e^{-\alpha\tau}). \tag{21}$$

Here we set $\tau = 0$ at $t = 0$. And we see that as

$$t \to \infty, \quad t \to \tau + \frac{\Delta}{m_0\alpha}\,. \tag{22}$$

So the potential creates a "force" that makes the particle decay, which also alters the rate at which proper time runs. If instead of being at rest, the particle was traveling at constant speed v, one can make a Lorentz transformation in order to bring it to rest. However because of the decay red shift, the proper time is no longer the time in the rest system of the particle, so one must make an "extended Lorentz Transformation", which is dictated by the extended Equivalence Principle.

5. – An extended Lorentz transformation and Schrödinger equation

In our example of a decaying particle above, the proper time, according to the extended equivalence principle, is given by eq. (8), where φ is determined by the external field, and may be a $f(x^\mu, \tau)$. The Hamiltonian for a particle in an external field will be given by eq. (16), and we see that this is consistent with the extended equivalence principle:

$$H = \sqrt{p^2 + m^2}(1+\varphi), \tag{23}$$

$$v = \frac{\partial H}{\partial p} = \frac{p(1+\varphi)}{\sqrt{p^2+m^2}}\,, \quad p = \frac{mv}{\sqrt{(1+\varphi)^2 - v^2}}\,,$$

$$\dot{\tau} = \frac{\partial H}{\partial m} = \frac{m(1+\varphi)}{\sqrt{p^2+m^2}} = \sqrt{(1+\varphi)^2 - v^2}$$

The equation for p, and for $\dot{\tau}$ come from rearranging the equation for v. So this form of the Hamiltonian is consistent with the extended equivalence principle. Gravity is coupled to the energy, rather than the mass, and the formula for $\dot{\tau}$ agrees with that of eq. (8).

The "extended" Lorentz transformation comes from using the invariance of the proper time. In other words, writing

$$\begin{aligned} \mathrm{d}x' &= a(\mathrm{d}x - v\mathrm{d}t), \\ \mathrm{d}t' &= b(\mathrm{d}t - e\mathrm{d}x), \end{aligned} \tag{24}$$

and solving for the coefficients a, b, e, by using the invariance of the proper time, as in the usual derivation,

$$\mathrm{d}\tau^2 = (1+\varphi(\tau))^2\mathrm{d}t'^2 - \mathrm{d}x'^2 = (1+\varphi(\tau))^2\mathrm{d}t^2 - \mathrm{d}x^2. \tag{25}$$

This yields the result

$$\begin{aligned} \mathrm{d}x' &= \gamma(\mathrm{d}x - v\mathrm{d}t), \\ \mathrm{d}t' &= \gamma(\mathrm{d}t - (v/\kappa^2)\mathrm{d}x), \\ \gamma &= \frac{\kappa}{\sqrt{\kappa^2 - v^2}}, \quad \kappa = (1+\varphi(\tau)). \end{aligned} \tag{26}$$

What we have done is introduce the extended gravitational Lorentz transformation, which is for a vector gravitational theory. Presumably in a complete theory, we would need a tensor theory. But in this simpler theory, all the main phenomena are present.

We can write a theory of variable mass particles, that can decay and change to accomodate binding energy changes. Classically there is a "decay red shift", so that a clock sitting on a decaying particle runs at a different rate than one on a stable particle at the same point and at the same time. And in a quantized theory there will be an uncertainty principle between the mass and proper time on a particle.

What would the quantum equivalent of our example of a decaying particle be? If we adopt eq. (16) for our quantum Hamiltonian, it will reduce to eq. (17) for the quantum case, with $m_{op} = (\hbar/i)\partial/\partial\tau$. Then the Schrödinger decay equation will be

$$(1+\varphi)\frac{\hbar}{i}\frac{\partial\psi}{\partial\tau} = i\hbar\frac{\partial\psi}{\partial t}\,. \tag{27}$$

or, if we introduce the variable z,

$$\mathrm{d}z = \frac{\mathrm{d}\tau}{1+\varphi(\tau)}\,, \quad z = \int \frac{\mathrm{d}\tau}{1+\varphi(\tau)}\,. \tag{28}$$

the equation and solution becomes

$$\frac{\partial\psi}{\partial z} = -\frac{\partial\psi}{\partial t}\,, \quad \psi = \psi_0(z-t). \tag{29}$$

This is the same as the classical solution of $\dot{\tau} = (1+\varphi)$, only now it says that ψ is a wave packet centered on the classical solution. There is a slight complication here, in

that eq. (27) is not Hermitian, but this will affect only the normalization of the solution. To see this, we symmetrize the problem,

$$\frac{\hbar}{i}\left(\frac{\partial\psi}{\partial\tau}+\frac{1}{2}\left(\varphi\frac{\partial\psi}{\partial\tau}+\frac{\partial}{\partial\tau}(\varphi\psi)\right)\right)=i\hbar\frac{\partial\psi}{\partial t}\,, \tag{30}$$

$$(1+\varphi)\frac{\partial\psi}{\partial\tau}+\frac{1}{2}\frac{\partial\varphi}{\partial\tau}\psi=-\frac{\partial\psi}{\partial t}\,.$$

The solution to this equation is just

$$\psi(z,t)=\frac{1}{\sqrt{1+\varphi(\tau)}}\psi_0(z-t), \tag{31}$$

as promised, where ψ_0 is the previous solution. If the particle is moving at a constant velocity v, instead of being at rest, then one can solve eq. (16) instead, by first making a Lorentz transformation on the system.

REFERENCES

[1] This argument is given in every text on gravity. In particular, see B. Shutz, in *Gravity, From the Ground Up* (Cambridge U. P.) 2003; or J. B. Hartle, *Gravity* (Addison Wesley) 2003.

[2] This formalism has been published piecemeal over the years. The key references are, all with GREENBERGER D. M., or GREENBERGER DANIEL M., as author:

"The Role of Equivalence in Quantum Mechanics", *Ann. Phys. (N. Y.)*, **47** (1968) 116;

"Theory of Particles with Variable Mass, I-Formalism", *J. Math. Phys.*, **11** (1970) 2329;

"Theory of Particles with Variable Mass, II-Some Physical Consequences", *J. Math. Phys.*, **11** (1970) 2341;

"Some Useful Properties of a Theory of Variable-Mass Particles", *J. Math. Phys.*, **15** (1974) 395;

"Wave Packets for Particles of Indefinite Mass", *J. Math. Phys.*, **15** (1974) 406;

"Some Remarks on the Extended Galilean Transformation", *Am. J. Phys.*, **47** (1979) 35;

"Comment on Non-Spreading Wave Packets", *Am. J. Phys.*, **48** (1980) 256;

"What is the Proper Role of Proper Time", article in *Physics as Natural Philosophy*, edited by SHIMONY A. and FESHBACH H. (MIT, Cambridge, Mass.) 1982, pp. 179–203;

"The Proper Role of Proper Time", in *Experimental Metaphysics: Quantum Mechanical Studies for Abner Shimony, Vol. I*, edited by COHEN R. S., HORNE M. and STACHEL J. (Kluwer, Dordrecht) 1997;

"The Inadequacy of the Galilean Transformation in Quantum Mechanics", *Phys. Rev. Lett.*, **87** (2001) 100405-1;

"The Disconnect Between Quantum Mechanics and Gravity", quant-ph. ArXiv: 1011.3719 (2010);

"Conceptual Problems Related to Time and Mass in Quantum Theory", Quant-ph. ArXiv: 1011.3709 (2010).

Proceedings of the International School of Physics "Enrico Fermi"
Course 197 "Foundations of Quantum Theory", edited by E. M. Rasel, W. P. Schleich and S. Wölk
(IOS, Amsterdam; SIF, Bologna) 2019
DOI 10.3254/978-1-61499-937-9-329

Some consequences of mass and proper time as dynamical variables

Daniel M. Greenberger

City College (CCNY) of the City University (CUNY) of New York
New York, NY 10031, USA

Summary. — We examine here some of the effects that are produced by considering mass and proper time as dynamical variables. First we consider Galilean invariance and point out that the Bargmann Theorem that masses cannot be superimposed in non-relativistic (NR) quantum theory is no longer valid. We also point out that while Galilean invariance is a consistent requirement of the NR Schrödinger equation as such, it provides a poor description of the NR limit of Lorentz invariance, as the proper time leaves a residue that is independent of c in this limit. Next we show that there is an inevitable uncertainty relation between mass and proper time and give several examples. Finally, we show that the classical limit is different for non-gravitational forces, and for gravitational forces that lead to the equivalence principle.

1. – The Lorentz transformation and the Galilean transformation

1˙1. *The Bargmann theorem in non-relativistic physics*. – The Lorentz Transformation (LT) is the boost of a system from its rest frame, to a frame moving with velocity v, using relativistic dynamics. The Galilean Transformation (GT) is also the boost of a system from its rest frame to a frame moving with velocity v, but using Newtonian dynamics. One would think that in the limit of slow-moving systems, where $v \ll c$, the

LT would smoothly flow into the GT, and in classical physics this is the case. But in quantum mechanics, there are residual phase effects that persist in the low velocity limit, yet they cannot be explained in the Newtonian limit. By "persist", we mean that they are present and independent of c, and have experimentally observable consequences. Now the Schrödinger equation is consistent with Newtonian physics, and the GT is a legitimate operation there. But the Schrödinger equation is also a legitimate limit of the Klein-Gordon equation, or the Dirac equation, in the non-relativistic limit, and one would expect a consistency in this limit between the LT and the GT.

But such is not the case, and the problem shows up most clearly in a well-known theorem due to Bargmann that says one cannot superimpose particles of different mass in non-relativistic quantum mechanics [1]. Of course, relativistically one can certainly superimpose particles of different mass, and in the non-relativistic limit, one can have a superposition between an atom in an excited state and in its ground state. Non-relativistically one says that the superposition is between two energy states, but if one believes relativity, it is also between two different mass states.

If one has a superposition between two mass states, say

$$\psi = \psi_1(r,t,m_1) + \psi_2(r,t,m_2), \tag{1}$$

in an inertial system at rest, then one can also describe this wave function in a reference frame moving at constant velocity. What Bargmann did was to make a series of four transformations into first, a frame translated by a distance a, then into a frame moving with velocity v, then a frame frame translated back by a distance $-a$, and finally into one moving at velocity $-v$. After this series of transformations, according to the GT, one has landed back where one started, at the initial point r, at the same time t, as the transformations commute with each other.

But there is a complication here. For the GT, besides carrying one into the boosted frame, also has an extra phase factor tagging along with it. This is allowed in Wigner's original discussion of symmetry operations, because while probability distributions are unaffected by unitary transformations, wave functions can get extra phases, that cancel out when one takes $|\psi|^2$. It is interesting that one does not need such a phase factor for the LT, or for 3-d rotations, but it is required to make the GT work properly. While this extra phase factor appears rather mysterious in the formalism, we shall see that it has a simple but important physical interpretation. We shall explain that soon, but for now it suffices to say that this extra phase factor that tags along in the GT is proportional to m, the mass of the system. The upshot is that when one makes the Bargmann series of transformations that end up back in the original system, at the original point, nonetheless there is an extra phase factor that tags along with the product of the transformations, and this phase factor is proportional to m, namely $e^{i\frac{m}{h}\alpha(v,a)}$. This has the effect of transforming eq. (1) into

$$\psi = e^{im_1\alpha/h}[\psi_1(r,t,m_1) + e^{i(m_2-m_1)\alpha/h}\psi_2(r,t,m_1)]. \tag{2}$$

So by merely rewriting a wave function in a different reference frame (actually the same frame after making a transformation equivalent to the unit transformation), one has picked up a relative phase factor between the terms, one that can actually be measured in an interference experiment. This would seem to be intolerable, and Bargmann's solution was to impose a superselection rule on the system, that one cannot superimpose particles of different mass. Furthermore this rule seems perfectly compatible with the fact that mass is preserved in Newtonian physics.

1˙2. *The problem with the Bargmann theorem.* – While the Bargmann theorem may at first glance seem innocuous enough, nonetheless, it ought to set off alarm bells, for a number of reasons. First, as we have mentioned, the theorem is certainly not true in relativistic physics, and even in the non-relativistic limit it ought to be true that $E = mc^2$. But also, how can a restriction that is not true for a larger symmetry group (the LT) be true for a more restricted symmetry group (the GT)? Usually things are the other way around. For example if you have the full rotational symmetry then angular momentum will be conserved. But if you only have cylindrical symmetry, then only L_z will be conserved. How can the limited GT imply a greater symmetry restriction on the system than the full symmetry LT? It does not seem to make sense. Would not it make more sense if instead, the extra phase coherence that appears in eq. (2) could be explained, rather than being outlawed by fiat, because it is inconvenient? And of course it can be explained, in a way that sheds some light on the GT.

In the Bargmann proof, one makes a series of transformations, ending up with the unit transformation. One can better make sense of this by using, not the GT directly, but a more general transformation, the Extended Galilean Transformation (EGT), which transforms a system according to

$$x \to x' = x - \xi(t), \qquad t \to t' = t. \tag{3}$$

Here ξ is an arbitrary $f(t)$, so the transformation is into a rigid accelerated system, rather than merely being restricted to one moving at constant velocity, so it includes the GT as a special case. It is non-relativistic, since t does not change, but it is general enough to incorporate rigid accelerations, and thus to discuss the equivalence principle in the non-relativistic limit.

In order to describe a wave function that is a superposition of different masses, eq. (1), we will have to look directly at the Schrödinger equation. Then under the transformation (3), we will have

$$\frac{\partial}{\partial x} = \frac{\partial}{\partial x'}\,, \qquad \frac{\partial^2}{\partial x^2} = \frac{\partial^2}{\partial x'^2}\,, \qquad \frac{\partial}{\partial t} = \frac{\partial}{\partial t'} - \dot{\xi}\frac{\partial}{\partial x'}\,, \tag{4}$$

and the free particle Schrödinger equation will become

$$-\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x'^2} = i\hbar\left(\frac{\partial\psi}{\partial t'} - \dot{\xi}\frac{\partial\psi}{\partial x'}\right). \tag{5}$$

Here we are assuming the standard Schrödinger equation, where m and τ are not considered to be dynamical variables. The problem with this equation is the last term has a momentum dependent potential, which is not what we would want in an accelerated system, but rather a static potential, so that we can say that a transformation to an accelerated system is equivalent to being in a gravitational field. We can solve this problem by making a unitary transformation,

$$\psi(x,t) = e^{if(x',t')}\varphi(x',t'). \tag{6}$$

The function f is uniquely determined by the requirements that we want the term in $\frac{\partial\varphi}{\partial x'}$ to vanish, and we also want the extra purely time dependent terms, that appear in the potential after the transformation, to vanish. This gives the result,

$$\begin{aligned} f(x',t) &= \frac{m}{\hbar}\left(\dot{\xi}x' + \frac{1}{2}\int \dot{\xi}^2 \mathrm{d}t\right), \\ \psi(x,t) &= e^{if(x',t)}\varphi(x',t), \\ -\frac{\hbar^2}{2m}\frac{\partial^2\varphi}{\partial x'^2} &+ m\ddot{\xi}x'\varphi = i\hbar\frac{\partial\varphi}{\partial t}\,. \end{aligned} \tag{7}$$

Here we dropped the prime on the t. Equation (7) for $\varphi(x',t)$ is the equation for the wave function in the accelerated system, and it contains a gravitational term in the potential, $V(x') = m\ddot{\xi}x'$, as one would expect from the equivalence principle.

So the extended Galilean transformation describes the equivalence principle, but it also says more, as we shall see when we analyze the phase term e^{if}.

In the special case of a uniform velocity transformation, the extended Galilean Transformation reduces to the usual GT. In that case we have

$$\begin{aligned} f &\xrightarrow[GT]{} \frac{m}{\hbar}(x'v + \frac{1}{2}v^2 t), \qquad \xi \to vt; \\ \psi(x,t) &= e^{im\alpha(x',\,t)/\hbar}\varphi(x',t). \end{aligned} \tag{8}$$

This is the mysterious phase factor $\alpha(= f)$ in the GT, and we see that it is proportional to m, as we promised.

To find the meaning of the phase change, first we calculate it for the series of four transformations in Bargmann's calculation, performed non-relativistically with the GT. Here we have

$$r_1 = r - a, \qquad r_2 = r_1 - vt, \qquad r_3 = r_2 + a, \qquad r_4 = r_3 + vt = r. \tag{9}$$

So the fourth transformation brings us back to r. For the wave functions, the translations follow the simple rule:

$$\psi_1(r_1) = \psi(r), \qquad \psi_3(r_3) = \psi_2(r_2), \tag{10}$$

while the boosts follow eq. (6), with f given by eq. (8),

$$\varphi'(r') = e^{-i\frac{m}{\hbar}(\pm vr' + \frac{1}{2}v^2 t)}\psi(r), \tag{11}$$
$$\varphi_4(r_4) = e^{-i\frac{m}{\hbar}(-vr_4 + \frac{1}{2}v^2 t)}\psi(r_3),$$
$$\varphi_2(r_2) = e^{-i\frac{m}{\hbar}(+vr_2 + \frac{1}{2}v^2 t)}\psi(r_1).$$

When we combine eqs. (10) and (11) with eq. (9), we finally get

$$\psi_4(r_4) = e^{i\frac{m}{\hbar}va}\psi(r). \tag{12}$$

If these four transformations had been made relativistically instead, we would have had

$$r_1 = r - a, \qquad t_1 = t; \qquad r_2 = \gamma(r_1 - vt_1), \qquad t_2 = \gamma(t_1 - vr_1); \tag{13}$$
$$r_3 = r_2 + a/\gamma, \qquad t_3 = t_2; \qquad r_4 = \gamma(r_3 + vt_3), \qquad t_4 = \gamma(t_3 + vr_3);$$

where $\gamma = 1/\sqrt{1 - v^2}$. Note that in r_3, one has to translate back by a/γ, because of the Lorentz contraction, in order to land back at the original point. This gives

$$r_4 = r, \qquad t_4 = t + va, \qquad \Delta t = t_4 - t = va. \tag{14}$$

So when the four transformations are finished, they land back at the original point, but *not* at the same time as they started. When one makes a closed loop of transformations in space, they do not form a closed loop in time. The additional time is nothing but the twin-paradox effect. The phase factor in eq. (2) is $e^{-i\Delta mc^2\Delta\tau/\hbar}$, and it agrees with the phase factor in eq. (12) obtained from the GT. So the mysterious phase factor that enters into the GT is nothing but the non-relativistic residue of the twin-paradox, which is independent of c, and so must show up, even in a non-relativistic calculation. It is a true physical effect, and the phase difference is meaningful. The problem is that the GT does not recognize the concept of proper time, and so the phase factor becomes an embarrassment, and must be expunged by fiat, which is the reasoning behind the Bargmann theorem. It is consistent within Newtonian theory, but experimentally it is untenable, and shows why the GT is not the non-relativistic limit of the LT. Another lesson to be learned by this is that there are proper time residue effects that show up in the non-relativistic limit, and are independent of c. They are correctly explained by the extended GT, which is consistent with the equivalence principle.

An interesting insight comes from considering m and τ as quantum operators [2]. Note that, just as when p is an operator, $P_{op} = \frac{\hbar}{i}\frac{\partial}{\partial x}$, so that

$$e^{iap/\hbar}\psi(x) = \psi(x + a), \tag{15}$$

then it will also be true when m is an operator, $m_{op} = \frac{\hbar}{i}\frac{\partial}{\partial\tau}$, that

$$e^{iam_{op}}\psi(\tau) = \psi(\tau + a), \tag{16}$$

and so eq. (7) becomes

$$\begin{aligned} e^{i(\dot{\xi}x+\frac{1}{2}\int\dot{\xi}^2\mathrm{d}t)\frac{m_{op}}{\hbar}}\psi(x,\tau;t) &= \psi(x,[\tau+(\dot{\xi}x+\frac{1}{2}\int\dot{\xi}^2\mathrm{d}t)];t) \\ &= \psi(x,[\tau+\delta\tau];t), \end{aligned} \tag{17}$$

and the system automatically compensates for the extra proper time involved.

2. – The mass-proper time uncertainty relation

The mass of a particle has been defined as the energy in its rest frame. That definition allows the mass to fluctuate. The mass of a particle that occurs in the free particle Klein-Gordon equation, or the free particle Dirac equation, we call the "nominal" mass of the particle. Its actual inertial or gravitational mass will not be that. For example if two hydrogen atoms pass each other by, the inertial masses of the system will be given to lowest order by that of the Hydrogen atoms, and not by the constituent particles.

Similarly, the proper time means the proper time in the rest system. For example, when an atomic clock is flown and undergoes a change in proper time, a mixture of the velocity and gravitational effects on the plane, it responds to the effect on its center of mass system, and not to the individual effects on its constituents.

Given the role of mass and proper time as conjugate dynamical variables, then, when quantized, the system will be subject to the uncertainty principle,

$$c^2\Delta m\cdot\Delta\tau \geq \hbar/2, \tag{18}$$

just as with any other set of conjugate variables. That means if one tries to measure m, or τ, for a system, they must be controlled by this inequality. And in fact, this is required for the consistency of the theory, in that if this were not true, one could violate the uncertainty principle for another set of variables. One can prove the uncertainty principle between m and τ by making a wave packet and using the standard proof, but instead, we shall show how it works, and how it contributes to the general consistency of the theory, by giving several examples.

There is one caution in applying this principle. Take a free particle. As it propagates its phase will be

$$e^{i\alpha} = e^{i(p\cdot r-\omega t)} = e^{i(p\cdot vt-\omega t)} = e^{im\gamma(v^2-1)t} = e^{-im\tau}. \tag{19}$$

In the non-relativistic limit, one factors out e^{-imt}, leaving

$$e^{-im\tau} = e^{-imt}e^{-im(t-\tau)}, \tag{20}$$

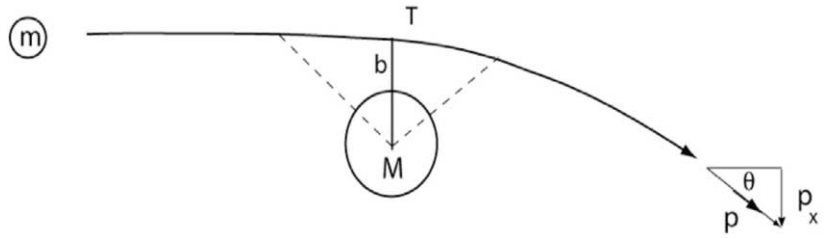


Fig. 1. – A gravitational scattering experiment of a small mass off a large one to determine the mass M, of the large mass. The gravitational field of the small mass prevents a clock on the large mass from being accurately read in the laboratory.

and it is this factor that obeys the uncertainty principle in this limit. So when we refer to $\Delta\tau$, we mean $\Delta(t-\tau)$, and this translates to

$$\mathrm{d}\tau = \sqrt{1+2\varphi - v^2}\mathrm{d}t \underset{NR}{\longrightarrow} \mathrm{d}\tau = (1+\varphi - v^2/2)\mathrm{d}t, \tag{21}$$

$$\Delta\tau \equiv [\Delta(t-\tau)] = (\Delta\varphi)t, \quad \text{or} \quad (\Delta v^2)t/2 \quad \text{or} \quad (\varphi - v^2/2)\Delta t.$$

The first example is an attempt to measure the mass of a heavy particle, M, by gravitationally scattering a much lighter particle, m, off of it, and measuring the deflection of m, described in fig. 1.

If the initial velocity of the light body and its mass are known accurately, then

$$\theta \sim p_x/p \sim \int F_x \mathrm{d}t/mv \tag{22}$$

Here F is the gravitational force exerted by M, b in the figure is the impact parameter, and the effective distance over which the impulse is exerted is within the dotted lines, so $\int \mathrm{d}t \sim b/v$. Then

$$F \sim \frac{GMm}{b^2}, \qquad \int F\mathrm{d}t \sim \frac{GMm}{bv} \sim p_x. \tag{23}$$

Therefore,

$$\theta \sim GM/bv^2, \qquad M \sim bv^2\theta/G \sim bvp_x/mG. \tag{24}$$

If the momentum transfer is measured to within Δp_x, then

$$\Delta M \sim bv\Delta p_x/mG. \tag{25}$$

However, the distance of closest approach cannot be known to within

$$\Delta b \cdot \Delta p_x \sim \hbar, \tag{26}$$

and therefore, because the particle m, while passing M, exerts a gravitational force on it so that the gravitational potential at M is unknown by

$$\Delta\varphi \sim \frac{Gm}{b^2}\Delta b. \tag{27}$$

Over the time of closest approach $T \sim b/v$, the clock on M will become uncertain by

$$\Delta\tau \sim T\Delta\varphi \sim \frac{Gm}{bv}\Delta b. \tag{28}$$

This will be true, no matter how accurately the clock on M was known before the experiment, or how accurate the lab clock t is known. And so

$$\Delta\tau \cdot \Delta M \sim \frac{Gm\Delta b}{bv}\frac{bv\Delta p_x}{mG} \sim \Delta b\Delta p_x \sim \hbar. \tag{29}$$

The quantity $\Delta\tau$ refers to how accurately the clock sitting on M can be known to an observer in the laboratory, and it needs no further assumptions than the uncertainty principle between x and p. One would get the same result if one wanted to determine the mass and proper time of a clock sitting on m.

A similar experiment is the famous one discussed by Bohr and Einstein. This was the one where Einstein proposed the experiment in fig. 2, which purported to produce a situation in which both t and E can be perfectly known for a system. It was refuted by Bohr by pointing out that Einstein had forgotten to include the uncertainties introduced by general relativity (really just the equivalence principle).

In this experiment, a box contains a gas of photons, and it is very accurately weighed (in the Earth's g field) before the experiment. Inside the box is a clock, which is set to open a trap door at a specific time, for a specific interval, to release a train of light. Afterward, the box is weighed again, and the energy emitted, E, is accurately determined. Ignoring the effects of gravity, it would seem that one has obtained independent, accurate measurements of both E and t, which would violate the uncertainty principle. Bohr argued that to accurately measure the mass of the box, m, the scale must be at rest. If the reading takes time T, then in order to accurately determine x, the scale reading, the uncertain impulse imparted by the reading process must be less than $g\Delta mT$, in order to get an accuracy Δp, in the momentum, and thus the mass m. Then

$$\Delta x > \hbar/\Delta p > \hbar/g\Delta mT. \tag{30}$$

But if x is uncertain, then the gravitational potential $\varphi = gx$ is uncertain, and

$$\Delta T/T \sim \Delta\varphi = g\Delta x, \tag{31}$$

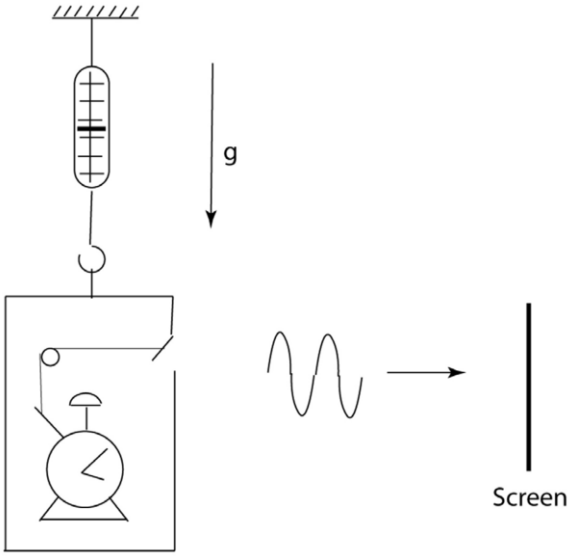


Fig. 2. – Einstein's experiment on weighing a box of photons, which was refuted by Bohr, who noted that Einstein had neglected equivalence in his argument. It can much more naturally be discussed in terms of $\Delta m \Delta \tau$ rather than in terms of $\Delta E \Delta t$. See text for particulars.

from which one sees that

$$\Delta x \cdot \Delta p \sim \frac{\Delta T}{Tg} \cdot g \Delta m T \sim \Delta m \cdot \Delta T \sim \hbar. \tag{32}$$

This uncertainty results from the inability of the laboratory observer to know the reading on the clock attached to the box, which is precisely the proper time in the box. So although Bohr discussed the experiment in terms of E and t, we see that the concepts involved are clearly Δm and $\Delta \tau$. One could convert the experiment to one measuring ΔE and Δt in the laboratory, by inserting a screen into the emitted beam and detecting the time received and the energy hitting the screen.

An important further point to be made here is that Bohr has often been criticized for bringing the gravitational shift, a concept foreign to non-relativistic theory, into the discussion. But we have already seen that there are many residual effects of relativistic theory that intrude themselves into the non-relativistic limit. The fallacy is not in including them, but rather is caused by arbitrarily excluding them.

We shall give an example that has nothing to do with gravity, but rather with the connection between velocity and proper time. Consider an attempt to measure the mass of a particle by using a mass spectrometer. A charged particle is sent into a region where a magnetic field exists, so it moves in a circle, and its mass is determined by measuring the radius of the circle as in fig. 3. There is no magnetic field below the horizontal line.

The particle enters from below, and there is a magnetic field B perpendicular to the

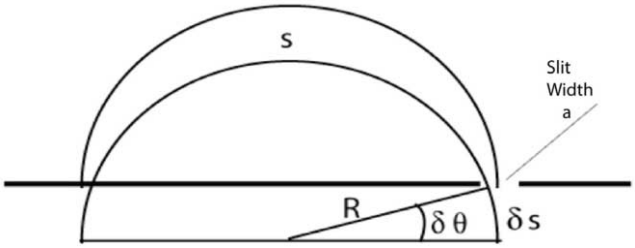


Fig. 3. – A non-gravitational example of the uncertainty principle between m and τ. There is a magnetic field perpendicular to the plane of the paper, and the particle enters from below, bends, and lands on a screen. The mass is determined by R and the diffraction at the screen gives an uncertainty to τ.

plane of the paper. The force on the particle is evB/c, and so

$$mv^2/R = evB/c, \qquad R = mvc/eB. \tag{33}$$

The particle hits the screen (the horizontal line) at $x = 2R$. The biggest uncertainty in x comes from the slit width, a, and so

$$\Delta m = \Delta R\, eB/cv = aeB/2cv. \tag{34}$$

As the particle passes through the slit, it undergoes diffraction, and gets a spread $\Delta\theta = \Delta p_x/p$. However in bending by $\Delta\theta$, the total distance traveled is changed by $2\Delta s = 2R\Delta\theta$.

The change in the distance traveled directly maps into a change in the time taken (with $c = 1$),

$$\Delta t = 2\Delta s/v = 2R\Delta\theta/v = 2R\Delta p_x/pv. \tag{35}$$

The change in proper time is given by

$$\tau = \sqrt{1 - v^2}\; t \sim (1 - v^2/2)\, t, \qquad \Delta\tau = \frac{v^2\Delta t}{2} = \frac{vR\Delta p_x}{p}\,, \tag{36}$$

and so

$$\Delta\tau \cdot \Delta m = \frac{vR\Delta p_x}{p}\,\frac{aeB}{2v} = \frac{R\Delta p_x}{mv}\,\frac{\Delta x eB}{2} = \frac{1}{2}\Delta p_x \Delta x \sim \frac{\hbar}{2}\,, \tag{37}$$

since $R = mv/eB$.

All these examples are vindicated by showing that they can be converted into different pairs of conjugate variables, which are known to obey the uncertainty principle. This interconnectivity shows the interdependence of all uncertainty relations, so that they must

be valid for all sets of conjugate variables, including mass and proper time. Thinking that one can invalidate one pair of uncertainties, without invalidating all of them, was precisely Einstein's problem in his argument with Bohr. However, it would be experimentally convincing if one could produce an experiment where it is natural to conduct the experiment by measuring the mass and proper time, while it would be difficult to convert it into an uncertainty relation for other sets of variables.

3. – The classical limit of the equivalence principle

3˙1. *The strange mass scaling in phase space.* – The way the weak equivalence principle manifests itself in quantum theory is somewhat peculiar, and the correspondence principle for external gravitational forces is different from that for non-gravitational forces. Classically, the trajectory of a body subject to an external gravitational field will be independent of its mass. This means (assuming that the body is captured in a closed orbit) that $r \neq f(m)$, $v \neq f(m)$, and $E = mf(r, v)$, so r and v are independent of the mass and the energy is linearly dependent on it. We also define the canonical momentum such that it is linear in $m, p = mf(r, v)$. However quantum mechanically, the situation is quite different. The de Broglie wavelength is defined such that it depends on the momentum, and therefore the mass. Therefore all these functions depend on the mass, in a very specific way. If we look at the Bohr orbits, we find that

$$\oint p\mathrm{d}q = nh, \qquad \oint v\mathrm{d}q = nh/m, \tag{38}$$
$$q, v = f(nh/m), \qquad E, p = m\, f(nh/m).$$

(Here we left out the irrelevant $1/2$ in $n + 1/2$). This is a dimensional argument, so it is also valid for the Schrödinger equation. As an example, take the case of a gravitational Bohr atom, where a mass m is bound to a much larger mass M:

$$\textit{electric}: \quad V = e^2/r, \qquad r_n = \frac{\hbar^2 n^2}{me^2}, \qquad E_n = \frac{me^4}{2\hbar^2 n^2}; \tag{39}$$
$$\textit{gravitational}: \quad e^2 \to GMm,$$
$$r_n = \frac{\hbar^2 n^2}{GMm^2} = \frac{1}{GM}\left(\frac{n\hbar}{m}\right)^2, \qquad E_n = \frac{G^2M^2m^3}{2\hbar^2 n^2} = \frac{G^2M^2}{2}m\left(\frac{m}{\hbar n}\right)^2.$$

Here, the integral $\oint v\mathrm{d}q$, which should not depend on the mass, is a function of $(\hbar n/m)$.

The really interesting question here is that if the quantum-mechanical solution depends on the mass, how can the mass dependence drop out in the classical limit? That's a real question, and the answer is that it drops out in a peculiar way. If one has two different particles, one of mass m_1, and the other of mass m_2, K times heavier than m_1, so $m_2 = Km_1$, and if m_1 is in some orbit characterized by n_1, then m_2 will be in an

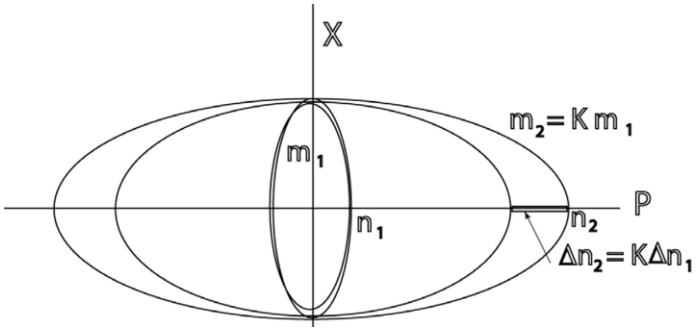


Fig. 4. – Because of the equivalence principle, there is a scaling of the momentum and the quantum number n in the high n limit, in order to preserve equivalence. This is different from the high n limit for non-gravity forces. Here the effect is shown in phase space.

orbit characterized by $n_2 = Kn_1$. Thus

$$r_1 = f\left(\frac{n_1\hbar}{m_1}\right), \qquad r_2 = f\left(\frac{n_2\hbar}{m_2}\right) = f\left(\frac{Kn_1\hbar}{Km_1}\right) = r_1. \tag{40}$$

So the mass does not drop out in the classical limit, rather it scales so that if one particle is in a given orbit, a heavier particle will be in a higher orbit, scaled proportionally to its relative mass. This is because there is no fundamental unit of length in the theory. Rather every different potential will have its own unit of length. In the gravitational Bohr atom, the unit of length will be $r(n = 1)$ in eq. (39). This scaling is only possible for $n \gg 1$, where the orbits will be close together, almost continuous. As n gets smaller, each case will have its own $r(n = 1)$, and weak equivalence breaks down.

If there were a fundamental length, it would enter as a new fundamental constant of nature. There is no reason to expect it to be the Planck length. The Planck length, λ_P, is a length constructed of known constants, $\lambda_P = \sqrt{G\hbar/c^3} \sim 10^{-34}$ cm. It is where gravity and quantum theory are expected to clash and various "quantum foam" effects appear. But this assumes that no new physics will intervene in the meantime. This same type of assumption has been made in the past, and it suffers a bad history. After the development of relativity, it was assumed that classical electricity would break down when the "classical radius of the electron" was reached, $r_C = (e^2/mc^2) \sim 10^{-13}$ cm, where the potential energy of the electron matched its rest mass. But before we ever got there came quantum mechanics, the Bohr radius at 10^{-8} cm, and pair creation at $\sim 10^{-11}$ cm. This argument never predicted the emergence of $\hbar$.

We like to use the following argument against assuming that the Planck length is the next true landmark on the length scale. Imagine that we knew about quantum theory and gravity, but not about Maxwell's equations, and we wanted to predict a fundamental velocity in nature. c^*. A good guess might be

$$c^* = Gm_n^2/\hbar = 10^{-28}\,\text{cm/sec} \tag{41}$$

This comes out closer to the speed of darkness, rather than light. Dirac noted that the fundamental constants tend to group themselves in sets that differ by multiples of 10^{20} from each other. Our c^* differs from c by about 10^{40}, which fits into this analysis, and seems to indicate to us that there is something really important missing from our knowledge of nature. This is reinforced by looking at the radius of the gravitational Bohr atom, eq. (38), consisting of two neutrons. This gives about $r_n \sim 10^{27}$ cm, around the size of the universe, with a preposterously small energy. This would tend to throw the entire interpretation of the theory into doubt, and sets off alarms to us about our complacency concerning what we think we know about gravity.

So the weak equivalence principle breaks down at low quantum numbers, and is implemented at high quantum numbers by the scaling phenomena we alluded to. At high quantum numbers, x is independent of the mass, but p, the momentum, varies linearly with the mass, so it also scales as K in this limit. The situation in phase space is depicted in fig. 4.

4. – The different correspondence principles for gravity and non-gravity forces: matrix elements in the classical limit

Because of the scaling we have indicated, the correspondence limit is different for the case of an external gravitational force, from that for a non-gravitational force. In the semiclassical limit, for a bound state, we can take the wave function as

$$\psi_E(x) = \frac{A}{\sqrt{p(x)}} e^{i\int_a^x p\mathrm{d}x/\hbar}. \tag{42}$$

Here $p(x)$ is given by the equation

$$\frac{p^2}{2m} + V(x) = E. \tag{43}$$

The situation is as depicted in fig. 5. This is the WKB wave function and is valid in the high n limit. We can determine the difference in energy between two neighboring levels from

$$\oint p_E \mathrm{d}x = nh, \qquad \oint p_{E+\delta E}\mathrm{d}x = \oint \left(p_E + \frac{\partial p}{\partial E}\delta E\right)\mathrm{d}x = (n+1)h. \tag{44}$$

From eq. (43), we have

$$p\ \delta p/m = v\ \delta p = \delta E, \tag{45}$$

so that

$$\begin{aligned} &\delta E \oint \frac{\partial p}{\partial E}\mathrm{d}x = \delta E \oint \frac{\mathrm{d}x}{v} = \delta E \oint \mathrm{d}t = \delta E\ T_n = h, \\ &\delta E = h/T_n = \hbar\omega_n. \end{aligned} \tag{46}$$

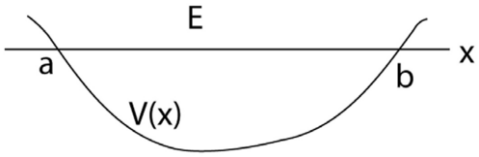


Fig. 5. – We are looking at a bound system with turning points a and b. The WKB wave function is good for high n orbits, but because of a scaling in the mass, the correspondence limit is different for gravitational and non- gravitational forces.

Thus the energy levels, in the region where $\Delta n \ll n$, are equally spaced, and

$$E_{n+\Delta n} \approx E_n + \Delta n\, \hbar\omega_n. \tag{47}$$

We can determine approximately the normalization constant in eq. (42) from

$$\int_a^b |\psi_E|^2 \,\mathrm{d}x = 1. \tag{48}$$

We integrate only over the allowed region because ψ decays exponentially to zero in the unallowed region. So

$$1 = \int \frac{|A|^2}{p}\mathrm{d}x = \frac{|A|^2}{m}\int_a^b \frac{\mathrm{d}x}{v} = \frac{|A|^2}{m}\frac{1}{2}\oint \frac{\mathrm{d}x}{v} = \frac{|A|^2}{2m}T_n, \tag{49}$$

$$A_n = \sqrt{\frac{2m}{T_n}}\,.$$

If one has some function $f(x)$, and one wants the matrix element f_{mn}, then

$$f_{mn} = \int \psi^*_{E(m)} f(x) \psi_{E(n)} \mathrm{d}x, \qquad m \sim n, \tag{50}$$

$$f_{mn} = \frac{2m}{T_n}\int \frac{f(x)}{p_n(x)} e^{-i\int_a^x (p_m - p_n)\mathrm{d}x/\hbar}.$$

Now we know from eqs. (42)-(44) that

$$p_m = p_{n+\Delta n} = p_n + \frac{(h\omega_n \Delta n)}{v}\,, \tag{51}$$

so

$$\int_a^x (p_m - p_n)\mathrm{d}x/\hbar = \omega_n \Delta n \int_a^x \mathrm{d}x/v = \omega_n \Delta n\; t(x), \tag{52}$$

where t is the time in going from a to x.

Thus,

$$f_{mn} = \frac{2}{T_n}\int_a^b f(x)e^{-i\Delta n\omega_n t(x)}\frac{\mathrm{d}x}{v} = \frac{1}{T_n}\oint f(x(t))e^{-i\Delta n\omega_n t}\mathrm{d}t. \tag{53}$$

Classically, we know that if we have some function $f(t)$, and the period is ω in a particular orbit (which corresponds to the nth quantum orbit, so $\omega \approx \omega_n$) then we can expand it in a fourier series,

$$f(t) = \sum f_\ell e^{i\ell\omega\, t}. \tag{54}$$

$$\begin{aligned}\oint f(t)e^{-ik\omega t}\mathrm{d}t &= \sum f_\ell \oint e^{i(\ell-k)\omega t}\mathrm{d}t \\ &= \sum f_\ell T\delta_{\ell k} = f_k T; \\ f_\ell &= \frac{1}{T}\oint f(t)e^{-i\ell\omega t}\mathrm{d}t.\end{aligned} \tag{55}$$

Comparison of this result with eq. (53) shows that the matrix element $f_{n+\ell,\,n}$ is equal to the corresponding classical Fourier component f_ℓ, or

$$f_{n+\ell,n} \to f_\ell, \qquad \ell \ll n, \qquad n \gg 1. \tag{56}$$

Thus we see that in the classical limit, it is the overlap between successive orbits that goes over to the classical result. For example, in the Hydrogen atom, the expansion of the ellipse, $r(t) = \sum r_\ell\, e^{i\ell\omega\, t}$, gives the terms of the dipole moment ex_{mn} that determine the intensity of spectral lines. Because of this, Bohr could use the correspondence principle to determine the intensities *before* quantum mechanics was invented.

But we also see, that the correspondence limit tells us that in the classical gravitational case, the classical fourier component $f_\ell(m_1)$ corresponds to the component $f_{K\ell}(m_2 = Km_1)$, for a particle $m_2 = Km_1$, that is K times heavier. So in order to preserve the Equivalence principle, this scaling property carries over into the classical correspondence limit of the theory. Thus we see that even in the correspondence limit, this limit is different for gravitational and non-gravitational forces.

We have tried to point out that one can certainly treat mass and proper time as independent conjugate dynamical variables, which will have experimental consequences. Some of these are:

1. That even in a non-relativistic experiment one can see fringes from a superposition of mass states, contradicting the Bargmann theorem. There should also be other experiments emphasizing the non-relativistic residues of relativistic experiments.

2. One should be able to detect fringes in proper time between the amplitudes of a particle that has been coherently recombined, especially if it is unstable.

3. One should be able to verify the $\Delta m\Delta\tau$ uncertainty relation experimentally in convincing ways.

4. One should be able to see the "decay red shift" separating an unstable particle from a stable alternative, and see fringes between them.

We are currently working to obtain convincing experimental tests of all these conclusions. We do not think quantum theory is quite consistent without them.

REFERENCES

[1] BARGMANN V., *Ann. Math.*, **59** (1954) 1.

[2] This formalism has been published piecemeal over the years. The key references are, all with GREENBERGER D. M., or GREENBERGER DANIEL M., as author:

"The Role of Equivalence in Quantum Mechanics", *Ann. Phys. (N. Y.)*, **47** (1968) 116;

"Theory of Particles with Variable Mass, I-Formalism", *J. Math. Phys.*, **11** (1970) 2329;

"Theory of Particles with Variable Mass, II-Some Physical Consequences", *J. Math. Phys.*, **11** (1970) 2341;

"Some Useful Properties of a Theory of Variable-Mass Particles", *J. Math. Phys.*, **15** (1974) 395;

"Wave Packets for Particles of Indefinite Mass", *J. Math. Phys.*, **15** (1974) 406;

"Some Remarks on the Extended Galilean Transformation", *Am. J. Phys.*, **47** (1979) 35;

"Comment on Non-Spreading Wave Packets", *Am. J. Phys.*, **48** (1980) 256;

"What is the Proper Role of Proper Time", article in *Physics as Natural Philosophy*, edited by SHIMONY A. and FESHBACH H. (MIT, Cambridge, Mass.) 1982, pp. 179–203;

"The Proper Role of Proper Time", in *Experimental Metaphysics: Quantum Mechanical Studies for Abner Shimony, Vol. I*, edited by COHEN R. S., HORNE M. and STACHEL J. (Kluwer, Dordrecht) 1997;

"The Inadequacy of the Galilean Transformation in Quantum Mechanics", *Phys. Rev. Lett.*, **87** (2001) 100405-1;

"The Disconnect Between Quantum Mechanics and Gravity", quant-ph. ArXiv: 1011.3719 (2010);

"Conceptual Problems Related to Time and Mass in Quantum Theory", Quant-ph. ArXiv: 1011.3709 (2010).

Proceedings of the International School of Physics "Enrico Fermi"
Course 197 "Foundations of Quantum Theory", edited by E. M. Rasel, W. P. Schleich and S. Wölk
(IOS, Amsterdam; SIF, Bologna) 2019
DOI 10.3254/978-1-61499-937-9-345

Atom interferometry and its applications

S. Abend, M. Gersemann, C. Schubert, D. Schlippert and E. M. Rasel

Institut für Quantenoptik, Leibniz Universität Hannover - Welfengarten 1, D-30167 Hannover, Germany

M. Zimmermann, M. A. Efremov and A. Roura

Institut für Quantenphysik, Universität Ulm - Albert-Einstein-Allee 11, D-89081 Ulm, Germany

F. A. Narducci

Naval Postgraduate School, Department of Physics - Monterey, CA 93943, USA

W. P. Schleich

Institut für Quantenphysik and Center for Integrated Quantum Science and Technology (IQ^{ST}), Universität Ulm - Albert-Einstein-Allee 11, D-89081 Ulm, Germany

Institute for Quantum Science and Engineering (IQSE), Texas A&M AgriLife Research, Hagler Institute for Advanced Study at Texas A&M University, Department of Physics and Astronomy, Texas A&M University - College Station, TX 77843-4242, USA

Summary. — We provide an introduction into the field of atom optics and review our work on interferometry with cold atoms, and in particular with Bose-Einstein condensates. Here we emphasize applications of atom interferometry with sources of this kind. We discuss tests of the equivalence principle, a quantum tiltmeter, and a gravimeter.

1. – Introduction

Based on the pioneering work [1, 2] by Mark Kasevich and Steve Chu starting in 1991, light-pulse atom interferometry has grown into an extremely successful tool for precision measurements. Indeed, ground-breaking experiments have been performed in the fields of inertial sensing and tests of the foundations of physics. Inertial sensing covers measurements of the local gravitational acceleration [3-23], or rotations, for example of the Earth [24-30], as well as gravity gradiometry [31, 32]. The present lecture notes aim at providing an introduction into, and an overview over this rapidly moving field. Moreover, our article complements the corresponding lectures by Daniel M. Greenberger.

In this section we intend to motivate this branch of physics located at the interface of atomic physics, quantum optics and solid state research, and to give a preview of coming attractions. In order to focus on the essential ideas we keep this section brief and postpone more detailed discussions to the later sections.

We start in sect. **1**˙1 by mentioning applications of interferometry with cold atoms ranging from tests of the foundations of physics to quantum sensors. We then outline in sect. **1**˙2 the realization of optical elements in atom optics such as beam splitters and mirrors leading us in sect. **1**˙3 to various sources for our interferometers. Here we emphasize especially the use of an atom chip as a trap and a mirror for laser light opening the avenue towards a quantum tiltmeter and a gravimeter. An outline of our lecture notes in sect. **1**˙4 concludes our introduction.

1˙1. *Applications of atom interferometry.* – The scope of testing fundamental physics with atom interferometry comprises on the one hand measurements of fundamental constants such as Newton's gravitational constant G [19, 33-35] and Sommerfeld's fine-structure constant α [36-40]. Indeed, the result for α obtained with photon-recoil measurements in recent years [38] has entered into the determination of the CODATA value. Moreover, a measurement of α has been reported this year [40] with an accuracy of $2.0 \cdot 10^{-10}$, which is even more accurate than the best measurements to date, based on measuring the anomalous magnetic moment of the electron [41].

On the other hand, testing the pillars of general relativity, for example, the universality of free fall (UFF) resulting from Einstein's equivalence principle [42], is of particular interest. The most elementary test of the UFF is to compare the measurements of local gravity with a classical and an atomic [3, 43] gravimeter.

More elaborate set-ups use two different quantum objects, for instance, two isotopes of the same atomic species, or two different atomic species, and measure their free-fall rate within the same device [18, 44-48]. Future experiments of this kind are expected to catch up to, or even overcome today's best classical tests of the UFF based on Lunar-Laser-Ranging [49], torsion balance experiments [50], or space missions using free-falling test masses [51].

In addition, atom interferometers can also test different models in particle physics in the search for unknown forces or dark energy [52-54]. Even more exotic experiments aim

for the detection of gravitational waves [55-58], new probes of the foundations of quantum mechanics, such as delayed-choice experiment [59], or for the creation of atomic Einstein-Podolsky-Rosen pairs [60, 61].

Especially in absolute gravimetry, the sensitivity of atomic sensors is competitive with classical devices [62]. The conventional sensors used for geodesy [63] can be categorized as absolute gravimeters, such as the falling-corner cube gravimeters [64, 65], and relative gravimeters like superconducting gravimeters [66-68], which have a changing bias over time.

State-of-the-art *atomic* gravimeters operate with Raman-type beam splitters and cold atoms, which are either dropped or launched from optical molasses —a technique invented for cesium fountain clocks [69, 70]. State-of-the-art laboratory grade examples of these gravimeters [14, 23, 71, 72] reach inaccuracies in the low μGal regime. The maturity of this technology has now arrived at a level that commercial products with a specified sensitivity of better than $10\,\mu$Gal [73-75] are available.

1˙2. *Optical elements for atoms.* – The coherent manipulation of matter waves is a central element in every matter wave interferometer [76]. Two methods to realize beam splitters and mirrors based on light pulses offer themselves: Raman [1] and Bragg diffraction [77, 78]. However, these techniques imply conceptual differences which have to be considered when constructing an atom interferometer aimed at measuring inertial effects [79].

Indeed, Raman diffraction, where an atomic Λ-scheme is driven, requires a phase-stable microwave coupling between two hyperfine ground states of an atom usually established by two phase-locked lasers. Working with two different internal states of an atom from an ensemble with a wide velocity distribution has the advantage of velocity filtering with blow-away pulses and state-selective detection [13, 80]. These state-labeling features are described in detail by Christian Bordé [81].

In contrast, Bragg diffraction involves only a single atomic ground state and allows us to construct with a single laser system a pure momentum, or recoil beam splitter. However, due to the transition frequency being in the radio-frequency (RF) range, the detection needs to be spatially resolved, and, in order to distinguish different diffraction orders [13, 82], requires a momentum distribution below recoil.

1˙3. *Sources for atom optics.* – Today's generation of atomic inertial sensors typically operates with cold atoms released or launched from an optical molasses. This approach was taken in our simultaneous, dual-species Raman-type interferometer with molasses-cooled ^{87}Rb and ^{39}K ensembles which measured the Eötvös ratio to $\eta_{\mathrm{Rb,K}} = (0.3 \pm 5.4) \cdot 10^{-7}$. The velocity distribution and finite size of these sources of atoms limit the efficiency of the beam splitters as well as complicate the analysis of systematic uncertainties.

These limitations can be overcome by the use of atomic ensembles with a typical average momentum well below the recoil of a photon, for example Bose-Einstein condensates (BECs) [83, 84]. The width of a momentum distribution corresponding to a BEC can be further reduced after reaching the regime of ballistic expansion, where all

mean field energy is converted to the kinetic energy, by the application of the delta-kick collimation (DKC) technique [85].

Atom-chip technologies offer the possibility to generate a BEC and perform DKC in a fast and reliable way, resulting in miniaturized atomic devices. BECs are very useful for Bragg and double Bragg diffraction [86, 87] leading to high diffraction efficiencies. Indeed, such beam splitters and mirrors can reach an efficiency of above 95% facilitating interferometry with high contrast.

Furthermore, BECs offer novel methods of coherent manipulation with high fidelity, to realize for example a tiltmeter [86]. A combination of double Bragg diffraction and Bloch oscillations gives rise to a relaunch procedure with an efficiency larger than 75% for the diffraction of atoms in a retro-reflected optical lattice [87]. The novelty of this method originates from the fact that it relies on a single laser beam, which is also used as a beam splitter, and thus does not lead to an increased complexity of the setup.

We realize a Mach-Zehnder interferometer (MZI) by dropping ensembles directly after release, or accelerating them upwards after a certain time of free fall. The interferometry is performed as in a fountain, such that the total time $2T$ of the interferometer can be extended. Here T is the time between the first beam-splitter pulse and the central mirror pulse.

We utilize an atom chip [87] for BEC generation and state preparation, including magnetic sub-state transfer, DKC and Stern-Gerlach-type deflection. A special feature of our setup is that the light field, which forms the MZI by Bragg diffraction, is reflected by the atom chip itself. In this way, the chip also serves as an inertial reference inside the vacuum chamber leading to a compact atom-chip gravimeter.

All atom-optics operations, the interferometry as well as the detection of the output states of the atom interferometer, are integrated into a volume of less than a cube of one centimeter side length. In the fountain mode, the MZI can be extended to have the total interferometer time $2T = 50\,\mathrm{ms}$ with a large contrast $C = 0.8$, which yields an intrinsic sensitivity $\Delta g/g = 1.4 \cdot 10^{-7}$. The state preparation comprised of DKC and Stern-Gerlach-type deflection makes an important contribution to this achievement by improving the contrast and reducing the detection noise. An estimation of systematic uncertainties for the current setup and their projection onto a future device prove that it is possible to reach sub-μGal accuracies with a fountain-type geometry.

1˙4. *Overview*. – Our lecture notes are organized as follows. In sect. **2** we introduce the basic tools of atom interferometry such as beam splitters, mirrors, and optical lattices to construct a MZI for atoms. We then turn in sect. **3** to tests of the equivalence principle. In particular, we present a dual-species atom interferometer for ^{87}Rb and ^{39}K to investigate the UFF. Next, we present in sect. **4** interferometers utilizing BECs on an atom chip. We introduce the technique of DKC and present a quantum tiltmeter as well as a gravimeter exploiting this technology. Finally, we conclude in sect. **5** by providing an outlook on future devices such as the very long base line atom interferometry (VLBAI) facility and atom interferometers in space.

2. – Tools of atom interferometry

An atom interferometer requires the realization of a beam splitter and a mirror for atom waves. The underlying processes have to be coherent and phase-stable in order to establish an interference pattern. Several options for such elements exist, and we analyze prominent ones exploited in current experiments below. In particular, in sect. **2**˙1 we discuss beam splitters and mirrors based on Bragg and Raman diffraction. Next, in sect. **2**˙2 we focus on the manipulation and accelerations of atoms by optical lattices. Concluding, we introduce in sect. **2**˙3 a common interferometer geometry based on beam splitters and mirrors.

2˙1. *Beam splitters and mirrors.* – Beam splitters and mirrors for matter waves can be realized with the help of mechanical gratings [88,89], or electromagnetic waves [1,2,90-92]. While single-photon electric or magnetic dipole transitions can implement a coherent electromagnetic coupling, the assessment in sect. **2**˙1.1 focuses on stimulated two-photon transitions. For this purpose, we first introduce the Rabi model which describes the atom-light interaction in an effective two-level system before we proceed to two-photon transitions in a three-level system. Here the absorption of a photon from a field with frequency ω_1 is followed by stimulated emission into a field with frequency ω_2, and *vice versa*.

Next, we consider in sect. **2**˙1.2 the momentum transfer due to the atom-light interaction which is at the very heart of the sensitivity of atom interferometers to inertial forces. We also discuss two standard approaches towards beam splitters. Raman diffraction is a widely used technique designed for atoms which have been laser-cooled in optical molasses without the application of additional cooling steps. Bragg diffraction is a powerful tool for delta-kick collimated Bose-Einstein condensates, since the velocity dispersion of the ensemble is of major relevance for the manipulation efficiency.

In sect. **2**˙1.3 we then concentrate on the generalization of a multi-photon coupling utilizing Bragg diffraction, and conclude in sect. **2**˙1.4, where we take into account the effects of the finite size of the atom cloud and the laser beam on the atom-light interaction.

2˙1.1. Rabi oscillations and two-photon coupling. We consider the atom as an effective two-level system consisting of the internal states $|\mathrm{g}\rangle$ and $|\mathrm{e}\rangle$ with energies $\hbar\omega_\mathrm{g}$ and $\hbar\omega_\mathrm{e}$, respectively, and a dipole moment $\mathbf{d}$. The time evolution of the state populations under the influence of a resonant ($\omega_0 = \omega_\mathrm{eg} \equiv \omega_\mathrm{e} - \omega_\mathrm{g}$) electromagnetic field $\mathbf{E} \equiv \mathbf{E}_0 \cos(\omega_0\tau + \phi)$ at time τ with frequency ω_0 and phase ϕ is determined by the Rabi frequency

$$\Omega_\mathrm{eg} \equiv \frac{\langle \mathrm{e}|\, \mathbf{d}\cdot\mathbf{E}_0\, |\mathrm{g}\rangle}{\hbar} = \Gamma\sqrt{\frac{I}{2I_\mathrm{sat}}}\,, \tag{1}$$

expressed in terms of the intensity I of the light field with the saturation intensity I_sat, and the natural linewidth Γ of the transition. Indeed, Ω_eg is assumed to be constant and is a measure of the coupling strength between the atom modeled by the two atomic states $|\mathrm{g}\rangle$ and $|\mathrm{e}\rangle$, and the electromagnetic field $\mathbf{E}$.

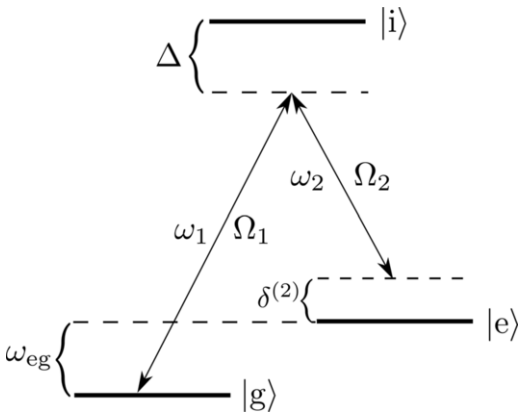


Fig. 1. – Three-level atom interacting with two light fields. A two-photon coupling between the atomic states $|\mathrm{g}\rangle$ and $|\mathrm{e}\rangle$ is established by two electromagnetic fields of frequencies ω_1 and ω_2, with Ω_1 and Ω_2 being the Rabi frequencies of the corresponding one-photon transitions. Here Δ is the common detuning of the two-photon transition from the intermediate state $|\mathrm{i}\rangle$, and $\delta^{(2)}$ is the two-photon detuning.

Off-resonant driving with a non-vanishing detuning $\delta \equiv \omega_{\mathrm{eg}} - \omega_0$ is taken into account in the effective Rabi frequency

$$\Omega_{\mathrm{eff}} \equiv \sqrt{|\Omega_{\mathrm{eg}}|^2 + \delta^2}\,, \tag{2}$$

which always leads to a faster oscillation of the probability

$$P_{\mathrm{e}}(\tau, \delta, \Omega_{\mathrm{eg}}) = \frac{1}{2}\left(\frac{\Omega_{\mathrm{eg}}}{\Omega_{\mathrm{eff}}}\right)^2 [1 - \cos(\Omega_{\mathrm{eff}}\tau)] \tag{3}$$

to find after an interaction time τ the atom in the excited state $|\mathrm{e}\rangle$ if the atom is initially prepared in the ground state $|\mathrm{g}\rangle$.

The reduced amplitude of the oscillation is determined by the ratio $\Omega_{\mathrm{eg}}/\Omega_{\mathrm{eff}}$ of the resonant and the effective Rabi frequency. For a vanishing detuning, that is $\delta = 0$, the amplitude of P_{e} is unity, whereas for a large detuning, $|\delta| \gg |\Omega_{\mathrm{eg}}|$, the amplitude of P_{e} tends towards zero.

Rabi oscillations can be driven efficiently only for long-lived states, that is for states in which Ω_{eff} is large compared to the inverse lifetime of the working states $|\mathrm{g}\rangle$ and $|\mathrm{e}\rangle$. A common method to avoid decays from $|\mathrm{e}\rangle$ to $|\mathrm{g}\rangle$ is to choose these states such that the selection rules forbid a single-photon transition between them. The two states could then be coupled via a two-photon transition, which requires an intermediate state $|\mathrm{i}\rangle$.

As depicted in fig. 1, light fields with frequencies ω_1 and ω_2 induce non-resonant transitions, where the transitions $|\mathrm{g}\rangle \longleftrightarrow |\mathrm{i}\rangle$ and $|\mathrm{i}\rangle \longleftrightarrow |\mathrm{e}\rangle$ are detuned by Δ and $\Delta + \delta^{(2)}$, respectively. Consequently, the frequency difference $\delta\omega \equiv \omega_1 - \omega_2$ is equal to

the frequency difference ω_{eg} between the working states plus the two-photon detuning $\delta^{(2)}$, that is $\delta\omega \equiv \omega_{\mathrm{eg}} + \delta^{(2)}$. Here and in the following, the superscript $^{(2)}$ indicates a two-photon process.

The intermediate state can now be short-lived itself, since it is only virtually populated, but enables sufficient coupling via simultaneous stimulated absorption and emission. The resulting two-photon Rabi frequency Ω_{12}, which drives the transition $|\mathrm{g}\rangle \longleftrightarrow |\mathrm{e}\rangle$ in the case $\Delta \gg \Omega_j$ with $j = 1, 2$, is governed by the product of both Rabi frequencies Ω_1 and Ω_2 as well as the common detuning Δ of the two-photon transition to the intermediate state $|\mathrm{i}\rangle$, that is

$$\Omega_{12} \equiv \frac{\Omega_1^* \Omega_2}{2\Delta} = \frac{\Gamma_1 \Gamma_2}{4\Delta} \sqrt{\frac{I_1}{I_{\mathrm{sat},1}} \frac{I_2}{I_{\mathrm{sat},2}}}\,. \tag{4}$$

Here I_j and $I_{\mathrm{sat},j}$ denote the intensity and the saturation intensity of the corresponding light beam, and Γ_j is the natural linewidth of the corresponding transition.

In this case we return to an effective two-level system where the probability

$$P_{\mathrm{e}}\left(\tau, \delta^{(2)}, \Omega_{12}\right) = \frac{1}{2}\left(\frac{\Omega_{12}}{\Omega_{\mathrm{eff}}^{(2)}}\right)^2 \left[1 - \cos\left(\Omega_{\mathrm{eff}}^{(2)} \tau\right)\right] \tag{5}$$

to find the atom in the excited state $|\mathrm{e}\rangle$ now depends on the two-photon Rabi frequency Ω_{12}, and the corresponding effective Rabi frequency

$$\Omega_{\mathrm{eff}}^{(2)} = \sqrt{|\Omega_{12}|^2 + (\delta^{(2)})^2}, \tag{6}$$

determined by the two-photon detuning $\delta^{(2)}$.

A fundamental loss mechanism of the coherent dynamics is spontaneous emission which is fortunately suppressed due to the fact that the two-photon transition is off-resonant by the detuning Δ relative to the intermediate state $|\mathrm{i}\rangle$. The rate R_{sp} of residual spontaneous decay then reads

$$R_{\mathrm{sp}} \equiv \frac{\sqrt{\Gamma_1 \Gamma_2}}{2\Delta} |\Omega_{12}| = \frac{(\Gamma_1 \Gamma_2)^{\frac{3}{2}}}{8\Delta^2} \sqrt{\frac{I_1}{I_{\mathrm{sat},1}} \frac{I_2}{I_{\mathrm{sat},2}}}\,. \tag{7}$$

For sufficiently large detuning it is possible to suppress spontaneous emission almost completely.

Moreover, the presence of an off-resonant light field has an influence on the atomic energy structure. Indeed, the one-photon ac-Stark shift causes an energy shift

$$\delta E_j^{\mathrm{ac}} = -\frac{\hbar |\Omega_j|^2}{4\Delta_j} \tag{8}$$

of the undisturbed atomic states $|\mathrm{g}\rangle\,(j=1)$ and $|\mathrm{e}\rangle\,(j=2)$, determined by the detuning Δ_j, and the Rabi frequency Ω_j of the corresponding transition [93] with $\Delta_1 \equiv \Delta$ and $\Delta_2 \equiv \Delta + \delta^{(2)}$.

Furthermore, for high-precision measurements such as the ones discussed in these lectures also the two-photon light shift has to be considered, which depends on the details of the internal atomic structure, as well as on the polarization of the light fields [94-96].

Finally, we consider two special cases of the Rabi dynamics given by eq. (5), namely "$\pi/2$"- and "π"-pulses, which are determined by their enclosed pulse areas. For fixed laser intensities I_1 and I_2, we define these pulses by their specific interaction times $\tau_{\pi/2} \equiv \pi/\left(2\,\Omega_{\mathrm{eff}}^{(2)}\right)$ and $\tau_\pi \equiv \pi/\Omega_{\mathrm{eff}}^{(2)}$, leading us to the probabilities

$$P_{\mathrm{e}}\left(\tau_{\pi/2}, \delta^{(2)}, \Omega_{12}\right) = \frac{1}{2}\left(\frac{\Omega_{12}}{\Omega_{\mathrm{eff}}^{(2)}}\right)^2 \tag{9}$$

and

$$P_{\mathrm{e}}\left(\tau_{\pi}, \delta^{(2)}, \Omega_{12}\right) = \left(\frac{\Omega_{12}}{\Omega_{\mathrm{eff}}^{(2)}}\right)^2, \tag{10}$$

where we have made use of eq. (5).

In the ideal case with $\delta^{(2)} = 0$, a $\pi/2$-pulse creates an equally weighted superposition of $|\mathrm{g}\rangle$ and $|\mathrm{e}\rangle$ when starting in one of the two working states. In contrast, a π-pulse inverts the states $|\mathrm{g}\rangle$ and $|\mathrm{e}\rangle$. Due to their functions in an interferometer, these pulses are called "beam splitter" and "mirror" for atoms, in complete analogy to their counterparts in optics for light beams.

2˙1.2. Bragg and Raman diffraction. So far we have only discussed the dynamics of *internal* atomic states induced by the atom-light interaction. However, the use of an atom interferometer for inertial sensing requires sensitivity to *external* degrees of freedom, in particular, the atomic center-of-mass motion relative to a reference frame. We satisfy this requirement when we recall that during the atom-light interaction the electromagnetic field does not only transfer energy, but also momentum to the atoms.

We now consider a two-photon process induced by two light fields with the wave vectors $\mathbf{k}_1$ and $\mathbf{k}_2$. In the case of counter-propagating fields, that is $\mathbf{k} \equiv \mathbf{k}_1 \approx -\mathbf{k}_2$, the momentum transfer between atom and field is maximal and approximately $2\hbar\mathbf{k}$. For co-propagating beams, that is $\mathbf{k}_1 \approx \mathbf{k}_2$, the momentum transfer is minimal and almost zero.

Furthermore, due to the fact that the dispersion relation of a free particle is parabolic, a non-zero momentum $\mathbf{p}_0$ of the atom and the resulting frequency shift have to be taken into account. Indeed, any offset $\mathbf{p}_0$ results in a Doppler shift

$$\omega_{\mathrm{D}} \equiv \frac{\mathbf{p}_0 \cdot \mathbf{k}_{\mathrm{eff}}}{m} \tag{11}$$

of the transition frequencies due to the motion of the atoms of mass m relative to the light fields. It vanishes only for atoms at rest.

These considerations also lead us to the definition

$$\omega_{\rm rec} \equiv \frac{\hbar|\mathbf{k}_{\rm eff}|^2}{2m} \tag{12}$$

of the recoil frequency associated with the light fields. Here, and in eq. (11) we have introduced the notation $\mathbf{k}_{\rm eff} \equiv \mathbf{k}_1 - \mathbf{k}_2$ to identify the effective momentum transfer $\hbar\mathbf{k}_{\rm eff}$ during a two-photon process.

In the general case of an n-th–order transition and counter-propagating light fields with $\mathbf{k} \equiv \mathbf{k}_1 \approx -\mathbf{k}_2$, the total momentum transfer

$$n\hbar\mathbf{k}_{\rm eff} \equiv n\hbar(\mathbf{k}_1 - \mathbf{k}_2) \approx 2n\hbar\mathbf{k} \tag{13}$$

is the sum of the momenta transferred by n photon pairs and can achieve large values.

In a quantum mechanical treatment of the atomic center-of-mass motion, an n-th–order two-photon transition couples the momentum eigenstates $|\mathbf{p}_0\rangle$, corresponding to the momentum $\mathbf{p}_0$ before the interaction, and $|\mathbf{p}_n\rangle$, representing the momentum $\mathbf{p}_n \equiv \mathbf{p}_0 + n\hbar\mathbf{k}_{\rm eff}$ after the interaction. Based on the considerations of sect. **2**˙1.1, we also have to include a coupling to the internal states. However, since a change of a momentum eigenstate does not necessarily require a change of an internal atomic state, different types of diffraction are possible [81].

We call an atomic scattering process a "Raman"-type diffraction if the atom-light interaction couples two different internal states of the atom, whereas we call it a "Bragg"-type diffraction if the internal state is unchanged. More sophisticated schemes employ double Raman diffraction [6, 97], or double Bragg diffraction [86, 98]. They lead to a larger momentum transfer, and hence provide us with an increased sensitivity as well as the elimination of certain noise sources due to the symmetric structure of the diffraction process.

2˙1.3. Multi-photon coupling by Bragg diffraction. Bragg diffraction of a matter wave is defined in complete analogy to the diffraction of an electromagnetic field by a crystal [99, 100]. Here the roles of light and matter are interchanged.

Indeed, when an atomic beam or ensemble is diffracted from two counter-propagating light fields of the frequencies ω_1 and ω_2 the Bragg condition reads

$$\delta E_{\rm kin} = n\hbar\delta\omega \equiv n\hbar(\omega_1 - \omega_2), \tag{14}$$

where

$$\delta E_{\rm kin} \equiv \frac{(\mathbf{p}_0 + n\hbar\mathbf{k}_{\rm eff})^2}{2m} - \frac{\mathbf{p}_0^2}{2m} \tag{15}$$

is the change of the kinetic energy of the atom associated with its change in momentum.

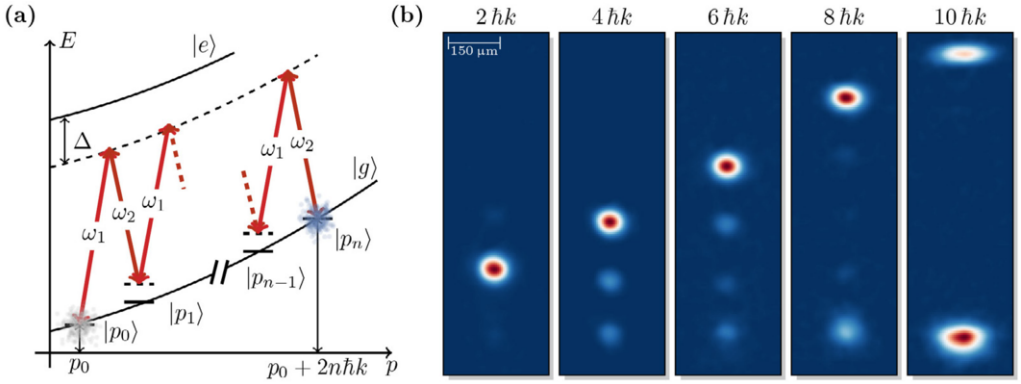


Fig. 2. – Momentum transfer of $2n\hbar k$ in n-th–order Bragg diffraction represented by the level scheme (a) and density plots (b) measured experimentally for $n = 1, 2, 3, 4$ and 5. This figure is an adaptation of figs. 4.7 and 5.12 in ref. [101].

For the case of first-order diffraction, $n = 1$, eqs. (14) and (15) give a very intuitive picture of the scattering process. An atom scatters two photons with momenta $\hbar\mathbf{k}_1$ and $\hbar\mathbf{k}_2$ from two traveling light waves if their energy difference $\hbar\delta\omega$ matches the energy δE_{kin} an atom has to absorb to climb the kinetic energy parabola, depicted in fig. 2(a).

When we use eqs. (14) and (15) and the definitions eqs. (11) and (12) of the Doppler shift ω_{D} and the photon recoil ω_{rec}, we arrive at the condition

$$\omega_1 - \omega_2 = n\,\omega_{\mathrm{rec}} + \omega_{\mathrm{D}}, \tag{16}$$

for the frequency difference which drives the diffraction process of the n-th order.

The case of scattering more than one photon pair at a time, $n > 1$, leads to the population of higher-order momentum states $|\mathbf{p}_n\rangle$ with $\mathbf{p}_n = \mathbf{p}_0 + n\hbar\mathbf{k}_{\mathrm{eff}}$, since ideally every momentum state in between is not on resonance and should not be populated, as shown in fig. 2(a).

In fig. 2(b) we present density plots for multi-photon Bragg diffraction with $n = 1, 2, 3, 4$ and 5, and note that the simultaneous scattering of n pairs of photons has been realized experimentally up to $n = 12$. This achievement allows us to construct a beam splitter [102] with a momentum transfer of $24\,\hbar k$, where $k = |\mathbf{k}|$. For ^{87}Rb the spacing between subsequent Bragg orders is only 15 kHz, which leads to populating multiple orders.

For the n-th–order Bragg transition the calculation of the transition probability P_n requires considerations [103] that go beyond the two-state assumption. The product of the Rabi frequency Ω_{eff} governed by the laser intensity I, and the duration τ of the atom-light interaction are the major ingredients. Indeed, the product of these parameters determines if a clean Rabi oscillation into a single momentum state is possible, or if multiple states are populated.

In the Bragg regime where a single order is dominantly populated, the generalized transition probability

$$P_n(\tau) \equiv \sin^2\left[\frac{1}{2}\int_0^\tau \mathrm{d}\tau' \Omega_n(\tau')\right] \tag{17}$$

from the initial momentum state $|\mathbf{p}_0\rangle$ to the n-th–order momentum state $|\mathbf{p}_n\rangle$ is determined by a new effective Rabi frequency Ω_n, as discussed in detail in ref. [103].

In order to perform an n-th–order transition, an increase in the laser power is required. An approximate solution obtained in ref. [103] presents conditions for a so-called "quasi-Bragg" regime, in which the probability to populate a single higher-order state is significantly larger compared to the one for a population of all other orders by applying short and intense pulses.

In order to achieve a large momentum transfer without the increase of laser power, one can use sequential pulses. Indeed, the same momentum transfer as in the case of a single n-th–order transition may be achieved at the cost of a larger total time of the beam splitting process, and a more complex waveform. For a sequence consisting of n_p sequential pulses the resonance condition to drive the n_s-th sequential transition with an n-th–order Bragg pulse reads

$$\omega_1 - \omega_2 = (2n_s - 1)n\omega_{\mathrm{rec}} + \omega_{\mathrm{D}} \tag{18}$$

with $n_s = 1, \ldots, n_p$. We emphasize that the frequency difference $\omega_1 - \omega_2$ has to be adjusted for each sequential transition.

In principle, there is no restriction on combining any number n_p of sequential pulses with any achievable Bragg order n, to obtain a particular transfer efficiency. For example, a sequence of beam splitters transferring $6\,\hbar k$ each, leads to a total splitting [104] of $102\,\hbar k$, or a sequence of first-order transitions results [105] in $90\,\hbar k$.

2˙1.4. Influence of atom cloud and beam size. In the single-atom picture, or in the case of a highly monochromatic ensemble, one can always find for fixed laser powers I_1, I_2 and a common detuning Δ, an interaction time τ, for which the amplitude of the transition probability given by eqs. (5) and (6) is unity. Here we assume that the two-photon detuning $\delta^{(2)}$ vanishes. In this case the beam-splitter efficiency is only limited by the loss of atoms due to spontaneous decay with the rate R_{sp}, expressed by eq. (7).

However, a non-zero temperature T_{a} of the atoms, and therefore a spread σ_v in the velocity $\mathbf{v}$ of the atomic ensemble needs to be taken into account, as it induces a broadening of the transition frequency due to the Doppler shift $\mathbf{k}_{\mathrm{eff}} \cdot \mathbf{v}$, eq. (11). Even for BECs, the beam-splitter efficiency may change drastically, when we employ for example higher-order Bragg diffraction [82].

We estimate the influence of such a velocity distribution by the use of a Gaussian distribution

$$f_{3\mathrm{D}}(\mathbf{v}) \equiv \frac{1}{(2\pi)^{3/2}\sigma_v^3} \exp\left[-\frac{(\mathbf{v}-\mathbf{v}_0)^2}{2\sigma_v^2}\right] \tag{19}$$

of velocities $\mathbf{v}$ across the atomic ensemble which is isotropic in all three spatial dimensions and has a width

$$\sigma_v \equiv \sqrt{\frac{k_\mathrm{B} T_\mathrm{a}}{m}} \tag{20}$$

determined by the temperature T_a and the Boltzmann constant k_B. Here $\mathbf{v}_0$ is an arbitrary offset velocity.

The total probability P_e to find the atom in the excited state $|\mathrm{e}\rangle$ then reads

$$P_\mathrm{e}(\tau) \equiv \iiint \mathrm{d}^3 v \, f_\mathrm{3D}(\mathbf{v}) P_\mathrm{e}\left[\tau, \delta^{(2)}(\mathbf{v}), \Omega_{12}\right], \tag{21}$$

where $P_\mathrm{e}[\tau, \delta^{(2)}(\mathbf{v}), \Omega_{12}]$ is the excitation probability given by eq. (5) for the atomic velocity $\mathbf{v}$, and $\mathrm{d}^3 v$ is the three-dimensional volume element in velocity space.

In order to evaluate the integral over $\mathbf{v}$, we recall that according to eq. (5) the probability $P_\mathrm{e}[\tau, \delta^{(2)}(\mathbf{v}), \Omega_{12}]$ is determined by the two-photon detuning $\delta^{(2)}$ which enters into the effective Rabi frequency $\Omega_\mathrm{eff}^{(2)}$ given by eq. (6). We assume here that the two-photon detuning $\delta^{(2)} \equiv \mathbf{k}_\mathrm{eff} \cdot \mathbf{v}$ is solely induced by the Doppler shift, and hence depends only on the projection of the velocity $\mathbf{v}$ onto the wave vector $\mathbf{k}_\mathrm{eff}$.

Thus, the three-dimensional integral, eq. (21), reduces to a one-dimensional integral over the x-component v_x of the velocity $\mathbf{v}$ if we align the x-axis with the direction of $\mathbf{k}_\mathrm{eff}$, giving rise to

$$P_\mathrm{e}(\tau) = \int \mathrm{d}v_x \, f_\mathrm{1D}(v_x) P_\mathrm{e}\left[\tau, \delta^{(2)}(v_x), \Omega_{12}\right], \tag{22}$$

with the one-dimensional Gaussian distribution

$$f_\mathrm{1D}(v_x) \equiv \frac{1}{(2\pi)^{1/2}\sigma_v} \exp\left[-\frac{(v_x - v_{x,0})^2}{2\sigma_v^2}\right]. \tag{23}$$

We emphasize that the x-component $v_{x,0}$ of the offset velocity $\mathbf{v}_0$ can be compensated by an adjustment of the frequency difference $\delta\omega$ of the two light fields.

Unfortunately, the velocity spread is not the only effect we need to account for. Indeed, the atomic ensemble has also a finite size and interacts with two laser beams having for example Gaussian intensity profiles

$$I_j(y, z) \equiv I_{0,j} \exp\left[-\frac{2(y^2 + z^2)}{w_j^2}\right] \tag{24}$$

in the (y, z)-plane, where $I_{0,j}$ is the amplitude and w_j is the radius of the beam for $j = 1, 2$.

The finite size of the laser beams causes a spatially dependence $\Omega_{12} = \Omega_{12}(\mathbf{r})$ of the Rabi frequency due to the dependence $I_j = I_j(\mathbf{r})$ of the intensity on the position $\mathbf{r}$ of the

atom within the two beams. As a result, for two collimated laser beams aligned along the x-direction, the excitation probability

$$P_{\mathrm{e}}\left[\tau, \delta^{(2)}, \Omega_{12}(\mathbf{r})\right] = P_{\mathrm{e}}\left[\tau, \delta^{(2)}, \Omega_{12}(y,z)\right] \tag{25}$$

only depends on the coordinates y and z perpendicular to $\mathbf{k}_{\mathrm{eff}}$.

Moreover, we model the spatial transverse distribution of the atomic ensemble by a Gaussian

$$s_{2\mathrm{D}}(y,z) \equiv \frac{1}{2\pi\sigma_y^2} \exp\left(-\frac{y^2+z^2}{2\sigma_y^2}\right) \tag{26}$$

of width $\sigma_y = \sigma_z$ centered at the maximum intensity of the beams ($y = z = 0$) in the (y,z)-plane.

When we combine eqs. (22) and (26), the total probability in three dimensions reads

$$P_{\mathrm{e}}(\tau) \equiv \iint \mathrm{d}y\,\mathrm{d}z \int \mathrm{d}v_x\, s_{2\mathrm{D}}(y,z) f_{1\mathrm{D}}(v_x) P_{\mathrm{e}}\left[\tau, \delta^{(2)}(v_x), \Omega_{12}(y,z)\right], \tag{27}$$

and is determined by the widths σ_v, σ_y, the laser beam radii w_j as well as the Rabi frequency Ω_{12} depending on the maximum laser intensity $I_{0,j}$.

Usually, the velocity distribution $f_{1\mathrm{D}}$ can be assumed to be time-independent if no external force is acting. However, the spatial distribution $s_{2\mathrm{D}}$ is always a function of time, since the cloud spreads due to the non-vanishing width σ_v of the velocity distribution. Hence, finding the three-dimensional probability P_{e} is more complicated if these assumptions are not valid. Even in the case of perfect monochromatic light fields, the efficiency of the coherent processes is fundamentally limited by the finite size and velocity of the cloud of atoms [82].

2˙2. *Optical lattices.* – In the preceding sections we have discussed the possibility of changing the atomic momentum with the help of the atom-light interaction leading to Bragg and Raman diffraction. However, there exists also the option of a *sequential* momentum transfer in an optical lattice. In particular, a large effect occurs due to Bloch oscillations in an accelerated optical lattice [106-109].

2˙2.1. Bloch theorem. In one space dimension we can obtain an optical lattice by retroreflecting a light field propagating in the x-direction and having the wave vector $\mathbf{k}$, from a mirror. This process leads to the formation of a standing light wave, and hence, to an effective periodic potential

$$V(x) \equiv 4V_{\mathrm{dip}} \sin^2(kx) = \frac{1}{2} V_0 [1 - \cos(2kx)] \tag{28}$$

for atoms with the amplitude $V_0 = 4V_{\mathrm{dip}}$, where V_{dip} is the magnitude of the atom-light interaction [110]. The factor of four results from the amplification of the electric field $\mathbf{E}$

by a factor of two due to the retro-reflection, and the quadratic scaling of the light field intensity $I \propto |\mathbf{E}|^2$.

The potential given by eq. (28) is periodic with the period $d \equiv \pi/k$ given in units of the wave number $k \equiv |\mathbf{k}| \equiv 2\pi/\lambda$, and the amplitude

$$V_0 \equiv \frac{\hbar\Gamma^2}{2\Delta}\frac{I}{I_{\mathrm{sat}}} \tag{29}$$

of the potential V can be expressed in terms of the intensity I, the saturation intensity I_{sat}, the natural linewidth Γ, and the detuning Δ. Here we have used eq. (1) and eq. (8) for the ac-Stark shift.

According to the Bloch theorem [111], the wave function

$$\psi_{\ell,q}(x) \equiv \mathrm{e}^{\mathrm{i}qx} u_{\ell,q}(x) \tag{30}$$

of an atom in a periodic potential V, eq. (28), is the product of a plane wave $\mathrm{e}^{\mathrm{i}qx}$ with the quasi-momentum $\hbar q$, and an amplitude $u_{\ell,q}(x) \equiv \langle x|\, u_{\ell,q}\rangle$ having the same period d as the original potential V. Here ℓ denotes the discrete band index.

The Bloch state $|u_{\ell,q}\rangle$ obeys the Schrödinger equation

$$\left[\frac{(\hat{p}+\hbar q)^2}{2m} + V(\hat{x})\right] |u_{\ell,q}\rangle = E_\ell(q)\, |u_{\ell,q}\rangle\,, \tag{31}$$

where the corresponding quasi-energy

$$E_\ell(q) = E_\ell\left(q + \frac{2\pi}{d}\right) \tag{32}$$

has a period $2\pi/d = 2k$ as a function of q. Therefore, following the convention of solid-state physics, the quasi-momentum $\hbar q$ can be restricted to the interval $(-\pi\hbar/d, +\pi\hbar/d] = (-\hbar k, +\hbar k]$, that is the first Brillouin zone [111].

In order to work in this interval, it is natural to consider atomic ensembles with a narrow momentum distribution, that is $m\sigma_v \ll \hbar k$. As an example, both a BEC [112], and a distribution of cold atoms prepared by a velocity filter in one dimension [106,107] fulfill this condition.

2˙2.2. Bloch oscillations. The Bloch states $|u_{\ell,q}\rangle$ following from eq. (31) and describing an atom in an optical lattice are stationary ones. Therefore, dynamics only occurs if an additional force F is acting, which can be either an external force, such as gravity, or an acceleration of the lattice itself.

When an atom in an optical lattice is suddenly exposed to a spatially uniform force F, the Bloch states $|u_{\ell,q}\rangle$ are no longer eigenstates [106,107] of the new Hamiltonian

$$\hat{H}_F \equiv \frac{\hat{p}^2}{2m} + V(\hat{x}) - F\hat{x}. \tag{33}$$

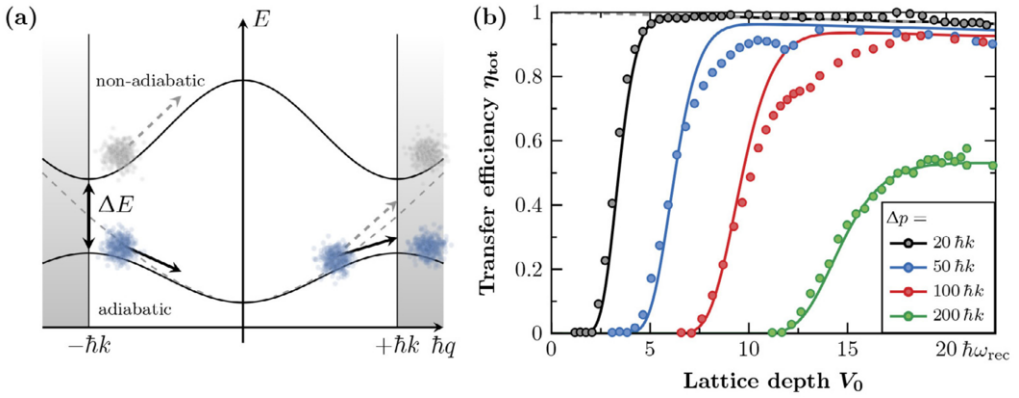


Fig. 3. – Bloch oscillations in the first Brillouin zone (a) caused by a force F acting on the center-of-mass motion of the atoms and transfer efficiency η_{tot} as a function (b) of the lattice depth V_0 for the different momentum transfers Δp. When the force is weak enough to avoid inter-band transitions, the atoms undergo an adiabatic acceleration for the time $\tau_{\mathrm{acc}} = 1\,\mathrm{ms}$ while they cross the individual Brillouin zones. Up to $\Delta p = 100\,\hbar k$ the fraction η_{tot} of atoms remaining in the target momentum state can be held above 0.9. The solid lines are given by the product $\eta_{\mathrm{tot}} = \eta_{\mathrm{sp}}\eta_{\mathrm{LZ}}$ normalized to the measured maximum transfer efficiency, where η_{sp} and η_{LZ} are defined by eqs. (40) and (42), accordingly. This figure is an adaptation of figs. 4.11 and 5.23 in ref. [101].

Indeed, for a small value of F, such that there are no inter-band transitions and the adiabatic assumption is valid, the wave function $\psi_{\ell,q(0)}$, eq. (30), evolves into

$$\psi_{\ell,q(\tau)}(x) = \exp\left\{-\frac{\mathrm{i}}{\hbar}\int_0^\tau E_\ell[q(t)]\mathrm{d}t\right\}\mathrm{e}^{\mathrm{i}q(\tau)x}u_{\ell,q(\tau)}(x), \tag{34}$$

and, apart from a phase factor and a linear shift of the quasi-momentum

$$\hbar q(\tau) \equiv \hbar q(0) + F\tau, \tag{35}$$

preserves its original form.

Because $\hbar q$ changes linearly in time τ, the wave function $\psi_{\ell,q(\tau)}$ has a temporal periodicity given by the Bloch period

$$\tau_{\mathrm{Bloch}} \equiv \frac{2\pi\hbar}{|F|d}, \tag{36}$$

which is the time after which the change $F\tau_{\mathrm{Bloch}}$ of the quasi-momentum is equal to the width $2\pi\hbar/d$ of the first Brillouin zone, displayed in fig. 3(a).

Moreover, since the quasi-energy $E_\ell = E_\ell(q)$, eq. (32), is a periodic function with period $2\pi/d$, the velocity

$$\langle v_\ell\rangle(q) \equiv \frac{1}{\hbar}\frac{\mathrm{d}E_\ell(q)}{\mathrm{d}q} \tag{37}$$

of an atom in the Bloch wave $u_{\ell,q(\tau)}$ has the same periodicity.

Furthermore, according to eq. (35) the quasi-momentum $\hbar q = \hbar q(\tau)$ is swept linearly in time τ. Hence, $\langle v_\ell \rangle[q(\tau)]$ is periodic in time with the Bloch period τ_{Bloch}, eq. (36). Its temporal average vanishes. This oscillation is called *Bloch oscillation* and fig. 3(a) shows its representation for the first Brillouin zone.

Bloch oscillations allow us to couple the momentum eigenstate $|p_0\rangle$ of a free particle to the momentum eigenstate $|p_{n_p} = p_0 + 2n_p\hbar k\rangle$, by sequentially transferring n_p times the momentum $2\hbar k$. The associated momentum transfer is the same as the one obtained when driving a sequence of n_p first-order Bragg pulses as discussed in sect. **2**˙1.3. For this purpose, $|p_0\rangle$ is adiabatically transferred to the lowest band of the Bloch lattice with $\ell = 0$ by increasing the lattice depth V_0. In order to avoid inter-band transitions this transfer has to be adiabatic, that is the change $\mathrm{d}V_0/\mathrm{d}\tau$ of the potential amplitude V_0, eq. (29), has to be much smaller than $\Delta E^2/\hbar$, where ΔE denotes the energy difference between the bands depicted in fig. 3(a),

Atoms in the fundamental band are then coherently accelerated by applying a linear chirp $2\pi\alpha$ to the frequency difference $\delta\omega \equiv \omega_1 - \omega_2$ of the two counterpropagating light waves with frequencies ω_1 and ω_2, that is [107]

$$\delta\omega(\tau) \equiv 2\pi\alpha\tau. \tag{38}$$

This chirp is equivalent to the effective force

$$|F| = \frac{2\pi m}{k_{\text{eff}}}|\alpha| \tag{39}$$

of eq. (33) with $k_{\text{eff}} \equiv 2k$.

The accelerated lattice generated by this chirp is very well controllable and allows us to efficiently couple the momentum states $|p_0\rangle$ and $|p_{n_p}\rangle$ by a sequential adiabatic transfer of the quasi-momentum $\hbar q$ of the lattice to the atoms. Finally, the atoms are unloaded from the lattice by slowly decreasing the potential amplitude V_0.

2˙2.3. Landau-Zener transitions. In order to estimate the efficiency of driving Bloch oscillations and the associated momentum transfer, we have to take into account two main loss mechanisms: Spontaneous emission and inter-band transitions.

Although the detuning Δ is large, still an appreciable fraction of atoms is lost due to spontaneous emission. We characterize the surviving fraction of atoms after the short acceleration time τ_{acc} by the parameter

$$\eta_{\text{sp}} \equiv 1 - R_{\text{sp}}\tau_{\text{acc}}\,, \tag{40}$$

where R_{sp} is defined by eq. (7).

The second loss mechanism results from the fact that the lattice is not infinitely deep and the momentum transfer is not fast enough. These deficiencies give rise to inter-band transitions. Indeed, for a lattice generated by two counter-propagating light waves with

the time-dependent frequency difference $\delta\omega$ given by eq. (38), the chirp α has to obey the adiabaticity criterion

$$|\alpha| \equiv \left| \frac{\mathrm{d}}{\mathrm{d}\tau} \frac{\delta\omega}{2\pi} \right| = n_p \frac{\omega_{\mathrm{rec}}}{\tau_{\mathrm{acc}}} \ll \frac{\Delta E^2}{\hbar^2}. \tag{41}$$

Here ΔE denotes the energy of the band gap shown in fig. 3(a) and we aim at transferring n_p pairs of photons during the time τ_{acc}.

We estimate the efficiency of the momentum transfer with the familiar Landau-Zener formula [113]

$$\eta_{\mathrm{LZ}} = \left[1 - \exp\left(-\frac{\pi}{2} \frac{\Delta E^2}{\hbar^2 \alpha} \right) \right]^{n_p}, \tag{42}$$

which provides us with the surviving fraction of atoms for a given chirp rate α and band gap energy ΔE. The power n_p in eq. (42) indicating the number of transitions reflects the fact that an inter-band transition may take place at each crossing of the Brillouin zone, that is whenever a photon pair is scattered.

The total fraction

$$\eta_{\mathrm{tot}} \equiv \eta_{\mathrm{sp}} \eta_{\mathrm{LZ}} \tag{43}$$

of surviving atoms after the acceleration is given by the product of η_{sp} and η_{LZ} defined by eqs. (40) and (42).

In fig. 3(b) we show by dots the measured efficiencies η_{tot} to transfer the momentum Δp as a function of the lattice depth V_0 for a fixed acceleration period $\tau_{\mathrm{acc}} = 1\,\mathrm{ms}$. The solid lines are based on eq. (43) normalized to the measured maximum transfer efficiency. For $\Delta p = 20\,\hbar k$ and $200\,\hbar k$ the solid line fits almost perfectly. However, for the data points with $\Delta p = 50\,\hbar k$ and $100\,\hbar k$, residual resonant tunneling is visible, which is not taken into account in the Landau-Zener formula, eq. (42).

In principle, there are no fundamental limits on the amount Δp of momentum that can be transferred by Bloch oscillations during a fixed time. It is a purely technical issue.

However, a limitation that is not easy to overcome is the spontaneous emission characterized by the rate R_{sp}, given by eq. (7). It effectively reduces the total fraction η_{tot} of atoms, since the laser detuning Δ and the laser power P cannot be increased arbitrarily. For a laser detuning $\Delta = 100\,\mathrm{GHz}$, even at the largest lattice depth $V_0 = 23\hbar\omega_{\mathrm{rec}}$, the relative contribution of spontaneous emission still stays at the few percent level. Up to an acceleration of $100\,\hbar k/\mathrm{ms}$, the transfer efficiencies per photon $\eta_{\mathrm{tot}}/\hbar k$ can be held at a level of $\eta_{\mathrm{tot}}/\hbar k > 0.999$. However, for larger values of the acceleration, the relative transfer efficiency starts to decrease due to a violation of the adiabaticity criterion, eq. (41).

2˙3. *Mach-Zehnder interferometer for gravity measurements.* – Measurements of gravity with atoms typically employ a MZI [1-4,114-116] which is similar to its original optical analogue and consists of three elements: two beam splitters and one mirror.

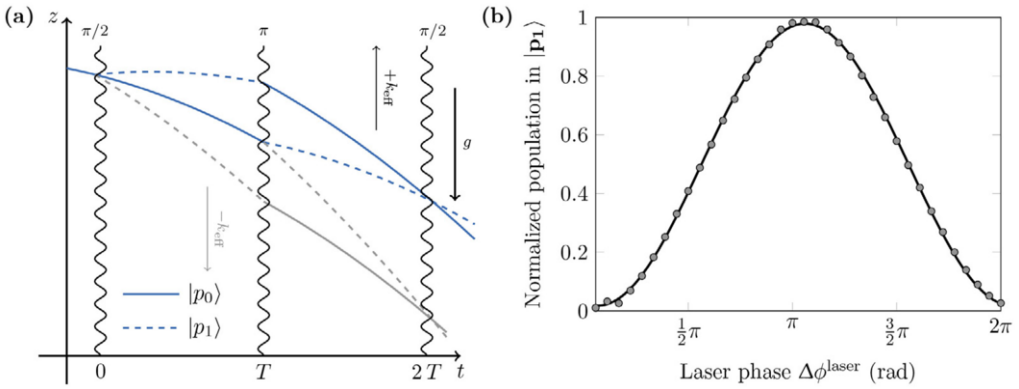


Fig. 4. – Mach-Zehnder interferometer (MZI) for atoms. The spacetime diagram (a) shows the atomic trajectories due to a three-pulse sequence consisting of i) an initial $\pi/2$-pulse to prepare a superposition of the two different momentum states $|p_0\rangle$ and $|p_1\rangle$ of the atomic center-of-mass motion, ii) a π-pulse to exchange the imprinted momenta, and iii) a final $\pi/2$-pulse to obtain the interference in the atomic population in an exit port. In (b) we depict a fringe scan for an adjusted laser phase $\Delta\phi^{\text{laser}}$ in the interval $[0, 2\pi]$. Figure (a) is reproduced from fig. 4.15(b) in ref. [101].

2˙3.1. Set-up. The first beam splitter generates a coherent superposition of two different momentum states $|\mathbf{p}_0\rangle$ and $|\mathbf{p}_1\rangle$ of the atomic center-of-mass motion. This superposition results in two spatially separated trajectories associated with each of the two momentum states, as depicted in fig. 4(a). After a time T, a mirror exchanges these two momentum states, leading to a redirection of the trajectories. Finally, after a total duration of $2T$, the two trajectories overlap again, and the second beam splitter is applied.

As discussed in sect. **2˙**1.1 we describe beam splitters and mirrors realized by the atom-light interaction in the Rabi formalism, and the corresponding pulse sequence for the MZI reads $\pi/2$-π-$\pi/2$. Here the two beam splitters ($\pi/2$-pulse) are separated from the mirror (π-pulse) by T. Usually, the momentum transfer $\hbar\mathbf{k}_{\text{eff}}$ induced by a beam splitter or mirror is parallel or antiparallel to the gravitational acceleration $\boldsymbol{g}$ as shown in fig. 4(a).

The interference signal of the interferometer is determined by the atomic population

$$P(\Delta\phi) = \frac{1}{2}[1 - C\cos(\Delta\phi)] \tag{44}$$

in the momentum state $|\mathbf{p}_1\rangle$ after the final beam splitter, where C and $\Delta\phi$ are the contrast and the total phase shift, respectively.

In the remainder of these lectures we consider only a perfectly closed interferometer, where the two wave packets propagated along the two trajectories perfectly overlap after the total interferometer time $2T$. In this case we obtain the maximal contrast $C = 1$.

2˙3.2. Contributions to phase shift. The total phase shift $\Delta\phi$ of the MZI reads [117,118]

$$\Delta\phi = \Delta\phi^{\text{laser}} + \Delta\phi^{\text{ac}} + \Delta\phi^{\text{2ph}} + \Delta\phi^{\text{inert}}, \tag{45}$$

where the laser phase

$$\Delta\phi^{\text{laser}} \equiv \phi_1 - 2\phi_2 + \phi_3 \tag{46}$$

is determined by the laser phases ϕ_1, ϕ_2, and ϕ_3 imprinted on the atom wave during the atom-light interaction due to the first, second, and third pulse. The phase $\Delta\phi^{\text{laser}}$ is of the form of a discrete second derivative. Unless ϕ_3 is used to modulate the signal, as depicted in fig. 4(b), the laser phase vanishes.

The single- and two-photon light shifts $\Delta\phi^{\text{ac}}$ and $\Delta\phi^{\text{2ph}}$ may lead to an offset shift, which in the first order depends on the difference in phase imprinted during the first and last pulse. In contrast to Raman diffraction, where the ratio of the intensities of the two frequency components needs to be properly adjusted [94,95], the phase $\Delta\phi^{\text{ac}}$ can be intrinsically suppressed in interferometers based on Bragg diffraction [96].

The two leading-order contributions to the phase shift

$$\Delta\phi^{\text{inert}} \equiv \Delta\phi^{\text{grav}} + \Delta\phi^{\text{rot}} \tag{47}$$

containing inertial effects, originate from the gravitational acceleration $\boldsymbol{g}$

$$\Delta\phi^{\text{grav}} \equiv \mathbf{k}_{\text{eff}} \cdot \boldsymbol{g}\, T^2, \tag{48}$$

and from the rotation of the Earth with frequency $\boldsymbol{\Omega}_{\text{E}}$

$$\Delta\phi^{\text{rot}} \equiv 2\mathbf{k}_{\text{eff}} \cdot (\boldsymbol{\Omega}_{\text{E}} \times \mathbf{v}_0) T^2, \tag{49}$$

representing the Sagnac effect. Here $\mathbf{v}_0$ is the velocity of the atoms before the first beam splitter.

2˙3.3. Influence of non-zero pulse duration. If the duration τ_{p} of a pulse is not negligibly short compared to the pulse separation time T, a modification of $\Delta\phi^{\text{grav}}$ given by eq. (48) is necessary [119-122]. For a Gaussian-shaped pulse of width $\sigma_{\tau_{\text{p}}}$ we find the new pulse separation time

$$T' \equiv T + \tau_{\text{p}} - \tau'_{\text{p}}\,, \tag{50}$$

where

$$\tau'_{\text{p}} \equiv \sqrt{2\pi}\sigma_{\tau_{\text{p}}} \tag{51}$$

is the duration of an equivalent box-shaped pulse covering the same area as the Gaussian-shaped one. Here we have assumed equally long π- and $\pi/2$-pulses, which, however, differ in intensity.

As a result the improved expression

$$\Delta\phi^{\mathrm{grav}} = \mathbf{k}_{\mathrm{eff}} \cdot \boldsymbol{g}\, T'^2 \Big[1 + \Big(1 + \frac{2}{\pi}\Big)\frac{\tau'_{\mathrm{p}}}{T'} + \dots\Big] \tag{52}$$

for $\Delta\phi^{\mathrm{grav}}$ represents a Taylor expansion in powers of τ'_{p}/T'.

Hence, the influence of the correction due to a non-zero pulse duration decreases for larger T, since the duration τ_{p} is independent of the separation T.

2˙3.4. Measurement of gravitational acceleration. The phase $\Delta\phi$ accumulated in an atom interferometer does not only depend on the laser phase $\Delta\phi^{\mathrm{laser}}$, but also on the time dependence of the laser frequency $\Delta\nu = \Delta\nu(\tau)$. In the following we restrict ourselves to a one-dimensional problem with the gravitational acceleration $g = |\boldsymbol{g}|$, and neglect for the time being the phase contributions $\Delta\phi^{\mathrm{laser}}$ and $\Delta\phi^{\mathrm{rot}}$.

During the free fall of the atom the resonance frequency of the transition changes due to the Doppler effect, eq. (11). This effect needs to be taken into account by a proper change of the detuning $\Delta\nu(\tau)$ after a certain free-fall time τ.

Indeed, the appropriate time dependence

$$\Delta\nu(\tau) = \frac{\delta\omega(\tau)}{2\pi} \tag{53}$$

of the chirp originates from the requirement for $\delta\omega$ to stay on resonance, eq. (18), where the Doppler frequency $\omega_{\mathrm{D}} = \omega_{\mathrm{D}}(0)$, eq. (11), is replaced by

$$\omega_{\mathrm{D}}(\tau) = \omega_{\mathrm{D}}(0) - k_{\mathrm{eff}} g \tau. \tag{54}$$

Chirping the laser frequency is not only a necessity to remain on resonance, but the chirp rate α also defines the acceleration $a = 2\pi\alpha/k_{\mathrm{eff}}$, compare to eq. (39), of the wavefronts of the Bragg lattice during free fall. In particular, the adjustment of $2\pi\alpha$ can be used to modulate the interferometer phase $\Delta\phi$ in a way that

$$\Delta\phi(\alpha, T) = \left(g - \frac{2\pi\alpha}{k_{\mathrm{eff}}}\right) k_{\mathrm{eff}} T^2. \tag{55}$$

It is the control of this chirp rate that allows us to realize an atomic gravimeter. Indeed, eq. (55) relates the output of the atom interferometer $P(\Delta\phi)$, eq. (44), depending on the gravitational acceleration g, to the two well-controlled parameters T and α. To extract g, we evaluate the dependence of the output signal $P(\Delta\phi)$ on one parameter with the other one fixed.

Figure 5 displays $P = P(\Delta\phi)$ *versus* the variation of the pulse separation time T for three different values of $2\pi\alpha/k_{\mathrm{eff}}$, namely 0, $3g/4$, and g. The closer the ratio $2\pi\alpha/k_{\mathrm{eff}}$ is

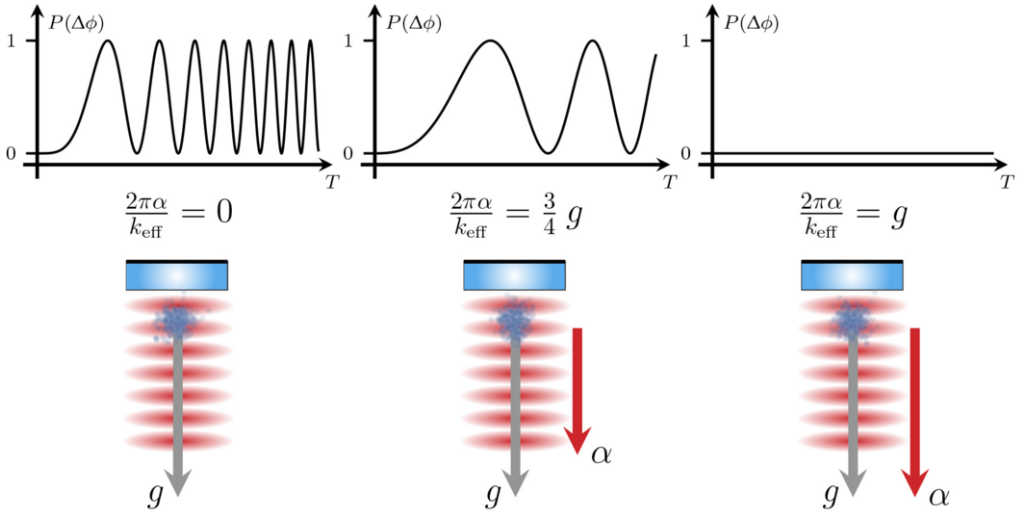


Fig. 5. – Local gravitational acceleration g obtained from the measurement of the chirp rate in an atomic gravimeter. The number of atoms in the exit port of an atom interferometer governed by $P = P(\Delta\phi)$ is counted for different chirp rates α of the laser frequency displayed here for the three lattice wavefront accelerations $2\pi\alpha/k_{\text{eff}} = 0$, $3g/4$, and g. For increasing pulse separation time T, a chirped sinusoidal oscillation is obtained. The chirp rate decreases with increasing α and finally vanishes for $2\pi\alpha/k_{\text{eff}} = g$. This picture is an adaptation of fig. 2.4 in ref. [95] and fig. 4.19 in ref. [101].

to the gravitational acceleration g, the slower is the oscillation chirp, and for $2\pi\alpha/k_{\text{eff}} = g$ the oscillations vanish completely.

When there is a significant mismatch between the chirp rate α and the free-fall rate g, a reduction in excitation efficiency due to a non-compensated Doppler shift appears. This effect was neglected in eq. (55) and fig. 5.

We conclude by noting that in a real experiment the ratio $2\pi\alpha/k_{\text{eff}}$ differs from the gravitational acceleration by less than one percent.

3. – Equivalence principle and atom interferometry

A prime application of an inertially sensitive atom interferometer is its use as a probe of Einstein's equivalence principle (EEP). We devote the present section to this topic and emphasize that the results summarized therein have originally been published in ref. [47].

First, we outline in sect. **3**˙1 a framework for testing the universality of free fall (UFF). We then present in sect. **3**˙2 an experiment based on a dual-species rubidium and potassium interferometer using atoms released from an optical molasses together with Raman diffraction to probe the UFF. Finally, we summarize our results and perform the associated data analysis in sect. **3**˙3.

3˙1. *Frameworks for tests of the universality of free fall.* – In spite of being 36 orders of magnitude weaker than the Coulomb interaction, gravitation dominates on a cosmological scale. Since astrophysical objects are electrically neutral gravitation governs the structure of our universe. Einstein's metric theory of gravity [123], that is general relativity, provides us with the tools necessary to understand a vast variety of astronomical phenomena and makes verifiable predictions, such as the existence of gravitational waves [124, 125]. As of today, however, a completely satisfactory microscopic theory of *quantum* gravity merging general relativity with quantum mechanics is still lacking.

The EEP is a cornerstone of general relativity, and unification attempts in general imply the violation of at least one of its three central assumptions: i) local position invariance, ii) local Lorentz invariance, and iii) the UFF. Ranging from the famous Pound-Rebka experiment [126] and gravity probe A [127], to the extremely sensitive torsion balance experiments [128] and lunar laser ranging [49, 129, 130], the EEP has been tested extensively.

Experiments employing matter wave interferometry have recently extended the landscape of classical UFF tests by entering the quantum domain. The UFF postulates the equality of inertial mass m_{in} and gravitational mass m_{gr}. Given two bodies A and B it can be tested by obtaining the Eötvös ratio

$$\eta_{\mathrm{A,B}} \equiv 2\frac{g_{\mathrm{A}} - g_{\mathrm{B}}}{g_{\mathrm{A}} + g_{\mathrm{B}}} = 2\frac{(\frac{m_{\mathrm{gr}}}{m_{\mathrm{in}}})_{\mathrm{A}} - (\frac{m_{\mathrm{gr}}}{m_{\mathrm{in}}})_{\mathrm{B}}}{(\frac{m_{\mathrm{gr}}}{m_{\mathrm{in}}})_{\mathrm{A}} + (\frac{m_{\mathrm{gr}}}{m_{\mathrm{in}}})_{\mathrm{B}}} \tag{56}$$

for their respective gravitational accelerations g_{A} and g_{B}.

Any non-zero measurement of $\eta_{\mathrm{A,B}}$ would imply a composition-dependent inequality of inertial and gravitational mass. Indeed, the values

$$\eta_{\mathrm{Earth,Moon}} = (-0.8 \pm 1.3) \cdot 10^{-13} \tag{57}$$

based on Lunar-Laser-Ranging [49, 129, 130] and

$$\eta_{\mathrm{Be,Ti}} = (0.3 \pm 1.8) \cdot 10^{-13} \tag{58}$$

employing torsion balances [128] have provided the best constraints on violations of the UFF for a number of years.

Alternatively, one can also compare pairs of freely-falling test masses. The best result on ground [131] corresponds to

$$\eta_{\mathrm{Cu,U}} = (1.3 \pm 5.0) \cdot 10^{-10}. \tag{59}$$

Moreover, a recent test in space [51, 132] obtained

$$\eta_{\mathrm{Ti,Pt}} = [-0.1 \pm 0.9(\mathrm{stat}) \pm 0.9(\mathrm{syst})] \cdot 10^{-14} \tag{60}$$

corresponding to an order of magnitude improvement of the bounds set by previous experiments.

The tests listed above employ classical test masses. In analogy to the first observation of a gravitation-induced phase in a neutron interferometer [133], the acceleration measured by the interference of a quantum object can be compared to that of a classical object [3, 43], or the differential gravitational phase shifts between two quantum objects [18, 44, 46, 48, 134] can be exploited to search for violations of the UFF. Beyond the comparison of two isotopes of strontium in 2014 [46], the latter approach was extended to comparing the free-fall acceleration of the two different elements rubidium and potassium [47].

Experiments using quantum objects are beneficial because they are generally subject to systematic effects different from those dominating classical tests. Moreover, unique properties of quantum objects, such as a macroscopic coherence length [135] and spin polarization [136-139] can be tested as features possibly coupling to EEP-violating effects. Furthermore, the set of available test masses is enhanced by those that can be laser-cooled and chosen as to maximize the sensitivity to violations. This approach can be clearly illustrated by dilaton models [140], and the so-called Standard Model Extension (SME) [141,142], which both provide consistent frameworks for parametrizing violations of the EEP.

Violations of the UFF can be naturally parameterized by writing the acceleration g_{X} of a species X as

$$g_{\mathrm{X}} = (1 + \beta_{\mathrm{X}})g, \tag{61}$$

where β_{X} is a small parameter which is species dependent, and vanishes in the absence of UFF violations.

For dilaton models one has

$$\beta_{\mathrm{X}} = D_1 Q'^1_{\mathrm{X}} + D_2 Q'^2_{\mathrm{X}}, \tag{62}$$

where D_1 and D_2 are fundamental violation parameters, whereas Q'^1_{X} and Q'^2_{X} are effective charges that mainly depend on the proton and neutron numbers of species X [140].

The Eötvös ratio for two species A and B is then given by

$$\eta_{\mathrm{A,B}} \approx \beta_{\mathrm{A}} - \beta_{\mathrm{B}} = D_1(Q'^1_{\mathrm{A}} - Q'^1_{\mathrm{B}}) + D_2(Q'^2_{\mathrm{A}} - Q'^2_{\mathrm{B}}), \tag{63}$$

and the differences of the effective charges, which determine the sensitivity to violations associated with D_1 or D_2 are listed in table I for several test pairs.

Similarly, for the SME one has

$$\beta_X = f_{\beta^{e+p-n}_{\mathrm{X}}}\beta^{e+p-n} + f_{\beta^{e+p+n}_{\mathrm{X}}}\beta^{e+p+n} + f_{\beta^{\bar{e}+\bar{p}-\bar{n}}_{\mathrm{X}}}\beta^{\bar{e}+\bar{p}-\bar{n}} + f_{\beta^{\bar{e}+\bar{p}+\bar{n}}_{\mathrm{X}}}\beta^{\bar{e}+\bar{p}+\bar{n}}, \tag{64}$$

where β^{e+p-n}, β^{e+p+n}, $\beta^{\bar{e}+\bar{p}-\bar{n}}$, and $\beta^{\bar{e}+\bar{p}+\bar{n}}$ parametrize the violations for various weighted combinations of elementary particles. Moreover, the sensitivity factors $f_{\beta^{e+p-n}_{\mathrm{X}}}$

Table I. – *Comparison of test masses* A *and* B *analyzed in the dilaton model. The charges* Q'^1_X *and* Q'^2_X *with* X *being either* A *or* B *are calculated according to ref.* [140]. *A larger absolute number corresponds to a larger anomalous acceleration, and thus a higher sensitivity to violations of the* EEP. *For* Ti *and* Cu *natural occurrence of isotopes is assumed. This table is a reproduction of table* 2.1 *in ref.* [95].

A	B	Ref.	$(Q'^1_A - Q'^1_B) \cdot 10^4$	$(Q'^2_A - Q'^2_B) \cdot 10^4$
^{9}Be	Ti	[128]	−15.46	−71.20
Cu	^{238}U	[131]	−19.09	−28.62
^{85}Rb	^{87}Rb	[18, 44]	0.84	−0.79
^{87}Sr	^{88}Sr	[46]	0.42	−0.39
^{6}Li	^{7}Li[(a)]	[143]	0.79	−10.07
^{39}K	^{87}Rb	[47]	−6.69	−23.69

[(a)] A UFF test comparing ^{6}Li *vs.* ^{7}Li has not yet been performed.

($f_{\beta_X^{\bar{e}+\bar{p}-\bar{n}}}$) and $f_{\beta_X^{e+p+n}}$ ($f_{\beta_X^{\bar{e}+\bar{p}+\bar{n}}}$) are charges related to the neutron excess, and the overall baryon number in a given normal matter (antimatter) nucleus [142].

The corresponding Eötvös ratio reads

$$\eta_{A,B} \approx \beta_A - \beta_B = \Delta f_{-n}\beta^{e+p-n} + \Delta f_{+n}\beta^{e+p+n} + \Delta \bar{f}_{-n}\beta^{\bar{e}+\bar{p}-\bar{n}} + \Delta \bar{f}_{+n}\beta^{\bar{e}+\bar{p}+\bar{n}}, \tag{65}$$

with

$$\begin{aligned}
\Delta f_{-n} &\equiv f_{\beta_A^{e+p-n}} - f_{\beta_B^{e+p-n}}, \\
\Delta f_{+n} &\equiv f_{\beta_A^{e+p+n}} - f_{\beta_B^{e+p+n}}, \\
\Delta \bar{f}_{-n} &\equiv f_{\beta_A^{\bar{e}+\bar{p}-\bar{n}}} - f_{\beta_B^{\bar{e}+\bar{p}-\bar{n}}}, \\
\Delta \bar{f}_{+n} &\equiv f_{\beta_A^{\bar{e}+\bar{p}+\bar{n}}} - f_{\beta_B^{\bar{e}+\bar{p}+\bar{n}}}.
\end{aligned} \tag{66}$$

These differences of sensitivity factors are listed in table II for relevant test pairs.

Tables I and II clearly show that the choice of test masses heavily influences the achievable impact on global violation bounds. Specifically, this information allows us to use novel test pairs that were previously unconstrained, or only weakly constrained, to improve the bounds on certain violation parameters [142, 146]. Here, a new, and independent result can have an enormous impact on the global model even if it does not reach the state-of-the-art sensitivity.

TABLE II. – *Comparison of test masses* A *and* B *analyzed in the Standard Model Extension. The sensitivity factors* Δf_{-n}, Δf_{+n}, $\Delta \bar{f}_{-n}$, *and* $\Delta \bar{f}_{+n}$ *are calculated according to ref.* [142]. *Relevant nuclide data is taken from ref.* [144]. *A larger absolute number corresponds to a larger anomalous acceleration, and thus higher sensitivity to violations of the* EEP. *For* Ti *and* Cu *natural occurrence of isotopes* [145] *is assumed. This table is a reproduction of table* 2.2 *in ref.* [95].

A	B	Ref.	$\Delta f_{-n} \cdot 10^2$	$\Delta f_{+n} \cdot 10^4$	$\Delta \bar{f}_{-n} \cdot 10^5$	$\Delta \bar{f}_{+n} \cdot 10^4$
^{9}Be	Ti	[128]	1.48	−4.16	−0.24	−16.24
Cu	^{238}U	[131]	−7.08	−8.31	−89.89	−2.38
^{85}Rb	^{87}Rb	[18, 44]	−1.01	1.81	1.04	1.67
^{87}Sr	^{88}Sr	[46]	−0.49	2.04	0.81	1.85
^{6}Li	^{7}Li[(a)]	[143]	−7.26	7.79	−72.05	5.82
^{39}K	^{87}Rb	[47]	−6.31	1.90	−62.30	0.64

[(a)] A UFF test comparing ^{6}Li *vs.* ^{7}Li has not yet been performed.

3˙2. *Simultaneous* ^{87}Rb *and* ^{39}K *interferometer*. – To test the UFF, we operate two Mach-Zehnder-type matter wave interferometers [1] with laser-cooled ^{87}Rb and ^{39}K as shown in fig. 6. To leading order, the phase shift in each interferometer reads

$$\Delta\phi_i = \left(g_i - \frac{\alpha_i}{k_{\mathrm{eff},i}}\right) k_{\mathrm{eff},i} T^2, \tag{67}$$

and can be deduced by counting the atoms in the exit ports. Here the effective wavefront acceleration $\alpha_i/k_{\mathrm{eff},i}$ introduced by linearly ramping the Raman laser frequency difference is utilized to null the phase shift induced by the gravitational acceleration. We emphasize that the definition of the chirp rate α_i differs by a factor of 2π from the definition of the chirp rate α in sect. **2**˙3.4.

An ensemble of $8 \cdot 10^8$ atoms ($3 \cdot 10^7$ atoms) of rubidium (potassium) atoms is collected from a two-dimensional magnetooptical trap, then sub-Doppler cooled [147-149] to $T_{\mathrm{Rb}} = 27\,\mu$ K ($T_{\mathrm{K}} = 32\,\mu$ K), and after being optically pumped to the $|F_i = 1\rangle$ manifold, released into free fall. Three two-photon Raman pulses separated in time by T are applied to coherently split, redirect, and recombine the atoms during free fall. For detection, the population in $|F_i = 2\rangle$ as well as the total population in $|F_i = 1\rangle$ and $|F_i = 2\rangle$ are obtained via state-selective fluorescence detection, yielding the normalized excitation probability. The overall cycle time is about 1.6 s

A global minimum of the interference fringes appears independently of the free evolution time T, when the condition $g_i - \alpha_i/k_{\mathrm{eff},i} = 0$ is fulfilled. We display this concept for the upward and downward directions of the momentum transfer([1]) in fig. 7.

([1]) The differential signal of the upward and downward directions of the momentum transfer allows us to suppress [31, 118] spurious phase shifts independent of the direction of k_{eff}.

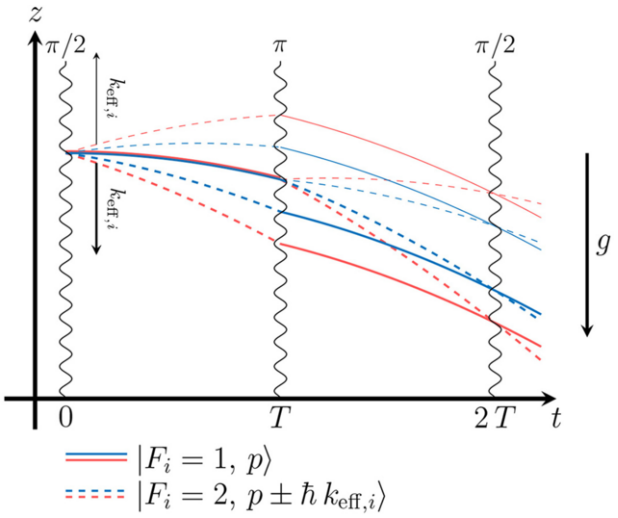


Fig. 6. – Spacetime diagram of a dual-species Mach-Zehnder atom interferometer in a constant gravitational field for the downward (thick lines) and upward (thin lines) direction of the momentum transfer. Stimulated Raman transitions at times 0, T, and $2\,T$ couple the states $|F_i = 1,\ p\rangle$ and $|F_i = 2,\ p \pm \hbar\, k_{\mathrm{eff},i}\rangle$, where i stands for Rb (blue lines) or K (red lines). The velocity change induced by the Raman pulses is not to scale compared to the gravitational acceleration. This figure is reproduced from fig. 5.1 in ref. [95].

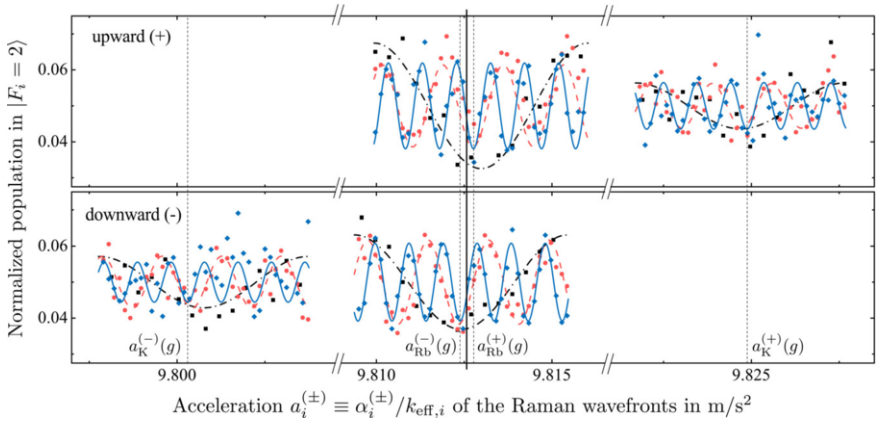


Fig. 7. – Determination of the differential gravitational acceleration of rubidium and potassium. Typical interference signals and sinusoidal fits as a function of the effective Raman wavefront acceleration are shown for pulse separation times $T = 8$ ms (black squares and solid black line), $T = 15$ ms (red circles and dashed red line), and $T = 20$ ms (blue diamonds and dotted blue line) both for the upward (+) and downward (−) directions of the momentum transfer. The central fringe positions $a_i^{(\pm)}(g)$ (dashed vertical lines), where i is Rb or K, are shifted symmetrically around $g_i = [a_i^{(+)}(g) - a_i^{(-)}(g)]/2$ (solid vertical line). The data sets are corrected for slow linear drifts and offsets. This figure is reproduced from fig. 5.2 in ref. [95].

3˙3. *Data analysis and result.* – Over a duration of about four hours we have tracked the central fringe position $a_i^{(\pm)}(g)$ by scanning across the minimum in 10 steps with alternating directions of the momentum transfer at a pulse separation time of $T = 20\,\mathrm{ms}$. Two scans yield $g_i = [a_i^{(+)}(g) - a_i^{(-)}(g)]/2$ and can be used to compute the Eötvös ratio, eq. (56).

Systematic effects affecting our measurement are listed in table III. The total bias $\Delta\eta_{\mathrm{tot}} = -5.4 \cdot 10^{-8}$ is subject to an uncertainty $\delta\eta_{\mathrm{tot}} = 3.1 \cdot 10^{-8}$. Considering the systematic and statistical uncertainty as well as the bias from table III, an overall result of

$$\eta_{\mathrm{Rb,K}} = (0.3 \pm 5.4) \cdot 10^{-7} \tag{68}$$

is obtained.

The column $\delta\eta^{\mathrm{adv}}$ points to possible future improvements (indicated in bold face) using a common optical dipole trap [150] as a source to further cool and collocate rubidium as well as potassium, and gain better control over their initial conditions.

The statistical uncertainty $\sigma_\eta = 5.4 \cdot 10^{-7}$ after 4096 s of integration time is dominated by technical noise in the potassium interferometer. This limitation can be improved in several ways.

For instance, the implementation of a sequence preparing a single m_F state would reduce the number of background atoms that currently lower the contrast. Moreover, selecting a narrower velocity class, as well as lower temperatures to begin with, would improve the beam splitting efficiency and consequently the contrast, too.

TABLE III. – *Systematic biases $\Delta\eta$ and comparison between the uncertainties $\delta\eta$ and $\delta\eta^{\mathrm{adv}}$ of the Eötvös ratio in the current, and in an advanced set-up. The improved values highlighted in bold face arise from the use of an optical dipole trap. The uncertainties are assumed to be uncorrelated at the level of the inaccuracy. This table is a reproduction of table* 5.1 *in ref.* [95].

Contribution	$\Delta\eta$	$\delta\eta$	$\delta\eta^{\mathrm{adv}}$
Second-order Zeeman effect	$-5.8 \cdot 10^{-8}$	$2.6 \cdot 10^{-8}$	$\mathbf{3.0 \cdot 10^{-9}}$
Wavefront aberration	0	$1.2 \cdot 10^{-8}$	$\mathbf{3.0 \cdot 10^{-9}}$
Coriolis force	0	$9.1 \cdot 10^{-9}$	$\mathbf{1.0 \cdot 10^{-11}}$
Two-photon light shift	$4.1 \cdot 10^{-9}$	$8.2 \cdot 10^{-11}$	$8.2 \cdot 10^{-11}$
Effective wave vector	0	$1.3 \cdot 10^{-9}$	$1.3 \cdot 10^{-9}$
First-order gravity gradient	0	$9.5 \cdot 10^{-11}$	$\mathbf{1.0 \cdot 10^{-12}}$
Total	$-5.4 \cdot 10^{-8}$	$3.1 \cdot 10^{-8}$	$\mathbf{4.4 \cdot 10^{-9}}$

4. – Atom-chip–based BEC interferometry

Today's generation of inertial sensors based on atom optics typically operates with cold atoms, released or launched from an optical molasses exemplified by the experiment discussed in the previous section. The velocity distribution and the finite size of these sources limit the efficiency of the beam splitters as well as the analysis of the systematic uncertainties. We can overcome these limitations by employing ensembles with a momentum distribution well below the photon recoil limit, which can be achieved with BECs.

After reaching the regime of ballistic expansion, where the mean-field energy has been converted into kinetic energy, the momentum distribution of a BEC can be narrowed down even further by the technique of delta-kick collimation (DKC) [85, 151-154]. Moreover, atom-chip technology offers the possibility to generate a BEC and perform DKC in a fast and reliable fashion, paving the way for miniaturized atomic devices. We devote sect. **4**˙1 to an introduction to DKC and note that the results reported in this section have originally been published in ref. [85].

The use of BECs allows us to implement Bragg and double Bragg diffraction with efficiencies above 95%, and thus to perform interferometry with high contrast. In sect. **4**˙2 we demonstrate a quantum tiltmeter using a MZI based on double Bragg diffraction with a tilt precision of up to 4.5 μrad. The results presented in this section have originally been published in ref. [86].

In sect. **4**˙3 we discuss an experiment combining double Bragg diffraction and Bloch oscillations, where we have implemented [87] a relaunch procedure with more than 75% efficiency in a retro-reflected optical lattice. We emphasize that here we rely on a *single* laser beam only, which also comprises the beam splitter resulting in a set-up of significantly reduced complexity.

Our relaunch technique allowed us to build a gravimeter on a small baseline with a comparably large interferometry time of $2T = 50\,\mathrm{ms}$ in the MZI for a fixed free-fall distance. At a high contrast of $C = 0.8$ the interferometer reaches an intrinsic sensitivity to gravity of

$$\Delta g/g = 1.4 \cdot 10^{-7}. \tag{69}$$

A key element of this result was the state preparation comprising DKC and Stern-Gerlach-type deflection, which improved the contrast and reduced the detection noise. The results presented in this section have originally been reported in ref. [87].

4˙1. *Delta-kick collimation.* – The desire to reach long expansion times serves as a motivation for DKC by a magnetic lens [155]. Recent experiments have shown expansion rates corresponding to a few nK in 3D in the drop tower [85] with QUANTUS-1, or even pK in 2D in a 10 m fountain [154]. These widths in momentum space are smaller than those of the coldest reported condensates [156]. For a detailed study of DKC using the QUANTUS-1 apparatus in the drop tower and also on ground we refer to refs. [157, 158].

After the release from the trap, the BEC starts to expand freely and falls away from the trap due to the gravitational acceleration. During the first milliseconds after release

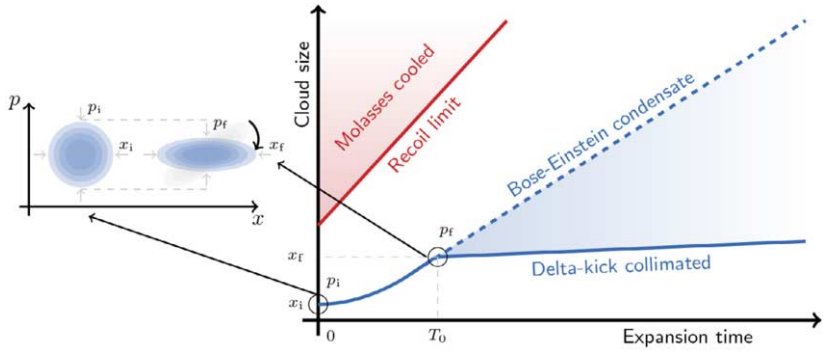


Fig. 8. – Principle of delta-kick collimation (DKC) explained in phase space. After release an ensemble of cold atoms has an initial distribution in the phase space spanned by position x and momentum p. After a time T_0 the cloud has expanded in space giving rise to a tilted ellipse which is then rotated in phase space by a collimating pulse. This figure is an adaptation of fig. 1.3(c) in ref. [101].

most of the mean-field energy is converted into kinetic energy. The required time for this conversion depends on the atomic density of the condensate, and hence, on the steepness of the trap from which the BEC is released. The final expansion rates in the ballistic regime can be precisely evaluated by time-of-flight measurements.

Even a BEC has a non-zero velocity spread leading after some expansion time to an increased cloud size, and possibly to a reduction in performance of an atom interferometer due to an increased detection noise, or larger contributions to systematic uncertainties. A condensate released from a shallow trap expands slowly enough to perform experiments with short free-fall times. Moreover, these condensates have a momentum width that is small enough to reach high Bragg diffraction efficiencies.

Indeed, for trapping frequencies of $f_{(x,y,z)} = (18, 46, 31)\,\mathrm{Hz}$ and $N = 10^4$ atoms the resulting expansion rate along the beam-splitting axis is $\sigma_v \approx 750\,\mu\mathrm{m/s} \approx 0.125\,\hbar k/m$, which corresponds to an effective temperature in the beam direction of about 5 to 10 nK. Thus, for this number of atoms the mean-field conversion in a time of $t_{\mathrm{exp}} \gtrsim 10\,\mathrm{ms}$ is acceptable for interferometry.

However, for larger densities of the condensate, the use of DKC becomes necessary. In fig. 8 we illustrate the essential idea of DKC with the time evolution of a phase-space distribution. The shearing in phase space caused by the free evolution of the condensate leads to a tilted ellipse that can be rotated by applying a harmonic potential for a suitable time τ_{DKC}, so as to align its major axis with the x-axis.

Indeed, when we consider the potential $V_{\mathrm{DKC}}(x) \equiv m\omega_{\mathrm{DKC}}^2 x^2/2$ applied for a short time τ_{DKC}, the resulting change Δp in the momentum p is approximately given by

$$\Delta p \cong F(x)\tau_{\mathrm{DKC}} = -\frac{\mathrm{d}V_{\mathrm{DKC}}}{\mathrm{d}x}\tau_{\mathrm{DKC}} = -m\omega_{\mathrm{DKC}}^2 x\,\tau_{\mathrm{DKC}}. \tag{70}$$

Assuming that most of the expansion takes place in the ballistic regime, and that the spatial size x_f when the DKC pulse is applied at time T_0 is much larger than the initial size x_i, such that $x_\mathrm{f} \approx p_\mathrm{i} T_0/m$, the induced momentum change reads

$$\Delta p = -p_\mathrm{i} \omega_\mathrm{DKC}^2 \tau_\mathrm{DKC} T_0. \tag{71}$$

The initial momentum width p_i considered here includes the contribution from the mean-field energy that was converted into kinetic energy.

In complete analogy to the power P of an optical lens the strength of a delta kick is defined by $S \equiv \omega_\mathrm{DKC}^2 \tau_\mathrm{DKC}$. For a given T_0 the optimal choice that leads to $\Delta p = -p_\mathrm{i}$ corresponds to $S = 1/T_0$, or equivalently, to

$$\omega_\mathrm{DKC}^2 \tau_\mathrm{DKC} T_0 = 1. \tag{72}$$

However, even in this case the minimum value attainable for the final momentum width is limited by Liouville's theorem, that is conservation of phase-space volume. This minimal value is determined by the ratio of the phase-space volume of the initial ensemble, and the spatial width x_f when the DKC pulse is applied. In principle this limitation can be reduced by increasing T_0 which leads to a larger x_f.

In a realistic implementation uncertainties in the relevant parameters lead to an uncertainty δp in the induced momentum change which should be smaller than the targeted final momentum width p_f, namely

$$\delta p = p_\mathrm{i} \delta(\omega_\mathrm{DKC}^2 \tau_\mathrm{DKC} T_0) < p_\mathrm{f}. \tag{73}$$

When we take into account the optimal choice, eq. (72), the requirement, eq. (73), can be expressed in terms of the relative errors for the individual parameters, that is

$$2\frac{\delta\omega_\mathrm{DKC}}{\omega_\mathrm{DKC}} + \frac{\delta\tau_\mathrm{DKC}}{\tau_\mathrm{DKC}} + \frac{\delta T_0}{T_0} < \frac{p_\mathrm{f}}{p_\mathrm{i}}\,. \tag{74}$$

For ground-based experiments, the waiting time T_0 and other quantities determining the errors are not only limited by technical means, but also by the free fall away from the chip, which reduces the trap frequencies of the potential, and restricts T_0 to times smaller than 6 ms. For a shallow trap the expansion after this time is not even in the ballistic regime, and the mean-field potential is still non-negligible. Thus, a release from the trap with a faster initial expansion is required to increase x_f prior to collimation.

We conclude our brief review of DKC by mentioning that the anharmonicities of the generated potential represent another source of errors. Indeed, they cause deformations when the condensates are too large.

4˙2. *Quantum tiltmeter based on double Bragg diffraction.* – An interesting extension of Bragg diffraction occurs when a retro-reflected pair of laser beams interacts with atoms that are at rest with respect to the mirror, such that the Doppler shift ω_D vanishes. In this case the transitions with opposing effective wave vectors are degenerate in frequency, and simultaneously diffract the atomic wave packets in both directions. The difference in momenta between both arms in an interferometer is then increased to $4\hbar k$. This symmetric diffraction called "double Bragg diffraction" was proposed in ref. [98] as a generalization of Bragg diffraction in complete analogy to double Raman diffraction [97], and experimentally demonstrated in ref. [86]. We emphasize that the traditional Bragg diffraction can be referred to as single or uni-directional Bragg diffraction, since only a single pair of laser beams drives the transition, while the other one is off resonant.

4˙2.1. Rabi oscillations. In fig. 9(a) we show the coupling scheme for first-order double Bragg diffraction. The transition frequency is given by the Bragg condition, eq. (14), with the recoil frequency ω_{rec}.

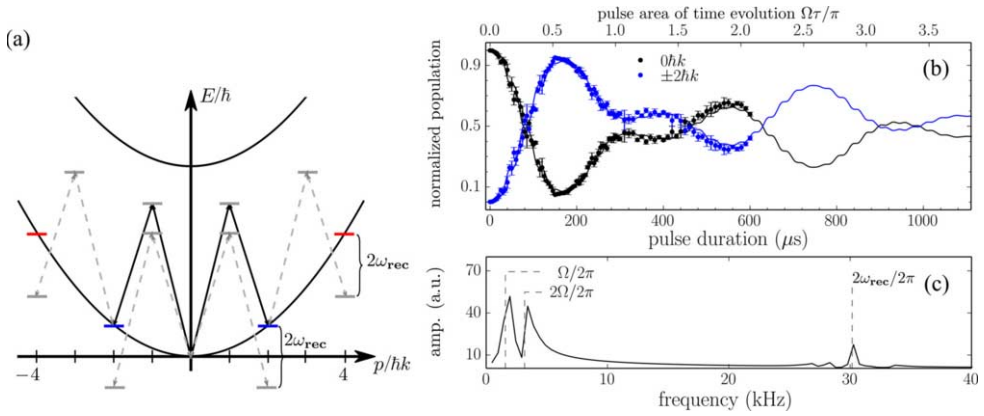


Fig. 9. – First-order double Bragg diffraction represented by the corresponding transitions (a), the comparison (b) between experimental observations and numerical simulations of the Rabi oscillations in the normalized populations as a function of the square-pulse duration, and their spectral decomposition (c). The energy diagram in (a) shows the resonant (solid lines) and off-resonant (dashed lines) light-induced transitions between the atomic momentum states $|0\hbar k\rangle$ (black), $|\pm 2\hbar k\rangle$ (blue) and $|\pm 4\hbar k\rangle$ (red). The experimental values (squares) and numerical simulations (solid curves) based on ref. [98] displayed in (b) are in good agreement. The frequency spectrum (c) of the simulated population N_1/N_{tot} displays components close to $2\omega_{rec}$ which stem from the off-resonant couplings depicted in (a) by dashed lines. The broad double-peaked structure at the Rabi, and twice the Rabi frequency, which is a consequence of the detuned three-level system with non-vanishing p_0 and/or δp, leads to the modulation of the oscillations observed in (b). This figure is reproduced from ref. [86] with permission of the authors, copyright American Physical Society (2016).

Due to the intrinsic symmetry of the diffraction process, a beam splitter of this kind offers several advantages for atom interferometry. Most prominently, the populations of the output ports no longer depend on the laser phase $\Delta\phi^{\text{laser}}$, since the wave functions of the center-of-mass motion in the two arms have the same laser phase imprinted during each pulse([2]). Moreover, we can choose the order of the diffraction process by matching the detuning δ with the kinetic energy gained during the scattering.

Rabi oscillations for double Bragg diffraction are more complicated than for single Bragg diffraction, since more states and transitions have to be taken into account. In the case of first-order double Bragg diffraction δ is chosen to correspond to the recoil frequency $\omega_{\text{rec}} \equiv (2\hbar k)^2/(2m\hbar)$ inducing resonant transitions between the momentum states $|0\rangle$ and $|\pm 2\hbar k\rangle$ depicted in fig. 9(a) by solid lines, together with off-resonant transitions to these, and higher momentum states indicated by dashed lines. Being off-resonant, the latter transitions are substantially suppressed.

When the width of the momentum distribution is small enough we can observe Rabi-type oscillations [86] as a function of the atom-light interaction time as shown in fig. 9(b). Here we also display numerical simulations [98] of the Rabi oscillations and the normalized atom populations are defined as

$$n_0 \equiv N_0/N_{\text{tot}} \qquad \text{and} \qquad n_j \equiv (N_{-j} + N_j)/N_{\text{tot}} \quad (j = 1, 2), \tag{75}$$

where N_j is the number of atoms in the momentum state $|2j\hbar k\rangle$ with $j = -2, -1, 0, 1$ and 2, and $N_{\text{tot}} \equiv \sum_{j=-2}^{2} N_j$.

4˙2.2. Tilt measurements. In most atom interferometers tilt variations can be significant and lead to an uncertainty in the quantity being measured. Hence, these devices are either designed to minimize the effect of tilts or do not allow to distinguish a tilt from other sources inducing a phase shift [13].

We have designed an atom interferometer to measure slight deviations from the horizontal direction with respect to gravity. In this quantum tiltmeter we diffract a delta-kick collimated BEC with small initial momentum and low expansion rate from laser beams and induce first- or higher-order double Bragg transitions.

Figure 10(a) illustrates the corresponding symmetric geometry emerging from a first-order (blue solid lines) double Bragg process, where the initial wave packet is split, redirected and recombined. A stepwise tilt of the whole apparatus changes the orientation α of the interferometer with respect to gravity g and induces a phase shift. As a result, the interference signal, that is the normalized populations at the exit ports, exhibits oscillations as a function of the change $\Delta\alpha$ in the tilt angle exemplified by fig. 10(b). This process can be extended to successive first-order (black dashed lines) and second-order (red dotted lines) double Bragg processes as displayed in fig. 10(c) and (d).

([2]) Needless to say, this feature is only present when we neglect off-resonant processes.

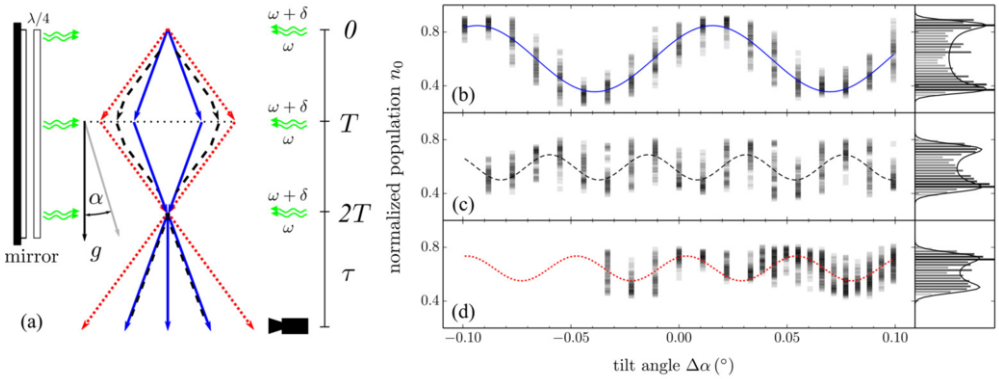


Fig. 10. – Three double Bragg interferometers employed as tiltmeters (a) and the interference signals corresponding to first-order (b), successive first-order (c), and second-order Gaussian pulses (d) as a function of the tilt angle $\Delta\alpha$. For each angular step in $\Delta\alpha$, the normalized population n_0 in the exit ports is measured 50 times. The blue solid, black dashed and red dotted lines represent sinusoidal fits of those datasets. Fitting the histograms of n_0 over a range of tilt settings corresponding to one, or two complete fringe periods, and fitting them with a theoretical distribution (black) which assumes that all noise sources combined with the tilt scan lead to an approximately uniform phase-shift distribution, yield contrasts of 43%, 29% and 23%, respectively. Further analysis reveals a tilt precision of 4.5 μrad, 5.9 μrad, and 4.6 μrad, respectively. This figure is reproduced from ref. [86] with permission of the authors, copyright American Physical Society (2016).

4˙3. *Sensitive atom-chip gravimeter on a compact baseline.* – Quantum sensors for gravimetry based on cold atoms have been with us for more than two decades [2, 3, 23], and can reach today accuracies competitive with falling corner cube gravimeters [43]. However, *compact* gravimeters using BECs have been demonstrated only recently [11, 87].

Due to their small extent and expansion rates, BECs which are delta-kick collimated have attracted attention for large-scale devices on ground [159], and in space missions [160]. In light of these experiments systematic uncertainties specific to BECs have been analyzed [161, 162], and novel techniques have been introduced [163]. These achievements will allow sensors relying on BECs to target sub-μGal accuracies in the near future, and to overcome current limitations set by cold atoms [14, 23, 71, 72].

The use of an atom chip for all preparation steps, and as a retro-reflector is the novelty of our approach [87] summarized in this section. Although our experiment is a proof-of-principle, it nevertheless represents an important pathway to the application of an atom-chip gravimeter for precision measurements important for example in geodesy.

4˙3.1. Relaunch of atoms in a retro-reflected optical lattice. In order to increase the observation time of a BEC on a small baseline, we have developed a simple but extremely valuable method to relaunch a BEC using Bloch oscillations in a retroreflected or dual lattice. Our procedure differs substantially from previous ones, which either rely on

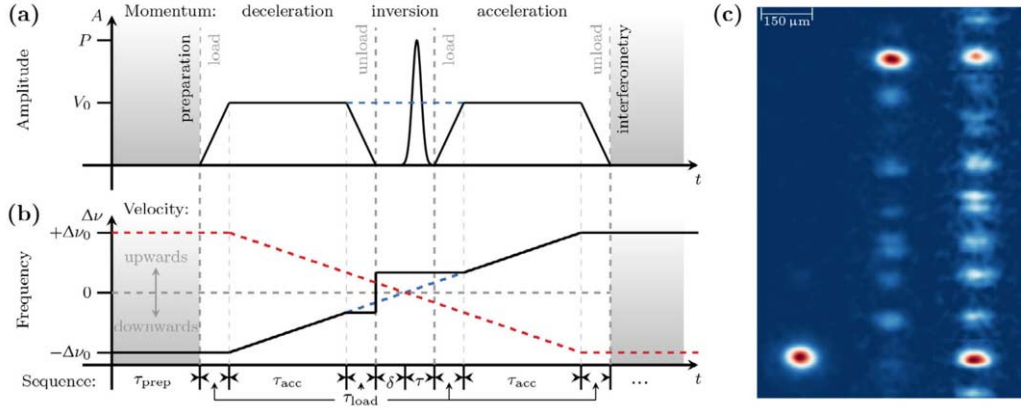


Fig. 11. – Relaunch of atoms in a retro-reflected optical lattice represented by an amplitude (a) and frequency modulation (b) of the lattice and compared to a single-lattice acceleration together with density plots (c) of the final momentum distribution. The relaunch sequence circumvents the zero-velocity crossing of the dual lattice by a double Bragg pulse inverting the momentum. The two frequency components of the dual lattice are depicted in red and blue. In (c) we compare the momentum inversion with a $16\,\hbar k$ double Bragg diffraction pulse (middle) to the one after the deceleration (left), and a simple sweep through resonance with Bloch oscillations (right). This figure is an adaptation of figs. 5.17 and 5.25(b) in ref. [101].

i) two crossed beams reflected from a mirror surface [164], ii) two opposing beams [17], iii) velocity selection from a molasses [12,13,165], or iv) the transfer of only a few photons from a standing wave [166,167]. Indeed, we employ a retro-reflected optical lattice which is a common configuration in atomic sensors.

The novelty, and at the same time the challenge of our method is that for the experiment we use only a *single* beam along the vertical axis with linearly polarized laser light of two co-propagating frequency components. They form in total four lattices: two moving lattices with opposite velocity, and two additional ones at rest.

We perform the relaunch procedure in three steps to avoid a zero-crossing of the velocities of the lattices: i) a lattice deceleration, ii) a momentum inversion pulse, and iii) a lattice acceleration.

Our sequence starts by loading the atoms adiabatically into one of the lattices where the Bloch oscillations for deceleration are performed until the atomic motion is almost stopped. The atoms are then adiabatically unloaded from the lattice with only a few $\hbar k/m$ of residual velocity.

After a small waiting time to carefully match the resonance condition to the velocity of the atoms, a higher-order double Bragg diffraction pulse is applied which inverts the momentum.

Finally, a second lattice acceleration sequence follows, which precisely speeds up the atomic ensemble to launch it on a parabolic trajectory with an adjustable apex close to the atom-chip surface.

In fig. 11(a) and (b) we show the amplitude and frequency modulation of the dual-lattice light used to perform the relaunch sequence. Compared to a single-lattice acceleration this scheme is rather complex, since the additional lattices from the retro-reflection are shifted out of resonance by the Doppler effect of the falling atoms. Unfortunately, this effect does not allow the acceleration of the lattices in a way that the atomic ensemble crosses the zero-momentum state without losing a major fraction of atoms as depicted in fig. 11(b).

These losses result from the fact that when the atoms are at rest, the two moving optical lattices are both on resonance with the atoms —one attempting to move the atoms upwards, and the other one to move them downwards. This feature reduces the fraction of atoms that are launched upwards to about one-half of the atom number achieved by double Bragg diffraction. Moreover, when the velocity of the atoms almost vanishes, non-adiabatic transitions arise due to parasitic acceleration in the non-resonant lattice. They remove atoms from the upward moving lattice, and further reduce the number of launched atoms to about one quarter.

Fortunately, a combination of Bloch oscillations in an optical lattice together with higher-order Bragg diffraction prevents these losses. In this scheme most of the momentum is transferred via Bloch oscillations with an efficiency close to unity to stop and launch the atoms. Since only a smaller fraction of momentum needs to be transferred by a single Bragg pulse, this sequence maintains a high overall efficiency.

4˙3.2. Experimental sequence of the atom-chip gravimeter. Bragg diffraction combined with the relaunch allows us to implement a sensitive gravimeter with the atom chip. Figure 12 shows the spacetime diagram of the fountain geometry.

Subsequent to the adiabatic rapid passage we also perform Stern-Gerlach-type deflection by a magnetic field pulse with a duration of $\tau_{\mathrm{SG}} = 7\,\mathrm{ms}$ using the Z-wire on the chip. In this way we remove the atoms remaining in magnetic sensitive states leading to an enhanced contrast.

The maximum value of the preparation time τ_{prep} is limited to 34 ms due to the end of the detection region 7 mm below the chip. The relaunch process has an overall duration of $\tau_{\mathrm{launch}} = 2.9\,\mathrm{ms}$. With a relaunch realized after the largest waiting time of $\tau_{\mathrm{prep}} = 33.2\,\mathrm{ms}$ the total time of flight τ_{ToF} after initial release of the atoms is greatly increased to $\tau_{\mathrm{ToF}} = 97.6\,\mathrm{ms}$.

The interferometer sequences start after the atoms have been launched on their fountain trajectories. The final waiting time τ_{sep} after $\tau_{\mathrm{ToF}} > 90\,\mathrm{ms}$ to separate the output ports can be reduced from $\tau_{\mathrm{sep}} \geq 20\,\mathrm{ms}$ to $\tau_{\mathrm{sep}} \geq 10\,\mathrm{ms}$ provided DKC is used. The remaining time $2T \equiv \tau_{\mathrm{ToF}} - \tau_{\mathrm{prep}} - \tau_{\mathrm{launch}} - \tau_{\mathrm{sep}} < 51\,\mathrm{ms}$ can be entirely used for the interferometry, which allows us to use a pulse separation time as large as $T = 25\,\mathrm{ms}$. The limit in T on our current baseline of 7 mm is reached with $T = 25\,\mathrm{ms}$. Any further extension beyond this value would result in a reduced contrast due to insufficient port separation.

State-of-the-art Raman-type gravimeters routinely operate with pulse separation times of 70 ms [72] or larger [23, 71]. To further increase the scale factor, not only first-order but also higher-order Bragg diffraction can be implemented in the MZI. At

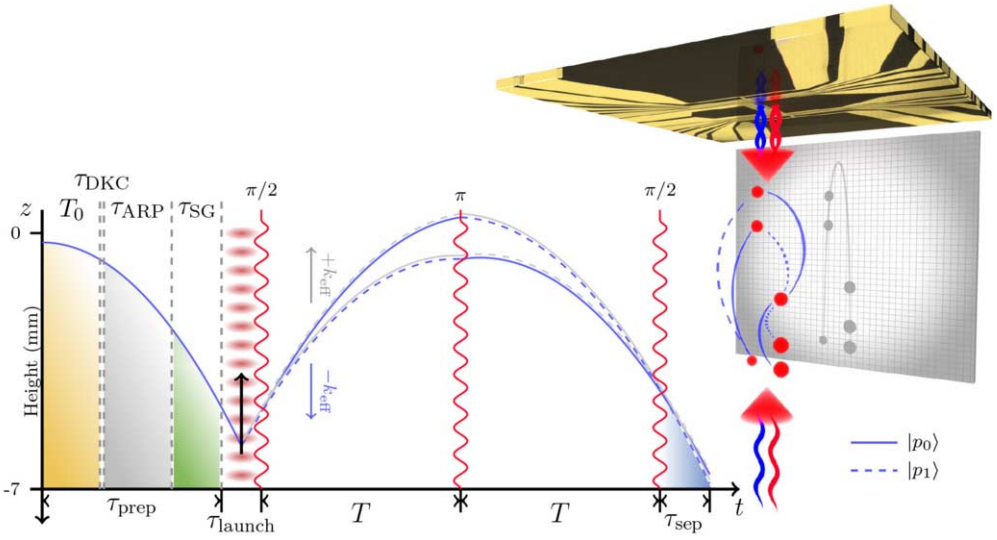


Fig. 12. – Atom-chip gravimeter with an extended free-fall time of up to $\tau_{\mathrm{ToF}} \approx 100\,\mathrm{ms}$ represented in spacetime (left) and space (right). The preparation of the atomic ensemble is performed during τ_{prep} before the relaunch. The elongated free-fall time allows us, after the expansion time T_0, to employ DKC (for the time τ_{DKC}) as well as adiabatic rapid passage (for the time τ_{ARP}) and Stern-Gerlach-type deflection (for the time τ_{SG}) to reduce the expansion rate, and to remove atoms remaining in magnetic sensitive states. The relaunch in the retro-reflected optical lattice is realized as described. The MZI features up to third-order Bragg diffraction, pulse separation times up to $T = 25\,\mathrm{ms}$, and a detection after a separation time of $\tau_{\mathrm{sep}} \geq 10\,\mathrm{ms}$. This figure is reproduced from fig. 6.3 in ref. [101].

the moment third-order Bragg diffraction has an efficiency of above 90%. With future improvements, already fourth-order Bragg diffraction will compensate a decrease by a factor of two in the pulse separation time T.

To perform higher-order Bragg diffraction, we have used shorter beam-splitter pulses, but at larger laser powers. More specifically, we have employed Gaussian-shaped pulses of widths $\sigma_\tau = 12.5\,\mu\mathrm{s}$.

The first $\pi/2$-pulse of the MZI follows with a short delay of one millisecond after the relaunch to maximize the time available for interrogation. To avoid π-pulses at the apex the timing of the MZI needs to be placed *asymmetrically* around the apex of the fountain. In this case, both lattices are on resonance, and losses due to double Bragg diffraction, and standing waves disturb the π-pulse. As a consequence, the Doppler detuning δ of the π-pulse should satisfy the condition $\delta > 100\,\mathrm{kHz}$, or correspondingly, the time difference to the apex should be of the order of 7–8 ms. Consequently, the separation time of the outputs is always larger than $\tau_{\mathrm{sep}} > 14\,\mathrm{ms}$.

A larger momentum transfer slightly reduces the free-fall time, because depending on the direction of the momentum transfer $\pm\hbar k_{\mathrm{eff}}$ the atoms are either kicked towards the atom chip, or downwards such that they leave the detection region faster. By choosing the

momentum transfer of the second and final lattice acceleration in the relaunch sequence according to the direction of the momentum transfer of the beam splitter, the height of the parabola can be maximized. As a consequence, the pulse separation time T can be held constant, independently of $\pm\hbar k_{\mathrm{eff}}$, as depicted in fig. 12. In both cases the free fall time τ_{ToF} of the atoms is slightly reduced due to the recoil imprinted during the beam splitting process.

4˙3.3. Analysis of the interferometer output. Since for pulse separation times larger than $T > 5\,\mathrm{ms}$ the background vibration level leads to a complete loss of the fringe pattern [168] we can no longer employ the common fringe fit method, or an Allan deviation analysis. Even when the laser phase and the chirp rate are identical in subsequent measurements, the output phase scatters over multiple 2π-intervals, as illustrated in fig. 13(a). As a consequence, at these high levels of sensitivity the readout of a gravity-induced phase shift is impossible without having additional information about the vibrations during the interferometry, such as seismic correlation.

We emphasize that the beam splitters still operate with high fidelity, and oscillations between the output ports are clearly visible. However, the simple peak-to-valley, or standard-deviation calculation over- or underestimates the contrast, and does not yield a useful noise analysis for the output signals.

One method to solve the problem of distinguishing between useful contrast C, and technical noise $\sigma_{\Delta\phi} \equiv \sigma_P/C$, rests on a histogram analysis revealing the contrast of an interferometer without relying on fringe visibility [169]. Here σ_P denotes the fluctuation in the measured population.

In this approach the output signals of a data set are first split into bins of equal size containing the normalized population P, and then the data points within each interval are counted. The resulting histogram shows a characteristic double-peak structure reflecting the sinusoidal dependence of the interference signal. This structure results from a simple noise model, assuming that for a completely random signal the probability to find an output state with a normalized population at top or bottom of a sinusoidal fringe pattern is larger than at the middle. We emphasize that this method requires sufficient statistics over several hundred experimental cycles with a stable signal.

We extract the contrast $C \equiv A/P_0$ from a fit of the distribution according to fig. 13(b), with the amplitude A of the signal and its mean P_0. As input for the fitting routine a kernel density estimation (KDE) of the data points is used, rather than the histogram itself([3]).

In a KDE each data point is weighted with a Gaussian function of a fixed width $\sigma_{\mathrm{KDE}} = 0.01$. All Gaussian functions are then added and the signal normalized such that in the end a continuous distribution properly reflects the density of data points without loss of information.

For the width σ_{kde} it is only of importance to choose a value smaller than the ex-

([3]) This evaluation yields a slightly better intrinsic sensitivity compared to the value published ref. [87].

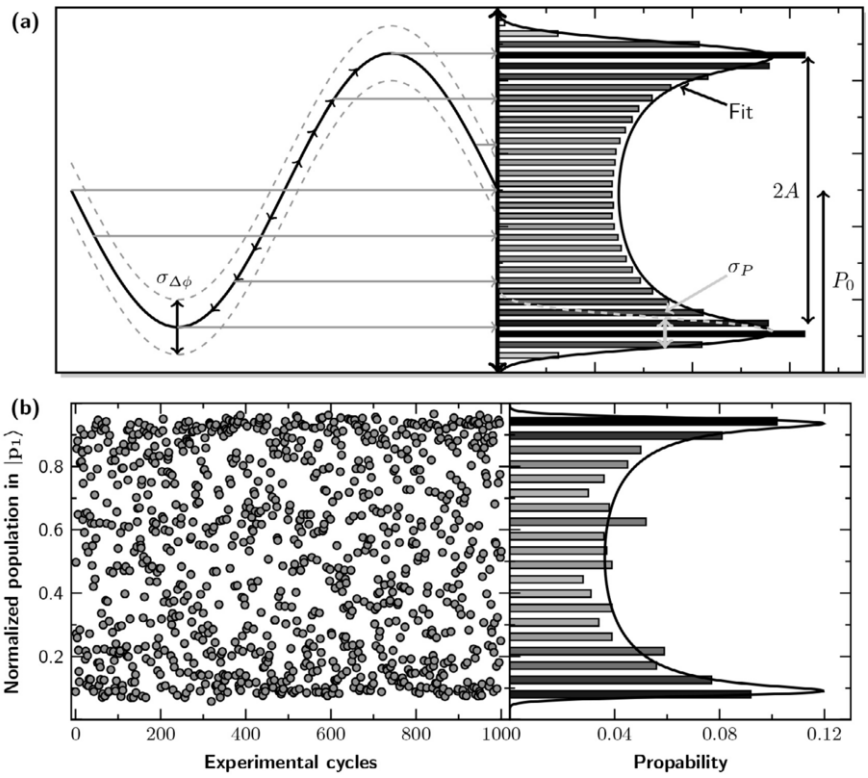


Fig. 13. – Extraction of the interferometer phase from a noisy signal with the help of the probability density. Vibrational noise completely washes out a sinusoidal signal (a), and only fluctuations are left to be measured in the normalized population between the two output ports. The corresponding histogram exhibits a characteristic double-peak structure from which we extract the contrast $C \equiv A/P_0$ given by the amplitude A of the sinusoidal signal and its mean P_0. The output (b) of a MZI in a fountain geometry using delta-kick collimated ensembles with a pulse separation time of $T = 25\,\mathrm{ms}$ and first-order Bragg diffraction resembles noise after roughly 1 000 measurements have been taken. Nevertheless, we can obtain a contrast of $C = 0.86$ and a technical noise $\sigma_{\Delta\phi}$ close to the shot noise from the histogram. This figure is an adaptation of figs. 6.11 and 6.17(c) in ref. [101].

pected technical noise. If the histogram is fitted for a small number n_{bin} of bins, the histogram would yield insufficient information to be properly fit, and would lead to a larger uncertainty in the extracted parameters.

In our fountain geometry the expected gain in contrast is between 5% and 10% and results from two factors: i) We can employ delta-kick collimated BECs with ultra-slow expansion and ii) the Stern-Gerlach-type deflection purifies the magnetic sub-states. With this configuration the pulse separation time can even be extended to $T = 25\,\mathrm{ms}$. At this time, the output ports are still separated due to the smaller final size of the clouds.

The measurements and evaluation depicted in fig. 13(b) are for our MZIs formed by first-order Bragg diffraction and $T = 25\,\mathrm{ms}$. The histogram analysis reveals that the contrast remains at $C = 0.86$. The technical noise level $\sigma_{\Delta\phi} = 14\,\mathrm{mrad}$ is extracted from the widths of the outer peaks in the fit to the density distribution, which is close to the calculated shot noise of 11 mrad for $N = 8\,000$ atoms. This remarkable result is due to the interplay between the high-fidelity Bragg diffraction and the DKC.

These ultra-slow expansion rates allow for even longer flight times, and also give rise to a boost in sensitivity. Indeed, the largest intrinsic sensitivity $\Delta g/g = 1.4 \cdot 10^{-7}$ was observed after a $\tau_{\mathrm{ToF}} = 97.6\,\mathrm{ms}$ at a noise of $\sigma_{\Delta\phi} = 14\,\mathrm{mrad}$. This achievement represents an important step towards compact but precise sensors.

5. – Outlook

Atom interferometry is a cornerstone of precision measurements with a wealth of promising applications. In particular, we expect atomic gravimeters based on BECs to reach sub-μGal accuracies in the near future. In sect. **5**˙1 we highlight the origins of measurement uncertainties, and present mitigation strategies for future devices.

The tools and methods for BEC interferometry outlined in these lecture notes have opened the path towards significantly enhanced scale factors due to an extended free evolution time. For this reason we focus in sect. **5**˙2 on very long baseline atom interferometers. Moreover, we devote sect. **5**˙3 to space-borne devices and analysize the potential for future gravity measurements as well as tests of fundamental physics such as the UFF.

5˙1. *Reduced systematic uncertainties in future devices.* – The main drive for sensors based on BECs is the reduction of systematic uncertainties. We now assess the potential of an atom chip gravimeter to reach sub-μGal accuracies. For this purpose we identify in table IV the origins of the largest contributions to the measurement uncertainty and suggest mitigation strategies.

To get a realistic estimate of the dominant uncertainties for a future experiment, our calculations [170] use a state-of-the-art flux of 10^5 atoms per second at a repetition rate of 1 Hz. For a free-fall distance of 1 cm the free-fall time increases to $\tau_{\mathrm{ToF}} = 135\,\mathrm{ms}$. If the needed detection separation time stays at $\tau_{\mathrm{sep}} > 15\,\mathrm{ms}$, a maximum pulse separation time of $T = 35\,\mathrm{ms}$ remains, which combined with a fourth-order beam splitter leads to the shot-noise limited intrinsic sensitivity [87]

$$(\Delta g/g)/\sqrt{\mathrm{Hz}} = 5.3 \cdot 10^{-9}. \tag{76}$$

The flux of 10^5 atoms/s achieved in the QUANTUS-2 experiment is sufficient [170] to reach this inferred sensitivity and a cycle time of roughly 1 s.

Table IV. – *Estimates of the major systematic uncertainties in the atom-chip gravimeter. As a result, the determination of local gravity with a relative accuracy* $\Delta g/g < 1 \cdot 10^{-9}$ *in less than* 100 *s seems possible* [87]. *This table is a reproduction of table* 6.9 *in ref.* [101].

Contribution due to	Mitigation strategy	Noise $(\Delta g/g)/\sqrt{\text{Hz}}$	Bias $\Delta g/g$
Intrinsic sensitivity	Next-generation source [170]	$5.3 \cdot 10^{-9}$	0
Mean-field shift	Tailored expansion and DKC [85, 154]	$1.5 \cdot 10^{-10}$	$6.4 \cdot 10^{-11}$
Launch velocity	Scatter 70 μm/s, stability 15 μm/s [118]	$1.5 \cdot 10^{-12}$	$3.1 \cdot 10^{-13}$
Wavefront quality	$\lambda/10$ chip-coating, $\oslash = 2$ cm beam [171]	$6.7 \cdot 10^{-10}$	$2.8 \cdot 10^{-10}$
Self gravity	Detailed modeling of chip mount [172]	$1.2 \cdot 10^{-12}$	$5 \cdot 10^{-10}$
Light-shifts	Suppressed in Bragg diffraction [96]	$1.4 \cdot 10^{-12}$	$1.4 \cdot 10^{-10}$
Magnetic fields	Three-layer magnetic shield [173]	$1 \cdot 10^{-10}$	$2.6 \cdot 10^{-10}$
Target estimation	Uncertainty after less than 100 s		$\approx 7.8 \cdot 10^{-10}$

In addition, we need to be able to detect atoms at the output ports at the shot noise limit, which corresponds to 4.5 mrad for this atom flux and cycle time assuming a contrast of $C = 0.7$. Suppressing the vibrational background noise is the crucial remaining noise source to be mitigated. A state-of-the-art vibration isolation would significantly improve the sensitivity, although maximum performance may only be reached at a vibrational quiet site [174].

The mean-field shift can be relaxed when we first lower the atomic density by a faster spreading of the wave packet during the 45 ms after release from the trap but before relaunch, and then stop it by DKC [85, 154]. For the final size of 300 μm at the first pulse, 10^5 atoms and 1% splitting-ratio stability, phase shifts introduced by the mean field [11] can be sufficiently suppressed below μGal, while expansion rates corresponding to nK temperatures preserving the beam-splitter fidelity are achievable.

Fluctuations in the launch velocity, which cause a bias due to the Coriolis effect or gravity gradients [118, 175], can be characterized to the required level and optimized by the tested release procedure. The measured scattering of 70 μm/s and the stability of 15 μm/s of the launch velocity is sufficient to suppress this shift.

The surface quality of the atom chip is crucial for preserving the high efficiencies and contrasts obtained for lower and higher-order Bragg diffraction and for Bloch oscillations. It must be significantly improved for a device of the next generation.

Indeed, a residual roughness of $\lambda/10$ typical for a standard mirror is assumed here. For a beam with a diameter of 2 cm the phase shifts resulting from the wavefront curvature are insignificant since BECs are smaller, and expand slower compared to thermal clouds [171, 176]. Furthermore, the possibility of analyzing the fringe patterns in the density profiles

at the exit ports [85,163,175] may allow the characterization of systematic errors arising from wavefront distortions.

The proximity of the atoms to the chip leads to a bias phase shift caused by the gravitational field [172] of the chip. A mass reduction of the chip mount by a factor of two, combined with a finite-element analysis of the mass distribution which calculates the self-gravity effect with an accuracy of 10%, is sufficient to reach the target level.

Compared to Raman diffraction the influence of light shifts is reduced in MZIs based on Bragg diffraction. Since the two-photon light shift scales [96] with the third power of the inverse of the atomic velocity, it is negligible in the fountain geometry.

Finally a three-layered shield instead of a single-layered one, resulting in a residual gradient below $10 \pm 3\,\mathrm{mG/m}$, should be sufficient to suppress any residual bias [173].

5˙2. *Very long baseline atom interferometry.* – Apart from high-contrast interferometry, delta-kick collimated BECs with effective temperatures below 1 nK enable extended free-evolution times, which can significantly increase the scale factor kT^2 for acceleration measurements. Ground-based setups require a large vacuum vessel to venture into the regime of seconds [9, 105, 177]. For example, a device with a height of 10 m implies a total free-evolution time $2T = 2.8\,\mathrm{s}$ when operated in the fountain mode [162]. This scenario would increase the scale factor by 25–300 compared to the current generation of gravimeters.

Indeed, with 10^5 atoms, first-order Bragg diffraction, and a cycle time of 5.3 s, the shot noise limit for full contrast would be at $0.2\,\mathrm{nm}/(\mathrm{s}^2\sqrt{\mathrm{Hz}})$, which is competitive with the superconducting gravimeter GWR iOSG [66]. In contrast to the latter, VLBAI also provides us with absolute measurements and possible further enhancements via large momentum transfer.

Environmental vibrations impose a typical limit in gravimeters, preventing the utilisation of larger scale factors. Therefore, a sophisticated vibration isolation, correlation with external sensors, or a combination of both is required [22, 169, 178].

Measurement schemes using gravity gradiometers, or testing the UFF intrinsically suppress the impact of vibration noise since the relevant quantity is encoded in the differential acceleration between two atom interferometers [18, 31]. Indeed, for gradiometry, ensembles from *two* different sources can be injected into interferometers, or two ensembles can be generated from a *single* source via a large momentum beam splitting process [179]. The latter approach implies a well-defined distance between the two interferometers which avoids noise contributions from a relative position jitter due to the outcoupling from two different sources and reduces systematic errors. Operating the interferometer with 10^5 atoms divided into two ensembles, first-order Bragg diffraction in the interferometer, a total interferometer time $2T = 1\,\mathrm{s}$, a cycle time of 4 s, and a baseline of 5 m between the interferometers would lead to a shot noise limit of $6.3 \cdot 10^{-10}/(\mathrm{s}^2\sqrt{\mathrm{Hz}})$ for gravity gradients. Further improvements are possible by upgrading the first-order Bragg diffraction to large momentum transfer.

A test of UFF with ^{87}Rb and ^{170}Yb may lead to the shot noise limit $0.1\,\mathrm{nm}/(\mathrm{s}^2\sqrt{\mathrm{Hz}})$

of measurements of differential accelerations, implying a statistical uncertainty $4 \cdot 10^{-14}$ in the Eötvös parameter after 24 h of integration, which is competitive with experiments on the ground achieving 10^{-13} [49, 130, 180], and in space reaching 10^{-14} [51], with the added benefit of different species. The assumptions are $2 \cdot 10^5$ ($1 \cdot 10^5$) atoms, eighth- (fourth-) order Bragg transitions at 780 nm (at 399 nm) for ^{87}Rb (^{170}Yb), a total free evolution time $2T = 2.6$ s, and a cycle time of 12.6 s [162]. Here, the transfer functions of the two interferometers will not be ideally matched due to the different wave vectors, requiring an auxiliary sensor to suppress vibration noise via cross correlation [178, 181].

Relevant systematics [162] originates from i) wavefront errors [118, 176] suppressed by the low and matched expansion rates to be traded off against residual mean-field contributions [11], ii) magnetic field inhomogeneities affecting ^{87}Rb [25, 56] and reduced by a magnetic shield, iii) rotations countered by a tip-tilt stage [71, 159, 182], and iv) gravity gradients coupling to the relative displacements of the two elements, which have to be assessed with the device itself [183]. In fact, the requirements on the relative position and velocity of the two initial wave packets due to gravity gradients can be substantially relaxed thanks to an effective compensation technique proposed in ref. [184], which has been experimentally demonstrated both for UFF tests [185] and gradiometry measurements [186].

5˙3. *Space-borne atom interferometers.* – The microgravity environment of a space-borne atom interferometer provides us with access to even longer free-evolution times. Moreover, no seismic noise disturbs the measurement. Indeed, since both the device and the atomic ensemble are in free-fall, the movement of the atoms with respect to the potentials for trapping and DKC is decreased, opening up a different parameter range for reducing residual expansion rates and mean-field contributions. The measurement may even benefit from a much smaller gravitational sag and the absence of a lattice launch. We also have the possibility of a signal modulation to suppress systematics if the apparatus is inertial pointing [160, 187].

A test of the UFF in space could profit from all these advantages and go beyond an accuracy of about 10^{-14}. The updated Space-Time Explorer and Quantum Equivalence Principle Space Test (STE-QUEST) scenario [188], based on a previous version of a dual-species interferometer with ^{87}Rb and ^{85}Rb [160, 173, 189, 190], proposes a dual-species interferometer with ^{87}Rb and ^{41}K and a target uncertainty in the Eötvös parameter of $2 \cdot 10^{-15}$.

This goal assumes 10^6 atoms of each species, beam splitters based on double Bragg diffraction [86, 97], a total interferometer time of $2T = 10$ s, a cycle time of 20 s, and a highly elliptical orbit for the clock comparison part of the mission with a perigee of ~ 2500 km, an apogee of ~ 33600 km and an orbital period of 10.6 h. Around perigee, the instrument observes a strong signal, whereas around apogee it almost vanishes. Additional measurements in between are utilized for calibration. The target uncertainty is reached after 1.2 years, limited by the small part of the orbit close to earth. A circular orbit in a dedicated mission could reduce this time to a few months.

* * *

We are most grateful to our colleagues H. Albers, H. Ahlers, S. Arnold, D. Becker, K. Bongs, H. Dittus, H. Duncker, W. Ertmer, A. Friedrich, N. Gaaloul, M. Gebbe, C. Gherasim, E. Giese, C. Grzeschik, T. W. Hänsch, J. Hartwig, O. Hellmig, W. Herr, S. Herrmann, E. Kajari, S. Kleinert, M. Krutzik, R. Kuhl, C. Lämmerzahl, W. Lewoczko-Adamczyk, J. Malcolm, N. Meyer, H. Müntinga, R. Nolte, A. Peters, M. Popp, J. Reichel, L. L. Richardson, J. Rudolph, M. Schiemangk, M. Schneider, S. T. Seidel, K. Sengstock, G. Tackmann, V. Tamma, T. Valenzuela, A. Vogel, R. Walser, T. Wendrich, A. Wenzlawski, P. Windpassinger, W. Zeller and T. van Zoest, who have worked with us very closely over the years on the topics presented in these lecture notes. We acknowledge the support by the CRC 1227 DQmat, the CRC 1128 geo-Q, and the QUEST-LFS. The German Space Agency (DLR) with funds provided by the Federal Ministry of Economic Affairs and Energy (BMWi) due to an enactment of the German Bundestag under Grant No. DLR 50WM1137 (QUANTUS-IV-Fallturm), and 50WM1641 (PRIMUS-III) was central to our work. We also acknowledge support by the Federal Ministry of Education and Research (BMBF), by "Wege in die Forschung (II)" of Leibniz Universität Hannover, and "Niedersächsisches Vorab" through the "Quantum- and Nano-Metrology (QUANOMET)" initiative. M.A.E. thanks the Center for Integrated Quantum Science and Technology (IQ^{ST}) for financial support. W.P.S. is most grateful to Texas A&M University for a Faculty Fellowship at the Hagler Institute for Advanced Study at the Texas A&M University as well as to Texas A&M AgriLife Research. F.A.N. acknowledges a generous grant from the Office of the Secretary of Defense (OSD) in the Quantum Science and Engineering Program (qSEP).

REFERENCES

[1] KASEVICH M. and CHU S., *Phys. Rev. Lett.*, **67** (1991) 181.
[2] KASEVICH M. and CHU S., *Appl. Phys. B*, **54** (1992) 321.
[3] PETERS A., CHUNG K.-Y. and CHU S., *Nature*, **400** (1999) 849.
[4] PETERS A., CHUNG K.-Y. and CHU S., *Metrologia*, **38** (2001) 25.
[5] ZHOU M.-K. *et al.*, *Phys. Rev. A*, **86** (2012) 043630.
[6] MALOSSI N. *et al.*, *Phys. Rev. A*, **81** (2010) 013617.
[7] LIN Z. *et al.*, *Chin. Phys. Lett.*, **28** (2011) 013701.
[8] SCHMIDT M. *et al.*, *Gyroscopy Navigat.*, **2** (2011) 170.
[9] ZHOU L. *et al.*, *Gen. Relativ. Gravit.*, **43** (2011) 1931.
[10] MCGUINNESS H. J., RAKHOLIA A. V. and BIEDERMANN G. W., *Appl. Phys. Lett.*, **100** (2012) 011106.
[11] DEBS J. E. *et al.*, *Phys. Rev. A*, **84** (2011) 033610.
[12] CHARRIERE R., CADORET M., ZAHZAM N., BIDEL Y. and BRESSON A., *Phys. Rev. A*, **85** (2012) 013639.
[13] ALTIN P. A. *et al.*, *New J. Phys.*, **15** (2013) 023009.
[14] BIDEL Y. *et al.*, *Appl. Phys. Lett.*, **102** (2013) 144107.
[15] HAUTH M. *et al.*, *Appl. Phys. B*, **113** (2013) 49.
[16] MOK C. *et al.*, *Phys. Rev. A*, **88** (2013) 023614.

[17] Andia M. *et al.*, *Phys. Rev. A*, **88** (2013) 031605.
[18] Bonnin A., Zahzam N., Bidel Y. and Bresson A., *Phys. Rev. A*, **88** (2013) 043615.
[19] Biedermann G. W. *et al.*, *Phys. Rev. A*, **91** (2015) 033629.
[20] Baryshev V. N. and Blinov I. Yu., *Meas. Techn.*, **57** (2015) 1333.
[21] Barrett B., Bertoldi A. and Bouyer P., *Phys. Scr.*, **91** (2016) 053006.
[22] Le Gouët J. *et al.*, *Appl. Phys. B*, **92** (2008) 133.
[23] Hu Z.-K. *et al.*, *Phys. Rev. A*, **88** (2013) 043610.
[24] Gustavson T. L., Bouyer P. and Kasevich M. A., *Phys. Rev. Lett.*, **78** (1997) 2046.
[25] Gauguet A., Canuel B., Léveque T., Chaibi W. and Landragin A., *Phys. Rev. A*, **80** (2009) 063604.
[26] Stockton J. K., Takase K. and Kasevich M. A., *Phys. Rev. Lett.*, **107** (2011) 133001.
[27] Berg P. *et al.*, *Phys. Rev. Lett.*, **114** (2015) 063002.
[28] Dutta I. *et al.*, *Phys. Rev. Lett.*, **116** (2016) 183003.
[29] Durfee D. S., Shaham Y. K. and Kasevich M. A., *Phys. Rev. Lett.*, **97** (2006) 240801.
[30] Wu S., Su E. and Prentiss M., *Phys. Rev. Lett.*, **99** (2007) 173201.
[31] McGuirk J. M., Foster G. T., Fixler J. B., Snadden M. J. and Kasevich M. A., *Phys. Rev. A*, **65** (2002) 033608.
[32] Lamporesi G., Bertoldi A., Cacciapuoti L., Prevedelli M. and Tino G. M., *Phys. Rev. Lett.*, **100** (2008) 050801.
[33] Snadden M. J., McGuirk J. M., Bouyer P., Haritos K. G. and Kasevich M. A., *Phys. Rev. Lett.*, **81** (1998) 971.
[34] Fixler J. B., Foster G. T., McGuirk J. M. and Kasevich M. A., *Science*, **315** (2007) 74.
[35] Rosi G., Sorrentino F., Cacciapuoti L., Prevedelli M. and Tino G. M., *Nature*, **510** (2014) 518.
[36] Cladé P. *et al.*, *Phys. Rev. A*, **74** (2006) 052109.
[37] Müller H., Chiow S.-w., Long Q., Vo C. and Chu S., *Appl. Phys. B*, **84** (2006) 633.
[38] Bouchendira R., Cladé P., Guellati-Khélifa S., Nez F. and Biraben F., *Phys. Rev. Lett.*, **106** (2011) 080801.
[39] Estey B., Yu C., Müller H., Kuan P.-C. and Lan S.-Y., *Phys. Rev. Lett.*, **115** (2015) 083002.
[40] Parker R. H., Yu Ch., Zhong W., Estey B. and Müller H., *Science*, **360** (2018) 191.
[41] Hanneke D., Fogwell S. and Gabrielse G., *Phys. Rev. Lett.*, **100** (2008) 120801.
[42] Damour T., *Class. Quantum Grav.*, **13** (1996) A33.
[43] Merlet S. *et al.*, *Metrologia*, **47** (2010) L9.
[44] Fray S., Diez C. A., Hänsch T. W. and Weitz M., *Phys. Rev. Lett.*, **93** (2004) 240404.
[45] Kuhn C. C. N. *et al.*, *New J. Phys.*, **16** (2014) 073035.
[46] Tarallo M. G. *et al.*, *Phys. Rev. Lett.*, **113** (2014) 023005.
[47] Schlippert D. *et al.*, *Phys. Rev. Lett.*, **112** (2014) 203002.
[48] Zhou L. *et al.*, *Phys. Rev. Lett.*, **115** (2015) 013004.
[49] Williams J. G., Turyshev S. G. and Boggs D. H., *Class. Quantum Grav.*, **29** (2012) 184004.
[50] Adelberger E. G., Gundlach J. H., Heckel B. R., Hoedl S. and Schlamminger S., *Prog. Part. Nucl. Phys.*, **62** (2009) 102.
[51] Touboul P. *et al.*, *Phys. Rev. Lett.*, **119** (2017) 231101.
[52] Hamilton P. *et al.*, *Science*, **349** (2015) 849.
[53] van Tilburg K., Leefer N., Bougas L. and Budker D., *Phys. Rev. Lett.*, **115** (2015) 011802.
[54] Elder B. *et al.*, *Phys. Rev. D*, **94** (2016) 044051.

[55] GRAHAM P. W., HOGAN J. M., KASEVICH M. A. and RAJENDRAN S., *Phys. Rev. Lett.*, **110** (2013) 171102.
[56] DIMOPOULOS S., GRAHAM P. W., HOGAN J. M., KASEVICH M. A. and RAJENDRAN S., *Phys. Rev. D*, **78** (2008) 122002.
[57] DIMOPOULOS S., GRAHAM P. W., HOGAN J. M., KASEVICH M. A. and RAJENDRAN S., *Phys. Lett. B*, **678** (2009) 37.
[58] HOGAN J. M. *et al.*, *Gen. Relativ. Gravit.*, **43** (2011) 1953.
[59] MANNING A. G., KHAKIMOV R. I., DALL R. G. and TRUSCOTT A. G., *Nat. Phys.*, **11** (2015) 539.
[60] PEISE J. *et al.*, *Nat. Commun.*, **6** (2015) 8984.
[61] ENGELSEN N. J., KRISHNAKUMAR R., HOSTEN O. and KASEVICH M. A., *Phys. Rev. Lett.*, **118** (2017) 140401.
[62] MIDDLEMISS R. P. *et al.*, *Nature*, **531** (2016) 614.
[63] TIMMEN L., *Absolute and relative gravimetry*, in *Sciences of Geodesy-I* (Springer) 2010, pp. 1–48.
[64] ZUMBERGE M. A., RINKER R. L. and FALLER J. E., *Metrologia*, **18** (1982) 145.
[65] NIEBAUER T. M., SASAGAWA G. S., FALLER J. E., HILT R. and KLOPPING F., *Metrologia*, **32** (1995) 159.
[66] PROTHERO JR. W. A. and GOODKIND J. M., *Rev. Sci. Instrum.*, **39** (1968) 1257.
[67] OKUBO S., YOSHIDA S., SATO T., TAMURA Y. and IMANISHI Y., *Geophys. Res. Lett.*, **24** (1997) 489.
[68] IMANISHI Y., SATO T., HIGASHI T., SUN W. and OKUBO S., *Science*, **306** (2004) 476.
[69] SALOMON C., DALIBARD J., PHILLIPS W. D., CLAIRON A. and GUELLATI S., *EuroPhys. Lett.*, **12** (1990) 683.
[70] PHILLIPS W. D. *et al.*, *Phys. Scr.*, **1991** (1991) 20.
[71] FREIER C. *et al.*, *J. Phys. Conf. Ser.*, **723** (2016) 012050.
[72] FANG B. *et al.*, *J. Phys. Conf. Ser.*, **723** (2016) 012049.
[73] BODART Q. *et al.*, *Appl. Phys. Lett.*, **96** (2010) 134101.
[74] `http://www.muquans.com/`.
[75] `http://www.aosense.com/`.
[76] CRONIN A. D., SCHMIEDMAYER J. and PRITCHARD D. E., *Rev. Mod. Phys.*, **81** (2009) 1051.
[77] KOZUMA M. *et al.*, *Phys. Rev. Lett.*, **82** (1999) 871.
[78] TORII Y. *et al.*, *Phys. Rev. A*, **61** (2000) 041602.
[79] ANTOINE C. and BORDÉ C. J., *J. Opt. B*, **5** (2003) 199.
[80] KASEVICH M. *et al.*, *Phys. Rev. Lett.*, **66** (1991) 2297.
[81] BORDÉ C. J., *Phys. Rev. A*, **140** (1989) 10.
[82] SZIGETI S. S., DEBS J. E., HOPE J. J., ROBINS N. P. and CLOSE J. D., *New J. Phys.*, **14** (2012) 023009.
[83] ANDERSON M. H., ENSHER J. R., MATTHEWS M. R., WIEMAN C. E. and CORNELL E. A., *Science*, **269** (1995) 198.
[84] DAVIS K. B. *et al.*, *Phys. Rev. Lett.*, **75** (1995) 3969.
[85] MÜNTINGA H. *et al.*, *Phys. Rev. Lett.*, **110** (2013) 093602.
[86] AHLERS H. *et al.*, *Phys. Rev. Lett.*, **116** (2016) 173601.
[87] ABEND S. *et al.*, *Phys. Rev. Lett.*, **117** (2016) 203003.
[88] KEITH D. W., EKSTROM C. R., TURCHETTE Q. A. and PRITCHARD D. E., *Phys. Rev. Lett.*, **66** (1991) 2693.
[89] CARNAL O. and MLYNEK J., *Phys. Rev. Lett.*, **66** (1991) 2689.
[90] RIEHLE F., KISTERS T., WITTE A., HELMCKE J. and BORDÉ C. J., *Phys. Rev. Lett.*, **67** (1991) 177.

[91] Rasel E. M., Oberthaler M. K., Batelaan H., Schmiedmayer J. and Zeilinger A., *Phys. Rev. Lett.*, **75** (1995) 2633.
[92] Berman P. R. (Editor), *Atom Interferometry* (Academic Press) 1997.
[93] Foot C. J., *Atomic Physics* (Oxford University Press) 2005.
[94] Berg P., *Ein rauscharmes Materiewellen-Gyroskop hoher Genauigkeit*, Dissertation, Leibniz Universität Hannover (2014).
[95] Schlippert D., *Quantum Tests of the Universality of Free Fall*, Dissertation, Leibniz Universität Hannover (2014).
[96] Giese E., Friedrich A., Abend S., Rasel E. M. and Schleich W. P., *Phys. Rev. A*, **94** (2016) 063619.
[97] Lévèque T., Gauguet A., Michaud F., Pereira Dos Santos F. and Landragin A., *Phys. Rev. Lett.*, **103** (2009) 080405.
[98] Giese E., Roura A., Tackmann G., Rasel E. M. and Schleich W. P., *Phys. Rev. A*, **88** (2013) 053608.
[99] Bragg W. L., *Nature*, **90** (1912) 410.
[100] Friedrich W., Knipping P. and Laue M., *Ann. Phys.*, **346** (1913) 971.
[101] Abend S., *Atom-chip Gravimeter with Bose-Einstein Condensates*, Dissertation, Leibniz Universität Hannover (2017).
[102] Müller H., Chiow S.-w., Long Q., Herrmann S. and Chu S., *Phys. Rev. Lett.*, **100** (2008) 180405.
[103] Müller H., Chiow S.-w. and Chu S., *Phys. Rev. A*, **77** (2008) 23609.
[104] Chiow S.-w., Kovachy T., Chien H.-C. and Kasevich M. A., *Phys. Rev. Lett.*, **107** (2011) 130403.
[105] Kovachy T. *et al.*, *Nature*, **528** (2015) 530.
[106] Ben Dahan M., Peik E., Reichel J., Castin Y. and Salomon C., *Phys. Rev. Lett.*, **76** (1996) 4508.
[107] Peik E., Ben Dahan M., Bouchoule I., Castin Y. and Salomon C., *Phys. Rev. A*, **55** (1997) 2989.
[108] Wilkinson S. R., Bharucha C. F., Madison K. W., Niu Q. and Raizen M. G., *Phys. Rev. Lett.*, **76** (1996) 4512.
[109] Raizen M., Salomon C. and Niu Q., *Phys. Today*, **50** (1997) 30.
[110] Grimm R., Weidemüller M. and Ovchinnikov Y. B., *Adv. At. Mol. Opt. Phys.*, **42** (2000) 95.
[111] Kittel C., *Einführung in die Festkörperphysik* (Oldenburg) 2005.
[112] Denschlag J. H. *et al.*, *J. Phys. B At. Mol. Opt. Phys.*, **35** (2002) 3095.
[113] Zener C., *Proc. Roy. Soc. A*, **137** (1932) 696.
[114] Bordé C. J., *C. R. Acad. Sci.*, **2** (2001) 509.
[115] Storey P. and Cohen-Tannoudji C., *J. Phys. II*, **4** (1994) 1999.
[116] Schleich W. P., Greenberger D. M. and Rasel E. M., *New J. Phys.*, **15** (2013) 013007.
[117] Kleinert S., Kajari E., Roura A. and Schleich W. P., *Phys. Rep.*, **605** (2015) 1.
[118] Louchet-Chauvet A. *et al.*, *New J. Phys.*, **13** (2011) 065025.
[119] Antoine C., *Phys. Rev. A*, **76** (2007) 033609.
[120] Cheinet P. *et al.*, *IEEE Trans. Instrum. Meas.*, **57** (2008) 1141.
[121] Stoner R., Butts D., Kinast J. and Timmons B., *J. Opt. Soc. Am. B*, **28** (2011) 2418.
[122] Bonnin A., Zahzam N., Bidel Y. and Bresson A., *Phys. Rev. A*, **92** (2015) 023626.
[123] Einstein A., *Ann. Phys.*, **49** (1916) 769.
[124] Abbott B. P. *et al.*, *Phys. Rev. Lett.*, **116** (2016) 061102.
[125] Abbott B. P. *et al.*, *Phys. Rev. Lett.*, **116** (2016) 241102.

[126] Pound R. V. and Rebka G. A., *Phys. Rev. Lett.*, **4** (1960) 337.
[127] Vessot R. F. C. *et al.*, *Phys. Rev. Lett.*, **45** (1980) 2081.
[128] Schlamminger S., Choi K.-Y., Wagner T. A., Gundlach J. H. and Adelberger E. G., *Phys. Rev. Lett.*, **100** (2008) 041101.
[129] Williams J. G., Turyshev S. G. and Boggs D. H., *Phys. Rev. Lett.*, **93** (2004) 261101.
[130] Müller J., Hofmann F. and Biskupek L., *Class. Quantum Grav.*, **29** (2012) 184006.
[131] Niebauer T. M., McHugh M. P. and Faller J. E., *Phys. Rev. Lett.*, **59** (1987) 609.
[132] Touboul P., Métris G., Lebat V. and Robert A., *Class. Quantum Grav.*, **29** (2012) 184010.
[133] Colella R., Overhauser A. W. and Werner S. A., *Phys. Rev. Lett.*, **34** (1975) 1472.
[134] Fray S. and Weitz M., *Space Sci. Rev.*, **148** (2009) 225.
[135] Göklü E. and Lämmerzahl C., *Class. Quantum Grav.*, **25** (2008) 105012.
[136] Leitner J. and Okubo S., *Phys. Rev.*, **136** (1964) B1542.
[137] Lämmerzahl C., *Class. Quantum Grav.*, **15** (1998) 13.
[138] Lämmerzahl C., *Quantum tests of space-time structures*, in *Proceedings of the International School of Cosmology and Gravitation, Course XV*, edited by Bergmann P., DeSabbata V., Gillies G. and Pronin P. (World Scientific, Singapore) 1998, p. 91.
[139] Lämmerzahl C., *Appl. Phys. B*, **84** (2006) 551.
[140] Damour T., *Class. Quantum Grav.*, **29** (2012) 184001.
[141] Kostelecký V. A. and Tasson J. D., *Phys. Rev. D*, **83** (2011) 016013.
[142] Hohensee M. A., Müller H. and Wiringa R. B., *Phys. Rev. Lett.*, **111** (2013) 151102.
[143] Hamilton P., Barter T., Kim G., Mukherjee B. and Müller H., *Bull. Am. Phys. Soc.*, **57** (2012) T5.00004.
[144] Audi G., Wapstra A. H. and Thibault C., *Nucl. Phys. A*, **729** (2003) 337.
[145] de Laeter J. R. *et al.*, *Pure Appl. Chem.*, **75** (2003) 683.
[146] Müller H., *Quantum mechanics, matter waves, and moving clocks*, in *Atom Interferometry, Proceedings of the International School of Physics "Enrico Fermi" Course CLXXXVIII*, edited by Tino G. M. and Kasevich M. A. (Società Italiana di Fisica, Bologna; IOS Press, Amsterdam) 2013, p. 339.
[147] Landini M. *et al.*, *Phys. Rev. A*, **84** (2011) 043432.
[148] Chu S., *Rev. Mod. Phys.*, **70** (1998) 685.
[149] Phillips W. D., *Rev. Mod. Phys.*, **70** (1998) 721.
[150] Zaiser M. *et al.*, *Phys. Rev. A*, **83** (2011) 035601.
[151] Chu S., Bjorkholm J. E., Ashkin A., Gordon J. P. and Hollberg L. W., *Opt. Lett.*, **11** (1986) 73.
[152] Ammann H. and Christensen N., *Phys. Rev. Lett.*, **78** (1997) 2088.
[153] Morinaga M., Bouchoule I., Karam J.-C. and Salomon C., *Phys. Rev. Lett.*, **83** (1999) 4037.
[154] Kovachy T. *et al.*, *Phys. Rev. Lett.*, **114** (2015) 143004.
[155] Smith D. A., Arnold A. S., Pritchard M. J. and Hughes I. G., *J. Phys. B At. Mol. Opt. Phys.*, **41** (2008) 125302.
[156] Leanhardt A. E. *et al.*, *Science*, **301** (2003) 1513.
[157] Wenzlawski A., *Matter-wave optics in microgravity: Laser technology and applications*, Dissertation, Universität Hamburg (2013).
[158] Krutzik M., *Matter wave interferometry in microgravity*, Dissertation, Humboldt-Universität zu Berlin (2014).
[159] Dickerson S. M., Hogan J. M., Sugarbaker A., Johnson D. M. S. and Kasevich M. A., *Phys. Rev. Lett.*, **111** (2013) 083001.
[160] Aguilera D. N. *et al.*, *Class. Quantum Grav.*, **31** (2014) 115010.

[161] Schubert C. *et al.*, preprint arXiv:1312.5963 (2013).
[162] Hartwig J. *et al.*, *New J. Phys.*, **17** (2015) 035011.
[163] Sugarbaker A., Dickerson S. M., Hogan J. M., Johnson D. M. S. and Kasevich M. A., *Phys. Rev. Lett.*, **111** (2013) 113002.
[164] Sugarbaker A., *Atom interferometry in a 10 m fountain*, PhD Thesis, Stanford University (2014).
[165] Mazzoni T. *et al.*, *Phys. Rev. A*, **92** (2015) 053619.
[166] Impens F., Bouyer P. and Bordé C. J., *Appl. Phys. B*, **84** (2006) 603.
[167] Hughes K. J., Burke J. H. T. and Sackett C. A., *Phys. Rev. Lett.*, **102** (2009) 150403.
[168] McDonald G. D. *et al.*, *Phys. Rev. A*, **88** (2013) 053620.
[169] Geiger R. *et al.*, *Nat. Commun.*, **2** (2011) 474.
[170] Rudolph J. *et al.*, *New J. Phys.*, **17** (2015) 065001.
[171] Tackmann G. *et al.*, *New J. Phys.*, **14** (2012) 015002.
[172] D'Agostino G., Merlet S., Landragin A. and Pereira Dos Santos F., *Metrologia*, **48** (2011) 299.
[173] Milke A. *et al.*, *Rev. Sci. Instrum.*, **85** (2014) 083105.
[174] Tang B., Zhou L., Xiong Z., Wang J. and Zhan M., *Rev. Sci. Instrum.*, **85** (2014) 093109.
[175] Roura A., Zeller W. and Schleich W. P., *New J. Phys.*, **16** (2014) 123012.
[176] Schkolnik V., Leykauf B., Hauth M., Freier C. and Peters A., *Appl. Phys. B*, **120** (2015) 311.
[177] Hardman K. S. *et al.*, *Phys. Rev. Lett.*, **117** (2016) 138501.
[178] Barrett B. *et al.*, *New J. Phys.*, **17** (2015) 085010.
[179] Asenbaum P. *et al.*, *Phys. Rev. Lett.*, **118** (2017) 183602.
[180] Wagner T. A., Schlamminger S., Gundlach J. H. and Adelberger E. G., *Class. Quantum Grav.*, **29** (2012) 184002.
[181] Barrett B. *et al.*, *Nat. Commun.*, **7** (2016) 13786.
[182] Lan S.-Y., Kuan P.-Ch., Estey B., Haslinger P. and Müller H., *Phys. Rev. Lett.*, **108** (2012) 090402.
[183] Hogan J. M., Johnson D. M. S. and Kasevich M. A., preprint arXiv:0806.3261 (2008).
[184] Roura A., *Phys. Rev. Lett.*, **118** (2017) 160401.
[185] Overstreet C. *et al.*, *Phys. Rev. Lett.*, **120** (2018) 183604.
[186] D'Amico G. *et al.*, *Phys. Rev. Lett.*, **119** (2017) 253201.
[187] Williams J., Chiow S.-w., Yu N. and Müller H., *New J. Phys.*, **18** (2016) 025018.
[188] Bongs K. *et al.*, *Proc. 50th Rencontre de Moriond*, **39** (2015) 137.
[189] Altschul B. *et al.*, *Adv. Space Res.*, **55** (2015) 501.
[190] Schuldt T. *et al.*, *Exp. Astron.*, **39** (2015) 167.

Proceedings of the International School of Physics "Enrico Fermi"
Course 197 "Foundations of Quantum Theory", edited by E. M. Rasel, W. P. Schleich and S. Wölk
(IOS, Amsterdam; SIF, Bologna) 2019
DOI 10.3254/978-1-61499-937-9-393

Atom-chip–based quantum gravimetry with BECs

S. Abend(*), M. Gersemann, H. Ahlers, M. Sahelgozin, J. Matthias, N. Grove, H. Heine, N. Gaaloul, W. Herr, C. Schubert, W. Ertmer and E. M. Rasel

Institut für Quantenoptik, Leibniz Universität Hannover - Hannover, Germany

M. Gebbe, H. Müntinga and C. Lämmerzahl

Zarm, Universität Bremen - Bremen, Germany

L. Timmen and J. Müller

Institut für Erdmessung, Leibniz Universität Hannover - Hannover, Germany

Summary. — We introduce two generations of quantum gravimeters using Bose-Einstein condensates generated in atom-chip–based set-ups. The first one is a prototype gravimeter implemented in QUANTUS-1, that allows us to demonstrate the first atom-chip–based gravity determination. The second device is a next generation quantum gravimeter QG-1, targeting sub-μGal uncertainties for mobile applications.

1. – Introduction

Today's inertially sensitive atom interferometry devices operate mostly with sources of laser-cooled atoms [1-3] and are now commercially available [4,5]. The finite temperature and size of these sources limit the efficiency of employed beam splitters and the analysis

(*) Email: abend@iqo.uni-hannover.de (corresponding author)

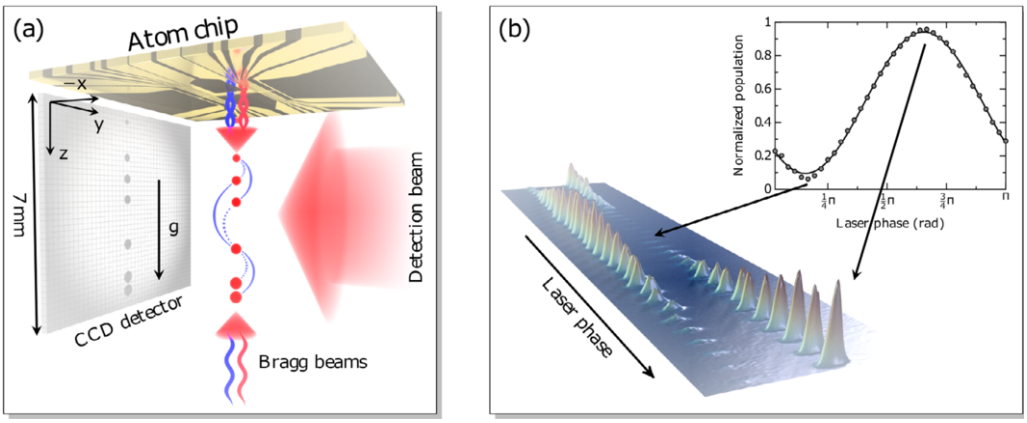


Fig. 1. – (a) Scheme of our atom-chip–based gravimeter set-up. (b) Exemplary laser phase scan and corresponding density profiles for $T = 0$ demonstrating high-contrast interferometry.

of systematic uncertainties to a few μGal ($1\,\mu\text{Gal} = 10^{-8}\,\text{m/s}^2$). These limits can be overcome by the use of ultracold sources such as a delta-kick collimated Bose-Einstein condensate (BEC) with an extremely narrow velocity distribution far below the photon recoil [6]. In the future, a detailed systematic analysis will allow to suppress the two most relevant sources of uncertainties, namely center-of-mass motion [7] and wavefront errors [8] exploiting the point-like nature of BECs [9,10]. Atom-chip technologies [11] offer the possibility to generate a BEC and perform delta-kick collimation in a fast and reliable away. We show two generations of experiments — a prototype and a next generation device — which use the combination of such an ensemble each generated in a robust and miniaturized atom-chip set-up together with the application of Bragg diffraction as a beam splitter for atom interferometry. Already today, the prototype allows us to perform inertially sensitive measurements at high contrast, while the next generation will significantly improve on the atom interferometer's sensitivity.

2. – Atom-chip–based gravimeter prototype in QUANTUS-1

In the QUANTUS-1 apparatus [12] we demonstrate the first atom-chip–based quantum gravimeter in detail described in ref. [13]. A specialty of our set-up depicted in fig. 1(a) is that the atom chip is used for the generation of BECs, state preparation, including magnetic sub-state transfer, delta-kick collimation and Stern-Gerlach–type deflection as well as itself as a retro-reflector for the beam splitting light fields. In that way, Mach-Zehnder interferometers (MZI) using Bragg diffraction to measure gravity can be formed and their output ports detected in a cube below the atom chip with a side length of less than one centimeter. The apparatus reliably provides us ^{87}Rb BECs of up to $1.5 \cdot 10^4$ atoms in the hyperfine state $F = 2$, $m_F = 2$ within 15 s mainly limited by the loading of the magneto-optical trap from background vapor. Using a simple experi-

mental implementation with freely-falling Bose-Einstein condensates out-coupled from a shallow $(46, 31, 18)$ Hz trap at 50 nK and transferred in $\tau_{\mathrm{sep}} \approx 10$ ms to the non-magnetic sub-state via an adiabatic rapid passage allows for MZI with pulse separation times up to $T = 5$ ms constrained by the end of the detection region 7 mm below the atom chip which limits the total time of free-fall to ToF $= 35$ ms minus the final separation of the output ports $\tau_{\mathrm{det}} \approx 15$ ms. Figure 1(b) shows the visualization of detected density profiles of a laser phase scan obtained for $T = 0$ in this devices which reveals a large interferometric contrast and low technical noise. The free-fall rate of the BECs is measured by chirping the difference of the laser frequencies with a rate α, such that the lattice motion precisely matches the acceleration of the atoms in a way that the total phase shift $\phi_{\mathrm{MZI}} = (k_{\mathrm{eff}} \cdot g - 2\pi\alpha)T^2$ vanishes independently of T for the special case of $\alpha = k_{\mathrm{eff}} \cdot g/(2\pi)$, where the atom number in one output port of the MZI assumes a minimum [1]. With this scheme a determination of local gravitational acceleration g with an uncertainty of $\Delta g/g = 1.3 \cdot 10^{-5}$ after roughly eight hours of integration has been demonstrated, which is limited by background vibrations acting on the non-isolated set-up. The intrinsic sensitivity, however, operating a MZI with first-order Bragg diffraction, an obtained interferometric contrast of $C > 0.75$ and the largest chosen pulse separation time of $T = 5$ ms is significantly better. Indeed the phase read out, which only includes all non-inertial noise contributions, is operated close to the shot noise limit for $N = 10\,000$ detected atoms and thus would allow to measure gravity to $\Delta g/g = 3.2 \cdot 10^{-6}$. Nonetheless, already in this simple prototype important techniques for the determination of gravity using BECs have been demonstrated, which are of valuable contribution to the next generation of atomic quantum sensors interesting for geodetic Earth observation.

3. – Next generation quantum gravimeter QG-1 for mobile applications

The quantum gravimeter QG-1 is designed to be a compact, transportable absolute gravimeter based on Bragg interferometry with ultracold ^{87}Rb atoms. From a technological point of view, QG-1 will in every relevant aspect improve on the QUANTUS-1 prototype, by the use of a specifically for gravimetry engineered sensor head depicted in fig. 2(a) in which the drop baseline is enlarged to approx. 30 cm enabling free evolution times up to $T = 100$ ms using freely-falling BECs from a next generation high-flux atom-chip source. The complete device consists only of the sensor head, placed on a vibration isolation platform and a single temperature stabilized 19 inch rack containing all sub-components which allows for an over all compact and transportable device capable of operation under non-laboratory conditions. The vacuum chamber itself is mounted inside a three-layer magnetic shield with the ion getter pump outside of the shield. Rather than using the atom chip itself as the retro-reflector, in the QG-1 the beam splitting light field is reflected from the atom chip at a 45° angle from the top of the apparatus onto a retro-reflecting mirror at the lower end of the vacuum system. Besides providing a superior optical quality, this allows for placing the inertial reference close to a seismometer and counter-acting the Earth's rotation via a piezo-controlled tip-tilt mirror

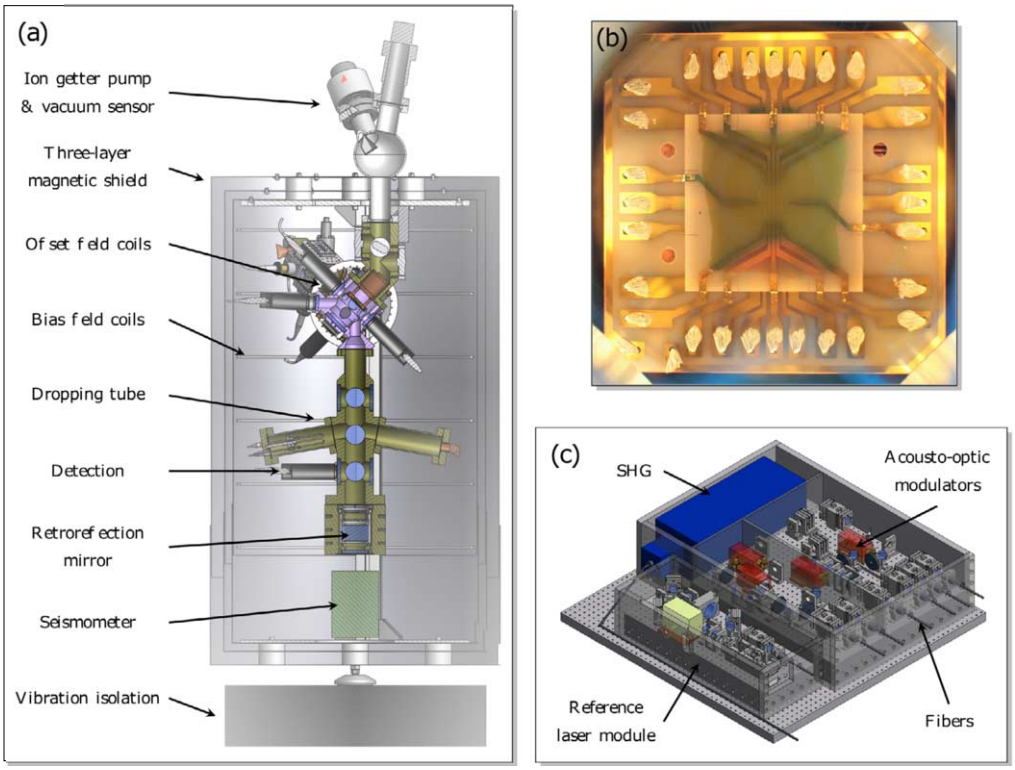


Fig. 2. – (a) Sensor head of the quantum gravimeter QG-1. (b) Photograph of the latest generation atom chip built at the IQ. (c) Scheme of the laser system and light distribution.

stage [14]. All necessary light fields for cooling, trapping, beam splitting and absorption detection of the final interferometer state are provided using a single frequency-doubled 1560 nm fiber laser system [15] with an output power of 3 W at 780 nm placed on a single 60×40 cm breadboard sketched in fig. 2(b). The atomic source system for the QG-1 device is similar to ref. [16] which combines a high-flux 2D+MOT with a next generation atom chip shown in fig. 2(b) achieving 10^5 condensed atoms per second, or even $4 \cdot 10^5$ condensed atoms in 1.6 seconds. With such a flux the QG-1 will be able to achieve a shot-noise limited sensitivity of 10–15 μGal/s allowing to reach the targeted uncertainty below 1 μGal in less than 200 s of integration. The major scientific motivation for the construction of the QG-1 device lies in establishing atom interferometers for geodetic field application in the cooperation with the group of J. Müller and L. Timmen from the Institut für Erdmessung - IfE in Hannover, who are long-term experts in the field of absolute gravimetry. The cooperation with the IfE in the framework of the Collaborative Research Center geo-Q [17] has the final goal of a gravity measurement campaign leading to a more accurate observation of the so-called Fennoscandian uplift [18, 19].

* * *

This work was supported by the Collaborative Research Center geo-Q of the Deutsche Forschungsgemeinschaft (SFB 1128), the German Space Agency (DLR) with funds provided by the Federal Ministry for Economic Affairs and Energy (BMWi) due to an enactment of the German Bundestag under grant numbers DLR 50WM1552-1557, as well as the Centre for Quantum Engineering and Space-Time Research (QUEST).

REFERENCES

[1] PETERS A. *et al.*, *Nature*, **400** (1999) 849.
[2] GAUGUET A. *et al.*, *Phys. Rev. A*, **78** (2008) 043615.
[3] HU Z.-K. *et al.*, *Phys. Rev. A*, **88** (2013) 043610.
[4] MUQUANS, `http://www.muquans.com/`.
[5] AOSENSE, `http://www.aosense.com/`.
[6] MÜNTINGA H. *et al.*, *Phys. Rev. Lett.*, **110** (2013) 093602.
[7] LOUCHET-CHAUVET A. *et al.*, *New J. Phys.*, **13** (2011) 065025.
[8] TACKMANN G. *et al.*, *New J. Phys.*, **14** (2012) 015002.
[9] DICKERSON S. M. *et al.*, *Phys. Rev. Lett.*, **111** (2013) 083001.
[10] SUGARBAKER A. *et al.*, *Phys. Rev. Lett.*, **111** (2013) 113002.
[11] FORTÁGH J. *et al.*, *Rev. Mod. Phys.*, **79** (2007) 235.
[12] VAN ZOEST T. *et al.*, *Science*, **328** (2010) 1540.
[13] ABEND S. *et al.*, *Phys. Rev. Lett.*, **117** (2016) 203003.
[14] LAN S.-Y. *et al.*, *Phys. Rev. Lett.*, **108** (2015) 090402.
[15] THERON F. *et al.*, *Appl. Phys. B*, **118** (2015) 1.
[16] RUDOLPH J. *et al.*, *New J. Phys.*, **17** (2015) 065001.
[17] GEO-Q, `http://www.geoq.uni-hannover.de/`.
[18] TIMMEN L. *et al.*, *Zeit. f. Phys.*, **133** (2008) 149.
[19] MÜLLER *et al.*, *Phys. Chem. Earth*, **53** (2012) 54.

This page intentionally left blank

Proceedings of the International School of Physics "Enrico Fermi"
Course 197 "Foundations of Quantum Theory", edited by E. M. Rasel, W. P. Schleich and S. Wölk
(IOS, Amsterdam; SIF, Bologna) 2019
DOI 10.3254/978-1-61499-937-9-399

Gravitational properties of light

D. Rätzel, M. Wilkens and R. Menzel

University of Potsdam, Institute for Physics and Astronomy
Karl-Liebknecht-Str. 24/25, 14476 Potsdam, Germany

Summary. — In this article, recent work by the authors on the gravitational properties of light is reviewed. In the first part, the gravitational field of a laser pulse of finite lifetime is investigated in the framework of linearized gravity. In the second part, the dependence of the differential cross section for gravitational photon-photon scattering on the polarization entanglement between the photons is investigated in Perturbative Quantum Gravity. These investigations are of conceptual interest regarding the properties of light and its constituents, the photons.

In general relativity, all energy-momentum leads to a gravitational effect, therefore so does the energy-momentum of light. However, the gravitational field of light shows interesting differences to the gravitational field of massive matter. This is because the speed of light c is also the speed of small deviations of the gravitational field, the characteristic speed if the curvature of the spacetime is small on the length scale of the wavelength of the deviations [1, 2]. Hence, light travels as fast as its effect on the gravitational field. It is the aim of the first part of this article to present the gravitational properties of a pulse of laser light in the framework of Linearized Gravity.

The first approach to the gravitational field of light in General Relativity can be found in [3] by Tolman, Ehrenfest and Podolsky, where the gravitational effect of an infinitesimally thin, cylindrical pulse of unpolarized light of finite lifetime was investigated. In particular, it was found that a parallel propagating test pulse is not affected if the test beam is co-propagating, but is deflected if it is counter-propagating.

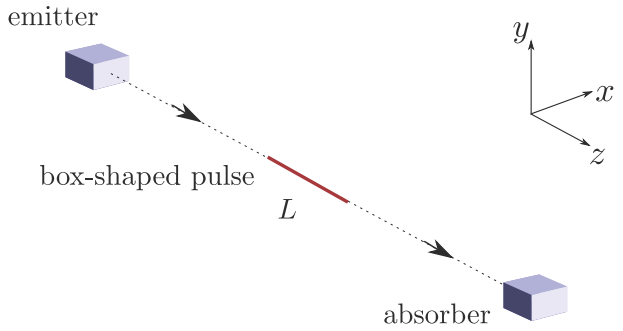


Fig. 1. – The laser pulse is modeled with finite extension L in the direction of propagation, but negligible extension $\Delta(z)$ in the transverse x/y-directions, $\Delta(z) \ll L$. It travels from the emitter to the absorber over a distance D along the z-axis. The figure was originally published in [8] under the CC Attribution 3.0 license `https://creativecommons.org/licenses/by/3.0/`.

In a series of subsequent articles, the gravitational field of light has been investigated using the full Einstein equations. Light was represented as a null-fluid of massless particles in [4], the Lorentz-boosted Schwarzschild-metric of a point mass in the limit $v \to c$, $m \to 0$ was derived by [5] and exact plane wave solutions of the coupled Maxwell-Einstein theory were derived in [6]. In [7], it was shown that the interaction between pulses running slower than the speed of light — *e.g.* in a wave guide — is non-zero.

In the following we will shortly present the model we use and the results. For a full derivation and a more extensive review of preliminary work by other authors see [8] where the work we are going to present in the first part of this article was originally published under the CC Attribution 3.0 license `https://creativecommons.org/licenses/by/3.0/`. In accordance with the established model, the pulse is represented as a "needle of null stuff", in our case consisting of coherently polarized electromagnetic radiation. Our model aims to catch the essential ingredients of a laser pulse, which are 1) its localizability, 2) its masslessness, and 3) its polarization. The spacetime events representing the emission and absorption of the pulse are also included (see fig. 1). However, the gravitational effects of emitter and absorber and their evolution during their emission is not considered here. We have shown in [8] that these effects are neglectable for small distances to the pulse trajectory. In [9], we have derived the full gravitational field of the whole emission process for a specific situation — the emission of two light pulses from a single atom. The gravitational field of light can be expected to be small, which justifies splitting the metric into a flat background metric and a perturbation $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$. The Einstein equations can then be written as a wave equation and solved using the method of retarded potentials. The only non-zero components turn out to be $h_{00} = h_{11} = -h_{10} = -h_{01}$. We define $h^{\mathrm{p}} = h_{00}$. In fig. 2, h^{p} is plotted in the x-z-plane for different times after the events of the emission of the pulse. The position of the pulse of length L corresponds to the bright spot in the main plot. It is propagating in the positive z-direction as fast as its effect on the metric — with the speed of light. For distances to the trajectory of the

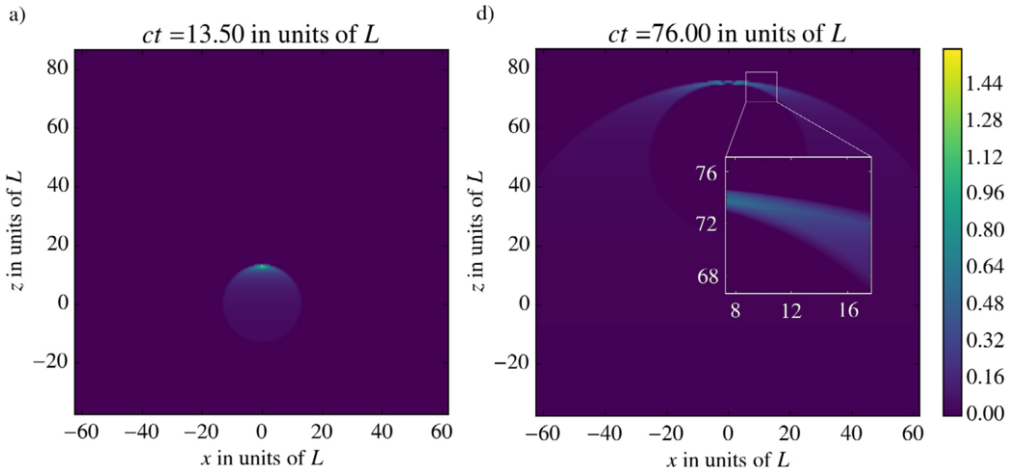


Fig. 2. – The plots show the metric perturbation $h^{\mathrm{P}} = h_{00} = h_{11} = -h_{10} = -h_{01}$ for a pulse of length L in the coordinates (ct, x, y, z) in the (x, y)-plane for different times t. h^{P} is normalized to units of κ and then the logarithm of the logarithm is taken. The figure was originally published in [8] under the CC Attribution 3.0 license `https://creativecommons.org/licenses/by/3.0/`.

pulse much smaller than the distance from the observer to the emitter, the gravitational field approaches the form of a plane fronted parallel propagating wave (pp-wave) [4].

In fig. 3, the metric perturbation is plotted for linearly and circularly polarized light and $\rho/r \ll 1$. The metric perturbation is modulated as the norm of the electric field strength ($|E|^2$) for linearly polarized light. For circularly polarized light no modulation arises. Since the energy momentum tensor is independent of the orientation of the polarization, the metric perturbation is independent of the orientation of the polarization. The laser pulse induces curvature only on spherical shells of thickness L (pulse length) that expand from the point of creation and annihilation. For the most part, these shells are causally disconnected from the pulse. Therefore, the curvature is due to creation/annihilation alone. In particular, the emission induces an attraction and the absorption induces a repulsion.

The world line $\gamma^\mu(\lambda)$ of free test particles is governed by the geodesic equation $\ddot{\gamma}^\mu = -\Gamma^\mu_{\rho\sigma}\dot{\gamma}^\rho\dot{\gamma}^\sigma$ with $g_{\mu\nu}\dot{\gamma}^\mu\dot{\gamma}^\nu = -1$ for massive test particles, and $g_{\mu\nu}\dot{\gamma}^\mu\dot{\gamma}^\nu = 0$ for massless test particles. The strongest laser pulses available today have a pulse power in the range of $10^{15}\,\mathrm{W}$. At a distance of 2.5 mm for a massive test particle at rest, this gives the acceleration $\ddot{\gamma}^x \approx -\frac{4GP}{c^3 x} \approx -10^{-18}\frac{\mathrm{m}}{\mathrm{s}^2}$. We can compare this acceleration to the acceleration experienced by a test particle in the Newtonian potential induced by a small spherical, massive object. At a distance of $r = 2.5\,\mathrm{mm}$, a mass of only $M = 10^{-13}\,\mathrm{kg}$ would be necessary to provide the same acceleration as the laser we considered above.

A periodically pulsed laser produces a periodically varying gravitational field. Close to the beamline, the corresponding tidal forces can be compared to those induced by a gravitational wave. For a laser of $10^{15}\,\mathrm{W}$, they are of the same order as those due to a

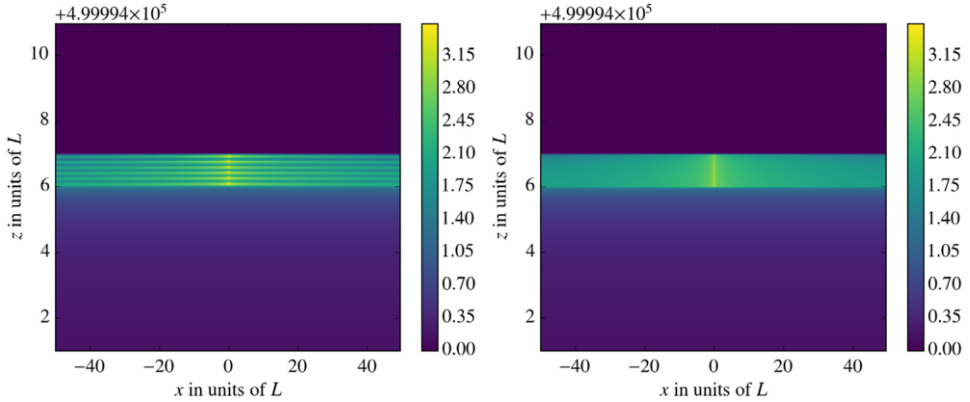


Fig. 3. – These plots show the double logarithm of the metric perturbation h^{P} for a linearly polarized pulse of length L and central wavelength $\lambda = \frac{2\pi c}{\omega} = \frac{2}{3}L$ in the x-y-plane at $t = 50000L/c$, after its emission at $z = 0$. h^{P} is normalized to units of $\kappa = 4GAu_0/c^4$ and then the logarithm of the logarithm is taken. The figure was originally published in [8] under the CC Attribution 3.0 license `https://creativecommons.org/licenses/by/3.0/`.

gravitational wave of angular frequency $\omega = 10^3$ Hz and amplitude $h_+ = 10^{-22}$. This is of the same order as the strain induced by gravitational waves detected by LIGO, however, only in a very small region close to the beamline. Hence, the detection of gravitational waves could be an ideal application for micrometer scale gravitational wave detectors as the one proposed in [10].

In agreement with the authors of [3], we find from the geodesic deviation equation that massless test particles moving in the same direction as the pulse are not deflected while counter-propagating pulses are deflected four times more strongly than massive particles at rest containing the same energy. The exact differential cross section (DCS) for the scattering of infinitely short light pulses based on the classical results for their gravitational field presented in [5] and [4] was derived in [11]. In the limit of one photon per pulse, this DCS can be seen as the classical DCS for photon-photon scattering. For small scattering angles, it coincides with the polarization averaged DCS from Perturbative Quantum Gravity (PQG) which was derived in [12, 13].

The independence of the gravitational field of a laser pulse from its helicity is one of the most interesting predictions of our work. In contrast in the framework of PQG the gravitational interaction between photons depends on their polarization. This result was already derived in [14] and [13].

In PQG, the initial two-photon state can also be entangled like $|\Psi\rangle_{\varphi,\rho} = \cos\varphi|p_1, 1; p_2, 2\rangle + e^{i\rho}\sin\varphi|p_1, 2; p_2, 1\rangle$, where $\xi_1, \xi_2 \in \{1, 2\}$ in the state $|p_1, \xi_1; p_2, \xi_2\rangle$ label the linear polarization directions (in plane $\xi_i = 2$) and $\varphi \in [0, \pi/2]$ parametrizes the entanglement of the state: for $\varphi = 0$ and $\varphi = \pi/2$, the state is not entangled, and for $\varphi = \pi/4$, it is maximally entangled. The parameter $\rho \in [-\pi/2, 3\pi/2)$ governs the relative phase of the superposed states $|p_1, 1; p_2, 2\rangle$ and $|p_1, 2; p_2, 1\rangle$. In particular,

$|\Psi\rangle_{\pi/4,0} = |\Psi^+\rangle$ and $|\Psi\rangle_{\pi/4,\pi} = |\Psi^-\rangle$ are known as the symmetric and the anti-symmetric Bell state, respectively. We find for the DCS

$$\frac{\mathrm{d}\sigma_{|\psi\rangle_e}}{\mathrm{d}\Omega} = \frac{8}{\sin^4\theta}\frac{l_P^4}{\lambda^2}\left[4(1+\sin(2\varphi)\cos\rho) + (1-\sin(2\varphi)\cos\rho)\left(\cos\theta+\cos^3\theta\right)^2\right], \tag{1}$$

where $l_P = \sqrt{\frac{G\hbar}{c^3}} \approx 1.6162 \times 10^{-35}\,\mathrm{m}$ is the Planck length and λ is the wavelength of the photons in the center of momentum frame. We published this result in [15] together with a more extensive derivation and interpretation. We found that polarization entangled photons gravitate more in the symmetric Bell state, $\rho = 0$ and $\varphi = \pi/4$, and less in the antisymmetric Bell state, $\rho = \pi$ and $\varphi = \pi/4$. Furthermore, the differential cross section for photon-photon scattering due to vacuum polarization in Quantum Electrodynamics shows the same dependence on the entanglement. The dependence on the polarization entanglement can be interpreted in the sense of quantum interference and in the sense of localized particles using Glauber's delayed coincidence measurement. The latter leads to the probability [16]

$$P_{i\to f} \propto \sum_f \left|\langle f|E_2^{(+)}(t,x)E_1^{(+)}(t',x')|\Psi\rangle_{\varphi,\rho}\right|^2 \propto \left[1+\sin(2\varphi)\cos\left(\frac{2}{\hbar}p\cdot(x'-x)+\rho\right)\right].$$

This shows that the probability to find two photons at two given points closer than $\pi\lambda/4 = \pi\hbar c/4E$ is increased for photons in the symmetric Bell state and decreased for photons in the antisymmetric Bell state when compared to not entangled photons, $\varphi = 0$ and $\varphi = \pi/2$. This naturally fits into the idea of particles interacting via forces that decay with the distance between these particles.

By investigating the gravitational properties of light we can hope to gain insight into the properties of photons. We have gone beyond a classical treatment of the gravitational properties of light by investigating photon-photon scattering in Perturbative Quantum Gravity. In particular, we have investigated the effect of entanglement in photon-photon scattering. The DCS for gravitational photon-photon scattering at a wavelength of 500 nm is of the order $10^{-124}\,\mathrm{m}^2$. Hence, the the effect is far from being directly experimentally accessible. However, as entanglement is an inherently quantum mechanical property, the effect of the entanglement of light on its gravitational properties exists in the overlap between quantum mechanics and gravity, which makes it a question of general physical interest.

REFERENCES

[1] WILL C. M., *Living Rev. Relat.*, **9** (2006) 3.

[2] MISNER C. W., THORNE K. S. and WHEELER J. A., *Gravitation* (W.H. Freeman) 1973, pp. 964.

[3] TOLMAN R. C., EHRENFEST P. and PODOLSKY B., *Phys. Rev.*, **37** (1931) 602. http://link.aps.org/doi/10.1103/PhysRev.37.602

[4] Bonnor W. B., *Commun. Math. Phys.*, **13** (1969) 163.
http://projecteuclid.org/euclid.cmp/1103841572
[5] Aichelburg P. and Sexl R., *Gen. Relativ. Gravit.*, **2** (1971) 303.
http://dx.doi.org/10.1007/BF00758149
[6] van Holten J., *Fortsch. Phys.*, **59** (2011) 284.
http://dx.doi.org/10.1002/prop.201000088
[7] Scully M. O., *Phys. Rev. D*, **19** (1979) 3582.
http://link.aps.org/doi/10.1103/PhysRevD.19.3582
[8] Rätzel D., Wilkens M. and Menzel R., *New J. Phys*, **18** (2016) 023009.
http://stacks.iop.org/1367-2630/18/i=2/a=023009
[9] Rätzel D., Wilkens M. and Menzel R., *Phys. Rev. D*, **95** (2017) 084008.
http://arxiv.org/abs/1607.01310
[10] Sabin C., Bruschi D. E., Ahmadi M. and Fuentes I., *New J. Phys.*, **16** (2014) 085003.
http://stacks.iop.org/1367-2630/16/i=8/a=085003
[11] Schücker T., *Class. Quantum Grav.*, **7** (1990) 503.
http://stacks.iop.org/0264-9381/7/i=3/a=024
[12] Barker B. M., Bhatia M. S. and Gupta S. N., *Phys. Rev.*, **162** (1967) 1750.
http://link.aps.org/doi/10.1103/PhysRev.162.1750
[13] Boccaletti D., de Sabbata V., Gualdi C. and Fortini P., *Nuovo Cimento B*, **60** (1969) 320.
http://dx.doi.org/10.1007/BF02710233
[14] Barker B. M., Bhatia M. S. and Gupta S. N., *Phys. Rev.*, **158** (1967) 1498.
http://link.aps.org/doi/10.1103/PhysRev.158.1498
[15] Rätzel D., Wilkens M. and Menzel R., *EPL*, **115** (2016) 51002.
[16] Glauber R. J., *Phys. Rev.*, **130** (1963) 2529.
http://link.aps.org/doi/10.1103/PhysRev.130.2529

International School of Physics "Enrico Fermi"
Villa Monastero, Varenna
Course 197
8 - 13 July 2016

In collaboration with the Wilhelm und Else Heraeus Foundation

Foundations of Quantum Theory

Directors

Ernst Rasel
Institut für Quantenoptik
Leibniz Universität Hannover
Welfengarten 1
D-30167 Hannover
Germany
tel: 0049 511 762 19203
rasel@iqo.uni-hannover.de
faber@iqo.uni-hannover.de

Wolfgang Schleich
Institut für Quantenphysik,
Center for Integrated Quantum Science and Technology (IQ^{ST})
Universität Ulm
Albert-Einstein-Allee 11
D-89081 Ulm
Germany
tel: 0049 731 502 3080
wolfgang.schleich@uni-ulm.de

Lecturers

Robert Boyd
Department of Physics
University of Ottawa
Ottawa ON K1N 6N5
Canada
tel: 001 585 275 2329
boydrw@mac.com

Giacomo Mauro D'Ariano
Istituto Nazionale di Fisica
della Materia
Unità di Pavia
Dipartimento di Fisica "A. Volta"
Via Bassi 6
I-27100 Pavia
Italy
tel: 0039 0382 98748
dariano@unipv.it

Scientific Secretary

Sabine Wölk
Department Physik
Universität Siegen
Emmy-Noether-Campus
Walter-Flex-Str. 3
D-57068 Siegen
Germany
tel: 0049 271 740 3799
woelk@physik.uni-siegen.de

Edward Fry
Department of Physics and Astronomy
4242 TAMU
College Station
TX 77843-4242
USA
tel: 001 979 845 7717
fry@physics.tamu.edu

Chris Fuchs
Department of Physics
University of Massachusetts
100 Morrissey Blvd.
Boston
MA 02125-3393
USA
tel: 001 617 287 6050
Christopher.Fuchs@umb.edu
QBism.Fuchs@gmail.com

Daniel Greenberger
City College of the City University
of New York
New York, 10031
USA
tel: 001 212 650 6940
dan.greenbergernyc@gmail.com

Nancy Thorndike Greenspan
7201 Glenbrook Road
Bethesda
MD 20814
USA
tel: 001 301 657 8745
nancy@nancygreenspan.com

Manfred Kleber
Physics Department
Technische Universität München
James-Franck-Str. 1
D-85748 Garching b. München
Germany
tel: 0049 89 289 12336
mkleber@tum.de

Gerd Leuchs
Max Planck Institut
for the Science of Light
Staudtstraße 2
D-91058 Erlangen
Germany
tel: 0049 9131 7133 101
gerd.leuchs@mpl.mpg.de
ulrike.bauer-buzzoni@mpl.mpg.de

Ralf Menzel
Institut für Physik und Astronomie
Universität Potsdam
Karl-Liebknecht Str. 24/25
D-14476 Potsdam-Golm
Germany
tel: 0049 331 977 1104
Menzel@uni-potsdam.de

Marlan Scully
Department of Physics
Texas A&M University
4242 TAMU
College Station
TX 77843-4242
USA
tel: 001 979 845 1534
scully@tamu.edu

Aephraim Steinberg
Department of Physics
University of Toronto
60 St. George St.
Toronto, ONT M5S 1A7
Canada
tel: 001 416 978 0713
steinberg@physics.utoronto.ca

Andrew White
Department of Physics
The University of Queensland
St Lucia
Brisbane 4072
Australia
tel: 0061 7 3365 7902
andrew.white@uq.edu.au
agx.white@gmail.com

Students

Sven Abend	Universität Hannover, Germany
Sanah Altenburg	Universität Siegen, Germany
Yael Avni	Weizmann Institute of Science, Israel
Stefano Bacchi	Università di Trieste, Italy
Ömer Bayraktar	Max Planck Institute for the Science of Light, Germany
Jonathan Ben-Benjamin	Texas A&M University, USA
Iva Bezdekova	Czech Technical University in Prague, Czech Republic
Pawel Blasiak	Polish Academy of Sciences, Poland
Marduk Bolaños Puchet	Universität Duisburg-Essen, Germany
Gudrun Bosch	Technische Universität Darmstadt, Germany
Maria Chiara Braidotti	Università di L'Aquila, Italy
Oleksandr Burlayenko	Kharkiv, Ukraine
Esteban Castro Ruiz	University of Vienna, Austria
Lorenzo Catani	University College London, UK
Robin Corgier	Leibniz Universität Hannover, Germany
Paulina Corona-Ugalde	University of Waterloo, Canada
Matthias Dahlmanns	University of Duisburg-Essen, Germany
Manuel Daiber	Ulm University, Germany
John DeBrota	University of Massachusetts, USA
Marco Erba	Università di Pavia, Italy
Gabriel Fagundes	University of the Basque Country, Spain
Adrien Feix	University of Vienna, Austria
Alexander Friedrich	Ulm University, Germany
Thomas Galley	University College London, UK
Matthias Germann	University of Basel, Switzerland
Flaminia Giacomini	University of Vienna, Austria
Marta Greselin	Università di Bergamo, Italy
Philipp Griebling	Technische Universität München, Germany
Kevin Günthner	Max Planck Institute for the Science of Light, Germany
Lucas Happ	Ulm University, Germany
Jannik Hoffmann	Universität Siegen, Germany
Patrick Huber	Universität Siegen, Germany
Radu Ionicioiu	Horia Hulubei National Institute of Physics and Nuclear Engineering, Romania
Pavel Jefremov	University of Bremen, Germany
Michael Jex	Czech Technical University in Prague, Czech Republic
Nandan Jha	Leibniz Universität Hannover, Germany
Tristan M. Kraft	Universität Siegen, Germany

Anton Krieger	Friedrich-Wilhelms-Universität Bonn, Germany
Fabiano Lever	Universität Siegen, Germany
Martin Malachov	Czech Technical University in Prague, Czech Republic
Natalia Salome Moller	Universidade Federal de Minas Gerais, Brasil
Nuria Munoz Gargante	Barcelona, Spain
Hai-Chau Nguyen	Max Planck Institute for the Physics of Complex Systems, Germany
Andrea Petrucci	Università di Roma 3, Italy
Luo Qi	Max Planck Institute for the Science of Light, Germany
Dennis Rätzel	University of Potsdam, Germany
Janis Hendrik Rosskamp	University of Bremen, Germany
Piotr Roztocki	Institut National de la Recherche Scientifique, Canada
Michele Sacerdoti	Università di Milano, Italy
Amirhossein Sadeghi	University of Vienna, Austria
Matteo Schiavon	Università di Padova, Italy
Stefania Sciara	Institut National de la Recherche Scientifique, Canada
Sorayya Shadi	Max Planck Institute for the Science of Light, Germany
Hendrik Siebeneich	Universität Siegen, Germany
Vyome Singh	University of St Andrews, UK
Michael Taheri	Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
Athanasios C. Tzemos	Academy of Athens, Greece
Murat Hazer Uygunol	Ege Üniversitesi Kampüsü, Turkey
Reinhold Walser	Technische Universität Darmstadt, Germany
Ramona Walter	Technische Universität München, Germany
Da-Wei Wang	Texas A&M University, USA
YiFan Wang	Universität zu Köln, Germany
Matthias Zimmermann	Ulm University, Germany

Observers

Thomas Driebe	DLR Space Administration, Germany
Wolfgang Ertmer	Leibniz Universität Hannover, Germany
Julian Fleniken	US Army Research Laboratory, USA
Rainer Forke	German Aerospace Research Center, Germany
Stefan Jorda	Wilhelm und Else Heraeus-Stiftung, Germany
Jürgen Mlynek	Humboldt-Univeristät zu Berlin, Germany
Sahar Sahebdivan	University of Vienna, Austria

Finito di stampare
nel mese di Gennaio 2019