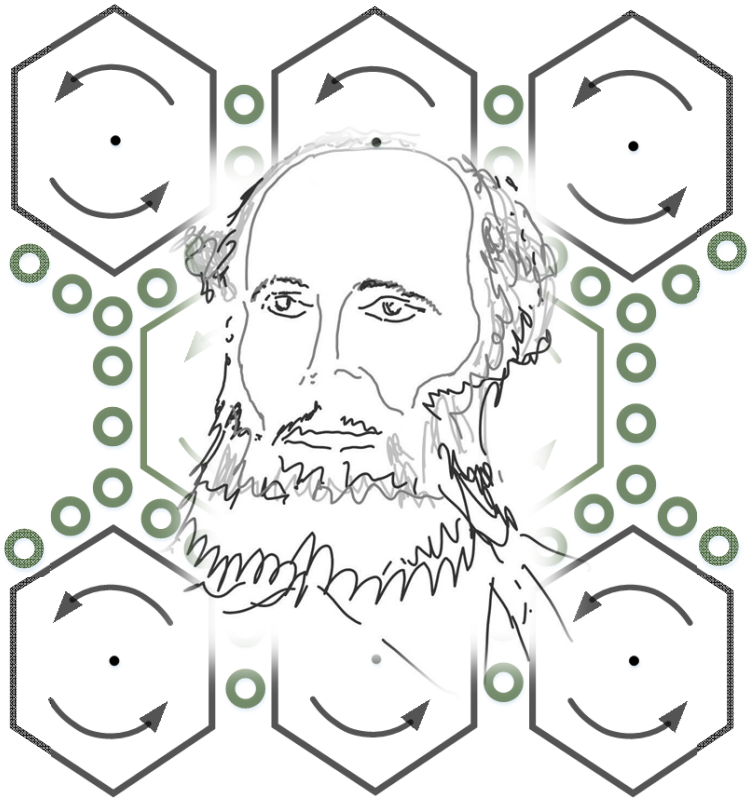
Quantum Information Theory for Engineers: An Interpretative Approach

Lecture Notes by C. Jansson

November 6, 2017



Contents

1 Introduction

What is the background of this tutorial? These notes comprise lectures on quantum computing and quantum information theory that I taught during the last ten years for students of electrical engineering. The well-known and excellently written book of Nielsen and Chuang[1] presents an extensive treatment of various issues and concepts in quantum computing and quantum information theory. There are also many other good introductions to quantum information theory. So, why these lecture notes?

Physics is known to be a hard science. The physical laws are difficult to interpret, the meaning of physical quantities and the relationship to reality is vague, and the necessary mathematics requires some effort and capabilities. In particular, quantum physics is perhaps recognized as the most difficult subject in the physics curriculum. It is not only the mathematics used in introductory courses about quantum theory that causes problems. It is the way of relating mathematical quantities and concepts, such as superposition, wave functions, spin, or entangled vectors, to the reality we observe.

A major goal of these notes is to present an alternative entrance to quantum information theory that is, hopefully, suitable for students studying engineering, but perhaps also to people interested in philosophy of physics.. To my knowledge engineers like machines, experimental set-up's and concrete things, but not so much abstract mathematical formalisms. These (hopefully enjoyable) notes try to take this mentality into consideration by describing many experiments, among them those exhibiting strange behaviours. Moreover, whenever possible, the mathematical formalism is described in an easily understandable, but sufficiently extensive, manner.

Is quantum mechanics really difficult? In many textbooks the *abstract principles of quantum mechanics* can be found in the following form, or in a slightly modified manner:

- **Principle 1 (quantum state space):** To every quantum object, quantum system, or quantum process a complex vector space (*Hilbert space*), called the *quantum state space*, is associated. The complex vectors $|\psi\rangle$ in this Hilbert space are called *quantum states*. Two vectors that differ only by a complex multiple represent the same quantum state.

- **Principle 2 (observable):** To every *observable* or dynamical variable of a quantum system a Hermitian operator $\hat{A}$ acting on a quantum state space is associated. The only possible results of a measurement are the eigenvalues of this Hermitian operator.

- **Principle 3 (measurement):** If an eigenvalue of an observable $\hat{A}$ is measured, then the state of the quantum system jumps to an eigenvector corresponding to this eigenvalue. Quantum mechanics is a probability theory. The expectation value of an observable $\hat{A}$ for a quantum system in quantum state $|\xi\rangle$ is $\langle\xi|\hat{A}|\xi\rangle$.

[1] Nielsen, Chuang [2010]

- **Principle 4 (Born's probability rule):** The transition probability for a given normalized quantum state $|\psi\rangle$ jumping to any normalized state $|\phi\rangle$ is the squared absolute value $|\langle\psi|\phi\rangle|^2$.
- **Principle 5 (time evolution):** The evolution of state vectors of an isolated quantum system is *unitary* with respect to time.
- **Principle 6 (composition rule):** Two or more quantum systems are represented by the *tensor product* of the corresponding quantum state spaces.

Basically, Principle 1 shows the arena for quantum mechanics. Principles 2, 3, and 4 state that quantum mechanics is a linear stochastic theory. Principle 5 says that the dynamics of isolated quantum systems is a unitary deterministic evolution, and in particular a reversible process. Principles 3 and 5 show the paradox that quantum mechanics is governed by two dynamics: the irreversible *state function collapse*, where the superposition of eigenstates reduces to a single eigenstate by measurement or observation, whereas the other dynamics is a continuous, deterministic and unitary transformation. Principle 6 tells us that different state spaces of quantum systems are combined to a composite system via the tensor product.

Representative for many other physicists we cite three well-known scientists. They write about these apparently weird quantum principles:

> *Nothing could be more arid than the principles of quantum mechanics. Its concepts and laws are cast in a blunt, inescapable mathematical form, without a trace of anything intuitive, a total absence of the obviousness we see in the things around us. And yet, this theory penetrates reality to a depth our senses cannot take us. Its laws are universal, and they rule over the world of objects so familiar to us. We, who inhabit this world, cannot make our own vision prevail over those arrogant laws, whose concepts seem to flow from an order higher than the one inspired by the things we can touch, see, and say with ordinary words.* Omnes[2] 1999

> *Might I say immediately, so that you know where I really intend to go, that we always have had (secret, secret, close the doors!) we always have had a great deal of difficulty in understanding the world view that quantum mechanics represents.* Feynman[3] 1982

> *Those who are not shocked when they first come across quantum mechanics cannot possibly have understood it.* Attributed to Bohr

Schrödinger cats, Wigner's friend, many worlds and many minds in non-relativistic quantum mechanics, and moreover time dilation, length contraction, the lack of simultaneity, worm-holes, non-causality, and several other

[2]Omnes [1999, page 163]
[3]Feynman [1982, page 471]

paradoxes in relativistic quantum mechanics do not make things easier (see Figures 1, 2[4], 3, and 4[5]).

Because the fundamental parts of physics are quantum theory and the theory of relativity, not surprisingly many physicists view their science as a collection of purely mathematical concepts without relation to reality. For instance, the positivist and cosmologist Tegmark writes:

> *I advocate an extreme "shut-up-and-calculate" approach to physics, where our external physical reality is assumed to be purely mathematical. This brief essay motivates this "it's all just equation" assumption and discusses its implications.* Tegmark[6] 2007

We shall spend a lot of time in finding out, however, that the previous principles are actually very intuitive and reasonable, and can be deduced from easily comprehensible fundamental rules. This road starts by justifying the use of complex numbers and ends by discussing some aspects of time. Although quantum mechanics seems to be a rather bizarre description of reality, we should keep in mind that it is known as our most fundamental and successful physical theory ever which is irreducible and random. Their predictions have been tested to an unprecedented accuracy.

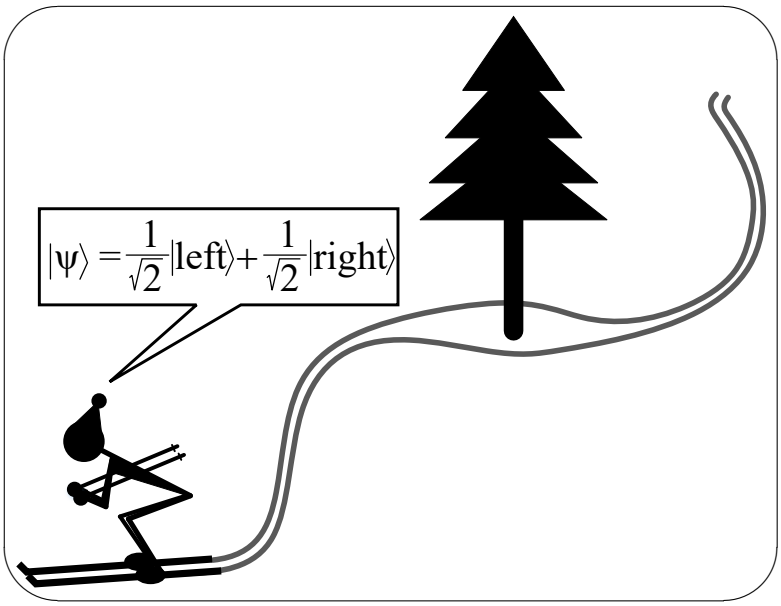


Figure 1: To show the strange behavior of quantum particles, in some books the quantum skier is used who goes both ways at once. They remark that although we do not observe such phenomena, they happen in the quantum world.

[4]Dhatfield [2008]
[5]Schirm [2011]
[6]Tegmark [2007]

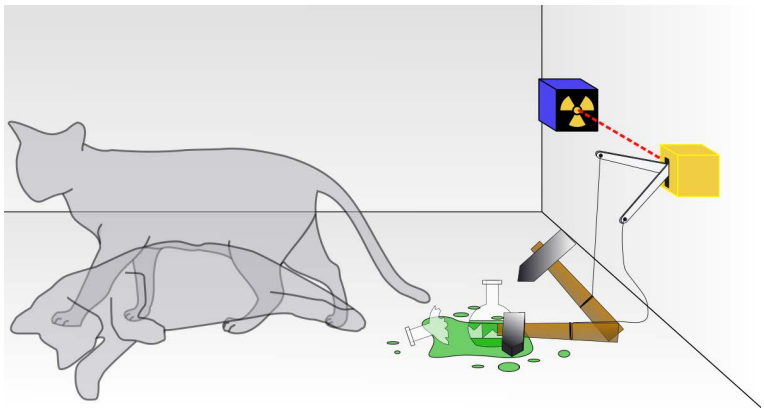


Figure 2: Schrödinger's cat is a famous illustration of the principles of superposition and entanglement in quantum theory, proposed by Erwin Schrödinger in 1935. Originally, his intention for this cat-killing box was to discredit non-intuitive implications of quantum mechanics.

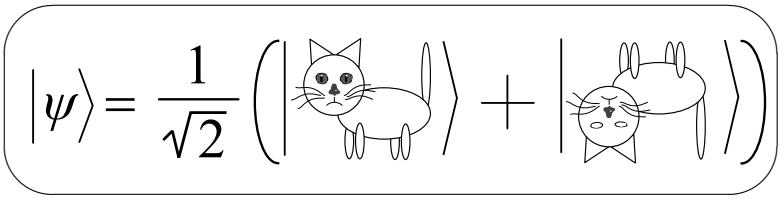


Figure 3: Superposition of alive and dead.

What is physics today? What is truth in physics? The expression "physics" comes from the ancient Greeks and means "knowledge of nature". This implies that we are and should be interested in true information about our world. What is true? This is a difficult question, and in fact von Weizsäcker[7] starts his book about the structure of physics with the question: What is the truth of physics?

On the gravestone of the well-known Russian mathematician, physicist, and philosopher Danilovich Aleksandrov (1912 - 1999) one finds the graving[8]: *"The truth is the only thing, which is worthy of our worshipping".* It should get the alarm bells ringing that the famous mathematician and physicist Roger Penrose, author of the excellent book "The Road to Reality, A complete Guide to the Laws of the Universe" said in an interview:

> *Physics is wrong, from string theory to quantum mechanics.*
> Roger Penrose, 2009, DISCOVER

In 2010 he said farewell to our celebrated *"big-bang theory"* and proposed the old *"steady-state model"*:

> *The scheme that I am now arguing for here is indeed unorthodox, yet it is based on geometrical and physical ideas which are very*

[7] von Weizsäcker [1988]
[8] Gessen [2013, page 151]

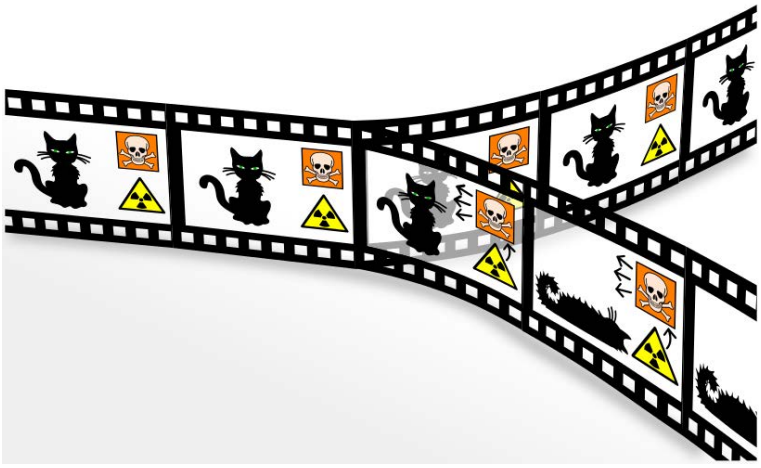


Figure 4: The *many worlds theory* claims that the universe splits into distinct worlds, in order to have only one unitary dynamics. This would imply the accommodation of each possible outcome.

> *soundly based. Although something entirely different, this proposal turns out to have strong echoes of the old steady-state model!* Penrose[9]

By the way, the widely glorified and seemingly experimentally verified message about the age of our universe would be wrong when believing Penrose. On the contrary, in Wikipedia, and not only there, you find the confirmation of the age of our universe in high precision:

> *In physical cosmology, the age of the universe is the time elapsed since the Big Bang. The best measurement of the age of the universe is 13.798 pm 0.037 billion years within the Lambda-CDM concordance model. The uncertainty of 37 million years has been obtained by the agreement of a number of scientific research projects, such as microwave background radiation measurements by the Planck satellite, the Wilkinson Microwave Anisotropy Probe and other probes.* Wikipedia, August 2012

All this becomes even more strange when knowing the *Wheeler DeWitt equation*: This field equation is one approach to the famous problem of quantum gravity, and turns out to be timeless. No time, no age of the universe? However, already Landau mentioned

> *Cosmologists are often in error, but never in doubt.* Lev Landau[10]

So who is right? Perhaps such contradictory statements support the well-known rule in physics:

[9] Penrose [2010, Preface]

[10] http://yquotes.com/quotes/lev-davidovich-landau/

The theory determines what we observe and measure.

This rule may point to circular reasoning. Measurements confirm the theory, which then determine the measurement results.

Among physicists there is no fundamental disagreement on how to use mathematical formalisms for practical applications. However, there are deep differences in the understanding and meaning of physical theories, in particular, of quantum theory but even so in the theory of relativity. For instance, almost all ancient famous scientists, among them Descartes, Newton, Gauss, Riemann, Lord Kelvin, and Maxwell, investigated different mechanical models of "ether". Actually, Maxwell derived his equations from a purely mechanical ether model, see the cover sheet of these notes. Einstein finished the ether in a lone hand. The Nobel honoree in physics Laughlin writes about *ether* in contemporary theoretical physics:

> *It is ironic that Einstein's most creative work, the general theory of relativity, should boil down to conceptualizing space as a medium when his original premise [in special relativity] was that no such medium existed [..] The word* **'ether'** *has extremely negative connotations in theoretical physics because of its past association with opposition to relativity. This is unfortunate [...] Relativity actually says nothing about the existence or nonexistence of matter pervading the universe, only that any such matter must have relativistic symmetry. [..] Subsequent studies with large particle accelerators have now led us to understand that space is more like a piece of window glass than ideal Newtonian emptiness. It is filled with 'stuff' that is normally transparent but can be made visible by hitting it sufficiently hard to knock out a part. The modern concept of the vacuum of space, confirmed every day by experiment, is a relativistic ether. But we do not call it this because it is* **taboo**. Laughlin[11]

Apparently, this taboo indicates that also many scientists behave politically correct, perhaps also in spite of a clearer understanding. Consequently, the truth content of physical theories is very controversial and doubtful.

In particular, for engineers interested in these aspects, we mention that there are plenty of books and articles with a critical attitude that clearly point out shortcomings, critics and incredible models of the present physics[12]. These books contain many further references. The history of science clearly demonstrates that doubtfulness is the chief impelling force, not faith. Unfortunately, as far as I know, almost all time in the education at schools and universities is spent with believing and learning, not casting doubt on the theories. Perhaps the literature mentioned above might help.

What can we understand? What should students understand? Understanding is a complex psychological process and hard to describe in an

[11] Laughlin [2005, pp. 120–121]

[12] Herbert [1985], Mirman [2001, 2006], Omnes [1999], Rothmann [2012], Selleri [1990], Sheldrake [2013], Smolin [2006, 2013], Unzicker [2010, 2012, 2013]

abstract form. Serious scientists as well as engaged students have frequently doubts about the proper understanding of their science. In particular, many students have profound doubts whether they have understood the lectures they have heard, or the textbooks they have read. Partially, these doubts are well-founded, but sometimes these doubts arise because certain scientific statements are fired from teachers or other students as simple or trivial. Let us consider three well-known examples.

One seemingly very simple statement is the *principle of inertia*, already described in physics courses in school. Is it in fact simple? Do we understand it? There is a nice story in YouTube[13] where Richard Feynman talks about his father:

> *My father taught me to notice things. One day I was playing with an express wagon, a little wagon with a railing around it. It had a ball in it, and when I pulled the wagon I noticed something about the way the ball moved. I went to my father and said,"Say, pop, I noticed something. When I pull the wagon, the ball rolls to the back of the wagon. And when I am pulling it along and I suddenly stop, the ball rolls to the front of the wagon. Why is that? ...*
>
> *"That, nobody knows", he said. "The general principle is that things which are moving tend to keep on moving, and things which are standing still tend to stand still, unless you push them hard. This tendency is called inertia, but nobody knows why it's true."*

Feynman was proud of the way he was educated by his father, who gave him **the difference between knowing the name of something and understanding something.** That the principle of inertia is in fact a deep physical problem and paradox is only rarely mentioned. An exception is the book of von Weizsäcker[14] who writes:

> *The law of inertia, which empirically enforces the occurrence of second derivatives in the equation of motion, is fundamental for classical mechanics. It, however, represents a causal paradox. Aristotle understood motion as a change of state, and thus force as the cause of motion. In classical mechanics, however, inertial motion is just the motion without any forces acting. In the seventeenth century one still felt the paradox therein; Descartes and, following him, Newton defined the state of a body in terms of its velocity such that only acceleration was seen as a change of state. But this is inconsistent, as two bodies with the same velocity but at different locations are in different states, as correctly put by the modern description in phase space; and during inertial motion the point in phase space varies. If one wants to think causally in a consistent way, one must radicalize Mach's ideas and interpret the inertial motion as being caused by the universe (the distant masses). This*

[13] `https://www.youtube.com/watch?v=Zjm8JeDKvdc`

[14] von Weizsäcker [2006, p. 29]

> *I have attempted in the ur theory, but now I doubt whether this is an adequate formulation.*

Another example is the theory of special relativity. It is widely accepted, widely tested, and widely understood, although it is a weird theory containing strange interpretations such as time dilation, Lorentz contraction, velocity addition, the relativity of simultaneity, moving frames, observers and many other mysterious quantities. However, the mathematics behind is simple since it requires only matrix-vector operations. But nobody should be shamed to admit having difficulties with this theory. Even their excellent co-founders Lorentz and Poincaré seemed not to have understood relativity, as Pais[15] writes:

> *In later years all three man, Einstein, Lorentz, and Poincaré, reacted to the special theory of relativity in ways which arouse curiosity. Why on the whole, was Einstein so reticent to acknowledge the influence of the Michelson-Morley experiment on his thinking? Why could Lorentz never quite let go of the aether? Why did Poincaré never understand special relativity?*

The book of Pais is very suitable for reading, contains many details about the history of the theory of relativity, and gives a very good insight into the thinking of their founders.

The third example is the well-known *Monty Hall problem*, a rather simple probabilistic problem, named after Monty Hall who presented it the first time in 1975. Even Erdös, one of the greatest experts in number and probability theory, could not solve this puzzle, and was unconvinced until a computer simulation, confirming the predicted result, was shown to him, see Vazsonyi[16].

There are various other examples in science, and sometimes this has unpleasant consequences. Griffiths[17] writes about Bohr:

> *It is interesting to note that Bohr was an outspoken critic of Einstein's light quantum (prior to 1924), that he mercilessly denounced Schrödinger's equation, discouraged Dirac's work on the relativistic electron theory (telling him, incorrectly, that Klein and Gordon had already succeeded), opposed Pauli's introduction of the neutrino, ridiculed Yukawa's theory of the meson, and disparaged Feynman's approach to quantum electrodynamics. Great scientists do not always have good judgment - especially when it concerns other people's work - but Bohr must hold the all-time record.*

When we don't impute some kind of badness, it seems that Bohr had difficulties to understand what other physicists published in his area of expertise.

My advice to students is: take the liberty to have doubts, for the moment skip the things you do not understand, and do not fear to make errors. Nobody is perfect, and science is not error-free, as we have already remarked.

Which problems may occur when teaching quantum physics?

[15] Pais [2005, p. 164]
[16] Vazsonyi [1999, pp. 17–19]
[17] Griffiths [2004, p. 23]

> *What is weird is not nature, not physics, not the universe, but the obsession of so many physicists with demonstrating their inability to understand physics by regarding nature as spooky.* Mirman[18]

For many students quantum physics is a hard science. In particular, if they are educated by the well-known "Shut up and calculate" approach, physics may become a very dubious pleasure, not really funny, and perhaps only suitable to earn some money, on a later date. If in textbooks classical concepts are used to generate contradictions and paradoxes such that quantum physics seems to be magical, then confusion arises (see Figures 1, 2, 3, and 4). Magic may be good for entertainment, but it is not really helpful for students to understand the physical concepts. Also for philosophers, interested in quantum physics and broader questions about nature, ontology, and human knowledge, magic is a bad adviser.

There is no experimental evidence whatsoever to support such magic, but naive pictures. Quantum paradoxes originate solely from erroneous interpretations of quantum theory. Only pseudo-realistic philosophies together with the misuse of the underlying mathematical concepts leads to paradoxical results.

Here, we try to avoid such magical descriptions, and instead try to maintain a more critical attitude:

> *If a man will begin with certainties, he shall end in doubts; But if he will be content to begin with doubts, he shall end in certainties.*
> Francis Bacon, Advancement of Learning.

What are the major differences to other textbooks on quantum information and quantum computation? In this lecture notes there are some interpretations and models that are different, or perhaps some may be new, when comparing with other books.

> *This work contains many things which are new and interesting. Unfortunately, everything that is new is not interesting, and everything which is interesting, is not new.* Lev Landau[19]

Also on the danger that Landau is right, I would like to mention the following points:

- We support the understanding of this difficult subject by **numerous experiments**, and use the ability of the engineer to think in terms of machines. This gives an **experimental entrance to quantum information theory**. The picture on the title page supports this point of view. It displays Maxwell and his mechanical notion of his famous equations describing electromagnetism.

- We argue why the **field of complex numbers $\mathbb{C}$ is the universal set of numbers** in science. In particular, it can be shown that classical mechanics and quantum mechanics can be embedded in the same

[18] Mirman [2001, p. 197]
[19] http://yquotes.com/quotes/lev-davidovich-landau/

mathematical framework when working with complex numbers; the real dynamical variables of classical mechanics are just the quantum mechanical average values of Hermitian operators. Consequently, classical mechanics works with real numbers and quantum mechanics with complex numbers.

- We explain the central topics of quantum mechanics and quantum computation only with the help of the simple and natural **Dirac-Feynman rules**, avoiding strange quantum mechanical principles whenever possible.

- Only a **minimal mathematical formalism** is necessary. Large parts of these notes can be already taught in school.

- These lecture notes are perhaps exceptional in a **unified treatment** of classical, random and quantum computation. In particular, **quantum computation** as well as **random computation** are described as a simple and natural mathematical **modification of classical reversible computation**, and in this way it is easy to understand for students having some knowledge in classical computation.

- We **avoid magic pictures and imaginations**, many paradoxes, and describe quantum computation as a linear stochastic process.

- We point out at various places the **conflict between quantum mechanics and the theory of relativity**.

- We treat quantum mechanics, classical probability theory and classical mechanics within the same mathematical framework which is based on *semimodules*. We obtain **unified definitions of states, observables and evolution operators that do not depend on the specific physical theory**. In particular, we introduce in our framework which describes various physical models the useful concepts *number representations*, *register representations*, and *vector representations*.

- We show that the apparent inconsistency between classical probability theory and quantum mechanics, as seen in slit experiments or Bell's inequality, can be resolved when carefully looking at the notion of outcomes in the classical theory and the possibilities in quantum mechanics.

- We derive the mathematical framework of quantum **two-state systems** and an **uncertainty principle** using only simple geometry at a macroscopic level. Microscopic properties like spin or polarization are not required.

- In physics time t is treated as an external background parameter. We avoid this concept, and use time only in terms of three modes, namely the trinity *future*, *present*, and *past*. Consequently spacetime vanishes. But we present an alternative approach to the Lorentz transform, the key to the theory of relativity.

- We explain the two fundamental problems in quantum theory, namely the meaning of **superposition** and **entanglement**, in terms of our time trinity.
- We investigate and try to explain the famous riddle of **inertia**.
- A fundamental assumption in physical theories is the dimension of the underlying space. In most cases this is the (3+1)-dimensional space-time. However, there are other well-known theories that use other dimensions, among them the 11-dimensional string theory or the 5-dimensional Kaluza-Klein theory. From the point of view of information theory we investigate in which dimensions a physical theory can be reasonable. Surprisingly, it turns out that, in some well-known theories, the underlying spaces do not satisfy even simple geometrical properties.
- We discuss very critically to what extent teleportation takes place in the experimental set-up.
- We give an interpretation of Heisenberg's uncertainty principle that is not common.

Which prerequisites are necessary? The following lecture notes go back to the quantum computation courses I taught during the last years. They are not written for experts or for brilliant students. They are mainly written for those readers who find the subject difficult. The text focuses on the major concepts of quantum information theory, while keeping mathematics to an indispensable minimum. We assume only a knowledge of the elementary facts of linear algebra, such as matrix-vector multiplication or eigenvalues and eigenvectors.

We abandon a separate section about linear algebra, since almost all textbooks about quantum mechanics or quantum computing contain such a section. In the afore mentioned book of Nielsen and Chuang[20] a very nice presentation is given in Section 2.1. Other descriptions, for instance, can be found in the books of Kaye, Laflamme, and Mosca[21], and in Plenio[22].

It should be clear that reading these notes only is not a good way to learn quantum information theory, especially without having any prior knowledge. I recommend some additional textbooks for students and engineers that are willing to broaden their knowledge about physics. Firstly, there are two books of the series *The Theoretical Minimum: What you Need to Know to Start Doing Physics*. The first one is written by Leonard Susskind and George Hrabovsky[23] and covers classical mechanics, the core of education in physics. The second one by Leonard Susskind and Art Friedman[24] explains quantum mechanics and its relationship to classical mechanics. Both volumes contain the necessary mathematical definitions, theorems, and prerequisites. They are

[20] Nielsen, Chuang [2010]
[21] Kaye et al. [2007, Chapter 2, pp. 799–800]
[22] Plenio [2002, Chapter 1]
[23] Susskind [2013]
[24] Susskind [2014]

based on Susskind's popular Stanford education courses that can be found in YouTube. Susskind, Hrabovsky and Friedman provide a toolkit for non-advanced students to learn physics at their own pace.

For a deeper understanding of physics, namely classical theories as well as quantum theory, we emphasize the older, but unforgettable, lecture notes of Feynman[25]. In contrast to many textbooks that teach mainly physical and mathematical formalisms in same compact manner, these notes explain physics from the very beginning by looking at many experiments and observations. It's just the way of how to think in physics.

Finally, for interested readers some further excellently written textbooks are mentioned[26].

What about the contents? I tried to write a small book guided by Landau's experience:

> *From thick books one can not learn anything new. A thick book is a cemetery where antiquated ideas are laid to rest.* Lev Landau[27]

The major goal of Section 2 of these lecture notes, is an elaborate presentation of some fundamental, strange sounding experiments of physics. These include experiments with photons, their polarization, reflection, multiple slit experiments, interaction free measurement, and delayed choice experiments, as well as Bell's inequality. These experiments are not only depicted, they are investigated using the Dirac-Feynman probability rules. Moreover, it is shown that these rules are almost inescapable and natural, leading to a description of quantum theory as a stochastic process with complex amplitudes.

In Section 3 the fundamental concepts of quantum computing and quantum information are described, including the similarity and differences between reversible classical computation, random computation, and quantum computation. Moreover, quantum teleportation as well as applications of quantum parallelism are considered.

In Section 4 we present some fundamental concepts and foundations of quantum theory. In particular, we try to give precise and unified definitions of base states, states, observables, compositions of systems, and transformations between systems. We replace the concept of an external time parameter by the trinity future, present and past and show its consequences. We develop an alternative to the theory of relativity.

In Section 5 we discuss detailed the fundamental concept of canonical quantization, known as the process of constraining continuous quantities to discrete ones. In particular, an introduction to quantum field theory is presented, including the quantization of the electromagnetic field.

Finally, an appendix is attached containing the theorem of Hurwitz, a very short introduction to symmetry and groups, and some "Keep in minds".

These lecture notes certainly do not meet the standard of a textbook, both because of structure and language. They are only an extended version of my

[25] Feynman Lectures [1963]

[26] Greiner [2005], Omnes [1994], Penrose [2005], Weinberg [2013], Zetteli [2009]

[27] http://yquotes.com/quotes/lev-davidovich-landau/

lectures. It was my intention to question certain interpretations of quantum information theory and physics, and to stimulate thought.

Feedback This text is free to download from the internet

- `http://www.ti3.tuhh.de/jansson/`.

I am deeply grateful for corrections, comments, and suggestions:

- `jansson@tuhh.de`.

Acknowledgements I am grateful to a number of colleagues and students. In particular, I wish to thank Ulrike Schneider for her assistance in preparing this lecture notes, including graphics, tables, much of the latex code, and reading. I wish to thank Kai Torben Ohlhus and David Sills for their critical reading of the manuscript, their feedback, and their suggestions. It is my pleasure to thank Fritz Mayer-Lindenberg. Many years ago I shared the lectures on quantum computation with him. My special thanks goes to Frerich Keil for many valuable and friendly conversations and hints regarding the literature. These notes have been tested especially in my last course 2016. I am grateful to some of the students for comments. In particular, I would like to mention Max Ansorge, Antonio Attimonelli, Annika Evers, Georg Felbinger, Lennart Manthey, Hendrik Meyer zum Felde, Hendrik Preuß, Nicolas Riebesel and Anselm Jonas Scholl.

Hamburg, Germany, August 2017

Christian Jansson

2 Experiments

The aim of this section is to give students an impression of quantum mechanics by describing and looking closely at several experiments. We present quantum mechanics from the point of view of probability theory and stochastic processes. We do this in a very elementary way, and demonstrate the need of a type of probability that does not satisfy the rules of Kolmogorov's *classical probability theory*, but requires squared magnitudes of complex numbers.

An understanding of nature and particularly physics can only be obtained when putting questions in the form of experiments. All of our knowledge about nature can be expressed in terms of a sequence of binary YES or NO questions, namely bits. They lead in a natural way to two-state systems. Two-state systems are the fundamental elements in quantum information theory.

We consider at the beginning the polarization of light which is one example of a two-state system. Other examples are the double-slit experiment and the Mach-Zehnder interferometer. Not surprisingly, it turns out that almost all fundamental principles of quantum mechanics can be derived on the basis of two-state systems.

This chapter is written in a manner such that also pupils of a grammar school, taught complex numbers, can understand the mathematical description of these quantum experiments. Even elementary concepts of matrix algebra are not used. Only the addition and multiplication of numbers is required.

2.1 Historical Remarks

The following remarks serve only as a rough orientation, and can be skipped in a first reading.

The basic frameworks of quantum mechanics, quantum field theory, and quantum information theory are presented in books frequently as a collection of postulates. These postulates and their consequences seem to be non-intuitive and mysterious. One reason for some paradoxes in quantum mechanics may be that in textbooks physics is described from a historical point of view, with less emphasis on reasoning or logic.

Schrödinger, 1926, introduced the concept of a wave function satisfying his linear partial differential equation in order to describe the motion of electrons in atoms. Almost at the same time Heisenberg introduced a seemingly completely different model of time-dependent operators acting on so-called "quantum states", the latter are time-independent. Later, both models were shown to be mathematical equivalent.

Based on Schrödinger's wave equation, in many textbooks a strange property of quantum mechanics, namely the *wave-particle duality*, is often verbalized. This duality says that microscopic particles, like photons, electrons or atoms, do neither behave as a point-particle nor as a wave, but as both. It depends on the experiment whether the microscopic object shows particle-like properties or wave-like properties.

The most frequently used experiment for depicting this wave-particle nature is the famous double-slit experiment, which we shall consider later in detail. If both slits are open a particle shows an interference property as we know from waves. But, if it is clear through which slit the particle goes, the interference disappears and a typical point-like statistic is observed.

Dirac suggested in 1932 a third formalism that avoids wave-particle duality. It was forgotten until 1941 when Feynman elaborated this idea. Now it is known as the *path integral formalism* or *sum over history formalism*. This formalism is described excellently in the Feynman lectures on physics, Volume 3, as well as in Feynman's famous book "QED the strange theory of matter and light", 1985, where he presents our best physical theory, *quantum electrodynamics*. **This theory describes all phenomena of the physical world, except gravitation and radioactive phenomena**[28].

The fundamental idea of Feynman's theory is so simple that its far-reaching consequences are astonishing:

> In this formalism complex numbers, called "probability amplitudes" are assigned to elementary events, paths, or other possibilities. These amplitudes are added for mutually exclusive events and are multiplied for independent events.
>
> The relation between classical probabilities and probability amplitudes is established by Born's rule, 1926: the squared magnitudes of the amplitudes are the probabilities. In fact, this idea might go back to Malus 1810, who used squared magnitudes for describing some optical experiments.

[28] Feynman [1985, page 8]

The addition and multiplication rule in quantum theory is similar to classical probability theory formalized by Kolmogorov in 1933. But in the classical theory we operate not with complex numbers. Instead only non-negative real numbers are used.

Summing up numbers along any events or paths is actually integrating, hence the word "path integral formalism". The path integral can be viewed as a function of the final event. This function satisfies Schrödinger's equation, and thus Schrödinger's and Heisenberg's formalisms can be derived from Feynman's theory. The important advantage of Feynman's approach is its nice interpretation as a linear stochastic process. It is rather easily understandable with a rudimentary knowledge of classical probability theory, and seems not to be mysterious. However, for many practical problems the Heisenberg and Schrödinger formalisms are more appropriate from a computational point of view.

Keep in mind: Feynman's path integral theory implies Schrödinger's as well as Heisenberg's formalism. It forms the basis of quantum electrodynamics, our best physical theory.

In this chapter, we try to get a feeling for the world of quantum mechanics as well as for the path integral formalism by looking closely at several astonishing experiments. A key point is the visualization of numbers as little arrows. We shall see that many physical features can be understood quite simply when doing this.

2.2 Numbers

Physical theories deal with systems of numbers. In probability theory non-negative numbers are used, in classical mechanics real numbers are used, and quantum mechanics works with complex numbers.

In particular, it can be shown that classical mechanics and quantum mechanics can be embedded in the same mathematical framework when working with complex numbers; the real dynamical variables of classical mechanics are just the quantum mechanical average values of Hermitian operators[29].

Since all physical models are formulated basically with the set of complex numbers or specific subsets, it raises many questions. Are there other sets of numbers, such as quaternions or octonions, that might describe physics and nature in a much better way than the numbers mentioned above? What do we call a number? Why is quantum mechanics, from most physicists viewed as the fundamental physical theory, based on the field of complex numbers? Would other number systems perhaps give new insights? Are there any properties that are common to all number systems? Of course, such basic questions are not easy to answer. In other words, if we go very deeply into these questions they might become unanswerable. However, some clarifications can be given.

The idea of numbers starts with simple counting, namely with integers. Integers can be visualized on a line as multiples of a unit arrow, which we denote by 1. Integers can be added and multiplied. With regard to arrows, they are added by attaching the head of one arrow to the tail of another, yielding a final arrow from the tail of the first to the head of the last one. The multiplication of integers is defined as the addition in succession. The advantage of visualizing numbers as arrows is an action-oriented view of numbers, and gives abstract symbols a pictorial form, well-suited for engineering education. Both, addition and multiplication, have the fundamental property that the result of the operation does not depend on the order of the integers. These operations are commutative as well as associative.

Already the Greeks generalized the concept of integers to rational numbers. They represented numbers as transformations of a unit arrow 1. The transformations are expansion and shrinkage, leading to the positive rational numbers. The well-known addition and multiplication of rational numbers can simply be visualized in terms of arrows. Addition is the same operation as for integers. Multiplication, in terms of arrows is as follows: we multiply 1/4 with 1/3 by shrinking the unit arrow 1 to 1/4, and then shrinking the resulting arrow by 1/3. As for integers the rational operations are commutative. The positive rational numbers are extended with the invention of zero and negative numbers, leading to a representation on the axis of real numbers. The negative numbers correspond to arrows that are rotated by the angle π. In particular, the unit arrow -1 is obtained by rotating 1. The well-known extended operations for rational numbers can be simply realized in terms of arrows as above, and they are commutative and associative. The concept of rational numbers was improved by inventing real numbers, that is, numbers or arrows on the real axis that cannot be represented by rational numbers. The real addition and real

[29] Strocchi [1966]

multiplication remain commutative and associative. All real numbers, say x and y, positive or negative, have a *length*, also called *magnitude* that satisfies the equation

$$|xy|^2 = |x|^2|y|^2. \tag{1}$$

In science it is customary to speak of a *number system*, if it shares the following basic indispensable properties:

> *Numbers* have a magnitude satisfying (1). They can be **added** and **multiplied**, where both operations are **associative** and **commutative**. Moreover, they have a **neutral element** 0 for addition and a **neutral element** 1 for multiplication.

The sets of numbers mentioned above satisfy these properties. Moreover, the numbers as well as their operations can be represented as arrows and geometrical operations with these arrows, respectively. The sets of natural numbers $\mathbb{N}$, integers $\mathbb{Z}$, rational numbers $\mathbb{Q}$, real non-negative numbers $\mathbb{R}_+$, real numbers $\mathbb{R}$, and complex numbers $\mathbb{C}$ are number systems. The set of purely imaginary numbers is not a number system, since the product of two imaginary numbers is a real number.

It seems that the process of constructing new systems of numbers, starting with the natural numbers, can be continued ad infinitum. So let's continue. What can we do with our arrows, in order to extend the field of real numbers. For obtaining negative numbers, we have rotated the positive numbers by the angle π. This suggests to rotate with an arbitrary angle in a plane, leading to an extension of the real line to a set of arrows living in a plane. If we rotate the unit arrow 1 with angle $\pi/2$, then we get an arrow which we call i. Since rotating with π is the same as rotating two times with $\pi/2$, we obtain $i^2 = -1$, provided we define the multiplication appropriately in terms of a rotation. In fact, this idea leads to the field of complex numbers.

Because of its importance, we give here a very short introduction to complex numbers. There are two ways to think about complex numbers. Firstly, the customary algebraic point of view is to describe *complex numbers* as two-dimensional vectors

$$z = (x, y) = x \cdot 1 + y \cdot i, \tag{2}$$

where both, x and y, are real numbers. Usually, the unit 1 and the dots are omitted. These vectors represent our arrows in the plane. The first component x is called the *real part*, and the second component y is called the *imaginary part* of z. Two complex numbers can be added

$$z_1 + z_2 = (x_1 + x_2, y_1 + y_2), \tag{3}$$

and multiplied

$$z_1 z_2 = (x_1 x_2 - y_1 y_2, x_1 y_2 + y_1 x_2). \tag{4}$$

Obviously, the addition is related to the addition of arrows, where we attach the head of one arrow to the tail of the other, see Figure 5. From formula (4)

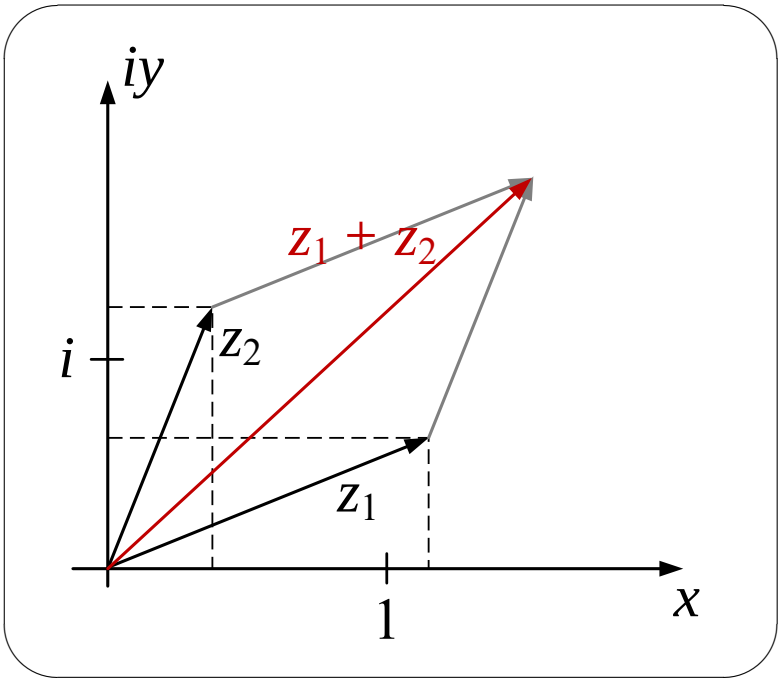


Figure 5: Addition of complex numbers $z_1 + z_2 = (x_1 + x_2) + i(y_1 + y_2)$ is commutative.

it is not clear that the multiplication corresponds to a rotation. This point will be explained below.

Both operations define a commutative field, where natural numbers, fractions and real numbers can be viewed as subsets of the field of complex numbers. The difference to the previous systems of numbers is the existence of complex numbers, living in a plane. Take for instance $i = (0, 1)$. Then the multiplication rule (4) yields $i^2 = (-1, 0)$ which can be identified with the negative number -1. In other words, squares of complex numbers can be negative.

The *complex conjugate* of z is denoted by

$$z^* = x - iy, \tag{5}$$

and is obtained by flipping z over the horizontally axis, see Figure 6.

It is important to notice that the complex conjugate star operator is defined for any number contained in the field of complex numbers. Hence, the complex conjugate is well-defined for integers, positive numbers, rational numbers or real numbers. In these cases the star operator is just the identity.

The squared *magnitude* of a complex number is the product with its complex conjugate:

$$|z|^2 = zz^* = x^2 + y^2. \tag{6}$$

Any complex number z can be written, using the famous *Euler formula*

$$\cos\phi + i\sin\phi = e^{i\phi}, \tag{7}$$

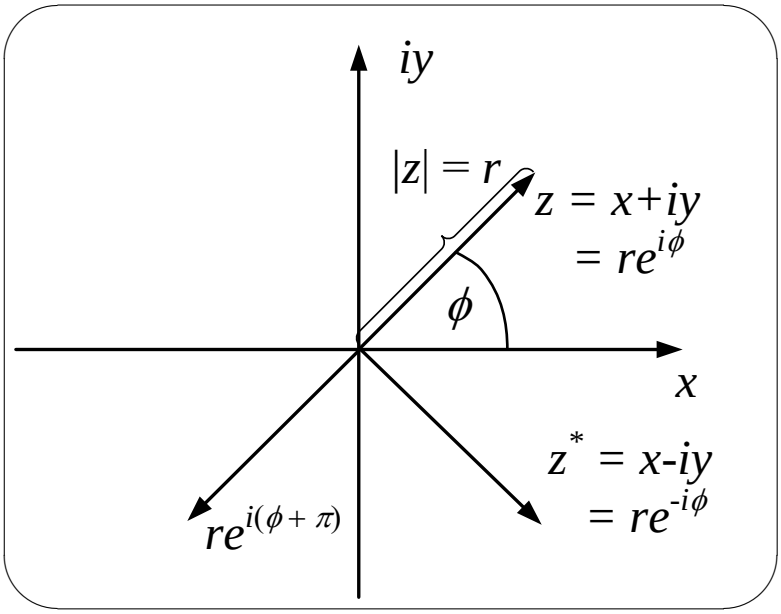


Figure 6: Complex and conjugate complex numbers z and z^*, respectively.

in its *polar form*

$$z = |z|e^{i\phi}, \tag{8}$$

where both, the magnitude $|z|$ and the phase angle ϕ are real non-negative numbers. In other words, any complex number is an arrow with a non-negative magnitude and a generalized sign, called the *phase*. The phase can be viewed as a rotational operator.

Given two complex numbers $z_1 = |z_1|e^{i\phi_1}$ and $z_2 = |z_2|e^{i\phi_2}$, we easily obtain with some trigonometric identities the multiplication rule

$$z_1 z_2 = |z_1|\,|z_2|e^{i(\phi_1+\phi_2)}. \tag{9}$$

Hence, the magnitudes of the complex numbers are multiplied, and the angles are added to yield the polar form of the product. In particular, the multiplication corresponds to a stretching operator and a rotation, see Figure 7.

Complex numbers are an extension of the field of real numbers, since they contain all real numbers and there is no real number which is the square root of −1. Moreover, rewriting a complex number as a pair of real numbers, namely the real and the imaginary part, the field of complex numbers can be viewed as a real vector space, that is, a set of vectors satisfying the two fundamental rules: we can add them, and we can multiply them with a real number, the latter is called *scalar multiplication*. But this does not mean that the field of complex numbers is algebraically isomorphic to an appropriate two-dimensional real space.[30] The fundamental complex operations make this set

[30]For instance, the *special unitary group* $SU(2)$, consisting of the two-dimensional unitary matrices with determinant equal to one, is not homeomorphic to the group $SO(4)$, consisting of the real orthogonal matrices with determinant equal to one, as it might be expected if complex numbers are just two-dimensional real vectors. It turns out that $SU(2)$ is homeomorphic to $SO(3)$.

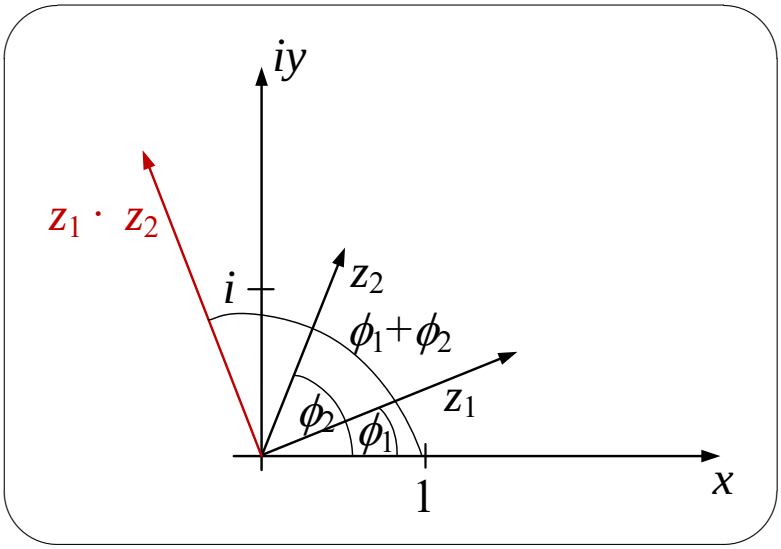


Figure 7: Multiplication of complex numbers $z_1 \cdot z_2 = r_1 r_2 e^{i(\phi_1+\phi_2)}$.

of numbers beautiful and unique, such that the property of viewing complex numbers as real two-dimensional vectors is only one aspect. There are various other deep properties of the field of complex numbers.

At a first glance, it seems to be possible to extend the field of complex numbers. For instance, we could work with arrows in a higher dimensional space. In fact this is possible, but at a high price. New systems like quaternions, octonions, or dual numbers, do not share our fundamental properties of numbers. Quaternions and octonions form only non-commutative fields. It is questionable whether it is right to speak of numbers, if they can be mathematically constructed, but loose the indispensable properties related to counting and arrows, and, in particular, violating commutativity. We would be very surprised, if on two meadows we have five cows, respectively, hence together $2 * 5 = 10$ cows. But if we look at five pairs of cows, all cows being on different meadows, we would obtain a number $5 * 2$ being different from 10. Hence, commutativity is likely to be necessary when thinking of numbers. This distinguishes numbers from operators or matrices that do not commute, in general.

It is a well-known fact that the field of complex numbers is the **largest commutative field** possessing the previous properties. This follows from a theorem of Hurwitz, see the appendix, and can be viewed as a basic reason that quantum mechanics, the most fundamental physical theory, uses complex numbers and not any other systems of numbers. Another point of view is that other number systems, like quaternions or octonions, are unlikely to appear as elementary numbers in physics, since they can be represented as matrices composed of complex entries.

Keep in mind: The field of complex numbers is the **largest commutative field** possessing the indispensable properties of numbers (like commutativity). This could be viewed as a basic reason that quantum mechanics, the most fundamental physical theory, requires complex numbers.

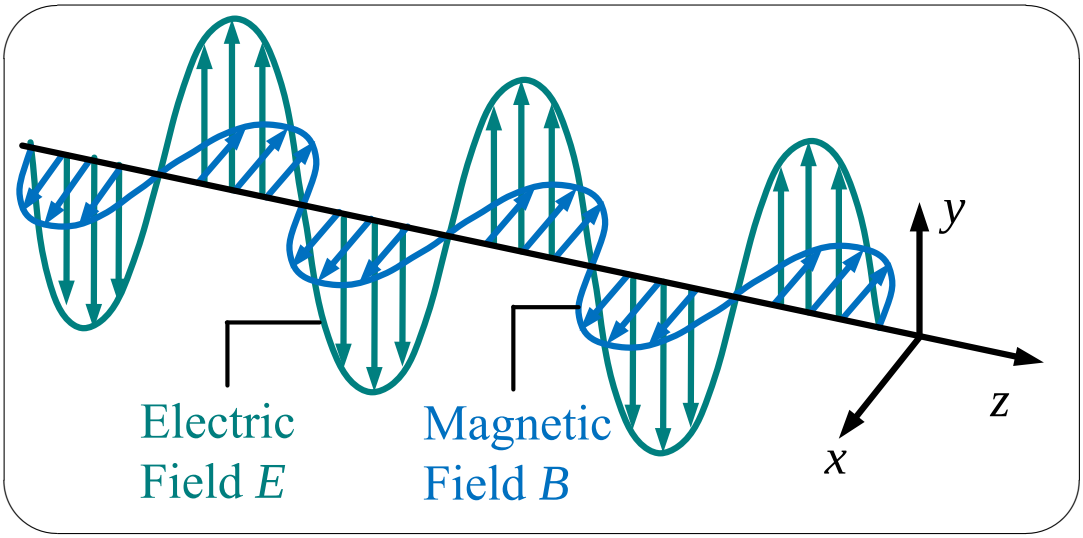


Figure 8: Propagation of an electromagnetic plane wave in direction z with horizontal x-axis and vertical y-axis.

2.3 Polarization of Light

The classical model of light is derived from Maxwell's equations as transverse electromagnetic waves. Mathematically, **Maxwell's linear equations** of electromagnetism imply **plane wave solutions**

$$E(r,t) = E_0 e^{i(k^T r - \omega t)}, \quad B(r,t) = \frac{1}{c}(k \times E), \tag{10}$$

where $r = (x,y,z)^T$ is the position vector, $k = (k_x, k_y, k_z)^T$ is the *wave vector* specifying the direction of propagation, ω is the *angular frequency*, and $E, E_0, B \in \mathbb{C}^3$. The electric field E is perpendicular to the magnetic field B, and both are perpendicular to the propagation direction k. Maxwell's equations are linear per excellence and allow the *superposition* of solutions.

What does matter in our polarization experiments is only the classical idea that light is an *electromagnetic wave*, where the electric field E and the magnetic field B are both oscillating orthogonally in a plane which is perpendicular to the direction of motion, see Figure 8. We are free to choose the coordinates of the direction of propagation as the z-axis together with the orthogonal xy-plane. We will choose the x-direction to be horizontal and the y-direction to be vertical.

By convention the notion "polarization" always refers to the polarization of the electric field. Light has *linear polarization* if the electric field E oscillates in one direction in the xy-plane. This plane can be stretched by any two orthogonal directions, usually called the horizontal x-axis and the vertical y-axis. Linearly polarized light can be viewed as the sum or *superposition* of a horizontal and a vertical component. In other words, light can be descibed by two base states, namely the *horizontal polarization* component and the *vertical polarization* component.

If the electric field E rotates in the plane perpendicular to the direction of motion the light has *circular polarization*. In the latter case the rotation may

be either clockwise or counter-clockwise. This property is called *chirality*.

Light carries momentum $p = \hbar k = 2\pi\hbar/\lambda$ and energy $\hbar\omega$ expressed by *Planck's constant* $\hbar$, wavelength λ, and angular frequency ω, respectively. Experiments demonstrate that light consists of a large number of small energy packets $\hbar\omega$, the *photons*. The number of photons determines the *intensity* of a light beam. These photons can be viewed naively as small arrows oscillating in the plane orthogonal to the direction of propagation. Beside the concept of polarization, photons have further properties: they travel at the speed of light, have no charge, have no mass, and have spin one. These properties, which we fortunately don't require, are derived in particle physics.

An *optical element* is a material machine that interacts with photons and changes the state of polarization. For instance, polarizers, rotators, phase shifters, or depolarizers are special optical elements. Most light sources emit unpolarized light, that is, a large number of photons, whose directions of oscillation are completely random. Light can be polarized when it passes through an optical element.

An important optical element is a *polarizer*. This element is designed to generate specific polarized light states, such as linearly or circularly polarized states. *Absorption polarizers* are characterized by the fact that they transmit light which is polarized along one axis, the *transmission axis*. The light is completely absorbed along the perpendicular axis. Absorption is an interaction between the photon and the polarizer that increases the energy of the polarizer. For instance, a *polaroid filter* consists of molecular chains, the polymers, that are aligned along an axis. The absorption takes place along this axis.

The *wire grid polarizer* has similar properties as the polaroid filter. Incident light, polarized parallel to the wires, causes the electrons to move inside the wires, see Figure 9. This results in a loss of some energy, and the remainder is reflected. Photons with polarization perpendicular to the wires do not interact with the electrons, and can travel through the grid.

Birefringent plates, like *calcit crystals*, provide a different kind of optical apparatus. They separate photons into two beams with perpendicular polarization. More precisely, a calcite transmits light polarized parallel to the *optical axis* along one path and transmits the remaining light polarized perpendicular to the optical axis along another path, see Figure 10. Hence, unlike polaroid filters, which absorb one of the components, calcits allow all photons to pass through, but along two different paths corresponding to two polarization states. Hence the incident light intensity is equal to the sum of the transmitted intensities. Birefringent plates are very flexibly applicable in experiments. They can be converted into optical elements similar to polaroids by blocking one of its exit beams.

The polarization of a photon is an example of a *two-state quantum system*. In quantum information theory this is called a *qubit*, a quantum bit. There are a lot of other two-state systems, for instance the spin of an electron. The spin of an electron can roughly be viewed as a small arrow as well, and a Stern-Gerlach device is comparable with a birefringent plate. Not surprisingly, the mathematical models of spin and polarization are the same. But for the purpose of demonstration during the lectures, light is much more appropriate. In

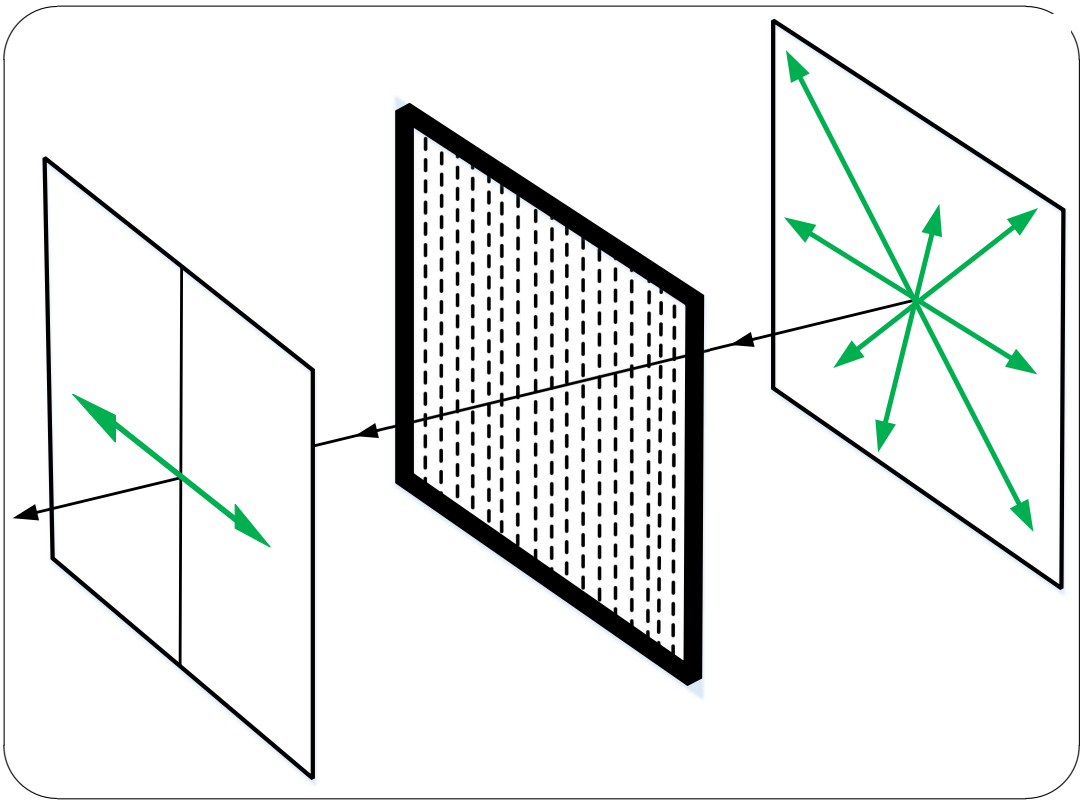


Figure 9: Unpolarized light interacting with a wire grid polarizer. The transmitted photons are polarized orthogonal to the wires.

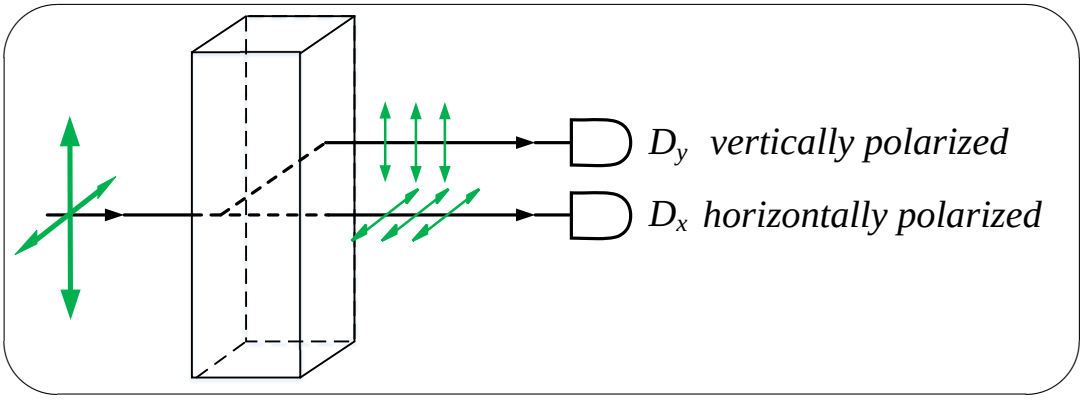


Figure 10: The birefringent plate splits incident light with respect to their optical axis into vertically polarized light and horizontally polarized light. If only one photon is in the experiment, either detector D_x or D_y clicks.

Figure 11: My experimental set-up costs about 90 Euros.

particular, because optical elements such as polaroid filters and calcite crystals are not expensive, see Figure 11.

In a similar manner the hydrogen molecule H_2^+ can be described as a two-state quantum system. It consists of two protons and one electron, and the Coulomb force implies that the electron can be close to the first proton or close to the second proton. From the mathematical point of view, all two-state systems are described within a two-dimensional Hilbert space.

Long before the term qubit was introduced, von Weizsäcker[31] described qubits as the smallest building blocks of physics. He denoted these blocks by the German word "ur". He proposed an *ur theory*, which is a quantum theory of binary alternatives, that can also be viewed as a theory of the real three-dimensional space.

The qubit is a fundamental building block, the basic unit in quantum information theory. Perhaps any system, small or large, can be described by combining these blocks via the so-called *tensor product*, which we consider later. On a quantum computer, a qubit plays the same role as a bit on a classical computer.

[31] von Weizsäcker [1988]

2.4 The Law of Malus

The interaction of photons with optical elements can change the polarization and the intensity of a beam of light, depending on the arrangement of the elements.

Let us start with a simple experiment using a polaroid filter. If we send a beam of unpolarized light through this polarizer, we observe that one half of the intensity of the beam is absorbed by the polarizer, and the second half is transmitted. Then we place another polarizer behind the first one, we rotate the second one and observe that the intensity after the second filter varies. We rotate until the intensity of the light beam is maximized. It can be seen that for an ideal polarizer in this position the intensity after the first polarizing filter is almost the same as the intensity after the second one. Thus both filters have the same transmission axis. In reality, however, there is no ideal polarizer and certain disturbances imply a slightly smaller intensity after the second polarizer. In the following we assume always ideal experiments.

If we rotate the transmission axis of the second filter through an angle α in small steps carefully, the light intensity decreases until the light vanishes completely at the angle of rotation $\alpha = \pi/2$, see Figure 12 and Figure 13. Hence, the intensity is zero if we send a beam of light through two filters with orthogonal transmission axes. A classical visualization could be to think of a polarizer in terms of a garden fence with crossbars parallel to the transmission axis.

Finally, we put a third filter between both polarizers. Since polarizers in general absorb light and decrease the intensity, we would expect that as before no light can be seen. But surprisingly and contrary to the previous visualization of a garden fence, this is not the case, see Figure 14. Which mathematical model is able to describe these observed unexpected intensities?

The experimental observations with different angles suggest a law that was already given in 1810 by Malus. The *law of Malus* states that the intensity I_1 of a beam of polarized light that has passed both filters is given by

$$I_1 = I_0 \cos^2 \alpha, \tag{11}$$

where I_0 is the intensity after the first polarizer, and α is the angle between the transmission axes of both polarizers, see Figure 12. Consequently, the intensity

$$J_1 = I_0 \sin^2 \alpha, \tag{12}$$

is absorbed.

Obviously, this law describes exactly the two experiments in Figure 12 and 13. For the third experiment in Figure 14 with three polarizers, we apply this law twice in succession and obtain

$$I_2 = I_1 \cos^2(\alpha - \beta) = (I_0 \cos^2 \alpha) \cos^2(\alpha - \beta). \tag{13}$$

For $\alpha = \pi/4$ and $\beta = \pi/2$, as displayed in Figure 14, we obtain $I_2 = 1/4 I_0$, which is consistent with the experimental results. Our garden fence imagination is wrong, but the law of Malus provides a correct description.

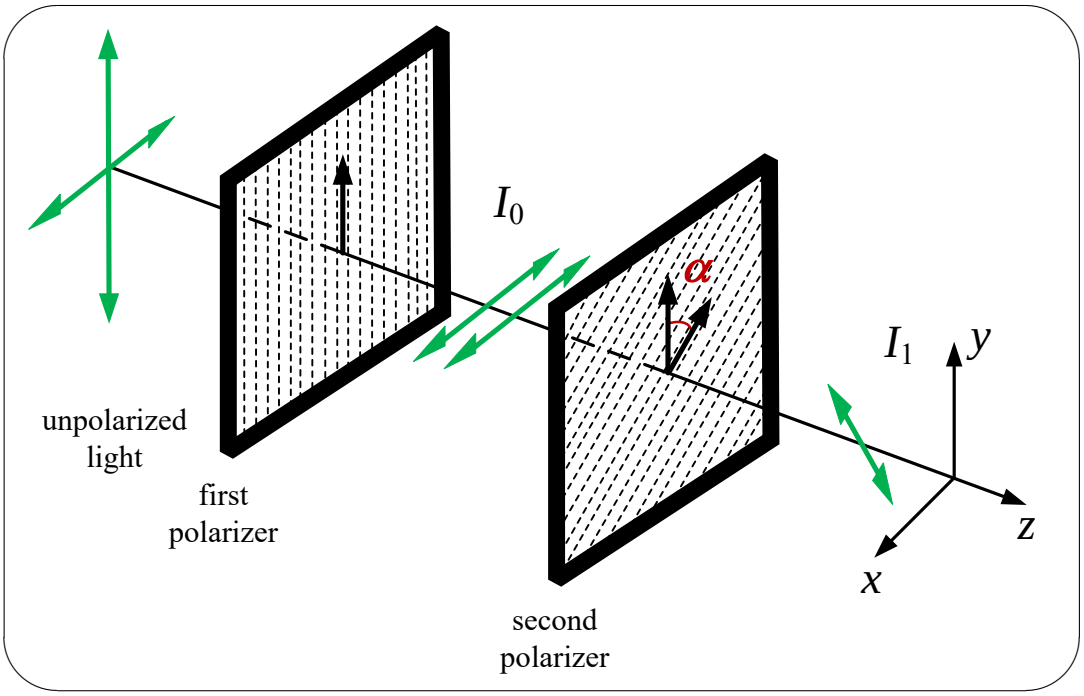


Figure 12: The law of Malus states that the intensity I_1 of light that has passed both polarizers is equal to $I_1 = I_0 \cos^2 \alpha$, where I_0 is the intensity after the first polarizer. If only one photon is in the experiment, then $\cos^2 \alpha$ is the probability that a photon passes the second polarizer, provided it has passed the first polarizer.

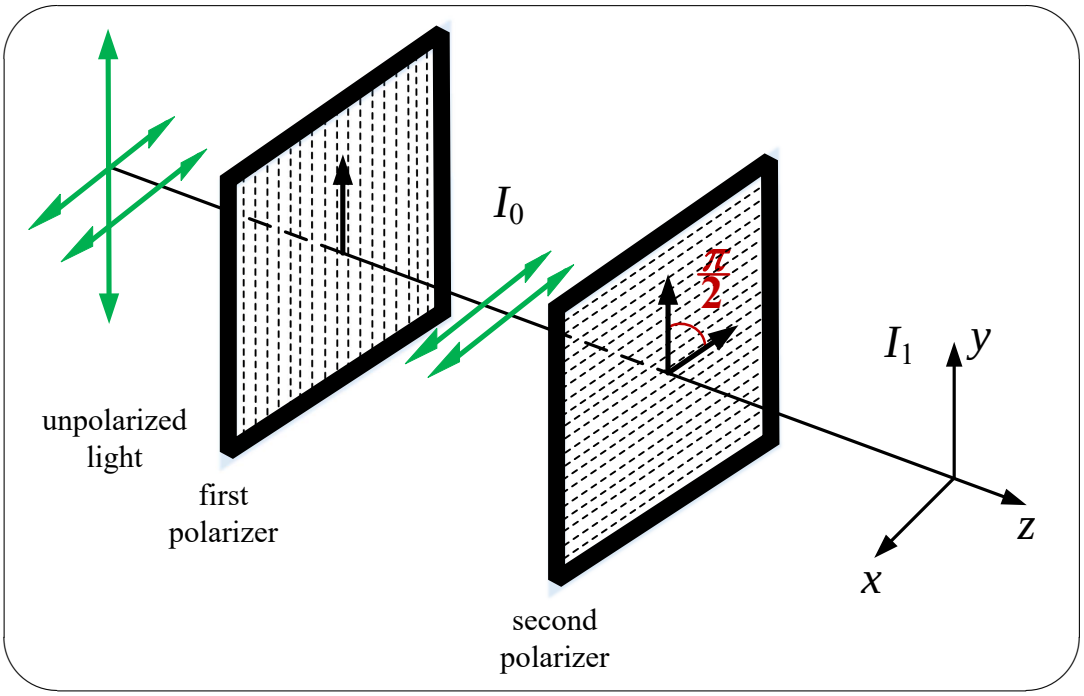


Figure 13: For angle $\alpha = \frac{\pi}{2}$ between both transmission axes the intensity after the second polarizer is zero.

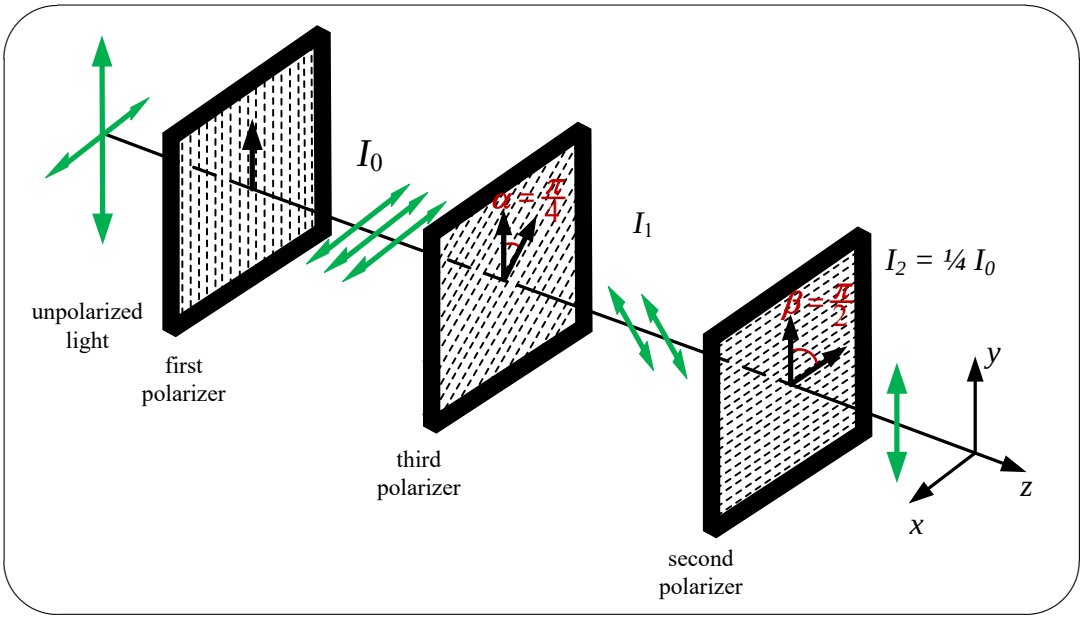


Figure 14: If we put a third polarizer between both polarizers with transmission axis rotated about $\alpha = \frac{\pi}{4}$, then the intensity is $I_2 = \frac{1}{4} I_0$.

The average value $\cos^2 \alpha$ for varying angle α is 1/2. Therefore, the transmission coefficient of a beam of unpolarized light, containing a uniform mixture of linear polarizations at all possible angles, should be 1/2. Almost all natural sources of light emit unpolarized light.

In actual experiments the light intensity can be reduced such that only one single photon is in the experiment. Then the intensity of a beam of light is given by the number of photons yielding the following interpretation:

- For a photon the squared quantity $\cos^2 \alpha$ can only be interpreted as a **probability**, namely the probability that it will be transmitted. The probability that it will be absorbed is $\sin^2 \alpha$.

Notice that $\cos \alpha$ has positive and negative values. Hence, the law of Malus requires squared magnitudes.

The law of Malus formulated in 1810 for intensities becomes Born's rule for single photons. This rule was formulated in 1926, and was awarded 1954 by the Nobel Prize in Physics. By the way, the rules of classical probability were formulated much later by Kolmogorov in 1933.

In terms of classical probability theory equation (13) has the following interpretation: the probability that a photon passes all three filters, when it has passed the first filter, is $\cos^2 \alpha \cdot \cos^2(\alpha - \beta)$ which is the classical product rule for independent events. This is in agreement with the experiments, at least for ideal polarizers.

Keep in mind: Simple experiments with light, already performed by Malus in 1810, demonstrate: when accepting that light is built up of photons, the interaction of photons with optical elements is a stochastic process, and the probabilities of the random outcomes are squared magnitudes of numbers.

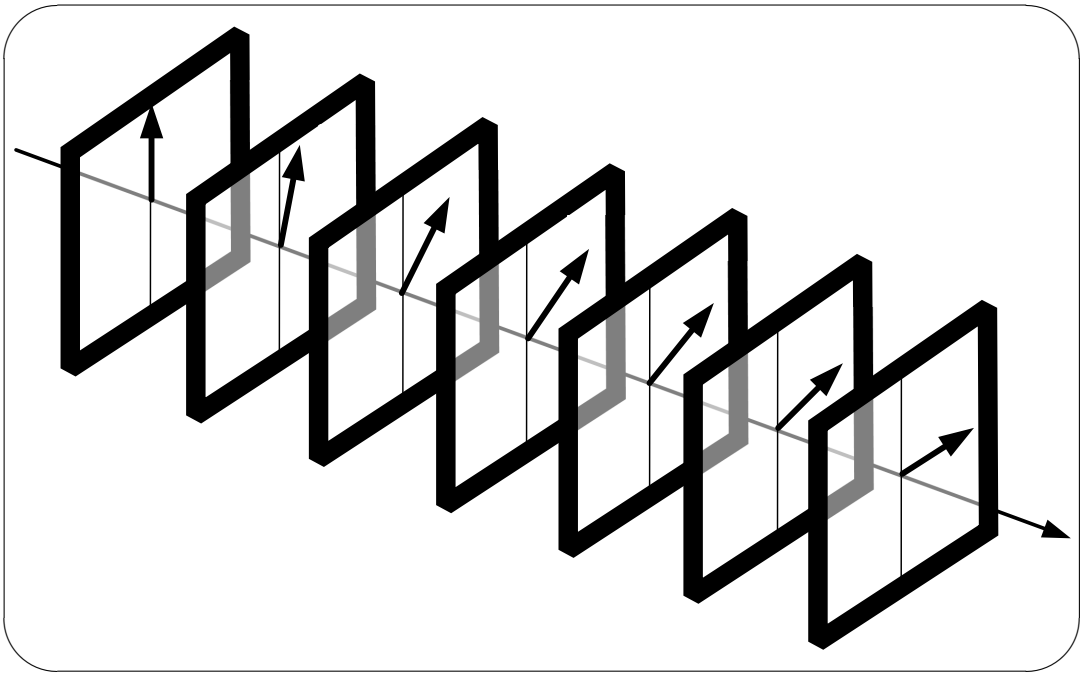


Figure 15: An almost lossless rotation from vertical polarization to horizontal polarization with polaroid filters.

The law of Malus has the fascinating property that it is perfectly symmetric. What is the symmetry of a law? *Symmetry* of geometrical objects is usual defined as a property of an object that does not change if something changes. Spheres do not change if they are rotated by an arbitrary angle. A square remains invariant when rotated by a right angle. Crystals, like snowflakes, have a beautiful symmetry. For physical laws the symmetry is rather similar. Equations are written down using certain letters. In our case of polarization, we use x and y to describe the horizontal and vertical direction, and we use α and β to describe angles related to transmission axes. But we are free to choose any other orthogonal directions using letters x', y' and angles α', β'. An equation describing a physical law is symmetric if it looks the same, independent of whether it is described by the primed or unprimed coordinates. Obviously, the law of Malus is the same in each coordinate system since the identity

$$\cos^2(\beta - \alpha) = \cos^2(\beta' - \alpha') \tag{14}$$

yields exactly the same probability. The law requires only the difference between two angles, rather than some absolute angles. All fundamental laws of physics are symmetric in this sense; they look the same in every reference frame.

Since in the law of Malus only the difference of the angles between both transmission axes is relevant, it follows that the polarization of the photon before both polarizers in series doesn't matter. The photon has completely forgotten the history. Moreover, we have seen that when placing a polarizer appropriately between two others, as displayed in Figure 14, the intensity

increases. We can easily generalize this experiment. Assume, we put a large number of polaroids, say N, with transmission axes differing only by a small angle, between two polaroids with perpendicular transmission axes, see Figure 15. Then the law of Malus (13) implies that the probability for a vertically polarized photon to pass all polaroids is $(\cos(\frac{\pi/2}{N}))^{2N}$. Then the photon is horizontally polarized. For $N = 18$ we roughly get the probability 0.9337. Thus this experimental set-up causes an almost lossless rotation from vertical to horizontal polarization. Of course, an analogous set-up for an almost lossless rotation from horizontal to vertical polarization is possible as well.

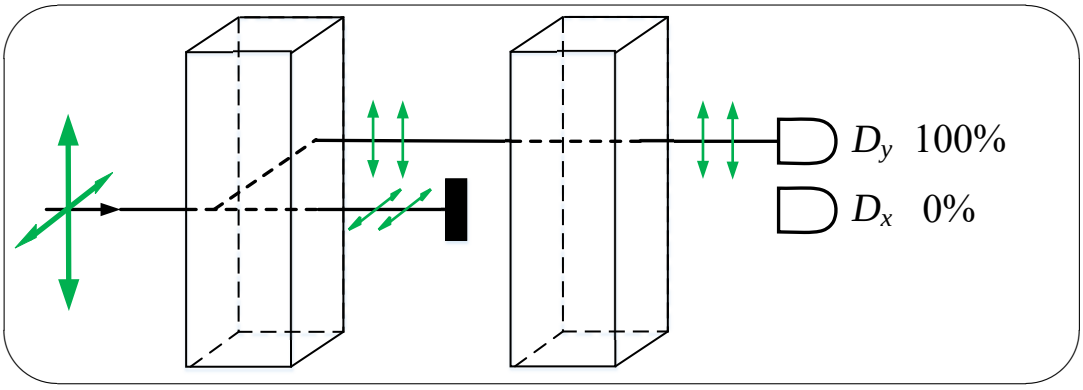


Figure 16: For two birefringent plates, where the horizontal light after the first plate is blocked, only the vertical detector D_y clicks. If we block the vertical light after the first plate, then only the horizontal detector D_x clicks.

2.5 Birefringent Plates

In the following we discuss some experiments with birefringent plates.

At first we introduce a useful notation which allows us to describe experiments in a neat form. It is called *Dirac's bracket notation*. For instance, if a photon, has passed a polarizer with transmission axis α, it is linearly polarized at this angle, and we write $|\alpha\rangle$ for this state. Moreover, we assign to any transition from one initial state $|\alpha\rangle$ to another state $|\beta\rangle$ a complex number (arrow)

$$\langle\beta|\alpha\rangle \in \mathbb{C}. \tag{15}$$

This complex number, written as a bracket, is called a *probability amplitude*. Notice that the initial state $|\alpha\rangle$ is on the right hand side of the bracket, and the final state $|\beta\rangle$ is on the left hand side. This chronological order is usual in quantum physics.

The probability of this transition is given by *Born's rule*

$$Prob(\langle\beta|\alpha\rangle) = |\langle\beta|\alpha\rangle|^2. \tag{16}$$

as the squared magnitude of the amplitude, and is called the *transition probability*. This is a generalization of the law of Malus who used only real amplitudes that are sufficient for linear polarization.

Now we discuss the experiments displayed in Figures 16, 17, 18, and 19. The experiments in Figures 16, 17 and 18 follow immediately from the law of Malus. Details are left as an exercise. The experiment displayed in Figure 19 is the most difficult example to explain. In this experimental set-up, both birefringent plates recombine both beams after the first polarizer. Therefore, the law of Malus or Born's rule yields the transition probability

$$|\langle\beta|\alpha\rangle|^2 = \cos^2(\beta - \alpha) \tag{17}$$

after the second polarizer.

The usual way to obtain this result is to calculate the probabilities for a photon travelling through the apparatus by using the classical *addition rule* for

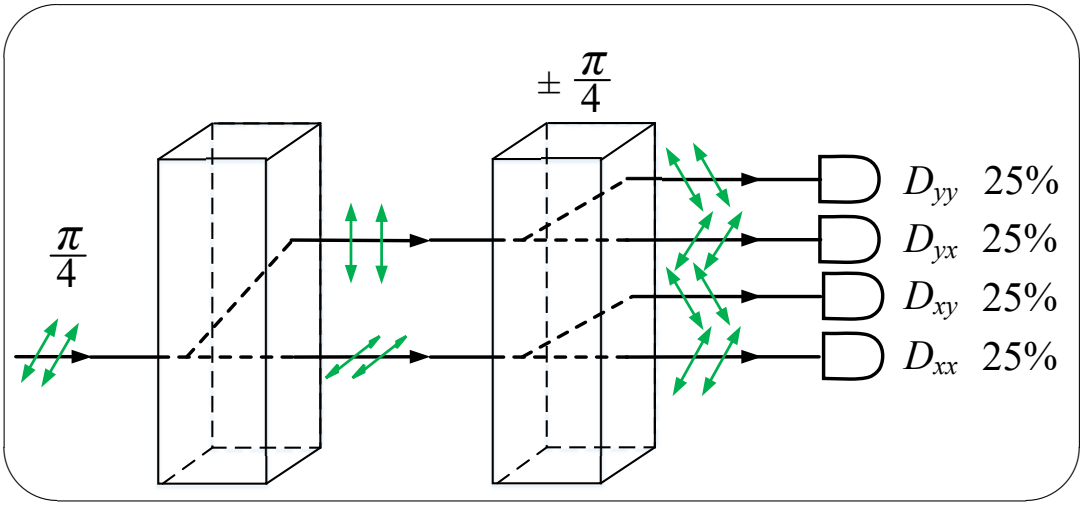


Figure 17: A beam of photons is polarized in state $|+\frac{\pi}{4}\rangle$ by an appropriately oriented polaroid filter. The photons pass horizontally or vertically the first birefringent plate, then they pass through the second plate with an optical axis rotated at an angle $+\frac{\pi}{4}$ or $-\frac{\pi}{4}$. Finally they are detected. It follows that the photons have forgotten their original polarization state $|+\frac{\pi}{4}\rangle$.

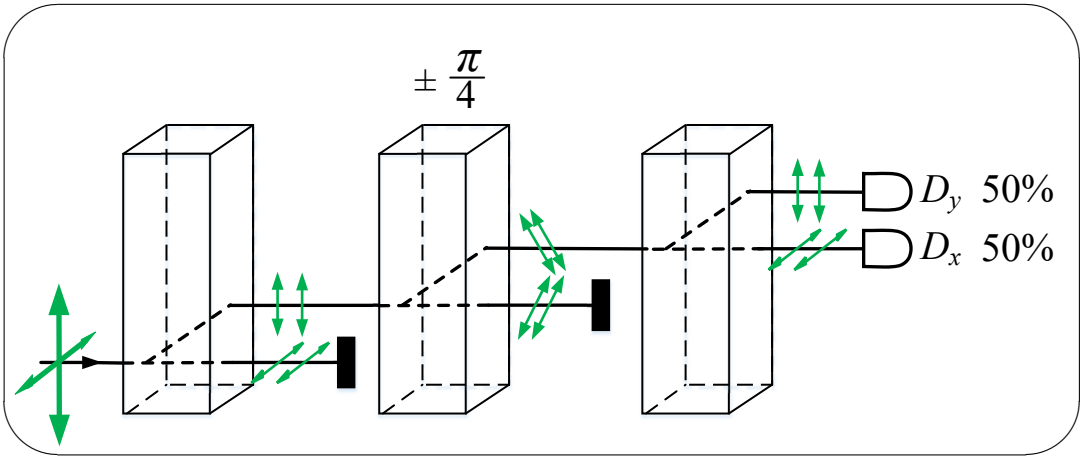


Figure 18: In this experimental set-up the horizontally polarized photons are blocked after the first two birefringent plates. Finally, after the third plate their polarization is measured. Obviously they have lost their original polarization.

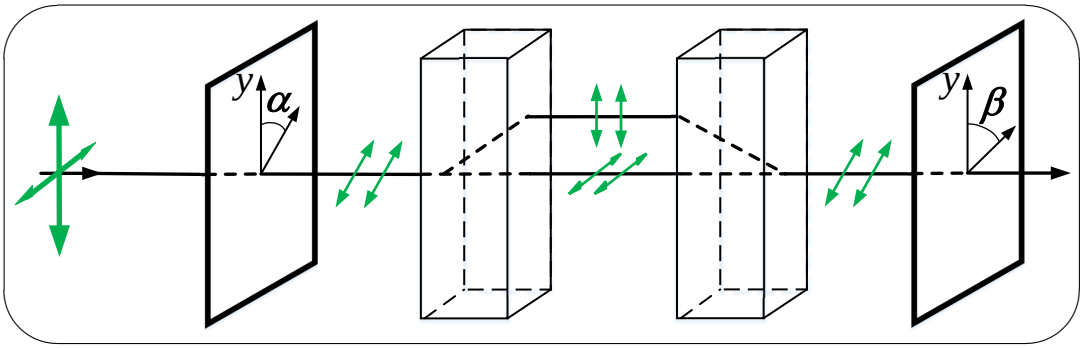


Figure 19: The first polarizer generates photons polarized at an angle α. The first birefringent plate splits into two beams of horizontally x-polarized and vertically y-polarized photons. These are recombined in a second birefringent plate which has an optical axis opposite to the first plate. According to the law of Malus the transition probability after the second polaroid is $\cos^2(\beta-\alpha)$.

mutually exclusive events and the *multiplication rule* for independent events. If we do this, we see that after the first polarizer the photon is linearly polarized at an angle α. The first calcite generates a horizontally and a vertically polarized beam. The equations (11) and (12) imply that the photon is in one of the beams with probabilities

$$\cos^2\alpha \quad \text{or} \quad \sin^2\alpha, \tag{18}$$

respectively. If the photon is vertically polarized it passes the second birefringent plate with probability one, and the second polarizer with transition probability

$$|\langle\beta|0\rangle|^2 = \cos^2\beta, \tag{19}$$

since the optical axis of the plate and the transmission axis of the second polaroid differ by an angle β. Similarly, we obtain the transition probability

$$\left|\left\langle\beta\left|\frac{\pi}{2}\right\rangle\right|^2 = \sin^2\beta, \tag{20}$$

if the photon is vertically polarized. Due to the split into a horizontally and a vertically polarized beam, we have two mutually exclusive paths

$$|\alpha\rangle \to |0\rangle \to |\beta\rangle \quad \text{and} \quad |\alpha\rangle \to \left|\frac{\pi}{2}\right\rangle \to |\beta\rangle. \tag{21}$$

The sub-paths are independent. Hence, the probabilities of the sub-paths are multiplied, yielding for each path the probability

$$\cos^2\alpha\cdot\cos^2\beta \quad \text{and} \quad \sin^2\alpha\cdot\sin^2\beta, \tag{22}$$

respectively. With the addition rule we get the transition probability after the second polarizer

$$\text{Prob}(\langle\beta,\alpha\rangle) = \cos^2\alpha\cos^2\beta + \sin^2\alpha\sin^2\beta, \tag{23}$$

which obviously is in disagreement with $\cos^2(\beta-\alpha)$ and also with experimental statistics. Classical probability theory does not work.

In order to obtain the correct result, we change the rules for calculating the probabilities by using amplitudes. Instead of adding the probabilities for mutually exclusive events, we add the non-squared amplitudes for these events and multiply the probabilities for independent events. We apply both rules to the amplitudes and then we square the magnitude of the final result. These are *Feynman's rules* for quantum phenomena. Then we get

$$\begin{aligned} \langle\beta|\alpha\rangle &= \langle\beta|0\rangle\langle 0|\alpha\rangle + \left\langle\beta\left|\tfrac{\pi}{2}\right.\right\rangle\left\langle\tfrac{\pi}{2}\right|\alpha\rangle \\ &= \cos\beta\cos\alpha + \sin\beta\sin\alpha \\ &= \cos(\beta-\alpha), \end{aligned} \tag{24}$$

yielding the desired result

$$|\langle\beta|\alpha\rangle|^2 = \cos^2(\beta-\alpha). \tag{25}$$

This prediction is in agreement with experiments. In particular, the probability amplitudes satisfy the *superposition principle*: the amplitudes for both paths are added.

Notice that all previous experiments did only require positive and negative numbers in terms of sine and cosine, but not complex numbers. It turns out, however, that complex numbers are necessary when we discuss circularly polarized light.

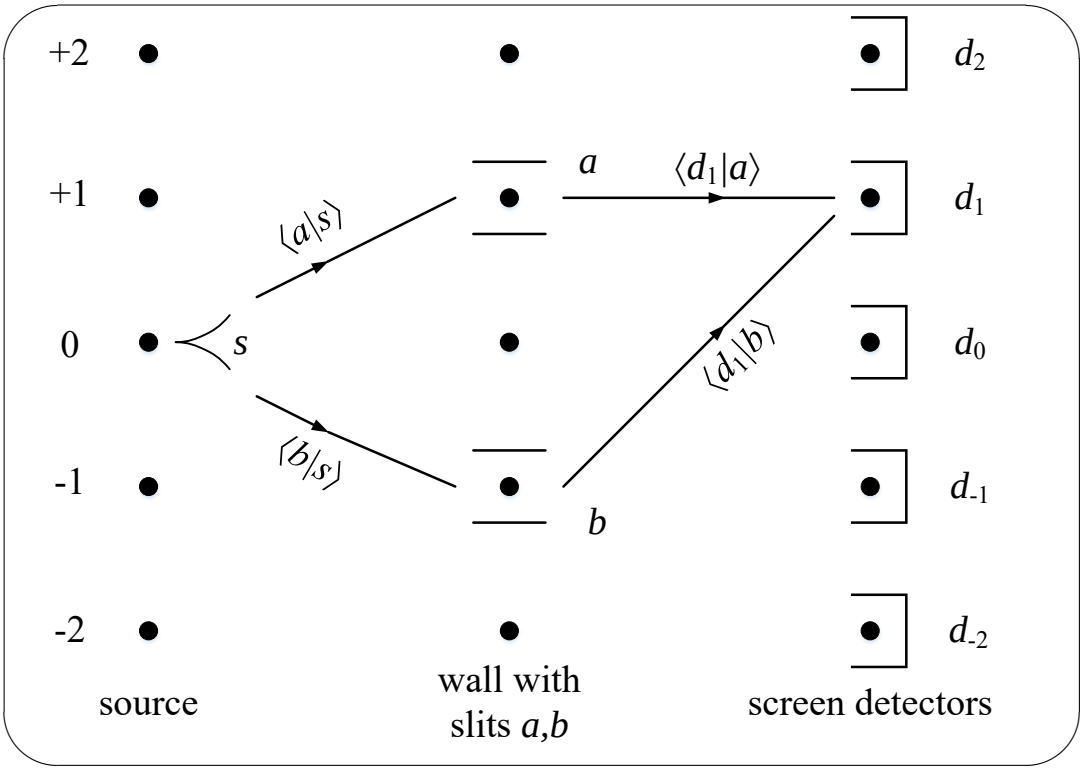


Figure 20: The double-slit experiment described for a discrete spacetime. The particle leaves source s, passes one of the two slits a or b, and is finally detected in d_1.

2.6 The Double-Slit Experiment

The double-slit experiment with its diffraction pattern has been called "The most beautiful experiment in physics[32]". The used experimental set-ups depend on the type of particles acting in the two-slit experiment. It can be done with photons or electrons, and becomes more difficult for increasing size of the particles. Until now, the largest molecules showing interference, are combined of 810 atoms.

We consider a source of particles, say photons, electrons or fullerene. For the sake of convenience, we assume a discrete space consisting of points

$$(m,t), \quad m = -2,-1,0,1,2, \quad t = t_0,t_1,t_2, \tag{26}$$

where m denotes five spatial points at three times t, as displayed in Figure 20. It makes no difference for the mathematical treatment, if we choose a much finer grid, for instance with 10^{100} points, leading to an approximation of spacetime much more finer than the accuracy of any measurements.

At time t_0 a particle leaves the source s, arrives at time t_1 at the wall with two slits a and b, and is detected at time t_2 in exactly one of the detectors d_m. We ask for the probability that the particle is detected at point m.

If we use classical probability theory, then the particle arrives at m either through slit a or through slit b. These are two mutually exclusive events that

[32] Crease [2002]

add up to the total probability

$$\begin{aligned}\text{Prob}\{\langle d_m|s\rangle\} &= \text{Prob}\{\langle d_m|a\rangle\}\text{Prob}\{\langle a|s\rangle\} \\ &\quad +\text{Prob}\{\langle d_m|b\rangle\}\text{Prob}\{\langle b|s\rangle\}.\end{aligned} \tag{27}$$

If the probability of passing slit a is the same as passing slit b, thus is $1/2$, we obtain

$$\text{Prob}(\langle d_m|s\rangle) = \frac{1}{2}\text{Prob}(\langle d_m|a\rangle) + \frac{1}{2}\text{Prob}(\langle d_m|b\rangle). \tag{28}$$

Since probabilities are always non-negative, their addition does not lead to any cancellation or destructive interference. This contradicts experimental results, since interference is observed if both slits are open, see Figures21[33] and 22 .

Now we try to explain this experiment by using Feynman's quantum probability rules: the amplitudes are multiplied for independent events, are added for mutually exclusive events, and are squared, finally. Firstly, we assume that slit b is closed. The probability amplitudes $\langle a|s\rangle = 1$, $\langle b|s\rangle = 0$, $\langle d_m|a\rangle = \psi_m$, and $\langle d_m|b\rangle = \varphi_m$ combine to

$$\begin{aligned}\text{Prob}\{\langle d_m|s\rangle\} &= |\langle d_m|a\rangle\langle a|s\rangle + \langle d_m|b\rangle\langle b|s\rangle|^2 \\ &= |\langle d_m|a\rangle \cdot 1 + \langle d_m|b\rangle \cdot 0|^2 \\ &= |\psi_m|^2.\end{aligned} \tag{29}$$

Thus we obtain a classical probability without any interference. Secondly, we assume that slit a is closed. Then $\langle a|s\rangle = 0$, $\langle b|s\rangle = 1$, and as above we get

$$\text{Prob}\{\langle d_m|s\rangle\} = |\varphi_m|^2. \tag{30}$$

Finally, we assume that both slits are open. The probability calculated by Feynman's rules with $\langle a|s\rangle = \langle b|s\rangle = \frac{1}{\sqrt{2}}$ is

$$\begin{aligned}\text{Prob}\{\langle d_m|s\rangle\} &= |\langle d_m|a\rangle\langle a|s\rangle + \langle d_m|b\rangle\langle b|s\rangle|^2 \\ &= |\tfrac{1}{\sqrt{2}}\psi_m + \tfrac{1}{\sqrt{2}}\varphi_m|^2 \\ &= \tfrac{1}{2}(\psi_m + \varphi_m)^*(\psi_m + \varphi_m) \\ &= \tfrac{1}{2}(\psi_m^*\psi_m + \psi_m^*\varphi_m + \varphi_m^*\psi_m + \varphi_m^*\varphi_m) \\ &= \tfrac{1}{2}\left(|\psi_m|^2 + |\varphi_m|^2\right) + \tfrac{1}{2}(\psi_m^*\varphi_m + \varphi_m^*\psi_m).\end{aligned} \tag{31}$$

Comparing with (28), (29), and (30), it follows that the first term in this sum corresponds to the classical probability, and the second term describes interference.

This can easily be seen as follows. For points m with $\psi_m = \varphi_m$ we obtain from (31)

$$\text{Prob}\{\langle d_m|s\rangle\} = 2|\psi_m|^2. \tag{32}$$

[33] Belzasar [2012]

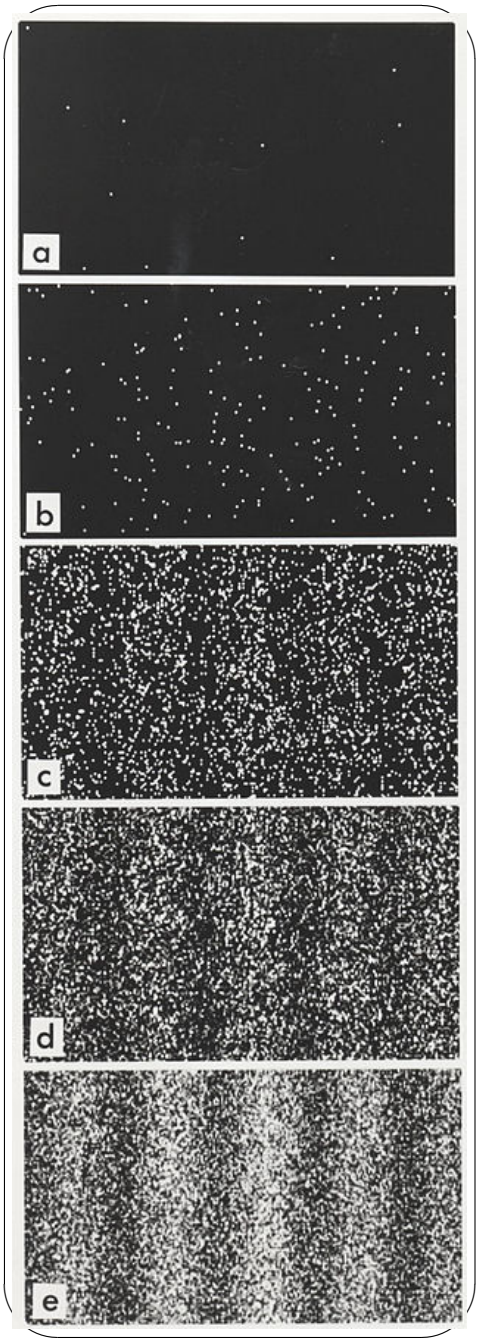


Figure 21: Results of a double-slit-experiment performed by Dr. Tonomura showing interference. The numbers of electrons are 11 (a), 200 (b), 6000 (c), 40000 (d), and 140000 (e). The electrons were shot one by one through the double-slit so that they could not interfere with each other.

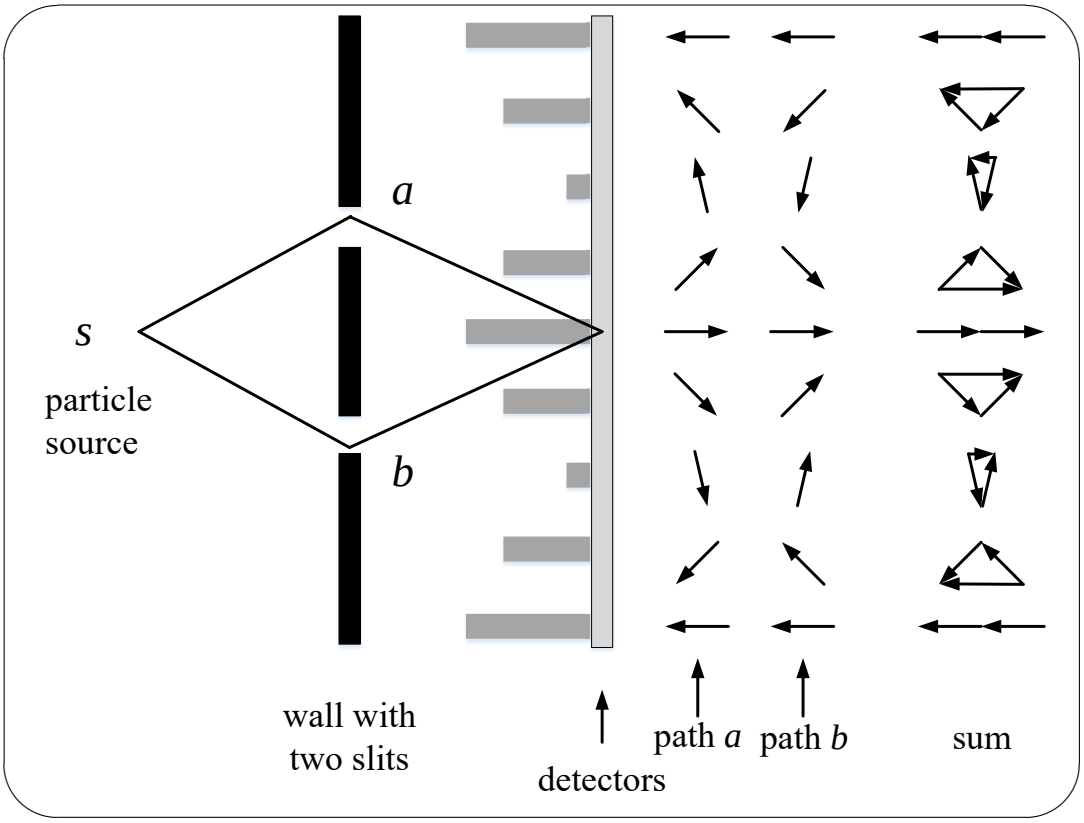


Figure 22: Schematic illustration of the double-slit experiment. The arrows represent the complex amplitudes for each path and their sum. Squaring the magnitude of the sum determines the corresponding probability. This leads to destructive and constructive interference, as displayed on the wall of detectors.

This doubles the classical probability, where only one slit is open. Hence, we have *constructive interference*. If $\psi_m = -\varphi_m$, the probability to find the particle at point m is

$$\text{Prob}\{\langle d_m|s\rangle\} = 0, \tag{33}$$

yielding *destructive interference*. For other combinations we obtain probabilities that are between both extreme cases, as displayed in Figure 22.

Until now we don't have the correct values for all amplitudes, such as ψ_m and φ_m. In order to calculate the amplitudes for going from one position to another we need the classical physical action of this process, usually denoted by S. In classical mechanics the action is a well-known concept. Roughly spoken, the action is the Lagrangian, that is the difference between the kinetic energy and the potential energy, integrated over a path where a particle may move. For a free particle (no forces act on it) the Lagrangian is just the kinetic energy. Classically, the path taken by the particle actually is a stationary point of the action S. This rule is called *Hamilton's principle of least action*.

In non-relativistic classical mechanics a free particle is an object where no forces act on it. It has only kinetic energy $p^2/2m$, and the action to go from position r_1 to position r_2 is the integral over time, taken along the path between

both positions. In first order the action is $S = p(r_2 - r_1)$. The amplitude $\langle r_2|r_1\rangle$ of a path between positions r_1 and r_2 is proportional to the complex number

$$\langle r_2|r_1\rangle = e^{iS/\hbar} = e^{ip(r_2-r_1)/\hbar}, \tag{34}$$

where $\hbar$ is Planck's constant.

In the relativistic case, the only difference is that the momentum p is related to the well-known *energy-momentum relation*

$$E^2 = (pc)^2 + (m_0c^2)^2 \tag{35}$$

with particles's rest mass m_0, the speed of light c, and assuming the flat Minkowsi spacetime in special relativity.

Equation (34) can be viewed as a complex plane wave with wave number equal to the momentum divided by Planck's constant. More precisely, the particle is at some time t at some place r such that actually we have a time-dependent amplitude $\langle r_2, t_2|r_1, t_1\rangle$. In the nonrelativistic case, when looking at the limit $t_2 \to t_1$, the amplitude is a function of time and it can be shown that it satisfies the well-known *Schrödinger's wave equation*.

Finally, we want to discuss the case that in the double-slit experiment we can get information about which slit the particle passes through. This information can be given by two detectors d_a and d_b that click when a particle passes slit a or b, respectively. Of course, detectors may fail and information might be wrong.

Looking at the experiment displayed in Figure 23, we have the event that a particle is detected at point m and the detector d_a or d_b clicks. We consider in the following this type of events, although we could easily incorporate events where a particle is detected in m, but neither one of d_a or d_b clicks.

The amplitude that a particle goes from source s via slit a to point m and detector d_a clicks is

$$\langle m|d_a\rangle\langle d_a|a\rangle\langle a|s\rangle. \tag{36}$$

But it may also happen that a particle arrives at m via slit b and detector d_a clicks. This possibility has the amplitude

$$\langle m|d_a\rangle\langle d_a|b\rangle\langle b|s\rangle, \tag{37}$$

and should happen rarely, provided the detectors work well. Other mutually exclusive possibilities do not occur, as can be seen from Figure 23. According to our addition rule we have to add both amplitudes

$$\langle (m, d_a)|s\rangle = \langle m|d_a\rangle\langle d_a|a\rangle\langle a|s\rangle + \langle m|d_a\rangle\langle d_a|b\rangle\langle b|s\rangle \tag{38}$$

for the outcome that a particle arrives at m from source s and detector d_a clicks. The corresponding probability is

$$\text{Prob}\{\langle (m, d_a)|s\rangle\} = |\langle (m, d_a)|s\rangle|^2. \tag{39}$$

If the detectors are perfect, then the probabilities

$$\text{Prob}\{\langle d_a|a\rangle\} = 1 \quad \text{and} \quad \text{Prob}\{\langle d_a|b\rangle\} = 0, \tag{40}$$

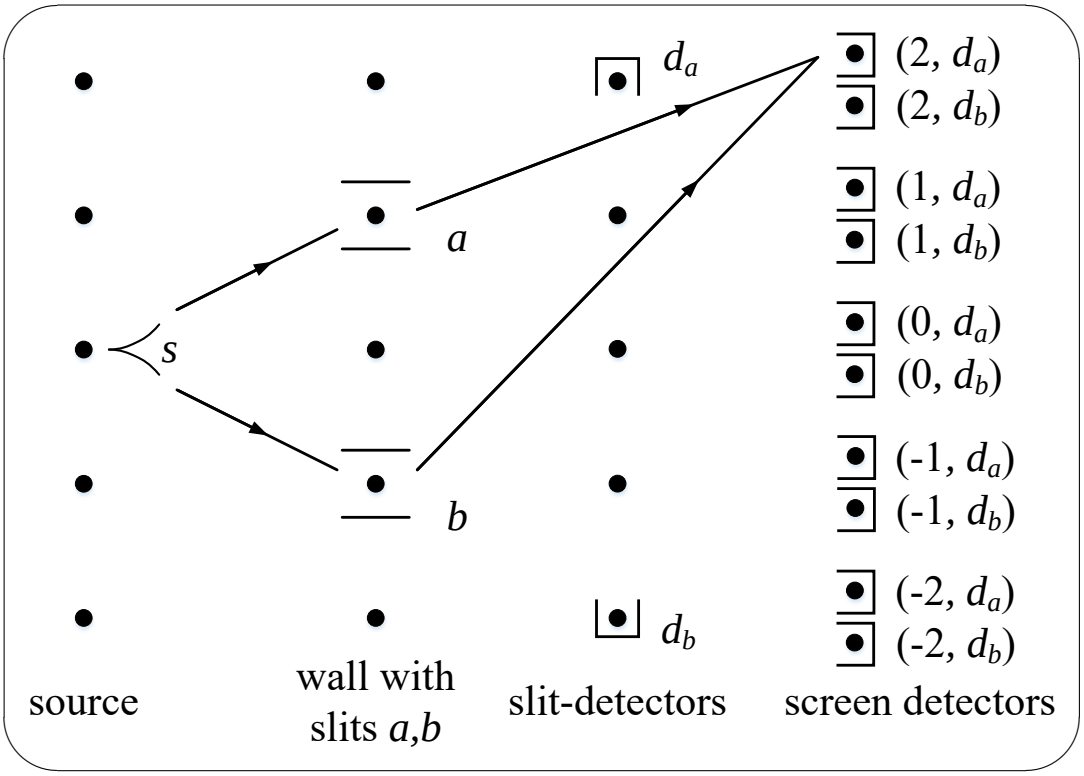


Figure 23: The double-slit experiment with slit-detectors. There are two paths for the event that a particle arrives at point 2 and detector d_a clicks. For the other points there are two paths as well.

yielding $\langle d_a|a\rangle = 1$ and $\langle d_a|b\rangle = 0$. Assuming $\langle a|s\rangle = \frac{1}{\sqrt{2}}$ gives

$$\text{Prob}\{\langle (m, d_a)|s\rangle\} = |\langle m|a\rangle\langle a|s\rangle|^2 = \frac{1}{2}|\psi_m|^2, \tag{41}$$

which is the classical probability as in (28).

With the same arguments as before, we obtain the amplitude

$$\langle (m, d_b)|s\rangle = \langle m|d_b\rangle\langle d_b|a\rangle\langle a|s\rangle + \langle m|d_b\rangle\langle d_b|b\rangle\langle b|s\rangle \tag{42}$$

for the event that a particle arrives at point m from source s and detector d_b clicks. For an ideal detector $\langle d_b|a\rangle = 0$, and assuming $\langle b|s\rangle = \frac{1}{\sqrt{2}}$ we obtain the classical probability

$$\text{Prob}\{\langle (m, d_b)|s\rangle\} = |\langle m|b\rangle\langle b|s\rangle|^2 = \frac{1}{2}|\varphi_m|^2. \tag{43}$$

This seemingly surprising result can be explained easily. The reason is that the outcomes have changed from spacetime points m to spacetime points (m, d_a) and (m, d_b) with detectors. A change of the outcome implies a change of the probabilities. That's all, nothing strange happened.

On the other hand, if both detectors don't work, say all amplitudes $\langle d_a|a\rangle$, $\langle d_a|b\rangle$, $\langle d_b|a\rangle$, and $\langle d_b|b\rangle$ are equal to some value α, then the total amplitude

becomes

$$\begin{aligned} \langle m|s\rangle &= \langle (m,d_a)|s\rangle + \langle (m,d_b)|s\rangle \\ &= \alpha\cdot\left(\tfrac{1}{\sqrt{2}}\psi_m + \tfrac{1}{\sqrt{2}}\varphi_m\right). \end{aligned} \tag{44}$$

Comparing with (31), it follows that this is the probability if both slits are open, except for the pre-factor. Thus we have interference as in the case without any detectors.

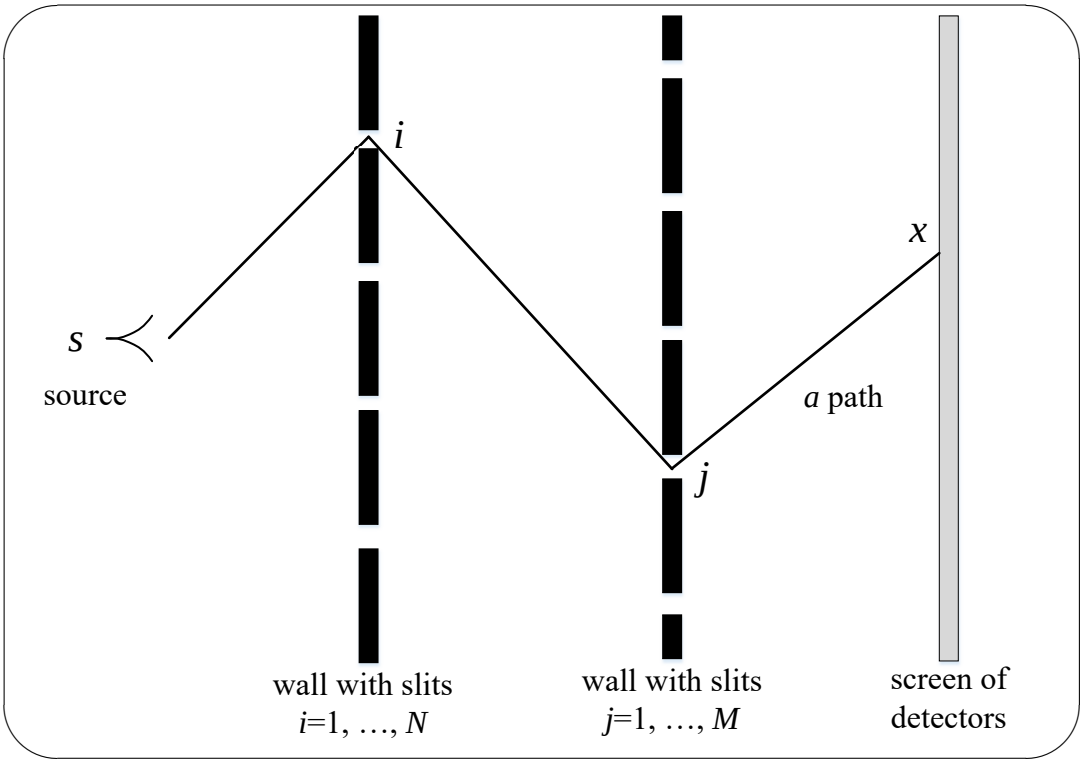


Figure 24: Two walls with multiple slits $i = 1, \ldots, N$ and $j = 1, \ldots, M$.

2.7 Diffraction at Multiple Slits

In this section we shall generalize our results for the double-slit experiment to multiple slits. Moreover, we shall derive computable probability amplitudes.

Let us consider the more complicated interference experiment as displayed in Figure 24.

There are several paths: the particle can go from the source s through some hole i in the first wall, then through some hole j in the second wall, and finally to a detector x at the screen, yielding the amplitude

$$\langle x|j\rangle\langle j|i\rangle\langle i|s\rangle. \tag{45}$$

All these paths from the source to any detector describe possibilities. The particle could move in a future action on this path from s to x *as well as* on another path from s to x. Therefore, the amplitude to arrive at a specific position x is the sum of all related amplitudes, that is,

$$\langle x|s\rangle = \sum_{\substack{i=1,\ldots,N \\ j=1,\ldots M}} \langle x|j\rangle\langle j|i\rangle\langle i|s\rangle. \tag{46}$$

Of course, this sum can easily be generalized to an arbitrary number of walls with an arbitrary number of slits with infinitesimal distances. Then the sum (46) becomes an integral, namely the *path integral*.

It will turn out that the single amplitudes depend on the geometry of the experimental set-up and the particles in the experiment. We will now derive the concrete amplitudes in the case of photons with diffraction at one wall only.

What is the idea behind calculating the concrete amplitudes? Photon's do not appear smoothly. Contrarily a photomultiplier clicks randomly, thus showing their discrete nature. A photon is a massless and chargeless particle travelling with the speed of light c in form of lumps of energy. The rule for obtaining the probability amplitude, and thus the chance of this path, is derived by computing the action S_{path} and forming the complex arrow $e^{\frac{i}{\hbar}S_{path}}$, where $h = 2\pi\hbar$ is the *Planck constant*.

Hence, our remaining problem is to find the kinetic energy of a photon. Photons move with the speed of light c, and thus should be described in accordance with the theory of special relativity. There, particles satisfy the *relativistic energy-momentum equation*

$$E^2/c^2 - (p_x)^2 - (p_y)^2 - (p_z)^2 = m^2c^2, \tag{47}$$

where the three-vector $p = (p_x, p_y, p_z)$ is the *relativistic momentum*

$$p = \gamma mv, \quad \text{where} \quad \gamma = \frac{1}{\sqrt{1 - \frac{v}{c}^2}}, \tag{48}$$

and

$$E = \gamma mc^2 \tag{49}$$

is the *relativistic energy*. Defining the *energy-momentum four vector* as $p^\mu = (E/c, p_x, p_y, p_z)$, it follows that the *relativistic energy-momentum equation* states that the quantity on the left hand side is relativistically invariant with respect to Lorentz transformations, and is equal to the squared rest energy. Especially this equation shows a beautiful symmetry in the relativistic spacetime.

A closer inspection exhibits a loophole for massless particles, since for $m = 0$ the numerators in equations (48) and (49) are zero. But if we demand that massless particles move at the speed of light $v = c$, then also the denominators vanish, and the relativistic energy-momentum equation is formally fulfilled when we set

$$v = c \text{ and } E = |p|c. \tag{50}$$

Actually, these are at most formal assignments. It is not clear where the energy and the momentum of a massless particle come from. In physics, however, a photon is defined as the energy lost by an electron when jumping from an excited state to a state of lower energy, yielding the equation

$$E = eV \text{ and } |p| = E/c. \tag{51}$$

where eV denote electronvolts. It is well-known that the energy determines the color of light. Moreover, we see that the energy is independent of the path. Photons have no position.

The action is the integral, or sum if you like, along the path. It follows that for any photon the action is equal to Et_{path}. The quantity t_{path} is just the parameter describing the required time for the photon to get from the starting

point of the path to the final point. Since the photon moves with constant velocity, the required time is proportional to the length of the path, the latter an important geometrical property. This geometrical quantity defines mainly our amplitude.

Classically, light is an electromagnetic wave. The quotient $\omega := E/\hbar$ is called *angular frequency* of light, and $\nu = \omega/2\pi$ is the *frequency* of light. Frequency and the speed of light are combined by the equation $c = \lambda\nu$, where λ is the *wavelength*. Thus a high frequency, or equivalently a short wavelength, determine high energy photons.

Using the previous quantities, the action S_{path}, related to a given path, is the energy multiplied by the time, that is,

$$S_{path} = \hbar\,\omega\,t_{path}. \tag{52}$$

Dividing the action by $\hbar$ yields a dimensionless number, namely the probability amplitude for the path

$$\langle \texttt{point}\ x \mid \texttt{start}\ s\rangle_{via\ path} = e^{\frac{i}{\hbar}S_{path}} = e^{i\omega t_{path}}. \tag{53}$$

In fact, this amplitude looks like a stopwatch that is started when the photon leaves the source. This stopwatch rotates very rapidly corresponding to the action for this path - this is about 36 000 times per inch for red light - until the photon reaches the photomultiplier. At this moment the watch is stopped, and the hand points in a certain direction, yielding the small arrow called the amplitude.

Suppose we have a wall with N slits as displayed in Figure 25. Then for a photon that passes slit $x_n = x_0 - nd,\ x_0 = 0,\ \ n = 0, 1, \ldots, N-1$ and arrives at some point x on the screen of detectors, the formula (53) takes the form

$$\psi_n(x) = \langle \texttt{point}\ x \mid \texttt{start}\ s\rangle_{via\ x_n} = e^{i\omega t}e^{i\omega\sqrt{(x-nd)^2+L^2}}. \tag{54}$$

There, t is the time that the photons needs to go from the source to slit x_n. Of course this time depends on the related slit. We assume that n is small, and the distance from the source to the wall of slits is large compared to the distance d between the slits. We set $c = 1$ such that $x = ct = t$. Thus time and distance are equal. The square root in the second exponential term is exactly the distance between the slits and the detectors. The action as an integral is additive, and the multiplication rule for the exponential function implies the product of both exponential terms.

If we assume additionally that the distance d between two slits is small compared to the distance L from the slits to the screen of detectors, we can write

$$\sqrt{(x-nd)^2+L^2} \approx L\left(1 + \frac{1}{2}\frac{(x-nd)^2}{L^2}\right) = L + \frac{x^2}{2L} - \frac{xnd}{L} + \frac{(nd)^2}{2L}, \tag{55}$$

where higher order terms of d are neglected. Therefore, for a photon passing slit x_n, we get the approximate amplitude

$$\psi_n(x) = e^{i\omega(t+L+\frac{x^2}{2L})}e^{-i\omega(\frac{xnd}{L}-\frac{(nd)^2}{2L})}. \tag{56}$$

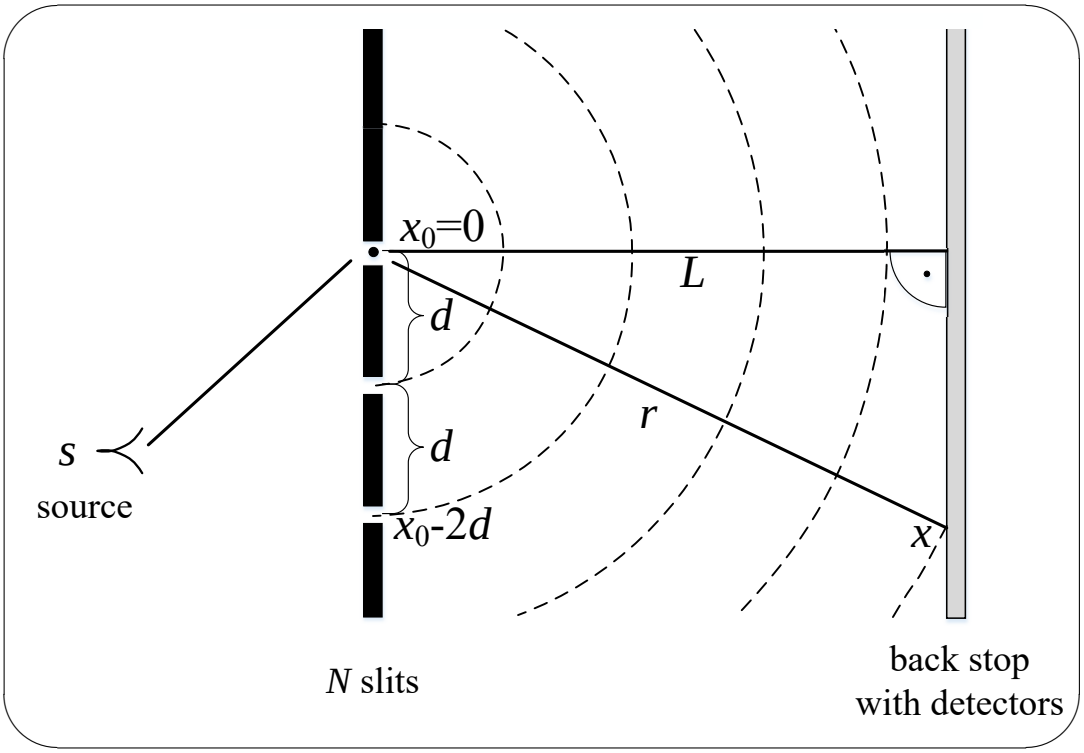


Figure 25: A wall with N slits, with distance d between two slits, and distance L from the slits to the back stop of detectors.

Superposition over all slits yields the final amplitude

$$\psi(x) = Ae^{i\omega(t+L+\frac{x^2}{2L})} \sum_{n=0}^{N-1} e^{-i\omega(\frac{xnd}{L}-\frac{(nd)^2}{2L})}, \tag{57}$$

for a photon passing any slit, which we don't know, and arriving at detector x. The constant A is a necessary normalization constant such that all squared magnitudes are classical probabilities and sum up to one. The function $\psi(x)$ of amplitudes depends mainly on the geometry of the experimental set-up, not so much on the photon itself. The contribution of the photon is its frequency only, or equivalently its energy.

The slit experiments are rather mysterious and strange. At the back-stop the particles arrive as small lumps demonstrating a point-particle like behaviour. They show a classical statistical point-like behavior whenever only one of the slits is open, or if we obtain information through which slit the particle has gone. But if the slits are open, and if it is not decidable through which slit the particle passes, then we obtain an interference pattern.

These results can be obtained also with the same mathematics that describes sound waves or water waves, except that we use complex numbers in quantum mechanics, not real numbers. If waves meet the wall of slits, then at each slit diffraction generates a radial wave traveling radially away from the slit, according to *Huygens' principle*, well-known in wave theory. These radial waves superpose, and summing them up returns an intensity distribution of a wave that shows the same interference as ψ. This is one reason for the well-known *wave-particle duality*. Nevertheless, the particles arrive as lumps at the

backstop in contrast to real waves, and we should have in mind what Feynman writes:

Keep in mind:

> *However, it must be emphasized that the wave function [of amplitudes] that satisfies the equation is not like a real wave in space; one cannot picture any kind of reality to this wave as was done to a sound wave.*

Feynman[34]

In other words, the wave function of amplitudes may mathematical look like a real wave, but actually it is not a wave. It implies merely a probability distribution.

Similar results can be obtained when performing diffraction experiments with particles that have mass. For free electrons and other free massive particles traveling with low speeds we can use the non-relativistic kinetic energy $\frac{1}{2}mv^2$ leading to the complex arrow $e^{\frac{i}{\hbar}S_{path}}$ that rotates with

$$\texttt{rate of rotation} = \frac{mv^2}{2\hbar}\texttt{turns per sec.} \tag{58}$$

When teaching quantum physics electron diffraction at multiple slits is usually discussed as thought experiments. An exception is Jönsson's article[35] that described already in 1961 a realization in the laboratory with electrons up to five slits. Besides the educational value of this precise experimental description, with this set-up neither single particle diffraction nor close individual slits can be observed. In 2013 Bach et al.[36] presented a full realization of this famous thought experiment.

If the particles are not free but forces act on it, then the Lagrangian is the difference between the kinetic and the potential energy. The kinetic energy depends only on the velocity, and the potential energy comes from the gravitational force acting on space points. Let us throw a ball in the air, then *Hamilton's principle of least action* says that the ball follows a parabola, which is the path of least action.

In classical mechanics a *force* is interpreted as something that pulls or pushes, even over long distances. In the theory of general relativity a force is determined through the curvature of spacetime. From our stochastic point of view, the difference between the action along a path with and without potential energy is the change of the phase angle of our amplitudes. In other words, the direction of the complex arrows is changed.

[34] Feynman Lectures [1963, Vol. 3, pp. 3–4]
[35] Jönsson [1974]
[36] Bach et al. [2013]

Keep in mind:

A modern stochastic point of view of what **forces** *are, is as follows: a force changes the action and thus phases, that is, the directions of probability amplitudes.*

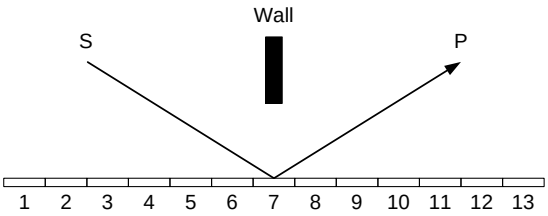


Figure 26: The classical view of the ray model: the mirror reflects light such that the angle of incidence is equal to the angle of reflection.

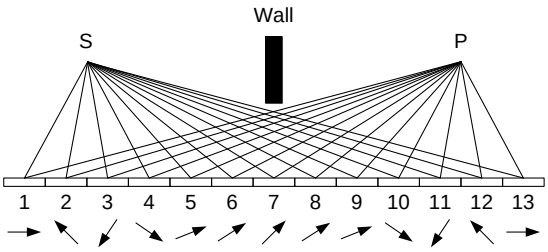


Figure 27: Feynman's view says that light has an amplitude equal in magnitude for each possible path from the source to the photomultiplier. In particular, it can be reflected from every part of the mirror, that is, from the middle as well as from the other parts.

2.8 Light Reflection

In this section we investigate Feynman's quantum probability predictions for the seemingly simple problem of how light is reflected by a mirror, see Figure 26. The experimental set-up is as follows: at a source, light of one color is emitted, and at another point there is a photomultiplier for detecting light. We use a very low light intensity such that some time passes between the clicks of the photomultiplier. In other words, only one photon is in the experiment at any time. To prevent a photon from going straight across to the detector without being reflected, a wall is placed in the middle.

The well-known *ray model of light* in optics says that the mirror reflects light in a way such that the angle of incidence is equal to the angle of reflection. Moreover, the length of the mirror as well as the right and the left end of the mirror have no influence on the light that reaches the detector. This model describes light in terms of rays and holds true in many practical situations.

In Feynman's formulation of quantum probability the photon has also the possibility to take any other path. Only paths through the wall are forbidden. Each possible path is furnished with a reasonable amplitude as displayed in Figures 27 and 28. Remember that we have seen that the action of the photon along any path corresponds to the length of this path, not to the path itself like the action of mass particles.

This is the way how the well-known *wave-particle duality* is resolved in Feynman's formulation: the photon has no complementary partner such as a

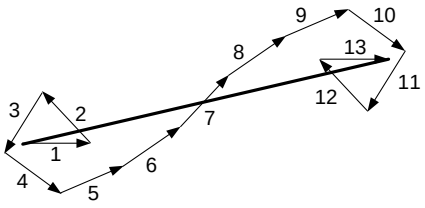


Figure 28: The amplitudes for all possible paths are added together. The major contribution to the final arrow's length is made by the paths of minimal action, that corresponds to the paths of minimal length.

the wave. Instead paths are equipped with arrows, the probability amplitudes, satisfying the general rule: For each path from the source to the photomultiplier draw an appropriate arrow, add all arrows with parallelogram addition. This is, the quantum rule of *superposition*. Then square the magnitude of the resulting arrow. This returns the probability of being detected by the photomultiplier. These are the same stochastic rules as stated in the sections before. It makes any wave pictures superfluous in Feynman's formalism.

The probability amplitude for the outcome that a photon starts at source s and is detected by the photomultiplier at point p is

$$\langle p|s\rangle = \sum_{all\ paths} e^{\frac{i}{\hbar}S_{path}}. \tag{59}$$

At a first glance, we guess that a photon should simultaneously take all possible paths from s to p, a strange visualization, that is mentioned in many textbooks. The action S_{path} for photons, however, depends only on the length of the path or equivalently on the time t_{path}. This is a geometrical property of the experimental set-up. Thus, the amplitudes don't require that the photons must actually interact with all paths. They just describe possibilities[37].

The small arrows in this sum are displayed in Figures[38] 27 and 28, where we have divided the mirror into little squares, with one path for each square. When we add all contributions for the paths, then, as seen in Figure 28, the final arrow length evolves mainly from arrows of the middle part of the mirror, whereas the contributions from the left and right part almost cancel out each other. All paths in the middle part have almost the shortest length, corresponding to the well-known classical principle of least action.

More precisely, for all paths from the source to the photomultiplier, the action S_{path} is very large compared to Planck's constant. Therefore, for nearby paths the amplitudes differ very much, since a relatively small change of the action is large compared to $\hbar$ thus yielding a completely different phase. This implies cancellation of the arrows in the sum.

There is only one exception, namely the paths that are infinitesimally close to the path of least action, also called the extremal paths. In this case the first

[37]The usual interpretation of this experiment can be found in the nice talk of Girvin in the KITP Public Lectures, see online `kitp.edu/online/plecture/girvin`.

[38]The figures in this section are modifications of related ones in the book Feynman [1985]

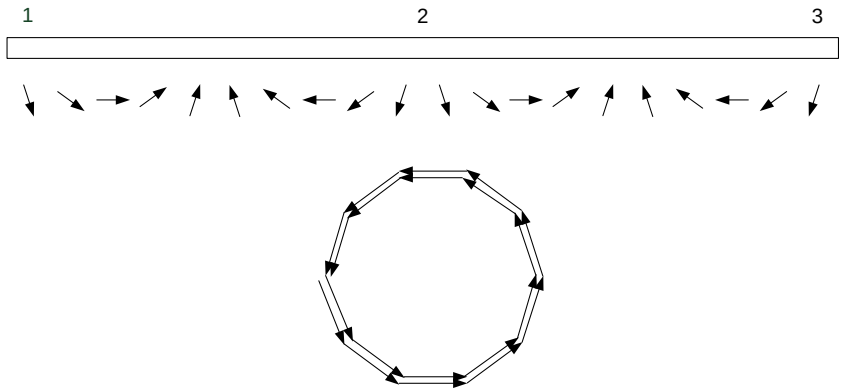


Figure 29: Considering only the piece of the left part of the mirror, the detector does not click, since the amplitudes add up to approximately zero.

variation of the action is zero. This implies that nearby paths have almost equal action, and thus have equal amplitudes in the first approximation. Exactly these paths are the important ones and contribute coherently. This occurs in the region where the arrows almost point in the same direction.

In other words, all paths distant from the classical path of least action interfere destructively. On the other hand, the paths in the neighbou rhood of the classical path interfere constructively. This is the reason why we observe mainly classical events, such as light travels in a straight line where the time is shortest. This fact was already discovered in 1650 by Fermat and is called *Fermat's principle of least time*. This principle is the success of the ray model in optics. Only the middle part of the mirror seems to be responsible for reflections. It is really surprising, however, that the stop watch rotates ten thousands of times until the photon reaches the photomultiplier, but the amplitude for this event is the final hand direction of the watch.

But that is not the whole story. The basic question is: how does the photon find the path of extremal action? Does the photon smell out all possible paths in order to find the right one. Or is this approach only a mathematical description far away from any reality? If this formalism has any reality, we should be able to show in an experiment that a photon sometimes chooses also other paths, shouldn't it?

It is simple to answer these questions using the following experiment. We cut off a large part of the mirror such that only three segments on the left side are left over, see Figure 29. Moreover, the amplitudes are displayed in greater detail. If we add all arrows, we see that they cancel out, and the probability to be detected in the photomultiplier is almost zero.

But if Feynman's theory is true then photons should be detected when we reduce the left part of the mirror in a manner such that no cancellation can occur, see Figure 30. Then the majority of arrows points to the right, and in total we obtain an amplitude that predicts a strong reflection. In fact, in agreement with our theory, the photomultiplier clicks sometimes. **This sounds crazy: in theory as well as in practice you cut off the important middle of the mirror, from the remaining part you scrape away appropri-**

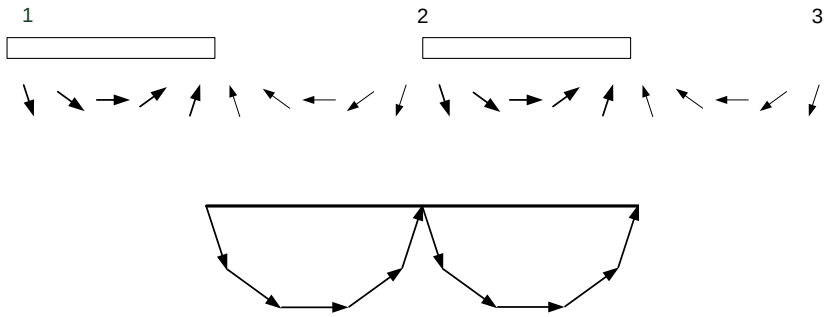


Figure 30: A striped mirror reflects a substantial amount of light, and is called a diffraction grating.

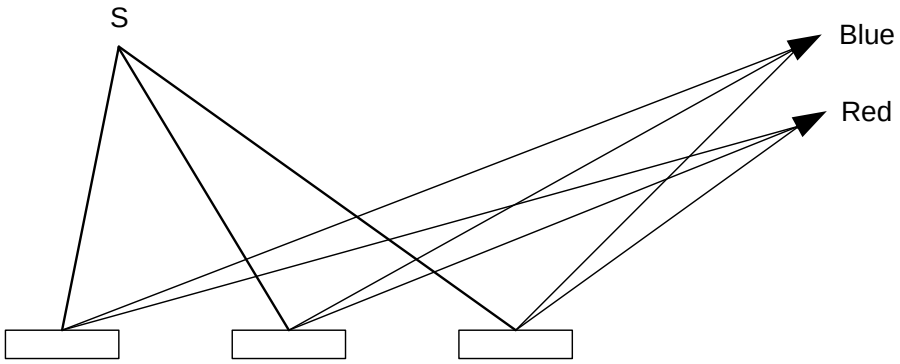


Figure 31: The angle of reflection depends on the color (wavelength) of light.

ate pieces, and then you observe reflection. Once more, the photon seems to walk on each possible path with a stopwatch.

Some further remarks should be given. The size of the experiment, the placements of source, photomultiplier and cut-out's of the mirror, and hence the direction of the arrows depend also on the colour of the light, hence on its energy or equivalently its frequency. This follows from the definition of the action that depends on the frequency or equivalently on the wavelength.

For other particles such as electrons or atoms Feynman's formalism works just as well. There the action is replaced by the integral over the difference between kinetic and potential energy along the considered path.

Keep in mind: Feynman's theory, a pure particle formalism without any waves, shows the close relationship between the deterministic classical mechanics, described by *Hamilton's principle of least action*, and quantum mechanics as a stochastic process with complex probability amplitudes. Both theories have in common that the particle prefers the path for which the action does not vary in the first approximation. The considered paths are not assumed to be smooth, they can exhibit a zig-zag curve as in *Wiener processes*.

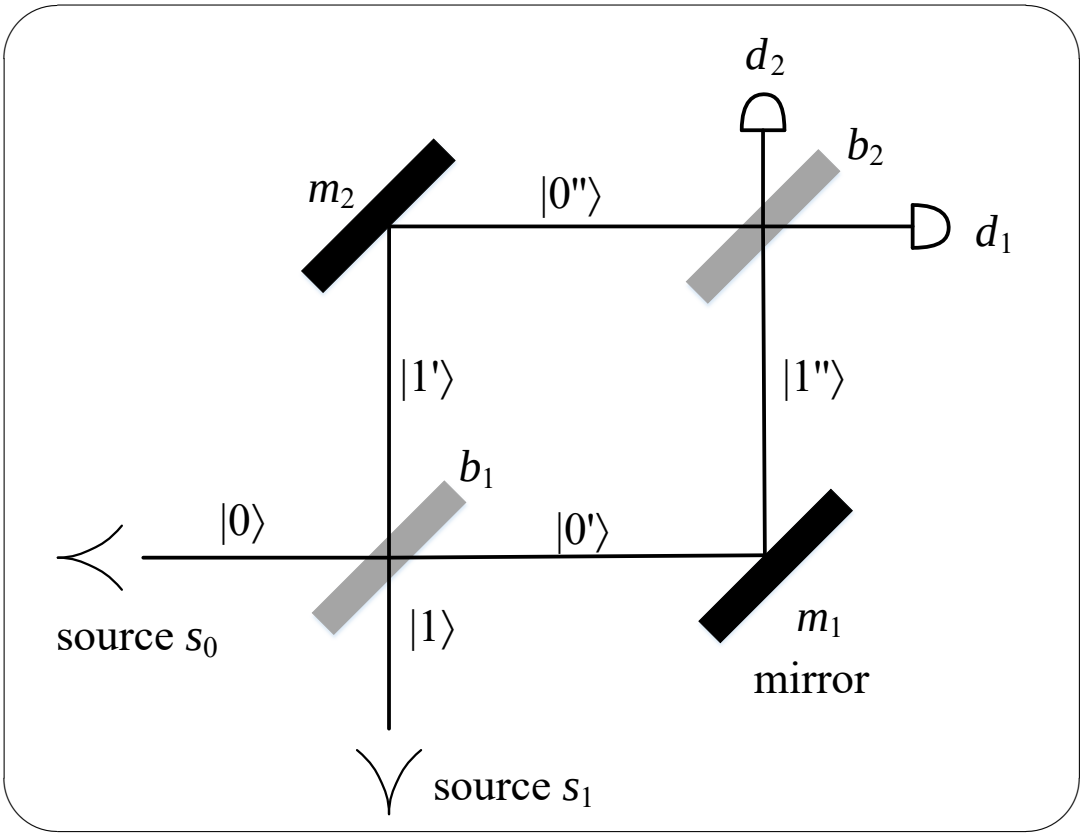


Figure 32: A Mach-Zehnder interferrometer.

2.9 Interferometer

The *Mach-Zehnder interferometer* is a device used, for instance, in electro-optic modulators, integrated circuits, electronic devices, and also in quantum information theory. It serves as a nice apparatus demonstrating the fundamental rules of quantum mechanics. The set-up, see Figure 32, is as follows. A beam of particles, say photons, is split by a beam splitter b_1, for instance a half-silvered mirror. The two resulting beams are reflected by two mirrors n_1 and m_2. Then they pass through a second beam splitter b_2 and finally enter two detectors d_1 and d_1. This experiment can be viewed as a special version of the double-slit experiment, where the photon or the electron can traverse on exactly two paths.

If a single photon enters the interferometer, either from path $|0\rangle$ or path $|1\rangle$, it interacts with the first beam splitter. We assign the passage through a beam splitter by the amplitude $\frac{1}{\sqrt{2}}$. Since the squared amplitude is $\frac{1}{2}$, it is a 50/50 percent beam splitter. The beam-splitter transforms the path of the photon into a two-state system, namely moving horizontally or vertically. This is also called a *path qubit*.

The number $i = e^{i\pi/2}$ is the probability amplitude when the photon is reflected, and the path is rotated by a right angle. Since $|i|^2 = 1$, the probability that a photon is reflected is one, as expected.

Now, let us calculate the amplitude that the photon, starting at source s_0,

is detected in d_1. The first path is from below yielding

$$\langle d_1|1''\rangle\langle 1''|0'\rangle\langle 0'|0\rangle = \frac{i}{\sqrt{2}} \cdot i \cdot \frac{1}{\sqrt{2}} = -\frac{1}{2}, \tag{60}$$

and from above we get

$$\langle d_1|0''\rangle\langle 0''|1'\rangle\langle 1'|0\rangle = \frac{1}{\sqrt{2}} \cdot i \cdot \frac{i}{\sqrt{2}} = -\frac{1}{2}.$$

Both paths represent two possibilities. There are no other paths arriving at detector d_1. Thus, adding both amplitudes we obtain the total probability amplitude

$$\langle d_1|0\rangle = -\frac{1}{2} - \frac{1}{2} = -1. \tag{61}$$

Hence, with probability $(-1)^2 = 1$ the photon will be detected in d_1.

The amplitude, for the photon to be detected in d_2, is the sum of

$$\langle d_2|0''\rangle\langle 0''|1'\rangle\langle 1'|0\rangle = \frac{i}{\sqrt{2}} \cdot i \cdot \frac{i}{\sqrt{2}} = -\frac{i}{2}, \tag{62}$$

and

$$\langle d_2|1''\rangle\langle 1''|0'\rangle\langle 0'|0\rangle = \frac{1}{\sqrt{2}} \cdot i \cdot \frac{1}{\sqrt{2}} = \frac{i}{2}. \tag{63}$$

Hence, the total amplitude is zero, yielding the correct probability zero that a photon is detected in d_2. Interference occurs in accordance with the experimental results.

If we would use the classical probability rules, each detector should click with probability 1/2. This contradicts the experimental results.

2.10 Delayed Choice Experiments

In order to provide a more precise understanding of wave-particle dualism and varying strange interpretations, Wheeler[39] proposed 1983 several types of *delayed choice experiments*, including variations on the double-slit experiment and the two-path interferometer. In particular, these experiments were planned for discussions related to questions such as: does a delayed choice to measure point-like or wave-like properties of an microscopic object change the past?

We describe delayed choice in terms of the experimental set-up of a *Mach-Zehnder interferometer*. In our experimental arrangement we assume that the second beam-splitter can be turned on and off randomly and at a speed such that the particle, electron or photon, has already passed the first beam-splitter. For example this can be done either by moving the second beam-splitter very quickly, or equivalently, by moving and removing the detectors into the two paths before or behind the second beam splitter.

For a better understanding we can also imagine to realize this experiment on a cosmic scale: a star emits photons some billions of light-years ago. The photons must pass a galaxy in their path towards earth. General relativity will make the light bend around the galaxy. A photon can take the left path and bend back toward earth, or it can take the right path and bend back toward earth. On earth, many billions of years later, we can decide about our experimental arrangements. If we use a photographic plate we can measure an interference pattern, since the photons can arrive on both paths. Or we can choose to put in detectors for determining the paths where the photons arrive. In the latter case the pattern will be a clump of photons at one detector, and a clump of photons at the other one. We have delayed the choice such that the particles have already passed one or the other side of the galaxy, or both sides of the galaxy in terms of a superposition a long time ago. However, it seems paradox that our late choice has an influence on the patterns.

More precisely, if the second beam-splitter in the interferometer is turned on with the detectors behind, we already know that wave-like interference occurs, since only one detector clicks, and Wheeler pointed out:

> *[this] is evidence . . . that each arriving light quantum has arrived by both routes.*

If the second beam-splitter is turned off such that the waves could not recombine, then each photon follows one path or the other with probability 1/2, and an almost equal number of photons reach both detectors. Wheeler writes in the same paper:

> *[either] one counter goes off, or the other. Thus the photon has travelled only one route.*

This point of view is in accordance with the *wave-particle duality*. Either one can measure wave-like properties or particle-like properties. But allowing

[39] Wheeler [1978, 1983]

that the experimenters could randomly switch the second beam splitter in the interferometer on and off after the photon has passed the first beam splitter, Wheeler writes:

> *Thus one decides the photon shall have come by one route or by both routes after it has already done its travel.*

This experiment is not based on any analysis on how particles evolve and behave with respect to time; it is what the mathematical formalism of quantum theory predicts. In Wheeler's words, since no signal traveling at a velocity less than that of light can connect these two events, he writes:

> *We have a strange inversion of the normal order of time. We, now, by moving the mirror in or out, have an unavoidable effect on what we have a right to say about the already past history of that photon.*

These thought-experiments have actually been carried out frequently. For instance see Jacques et al.[40]. In the actual experiments, activation and de-activation is decided by a random number generator, not by a person. To summarize, quantum mechanics emerges triumphant and observations depend on the experimental setup.

Keep in mind: Nature seems to behave in accordance with the predictions of quantum mechanics, but seems to agree less with the theory of relativity.

Jacques et al.[41]

[40] Jacques et al. [2006]
[41] Jacques et al. [2006]

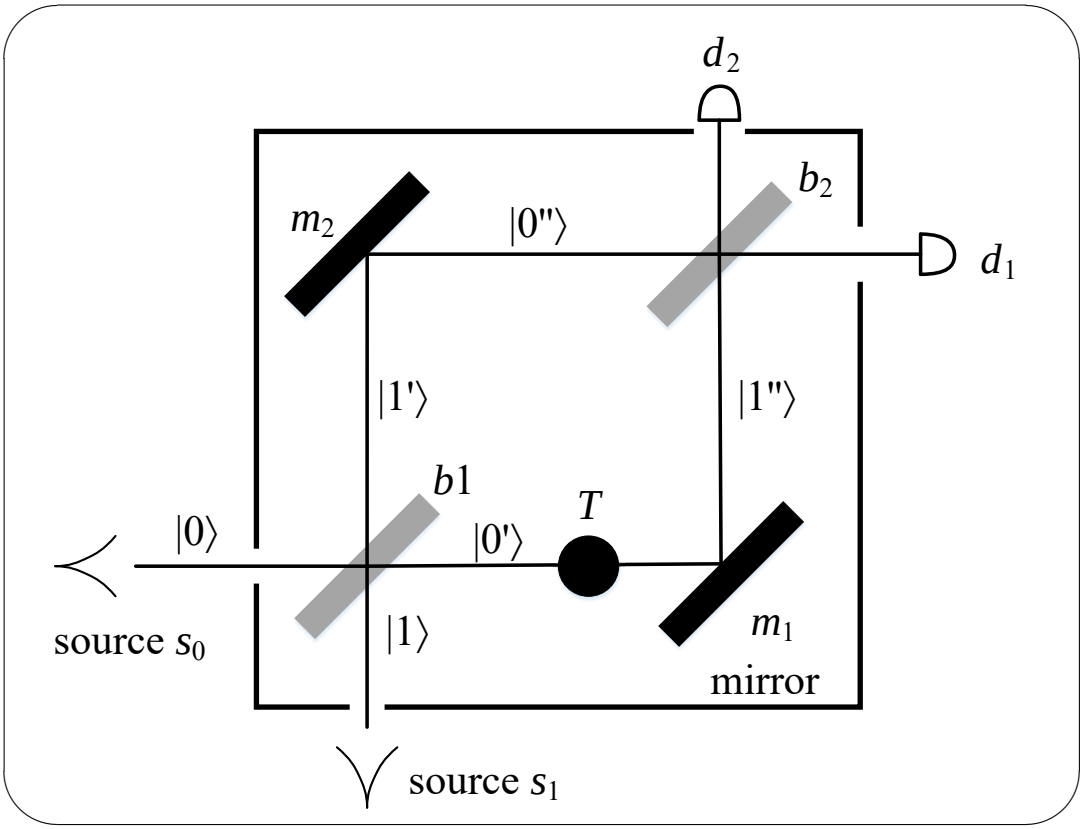


Figure 33: A mantled Mach-Zehnder interferrometer with an object described by amplitude T on the lower path.

2.11 Interaction-Free Measurement

An interaction-free measurement is a type of measurement in quantum mechanics that detects or locates an object without an interaction occurring between it and the measuring device. In other words, the object is not touched locally, but its existence is proved.

A simple experimental set-up, proposed by Elitzur and Vaidman, is given in terms of a modified *Mach-Zehnder interferometer*, see Figure 33. In this interferometer on one of the routes, an object might be placed. We take the lower one. Elitzur and Vaidman speak of an ultra-sensitive bomb that explodes when touched by a photon. The interferometer is mantled such that we cannot decide whether the bomb is put into the experiment or not.

Now we describe the object with a transmission amplitude T. If the object is not present, the photon passes with probability one, and the amplitude is set $T = 1$. Otherwise, if the photon is on the object's path , it will be absorbed, and thus cannot pass. Therefore, we set in this case $T = 0$. All other amplitudes are defined as in Section 2.9.

Now, let us calculate the amplitude that the photon is detected in d_1. The first path is from below where the object might be placed. Then we obtain its amplitude

$$\langle d_1|1''\rangle\langle 1''|T\rangle\langle T|0'\rangle\langle 0'|0\rangle = \frac{i}{\sqrt{2}} \cdot i \cdot T \cdot \frac{1}{\sqrt{2}} = -\frac{1}{2}T. \tag{64}$$

For the upper path from above we get

$$\langle d_1|0''\rangle\langle 0''|1'\rangle\langle 1'|0\rangle = \frac{1}{\sqrt{2}} \cdot i \cdot \frac{i}{\sqrt{2}} = -\frac{1}{2}.$$

As before, both paths are mutually exclusive, and there are no other paths arriving at detector d_1. Adding both amplitudes yields the total probability amplitude

$$\langle d_1|0\rangle = -\frac{1}{2}(T+1). \tag{65}$$

Hence, with probability $\frac{1}{4}(T+1)^2$ the photon will be detected in d_1.

The amplitude for the photon to be detected in d_2 is the sum of the amplitudes

$$\langle d_2|1''\rangle\langle 1''|T\rangle\langle T|0'\rangle\langle 0'|0\rangle = \frac{1}{\sqrt{2}} \cdot i \cdot T\frac{1}{\sqrt{2}} = \frac{i}{2}T, \tag{66}$$

and

$$\langle d_2|0''\rangle\langle 0''|1'\rangle\langle 1'|0\rangle = \frac{i}{\sqrt{2}} \cdot i \cdot \frac{i}{\sqrt{2}} = -\frac{i}{2}. \tag{67}$$

Hence, the total amplitude is

$$\langle d_2|0\rangle = \frac{i}{2}(T-1), \tag{68}$$

yielding the probability $\frac{1}{4}(T-1)^2$ that a photon is detected in d_2.

If the object is not put into the interferometer, the transmission amplitude $T = 1$, and the probability that detector d_1 clicks is $\frac{1}{4}(T+1)^2 = 1$. As in Section 2.9, only detector d_1 clicks.

On the other hand, if the object is present, then $T = 0$, and with probability $\frac{1}{4}(T+1)^2 = \frac{1}{4}$ detector d_1 clicks, with probability $\frac{1}{4}(T-1)^2 = \frac{1}{4}$ detector d_2 clicks, and with probability $\frac{1}{2}$ the photon touches the object and is absorbed.

Therefore, if the object is in the interferometer, the detector d_2 will click in 25% of the cases, detecting the object without any interaction or having any contact with the object.

At a first glance this seems to be surprising, but it isn't. The object, when present with amplitude $T = 0$, can be viewed as a third detector, and classical probability implies the same result, because it is clear which path the photon has taken. So the photon is either on the lower path or on the upper path, with the same probability 1/2. Then the photon moves either on the path to detector d_1 or detector d_2 with probability 1/2. Therefore, with probability 1/2 the photon is absorbed by the object, and if not it will be detected in the detectors d_1 or d_2 with probability $1/2 \cdot 1/2 = 1/4$, respectively. This is simple classical probability theory.

Hence, what is surprising and important, is not the case when the object is present, but the interference when the object is absent, as already investigated in Section 2.9. Then the photon may be on the lower as well as on the upper path. In particular, we see how classical probability emerges from the quantum probability rules.

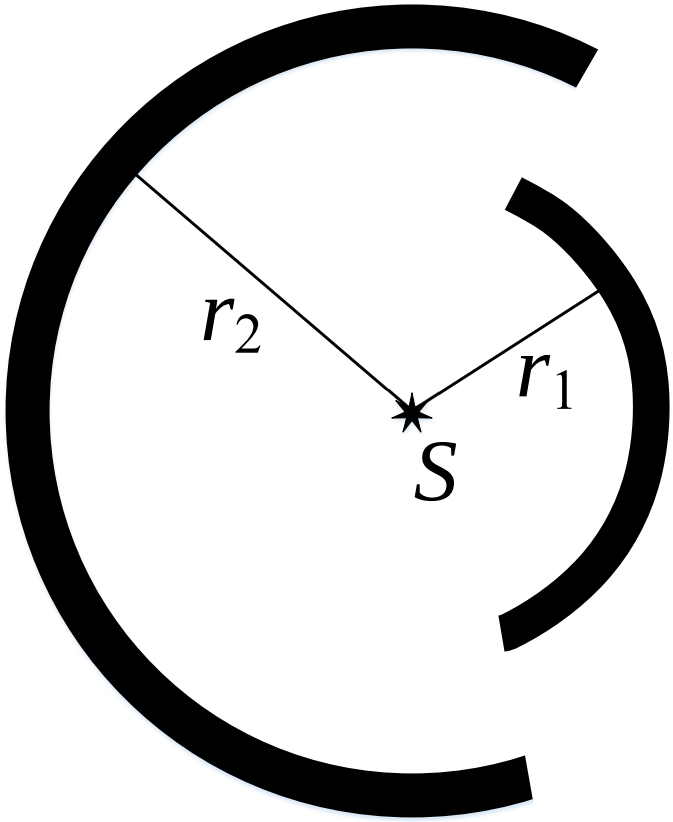


Figure 34: Renninger's experiment.

2.12 Renninger's Negative-Result Experiment

Already 1960, Renninger[42]invented an experiment that is closely related to interaction-free measurements, but is well-known under the name *Renninger's negative-result experiment*[43]. He was the first pointing out the possibility of receiving information by observing that **"nothing happens"**.

He considered the following two-state system, see Figure 34: around a light source two spherical scintillation screens with radii $r_1 < r_2$ are placed. The outer radius is very large compared to the inner radius.

The light source emits in sufficiently large time intervals a photon that can move in all directions. Usually quantum mechanics describes the state of the photon by a radially symmetrical wave function ψ depending on both screens. If the photon is not detected at the inner screen, we have the information that the photon is moving outside the inner radius. But this state must correspond to another wave function ψ' reflecting the possibility to interact only with the outer sphere. Without any disturbance of the photon or any interaction we

[42]In 2012 I have got an acute Leukemia. In the hospital of St. Georg in Hamburg I was cured mainly by two doctors, Prof. Dr. M. Zeis and Dr. H. Hauspurg, with several chemotherapeutics and a bone marrow transplantation. I am indebted to both doctors. The last-mentioned doctor Dr. H. Hauspurg was a grandson of Mauritius Renninger.

[43]Renninger [1960]

have a reduction of the photon's wave function $\psi \to \psi'$.

Renninger's experiment shows that knowledge is gained, although not any detection has happened. This means that the sheer possibility to interact with any detector, although it does not click, is sufficient for the collapse of the wavefunction $\psi \to \psi'$. Some physicists say that this experiment shows a link between the mystery of matter and the observer's mind; see the discussions about the many mind interpretations of quantum mechanics in several books.

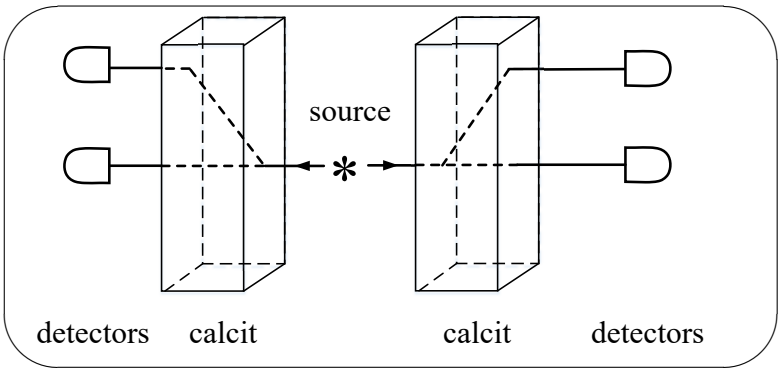


Figure 35: Certain excited atoms can be made to emit a pair of photons that move away in different directions with different colours, say the red one to the left hand side and the blue one to right hand side. When the pair interacts with two calcite crystals having the same polarization axis then we observe: If the red photon is horizontally polarized then the blue one is vertically polarized, and vice versa.

2.13 The EPR Paradox and Bell's Inequality

The quantum behaviour of systems of interacting particles was and is the subject of serious discussions. Quantum physics allows interactions between particles that are spatially separated. They seem to influence each other in a non-local manner. The underlying phenomenon is now known as *quantum entanglement*.

In 1935 Einstein, Podolski and Rosen published a paper in which they argue that quantum theory is incomplete and should be extended with hidden variables. This subject of debate is known as the EPR paradox, where its name corresponds to the initials of the authors.

We consider an atomic system where transitions from an excited state to the ground state occur such that a pair of photons is emitted. Both photons may have different frequencies, hence different colours, say a red one moving to the left side and a blue one moving to the right side, see Figure 35.

The experimental set-up is as follows: the photons move far away in opposite directions. After the generation of both photons their polarization is not specified, thus random. The optical axes, say α and β, of the calcite crystals on the left and the right-hand side are not fixed until the photons are emitted. The intensity of the light is low enough such that only one photon pair is in the experiment at the same time.

The experimental results show that their polarization is *entangled*: they are always polarized at right angles, that is, if the red one is horizontally polarized then the blue one is vertically polarized. To be exact: if the red one is linearly polarized at any angle α, the blue one is linearly polarized at the complimentary angle $\bar{\alpha} = \alpha - \pi/2$, and vice versa.

Consequently, we can remove one of the calcites with its detectors, say the left one, because the polarization is always perpendicular between both entangled photons. Thus, the remaining calcite on the right-hand side tells

us not only the polarization of the blue photon, but also the polarization of the distant red photon. Notice that before the blue photon is measured by the calcite it is randomly polarized. Whatever the actual polarization before interacting with the calcites is, both photons have opposite polarization. Thus, the central question emerges: why has the measurement of the blue photon changed the polarization of the red distant one?

It seems to be natural that both photons are glued together by some hidden variables. In their EPR paper[44] they formulated this problem as follows:

> *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.*

This leads to some important questions. Would it be possible that there exists an element of reality, that is, a hidden variable or a hidden rule which is incorporated in each pair of photons and explains the experimental results. Can we develop a classical *hidden-variable theory* that reproduces these experimental results and thus agrees with quantum predictions, by adding some extra degrees of freedom? Or is it possible to make verifiable and falsifiable predictions that allow to distinguish clearly between hidden variable theories and quantum mechanics? In 1960 Bell investigated this question and found an important answer that will be discussed at the end of this section.

At a first glance, there seems to be a simple surprising answer to all these questions which is purely classical. Suppose, someone goes to skiing. He picks a pair of gloves out of a box without looking at them. But he knows that in the box there were only two pairs of gloves, a black pair and a blue pair. At the alpine resort he unpacks his black gloves. Then he instantaneously knows that his blue gloves are at home. Obviously, this entangled situation does not contradict special relativity, since no signal traveled faster than light.

Of course, this is not an adequate solution to the EPR paradox. Since the polarization of the pair of photons is not known in advance, this means figuratively that the colors of the gloves would be not known. Moreover, we can measure the photons along arbitrary different optical axes. The latter possibility is one of the keys in Bell's work.

If we arrange the optical axes of the left and the right calcite at different angles α and β, respectively, and measure their polarization then we observe a random correlation between both measurements. If we measure horizontal polarization $\alpha = +$ then $\beta = +$ or $\beta = -$, and if we measure vertical polarization $\alpha = -$ then $\beta = +$ or $\beta = -$. In particular, when α and β differ by the angle $\pi/4$, the blue photon is observed with probability $1/2$ in state $\beta = +$, and with probability $1/2$ in state $\beta = -$.

The experimental set-up in Figure 36 can be used to investigate three separate measurements on similarly prepared pairs of entangled photons. In the first type of random experiments the optical axis of the left calcite is α, and on the right calcite the angle is β. In the second type the optical axis of

[44] ERP [1935]

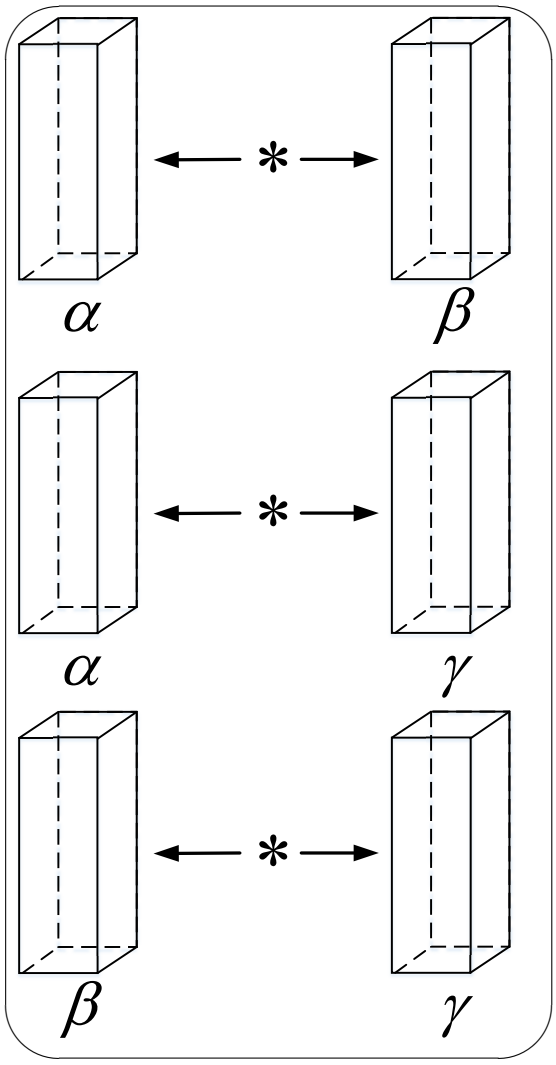


Figure 36: Three separate random experiments on entangled photons. The optical axes of the calcite crystals are α, β and γ.

the left calcite is α, and on the right calcite is γ. Finally, for the third type the optical axis of the left calcite is β, and on the right calcite is γ, see Figure 36. This experiment is performed N times by randomly choosing the optical axes of both calcites, the angles α and β for the left calcite and the angles β and γ for the right calcite. It was actually performed by Aspect et al.[45].

When we investigate this experiment, firstly we have to define the *sample space*, that is, the set of all possible outcomes. Obviously, there are eight ± assignments that define eight mutually exclusive events. These eight events can be visualized on an octahedron, as displayed in Figure 37. On each face there are three symbols which are either "+" or "−".

Now, we leave the world of quantum physics, and use classical stochastic theory. We can throw the octahedron like a dice and apply classical probability theory. If we do this, say $N = 10$ times, we obtain any pattern such as displayed

[45] Aspect et al. [1982]

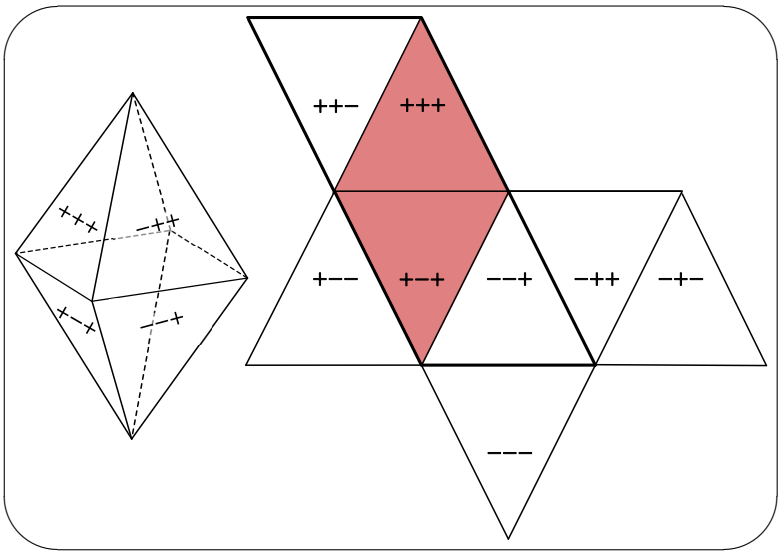


Figure 37: A dice in form of an octahedron.

α	β	γ
+	−	+
−	−	+
+	−	+
+	+	−
+	−	+
−	+	−
+	+	+
−	+	+
−	−	+
−	−	−

Table 1: A possible random pattern when throwing an octahedron.

in Table 1.

We denote by $N(++\,+)$ the number of throws where the octahedron shows the elementary event $(+++)$. In the same way, the other numbers like $N(++-)$ and so on are defined. The eight faces form eight mutually exclusive elementary events. The two events described by the sets $(++\,\pm) = \{(+++), (++-)\}$ and $(\pm\,-\,+) = \{(+-+), (--+)\}$ are subsets of the sample space and form two mutually exclusive non-elementary events.

If we throw the dice many times, it follows immediately from set theory, see Figure 37, that the following inequality

$$N(++\,\pm) + N(\pm\,-\,+) - N(+\,\pm\,+) \geq 0 \tag{69}$$

holds true, where $(+\,\pm\,+) = \{(+++), (+-+)\}$. The quantities $N(++\,\pm)$, and $N(\pm\,-\,+)$ are the number of throws with face in one of the two disjoint sets, respectively. The last number corresponds to faces in $\{(+++), (+-+)\}$. This is one form of *Bell's inequality*.

If we define the classical probability of an event as the number of cases favorable for the event, divided by the number of total outcomes N, then this inequality takes the form:

$$\text{Prob}(++\pm) + \text{Prob}(\pm -+) - \text{Prob}(+\pm+) \geq 0. \tag{70}$$

Let us now investigate the question whether quantum mechanics is consistent with Bell's inequality. Without loss of generality we can assume that the angle α is zero. We denote by

$$N(++\pm) = N(\alpha = +, \beta = +, \gamma = \pm) \tag{71}$$

the number of measurements of the first experiment with the result that the left calcite shows $\alpha = +$, the right calcite shows $\beta = +$, and $\gamma = \pm$ is undefined. In the same way the other numbers, such as $N(+\pm-) = N(\alpha = +, \beta = \pm, \gamma = -)$, are defined.

From the law of Malus it follows immediately that the probability of the experiment with outcomes $\alpha = +$, $\beta = +$, and $\gamma = \pm$, is $\cos^2(\beta)$ since $\alpha = 0$. For the experiment with $\alpha = \pm$, $\beta = +$, and $\gamma = +$ the probability is $\cos^2(\beta - \gamma)$. Therefore, for $\alpha = \pm$, $\beta = -$ and $\gamma = +$ we obtain the complementary probability $\sin^2(\beta - \gamma)$. Finally, the probability for the experiment with $\alpha = +$, $\beta = \pm$ and $\gamma = +$ is $\cos^2(\gamma)$. From inequality (70) we get

$$\cos^2(\beta) + \sin^2(\beta - \gamma) - \cos^2(\gamma) \geq 0. \tag{72}$$

This inequality must be fulfilled for all angles β and γ. Otherwise, any *hidden-variable theory* that satisfies the classical rules of probability theory would not reproduce the quantum law of Born and Malus in optics.

It is convenient to choose the angles $\beta = 3\gamma$, then inequality (73) becomes

$$\cos^2(3\gamma) + \sin^2(2\gamma) - \cos^2(\gamma) \geq 0. \tag{73}$$

But the function on the left hand side is negative for the angles γ between 0 and $\pi/6$. Hence, Bell's Theorem is violated, and hidden variable theories are invalid. The correct quantum predictions for polarization experiments with photon pairs are clearly supported.

Keep in mind: Bell's Theorem excludes specific hidden variable theories and confirm quantum mechanics.

2.14 Basic Rules of Quantum Mechanics

Now having in mind these experiments, we summarize the fundamental rules and conclusions of quantum mechanics that were elaborated mainly by Feynman.

In all these experiments we have assigned complex numbers to transitions or possibilities such as "a horizontally polarized photon passes a polarizing filter with transmission axis at an angle α", or "a photon passes both beam splitter in an interferometer and ends up in detector d_1", or "a particle moves on a path in phase space through the slits of the wall and is detected in position x". Hence, a *transition* is a change from an initial quantum state to a final one. In other words, it can be represented as a specific set of initial and final conditions. The complex numbers which are assigned to transitions are called *probability amplitudes*. We refer to transitions also as possibilities. Given any transition, a *path* is a sequence of possibilities that connects the initial and the final conditions. For example, "a particle moves on a path in phase space choosing the possibility slit 1 and is detected in position x". Another path is that of moving through slit 2.

We use here a slightly different notation compared to quantum mechanics. In the Feynman lecture Volume III[46] and in many other textbooks on quantum mechanics, *events* are what we call transitions. This notation of events in quantum mechanics is in conflict with events in classical probability theory. There, events are subsets of the sample space, the latter is defined as the set of all outcomes of an experiment. When we speak of events we mean the events of classical probability theory, and pairs of initial and final conditions are called *transitions*.

The probability amplitudes that are assigned to transitions satisfy the following fundamental rules:

1. *Rule of Born and Malus*: The probability of any transition is the squared magnitude of a complex probability amplitude.

2. *Addition rule*: If a transition occurs in several mutually exclusive alternatives, the probability amplitude for this transition is the sum of all probability amplitudes for the alternative paths. This rule is responsible for *interference*.

3. *Multiplication rule*: For transitions, that occur in a series of steps which happen independently, the probability amplitudes are multiplied for each of these steps.

4. *Action rule*: For any alternative path between initial and final condition there exists some action S_{path} that determines the related probability amplitude $e^{\frac{i}{\hbar}S_{path}}$.

These are the most fundamental rules of quantum mechanics. All other quantum mechanical rules emerge in a rather natural way. Moreover, these

[46] Feynman Lectures [1963, volume 3, 2005 ed., page 1–13]

rules are very reasonable in contrast to the principles of quantum mechanics presented in Section 1. A remarkable feature of Feynman's formulation is on the one hand its simplicity, and on the other hand its universal applicability. They were never falsified when correctly applied. This is in contrast to many other physical frameworks. Even strange claims last until now. For example, Dyson writes:

> *Thirty-one years ago [1948], Dick Feynman told me about his "sum over histories" version of quantum mechanics. "The electron does anything it likes," he said. "It just goes in any direction at any speed, forward or backward in time, however it likes, and then you add up the amplitudes and it gives you the wave-function." I said to him, "You're crazy." But he wasn't.*[47]

The addition and multiplication rule have a noticeable affinity to the rules of classical probability theory. They use the self-evident terms *mutually exclusive* and *independent*. Two events are mutually exclusive if they are disjoint in the sense that they cannot occur at the same time. When throwing a dice the events $\{1, 2, 3\}$ and $\{4, 5, 6\}$ are mutually exclusive, whereas the events $\{1, 2, 3, 4\}$ and $\{4, 5, 6\}$ occur at the same time when the dice shows a 4. If the occurrence of one event does not affect the probability of another event, then the events are independent. When throwing two dice both events are independent, except we have glued both dice together.

A striking feature of the Dirac-Feynman rules is the direct link to classical mechanics via the action rule. In particular, it allows the derivation of many physical theories, from Schrödinger's formulation of quantum mechanics to quantum electrodynamics. Classical theories like classical mechanics or Maxwell's theory of electromagnetism appear in a quantized form when using these rules. In classical mechanics, the state of a particle is determined by its position and its momentum, and the dynamics can be formulated in terms of the least action principle. When switching to quantum mechanics the deterministic change of position and momentum is no longer fulfilled. We can only calculate probabilities for switching between classical states. But both frameworks, the classical and the quantum one, are closely linked by the action. In particular, the geometry as well as symmetries in classical physics transform into quantum physics.

Keep in mind: These four Dirac-Feynman rules (Born's rule, addition and multiplication rule, and action probability amplitudes) are the fundamental cornerstones of quantum electrodynamics, our best known physical theory that describes all physical phenomena, except gravitation and radioactive phenomena.

Feynman[48]

Perhaps, these rules could be viewed not only as a concrete physical theory in itself, but more as an abstract framework how to obtain specific theories. In

[47] `https://en.wikiquote.org/wiki/Freeman_Dyson`
[48] Feynman [1985, page 8]

other words, these are metaphysical rules of an *abstract quantum theory*. We use the word "metaphysical" in the sense of physical rules that deal with first fundamental principles outside of concrete theories.

3 Introduction to Quantum Information Theory

Quantum computing and quantum information theory was pioneered mainly by Feynman[49] 1982. His idea was to build computers with the use of quantum mechanics in hopes of achieving more efficiency compared with classical computers. In particular, he tried to utilize the exponential *parallelism* that is inherent in quantum mechanics. In the meantime this speculation seems to be justified. The most spectacular success is *Shor's factoring algorithm* which factors on a quantum computer each n-digit integer with computational costs of order n^2. In contrast, the best known algorithm executed on a classical computer requires exponential time. The exponential speed of Shor's algorithm would have dramatic consequences in *cryptography*.

The major goal of this section is to guide engineers and other scientists to quantum computing, starting with classical computing via reversible computing and probabilistic computing.

[49]Feynman, Richard P. "Simulating physics with computers." International journal of theoretical physics 21.6 (1982): 467-488., Feynman, Richard P. "Quantum mechanical computers." Foundations of physics 16.6 (1986): 507-531.

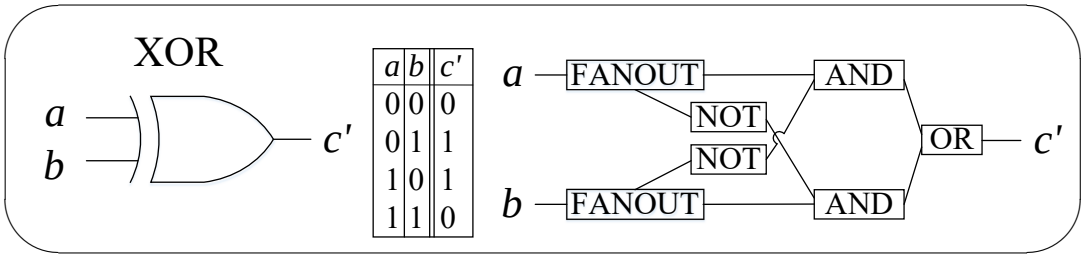


Figure 38: On the left hand side the symbol for the XOR function $\mathrm{XOR}(a,b) = a \oplus b$ is displayed, together with its truth table in the middle. The network representing XOR, using the universal set of gates {AND, OR, NOT, FANOUT}, is given on the right hand side. The lines in the diagram are the wires that carry the bits. The rectangles represent the logical gates.

3.1 Classical Boolean Circuits

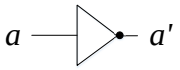
In computer science the most basic information unit is a *bit*. A bit is described as a quantity that can assume only two values or states, either '0' or '1'. Physically, it can be represented as a *wire* carrying one of two different voltages. Information is stored on classical computers in terms of bit strings of any length n, called *register*. In the Boolean circuit model logical *gates*, like AND, OR, NOT, or NAND are used for computational tasks. *Circuits* are networks composed of *wires* which carry the bit values, and gates which operate on the bits. This Boolean circuit model is frequently used in the classical theory of computation.

Computational tasks as well as all manipulations of information can be modeled in terms of *Boolean functions* $f : \{0,1\}^n \to \{0,1\}^m$, where $x = (x_1, ..., x_n) \in \{0,1\}^n$ denotes a vector with n components that are either 0 or 1. It is a well-known fact that any Boolean function is computable by a circuit using only AND, OR, NOT and FANOUT gates. Therefore, this set of gates is called *universal*. Some classical gates together with their truth tables are displayed in Table 2. Another universal set of gates is FANOUT and the NAND gate, the latter is defined as $NAND(a,b) = NOT(AND(a,b))$.

For illustration purposes we have displayed a circuit in Figure 38 that computes the XOR Boolean function, and the CNOT gate (*controlled-NOT*) in Figure 39. The latter has two input bits and two output bits. The first bit, the control bit, is always passed unchanged, the second bit switches its value if and only if the first control bit has value one.

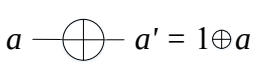
Classical gates	Truth tables	Reversible gates

NOT

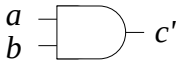


a	a'
0	1
1	0

NOT



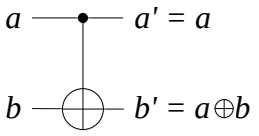
AND



a	b	c'
0	0	0
0	1	0
1	0	0
1	1	1

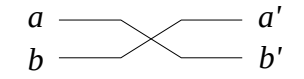
a	b	a'	b'
0	0	0	0
0	1	0	1
1	0	1	1
1	1	1	0

CNOT

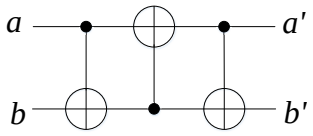


a	b	a'	b'
0	0	0	0
0	1	1	0
1	0	0	1
1	1	1	1

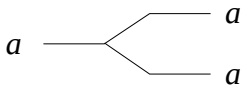
EXCHANGE (SWAP)



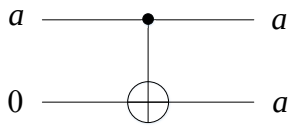
EXCHANGE (SWAP)



FANOUT



FANOUT



a	b	c	a'	b'	c'
0	0	0	0	0	0
0	0	1	0	0	1
0	1	0	0	1	0
0	1	1	0	1	1
1	0	0	1	0	0
1	0	1	1	0	1
1	1	0	1	1	1
1	1	1	1	1	0

CCNOT (Toffoli gate)

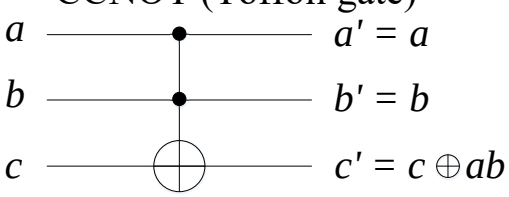


Table 2: Frequently used gates with truth tables. There, $a \oplus b$ means binary addition or equivalently the XOR operation, whereas ab is the binary multiplication that corresponds to AND.

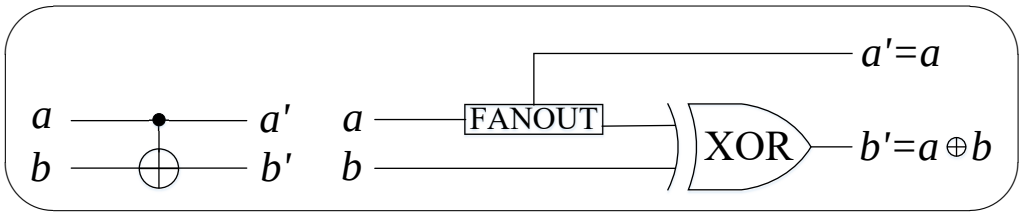


Figure 39: On the left hand side the symbol for the CNOT function is displayed, followed by a network representing CNOT via the FANOUT and XOR functions. The second bit switches its value if and only if the first bit has the value one. The symbol $\oplus$ denotes binary addition.

3.2 Reversible Computation

In computers more and more gates and circuits are packed into smaller and smaller volumes. This requires more energy and may cause overheating. Thus, it is an important question how energy-saving computers can be built. This problem and similar issues were investigated since the 1960s, mainly by Landauer and Bennet. They asked whether there are Boolean circuits that are both universal and reversible, since irreversible gates must dissipate some energy in contrast to reversible ones. For instance, the AND gate maps the information of two input bits with one of four values 00, 01, 10, or 11 into one output bit that has one of two values 0 or 1. Hence, the disorder increases, the entropy change is $\ln(2)$ units and generates, according to thermodynamics, a heat of $kT\ln(2)$ at temperature T, where k is the Boltzman constant. A Boolean gate is said be *reversible* if it has the same number of input bits and output bits, hence forms a bijection. Especially, a reversible gate does not change the entropy. Although it turned out later that energy dissipation is not a major practical problem, reversible computation is an excellent preparation for the quantum circuit model.

In the following we describe three reversible gates that can be used to implement a universal reversible machine. Obviously the NOT gate, see Table 2, is reversible. It has one input bit, one output bit, and acting twice with NOT returns the identity.

The next one is the CNOT gate, speak controlled-Not. There are two input bits a and b and two output bits a' and b', see Figure 39. The input bit on the control wire does not change (a'=a), and the second input bit is changed if and only if the first input bit is 1. CNOT is a bijection between input and output, and thus is reversible. This gate can simulate FANOUT and EXCHANGE as shown in Table 2.

An important generalization of CNOT is CCNOT, speak controlled-controlled NOT. It is displayed in Table 2. It has three input bits and three output bits, the first two inputs are passed directly, and the third input bit is negated if and only if the first two control inputs are both equal to 1.

These three reversible gates have the nice property that they are their own inverses, that is, performing them twice consecutively restores the input. The gates NOT and CNOT alone are insufficient to generate each Boolean function. But the CCNOT gate is very flexible. In Figure 40 it is shown

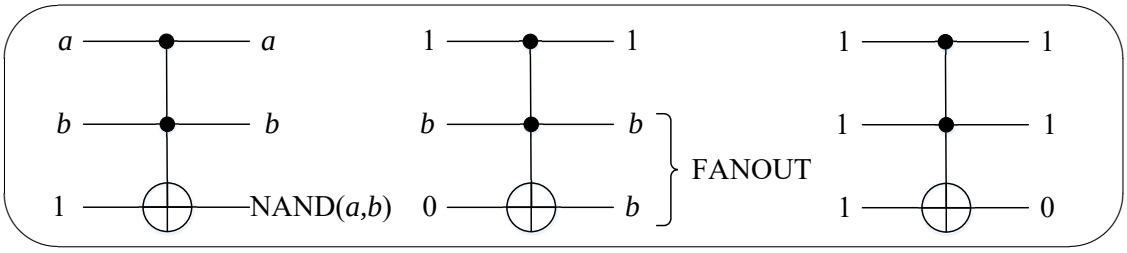


Figure 40: CCNOT can simulate NAND, FANOUT, and generate 0's from 1's. Notice that simulations may produce unneeded ancilla bits. These are a and b for NAND, the top 1 in FANOUT, and two 1's of the first two wires in the third gate.

how CCNOT simulates NAND, FANOUT, and generates on the third wire a 0 from 1. Thus CCNOT is universal, and any Boolean circuit can be transformed into a reversible one using CCNOT gates only. As a reminder for these three important reversible gates we have displayed them with their common features together in Figure 41.

We have seen above that reversible simulations produce undesirable *ancilla bits*. For example in the simulation of FANOUT with CCNOT the first wire is not used. However, any Boolean function $f : \{0,1\}^n \to \{0,1\}^m$ can be realized as a reversible circuit using at most $m+n$ wires, starting with the input bits and m additional bits being 0. This follows immediately from the reversible function

$$\widetilde{f} : \{0,1\}^{m+n} \to \{0,1\}^{m+n},\ (a,b) \to (a, b \oplus f(a)), \tag{74}$$

where a has n and b has m components. If we set $b = 0$ then we obtain the values $f(a)$ on the last m bits. It is a reversible circuit, since

$$\widetilde{f}\widetilde{f}(a,b) = \widetilde{f}(a, b \oplus f(a)) = (a, (b \oplus f(a)) \oplus f(a))) = (a,b). \tag{75}$$

Hence, $\widetilde{f}$ is its own inverse just as NOT, CNOT and CCNOT.

As a small application we have displayed in Figures 42 and 43 reversible circuits of a half adder and a full adder, respectively. These figures show the characteristic structure of reversible circuits as *acyclic networks*.

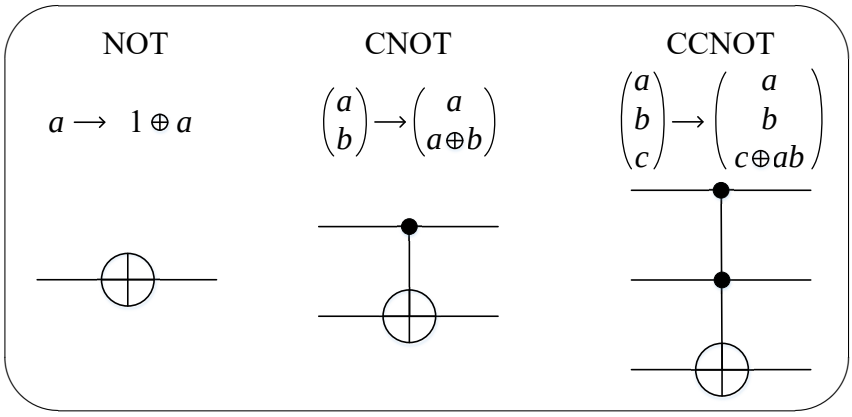


Figure 41: The three fundamental, reversible gates NOT, CNOT, CCNOT and their common features. Acting twice, these gates return the identity.

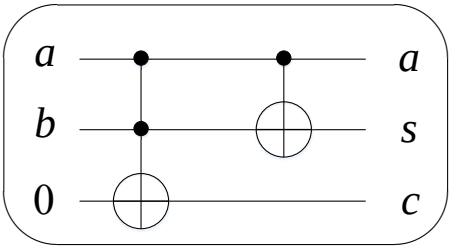


Figure 42: This half adder adds the bits a and b with sum s on wire b and the carry on the third wire.

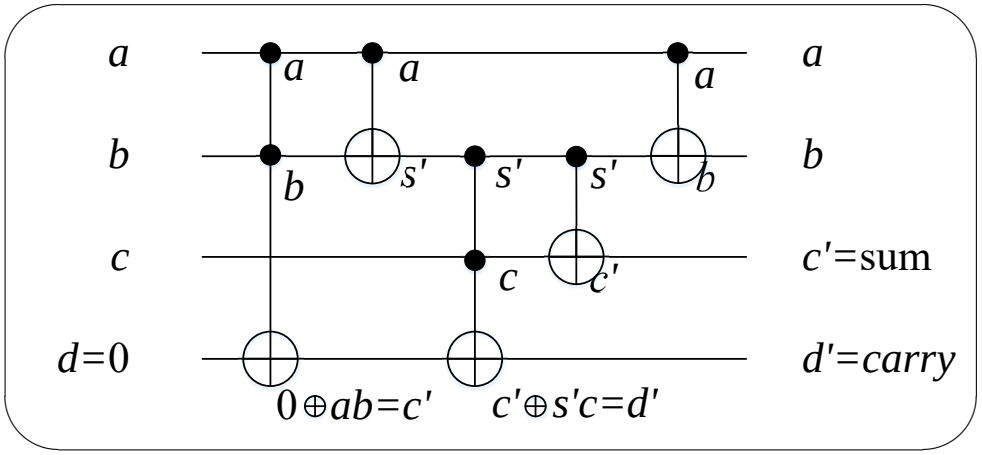


Figure 43: This machine represents a full adder. It adds the sum of the bits on the two wires a and b, yielding the sum on line b with value s' and the carry on the fourth line d with value c', see Figure 42. Then it adds s' to c, which is the carry from any previous addition. Then we get the final sum c' and the final carry d'. Now we apply the CNOT gate twice. This restores the input line b. In summary, the first two lines contain the input, the other two lines contain the output.

3.3 The Linear Algebra Formalism of Reversible Computation

In this section we present classical reversible computation in terms of a formalism that is completely based on linear algebra. In quantum mechanics the theory of linear algebra is described by using *Dirac's "bra-ket" notation*: each vector in a linear space is written in the form

$$|x\rangle \tag{76}$$

where x is a label for the vector, and the notation $|\cdot\rangle$ denotes a column vector called "ket". The conjugate transpose of this vector is written as

$$\langle x|, \tag{77}$$

and called "bra". It follows that the inner product of two vectors $|x\rangle$ and $|y\rangle$

$$\langle x|y\rangle = \sum_i x_i^* y_i \tag{78}$$

is represented as a "bracket". It consists of the bra part $\langle x|$ and the ket part $|y\rangle$.

It is very natural to think of a classical bit as a two-dimensional vector

$$|\psi\rangle = \psi_0|0\rangle + \psi_1|1\rangle, \quad \psi_0, \psi_1 \in \{0,1\}, \quad \psi_0 + \psi_1 = 1, \tag{79}$$

such that the bit value 0 is represented in the form

$$|\psi\rangle = |0\rangle \quad \Leftrightarrow \quad |\psi\rangle = \begin{pmatrix}\psi_0\\ \psi_1\end{pmatrix} = \begin{pmatrix}1\\ 0\end{pmatrix}, \tag{80}$$

and the bit value 1 is written as

$$|\psi\rangle = |1\rangle \quad \Leftrightarrow \quad |\psi\rangle = \begin{pmatrix}\psi_0\\ \psi_1\end{pmatrix} = \begin{pmatrix}0\\ 1\end{pmatrix}. \tag{81}$$

This is the *vector representation* of a bit.

A string of two bits allows the states 00, 01, 10, and 11. We can represent four states as four-dimensional vectors

$$\begin{aligned} |\xi\rangle &= \xi_{00}|00\rangle + \xi_{01}|01\rangle + \xi_{10}|10\rangle + \xi_{11}|11\rangle, \\ &\xi_{ij} \in \{0,1\}, \sum_{i,j=0}^{1} \xi_{ij} = 1. \end{aligned} \tag{82}$$

Thus exactly one ξ_{ij} is equal to 1, the other ones are 0, and it follows that

$$|00\rangle = \begin{pmatrix}1\\0\\0\\0\end{pmatrix}, |01\rangle = \begin{pmatrix}0\\1\\0\\0\end{pmatrix}, |10\rangle = \begin{pmatrix}0\\0\\1\\0\end{pmatrix}, |11\rangle = \begin{pmatrix}0\\0\\0\\1\end{pmatrix}. \tag{83}$$

Notice that these are the four-dimensional canonical orthonormal unit vectors in the complex vector space $\mathbb{C}^4$. If we write both bits in the form (79), that is,

$$\begin{array}{c} |\psi\rangle = \psi_0|0\rangle + \psi_1|1\rangle, \quad |\varphi\rangle = \varphi_0|0\rangle + \varphi_1|1\rangle, \\ \psi_0, \psi_1, \varphi_0, \varphi_1 \in \{1, 0\}, \quad \psi_0 + \psi_1 = 1, \quad \varphi_0 + \varphi_1 = 1, \end{array} \tag{84}$$

then comparing with (82) we obtain

$$\xi_{00} = \psi_0\varphi_0, \xi_{01} = \psi_0\varphi_1, \xi_{10} = \psi_1\varphi_0, \xi_{11} = \psi_1\varphi_1. \tag{85}$$

Hence, the coefficients ξ_{ij} are just the products of the coefficients ψ_i and φ_i.

Products between quantities like vectors or matrices play an important role in physics and mathematics. An important product is the *Kronecker product* of two matrices of arbitrary size. For an $m \times n$ matrix $\hat{A} \in \mathbb{C}^{m\times n}$ and a $p \times q$ matrix $\hat{B} \in \mathbb{C}^{p\times q}$ the Kronecker product takes the form

$$\hat{A} \otimes \hat{B} = \begin{pmatrix} A_{11}\hat{B} & \dots & A_{1n}\hat{B} \\ \vdots & & \vdots \\ A_{m1}\hat{B} & \dots & A_{mn}\hat{B} \end{pmatrix}. \tag{86}$$

For example, if

$$\hat{A} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{pmatrix}, \quad \hat{B} = \begin{pmatrix} B_{11} & B_{12} \end{pmatrix}, \tag{87}$$

then

$$\hat{A} \otimes \hat{B} = \begin{pmatrix} A_{11}B_{11} & A_{11}B_{12} & A_{12}B_{11} & A_{12}B_{12} \\ A_{21}B_{11} & A_{21}B_{12} & A_{22}B_{11} & A_{22}B_{12} \\ A_{31}B_{11} & A_{31}B_{12} & A_{32}B_{11} & A_{32}B_{12} \end{pmatrix}. \tag{88}$$

In other words, in the Kronecker product each coefficient of $\hat{A}$ is multiplied with all other coefficients of $\hat{B}$. The Kronecker product has many nice properties, among them:

$$\begin{array}{c} \hat{A}, \hat{B} \text{ invertible} \quad \Leftrightarrow \quad \hat{A} \otimes \hat{B} \text{ invertible, and} \\ (\hat{A} \otimes \hat{B})^{-1} = \hat{A}^{-1} \otimes \hat{B}^{-1}. \end{array} \tag{89}$$

The Kronecker product is a special case of the *tensor product* which applies not only to matrices, but in a more general context to vector spaces, algebras and tensors. Therefore, the operation (86) is also called tensor product in many textbooks, and we use this name in most cases.

Using the notation (86) of the tensor product, we can rewrite (83) as tensor products of vectors:

$$\begin{aligned} |00\rangle &= \tbinom{1}{0} \otimes \tbinom{1}{0} = (1\,0\,0\,0)^T, \\ |01\rangle &= \tbinom{1}{0} \otimes \tbinom{0}{1} = (0\,1\,0\,0)^T, \\ |10\rangle &= \tbinom{0}{1} \otimes \tbinom{1}{0} = (0\,0\,1\,0)^T, \\ |11\rangle &= \tbinom{0}{1} \otimes \tbinom{0}{1} = (0\,0\,0\,1)^T. \end{aligned} \tag{90}$$

A *register* is a collection of n bits $|\psi^k\rangle = \binom{\psi_0^k}{\psi_1^k}$, $\psi_i^k \in \{0,1\}$ for $i = 0,1$ and $k = 0,\ldots,n-1$. In the same way as for $n = 2$ bits we represent a register as a vector

$$\begin{aligned} |\xi\rangle &= |\psi^0 \ldots \psi^k \ldots \psi^{n-1}\rangle \\ &= |\psi^0\rangle \otimes \ldots \otimes |\psi^k\rangle \otimes \ldots \otimes |\psi^{n-1}\rangle \\ &= \tbinom{\psi_0^0}{\psi_1^0} \otimes \ldots \otimes \tbinom{\psi_0^k}{\psi_1^k} \otimes \ldots \otimes \tbinom{\psi_0^{n-1}}{\psi_1^{n-1}} \\ &= \begin{pmatrix} \vdots \\ \psi_{i_0}^0 \cdot \ldots \cdot \psi_{i_k}^k \cdot \ldots \cdot \psi_{i_{n-1}}^{n-1} \\ \vdots \end{pmatrix}. \end{aligned} \tag{91}$$

This is the vector representation of a register with 2^n components, where exactly one component has the value 1 and the others are 0.

Let us consider an example with three bits. In this case we obtain a vector with $2^3 = 8$ components:

$$\begin{aligned} |\xi\rangle &= |0\ 1\ 0\rangle \\ &= \tbinom{1}{0} \otimes \tbinom{0}{1} \otimes \tbinom{1}{0} \\ &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \\ &= (0\,0\,1\,0\,0\,0\,0\,0)^T. \end{aligned} \tag{92}$$

A register of three bits, where each bit has the value 1, can be written as

$$\begin{aligned} |\xi\rangle &= |1\ 1\ 1\rangle \\ &= \tbinom{0}{1} \otimes \tbinom{0}{1} \otimes \tbinom{0}{1} \\ &= \begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \\ &= (0\,0\,0\,0\,0\,0\,0\,1)^T. \end{aligned} \tag{93}$$

Summarizing, each register of n bits can be represented as a canonical unit vector of $\mathbb{C}^{2^n}$. We call this process of identifying vectors with registers *vectorization*. In the following, when we speak of registers, we distinguish not between the original definition of a register and its equivalent vector form. This will cause no confusion.

Vice versa, the orthonormal basis of canonical unit vectors in the space $\mathbb{C}^{2^n}$ represents all bit configurations of a register. Formally, we can write (91) as a *superposition*

$$|\xi\rangle = \sum_{i_0,\ldots,i_k,\ldots,i_{n-1}=0}^{1} \psi_{i_0}^0 \cdot \ldots \cdot \psi_{i_k}^k \cdot \ldots \cdot \psi_{i_{n-1}}^{n-1} |i_0 \ldots i_k \ldots i_{n-1}\rangle. \tag{94}$$

Of course, this superposition is trivial since exactly one coefficient is 1 and the others are 0. But this notation helps us to understand the more complicated random registers later.

In computing it is sometimes useful to allow in addition an *empty register*, which is represented as the zero vector 0 in $\mathbb{C}^{2^n}$. This vector should not be mixed up with the register $|0\dots0\rangle$ where each bit has the value 0. Obviously, the latter register is the unit vector with a 1 in the first component, and zero otherwise. The empty register is important when defining creation and annihilation operators. We will consider these operators later.

We know already that any logical reversible gate or circuit transforms a register to another register, and hence a canonical unit vector onto any other canonical unit vector. Therefore, only the components of vectors are permuted, and hence reversible circuits can be described by *permutation matrices* $\hat{U}$. These are matrices that have in each row and column exactly one 1, and zero coefficients otherwise. In other words, a reversible gate encodes a specification how to permute 2^n possible bit strings. From a mathematical point of view we can write

$$|\xi'\rangle = \hat{U}|\xi\rangle, \quad \xi'_j, \xi_j, U_{ij} \in \{0,1\}, \quad \hat{U} = (U_{ij}) \quad \text{permutation matrix}, \tag{95}$$

and $\hat{U}|\xi\rangle$ is the matrix vector multiplication.

It is now easy to write down the matrix representations of the previously defined gates that are displayed in Figures 44, 45, 46, and 47.

By matrix-vector multiplication, for instance, we obtain as expected

$$\begin{aligned} \text{CCNOT}(|110\rangle) &= \begin{pmatrix} 1&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0\\ 0&0&0&1&0&0&0&0\\ 0&0&0&0&1&0&0&0\\ 0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&1\\ 0&0&0&0&0&0&1&0 \end{pmatrix} \cdot \begin{pmatrix} 0\\0\\0\\0\\0\\0\\1\\0 \end{pmatrix} \\ &= \begin{pmatrix} 0\\0\\0\\0\\0\\0\\0\\1 \end{pmatrix} = |111\rangle. \end{aligned} \tag{96}$$

$$\text{NOT} = \begin{array}{c|c} & \begin{array}{cc} 0 & 1 \end{array} \\ \hline \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{array}$$

Figure 44: Matrix representation of NOT.

$$\text{CNOT} = \begin{array}{c|c} & \begin{array}{cccc} 00 & 01 & 10 & 11 \end{array} \\ \hline \begin{array}{c} 00 \\ 01 \\ 10 \\ 11 \end{array} & \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

Figure 45: Matrix representation of CNOT.

$$\text{CCNOT} = \begin{array}{c|c} & \begin{array}{cccccccc} 000 & 001 & 010 & 011 & 100 & 101 & 110 & 111 \end{array} \\ \hline \begin{array}{c} 000 \\ 001 \\ 010 \\ 011 \\ 100 \\ 101 \\ 110 \\ 111 \end{array} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

Figure 46: Matrix representation of CCNOT.

$$\text{SWAP} = \begin{array}{c|c} & \begin{array}{cccc} 00 & 01 & 10 & 11 \end{array} \\ \hline \begin{array}{c} 00 \\ 01 \\ 10 \\ 11 \end{array} & \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

Figure 47: Matrix representation of SWAP.

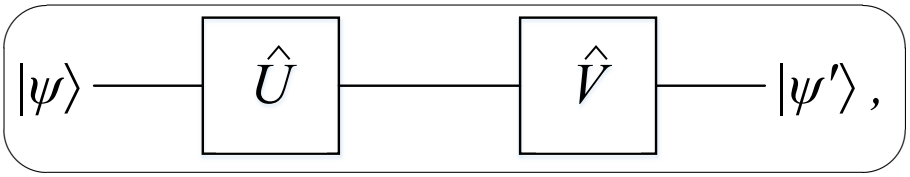


Figure 48: Two gates in series.

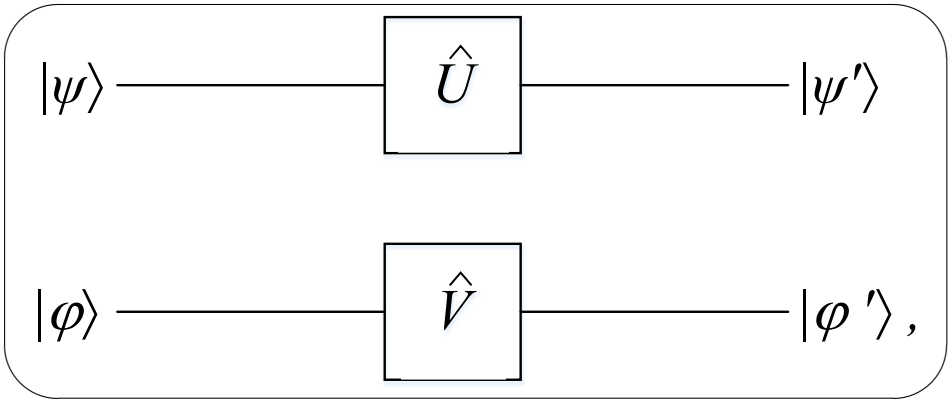


Figure 49: Two gates parallel.

3.4 Composition of Gates

Given two reversible classical gates described by the permutation matrices $\hat{U}$ and $\hat{V}$, we can arrange them in a *series* consecutively, *parallel* or as a *controlled gate*.

In the first case in Figure 48, when gate $\hat{U}$ is followed by $\hat{V}$, we obtain from (95) the register $\hat{U}|\psi\rangle$ by matrix-vector multiplication, and then

$$|\psi'\rangle = \hat{V}(\hat{U}|\psi\rangle) = (\hat{V}\cdot\hat{U})|\psi\rangle. \tag{97}$$

From linear algebra the identity above is well-known. Hence, the matrix representation of the gate corresponding to gates in series is the matrix product. We notice that the product of permutation matrices is a permutation matrix, thus representing a reversible classical gate.

It is useful to realize that in the matrix product the *Feynman's rules* are

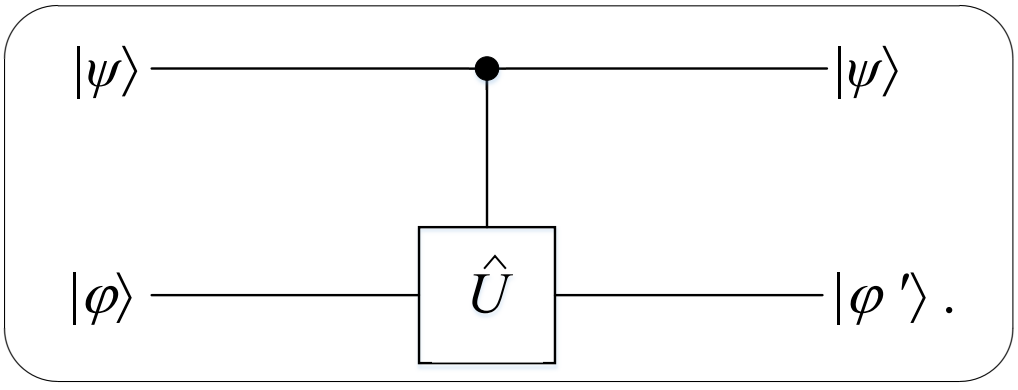
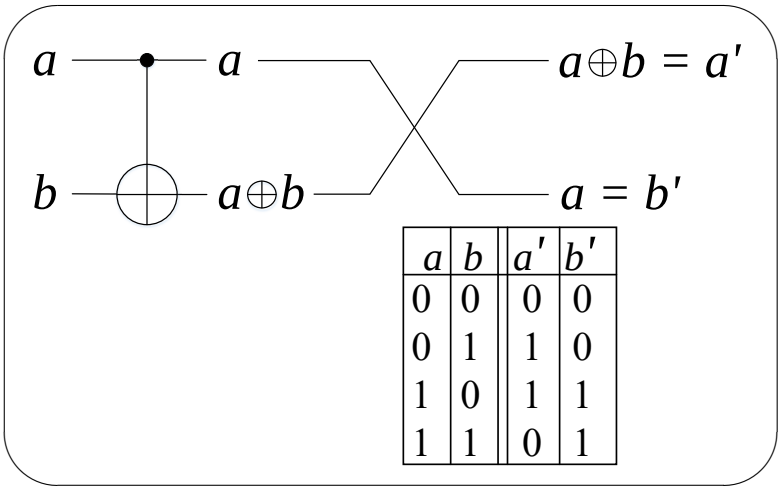


Figure 50: $\hat{U}$ as a controlled gate.

hidden. The input ψ_i evolves to the output ψ'_j via the intermediate values $V_{jk}U_{ki}$. The deterministic values 0 and 1 can be viewed as extreme probabilities. Both gates are independent, so the (extreme) probabilities zero and one are multiplied. The transition occurs in mutually exclusive ways k, and we have to sum up over all possible values k. Hence, matrix-matrix multiplication reflects Feynman's rules of multiplying probabilities for independent ways, and adding probabilities for mutually exclusive ways.

Let us consider the example of the gate where at first CNOT is applied, and then SWAP, see Figure 51.



a	b	a'	b'
0	0	0	0
0	1	1	0
1	0	1	1
1	1	0	1

Figure 51: This gate is the CNOT gate followed by the SWAP gate, and its truth table.

If we multiply the SWAP matrix with CNOT, then we get

$$\text{SWAP} \cdot \text{CNOT} = \text{SC} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{98}$$

Since

$$\begin{aligned}
\mathrm{SC}|00\rangle &= \mathrm{SC}\begin{pmatrix}1\\0\\0\\0\end{pmatrix} = \begin{pmatrix}1\\0\\0\\0\end{pmatrix} = |00\rangle,\\
\mathrm{SC}|01\rangle &= \mathrm{SC}\begin{pmatrix}0\\1\\0\\0\end{pmatrix} = \begin{pmatrix}0\\0\\1\\0\end{pmatrix} = |10\rangle,\\
\mathrm{SC}|10\rangle &= \mathrm{SC}\begin{pmatrix}0\\0\\1\\0\end{pmatrix} = \begin{pmatrix}0\\0\\0\\1\end{pmatrix} = |11\rangle,\\
\mathrm{SC}|11\rangle &= \mathrm{SC}\begin{pmatrix}0\\0\\0\\1\end{pmatrix} = \begin{pmatrix}0\\1\\0\\0\end{pmatrix} = |01\rangle,
\end{aligned} \tag{99}$$

the SC matrix corresponds to the truth table in Figure 51.

Now we investigate the second case where gates are arranged in parallel, see Figure 49. There we must bring together registers with m and n bits forming a new register with $m+n$ bits. Let the gate $\hat{U}$ act on the first m bits $|\psi\rangle$, and let $\hat{V}$ act on the remaining n bits $|\varphi\rangle$.

In (91) and (94) we have shown how registers are constructed using tensor products. It follows that the new register $|\xi\rangle$ with $m+n$ bits can be represented as the tensor product of the two old registers. Let the old registers be

$$|\psi\rangle = \sum_i \psi_i |i\rangle, \quad |\varphi\rangle = \sum_j \varphi_j |j\rangle, \tag{100}$$

where $i = (i_0, \ldots, i_{m-1}), j = (j_0, \ldots, j_{n-1})$ are the registers corresponding to the bit values $i_k, j_k \in \{0, 1\}$.

Then their tensor product has the form

$$|\xi\rangle = |\psi\rangle \otimes |\varphi\rangle = \sum_{i,j} \psi_i \varphi_j |ij\rangle, \tag{101}$$

where

$$|ij\rangle = |i_0 \cdots i_{m-1} j_0 \cdots j_{n-1}\rangle. \tag{102}$$

is a canonical unit vector with 2^{m+n} components. Since the gates $\hat{U}$ and $\hat{V}$ act parallel and independent on $|\psi\rangle$ and $|\varphi\rangle$, respectively, the combined parallel output is the tensor product

$$(\hat{U}|\psi\rangle) \otimes (\hat{V}|\varphi\rangle). \tag{103}$$

It is a well-known fact and a good exercise that

$$(\hat{U}|\psi\rangle) \otimes (\hat{V}|\varphi\rangle) = (\hat{U} \otimes \hat{V})(|\psi\rangle \otimes |\varphi\rangle), \tag{104}$$

proving that gates in parallel are described by their tensor product, that is,

$$|\xi'\rangle = (\hat{U} \otimes \hat{V})|\xi\rangle. \tag{105}$$

Please note that the tensor product of permutation matrices is a permutation matrix, hence representing a classical gate.

Since

$$(\hat{U} \otimes \hat{V})_{kl,ij} = \hat{U}_{ki} \cdot \hat{V}_{\ell j}, \tag{106}$$

formula (105) means that a combined input $\psi_i \varphi_j$ is transformed into a combined output $\psi'_k \varphi'_\ell$ with the coefficient $\hat{U}_{ki}\hat{V}_{\ell j}$. Both gates $\hat{U}$ and $\hat{V}$ are independent, and the latter product can be interpreted as the product of the extreme probabilities zero and one. This is the product rule for independent extreme probabilities. For example, let $\hat{U}$ = SWAP and $\hat{V}$ = NOT be combined in parallel, then

$$\begin{aligned} \text{SWAP} \otimes \text{NOT} &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}. \end{aligned} \tag{107}$$

This is a permutation matrix, and it is easy to see that the truth table of SWAP ⊗ NOT implies the permutation matrix on the right hand side of (107), and vice versa.

Finally, we consider how controlled reversible gates can be modelled with the *direct sum* of matrices, which is defined as

$$\hat{U} \oplus \hat{V} = \begin{pmatrix} \hat{U} & \hat{0}_{m\times q} \\ \hat{0}_{p\times n} & \hat{V} \end{pmatrix}. \tag{108}$$

$$\mathrm{CNOT} = \hat{1} \oplus \mathrm{NOT} = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{array}\right).$$

Figure 52: CNOT as the direct sum of the identity and NOT.

$$\mathrm{CCNOT} = \hat{1} \oplus \mathrm{CNOT} = \left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right),$$

Figure 53: CCNOT as the direct sum of the identity and CNOT.

There $\hat{U}$ is an $m \times n$ matrix, $\hat{V}$ is a $p \times q$ matrix, and $\hat{0}$ denotes a zero matrices with appropriate dimensions. The direct sum implements an "if then else" operation, also called a *controlled operation*.

The most important example is CNOT, see Figure 41 and 52. The universal gate CCNOT is the direct sum of the identity and CNOT. More generally, any gate $\hat{U}$ can be implemented in a controlled gate $C\hat{U}$ which is defined as

$$C\hat{U} = \hat{1} \oplus \hat{U} = \begin{pmatrix} \hat{1} & \hat{0} \\ \hat{0} & \hat{U} \end{pmatrix}. \tag{109}$$

To summarize, we introduced three basic rules (matrix product, tensor product, and direct sum) that compose new reversible gates from given ones and compute the related permutation matrices.

3.5 Randomized Computation

Randomized computation refers to the capability to operate with random bits, shortly called *rbits*,

$$\begin{aligned} &|\psi\rangle = \psi_0|0\rangle + \psi_1|1\rangle = \binom{\psi_0}{\psi_1}, \\ &\psi_0, \psi_1 \geq 0, \quad \psi_0 + \psi_1 = 1, \end{aligned} \tag{110}$$

instead of bits as defined in (79). The only difference is that the zero-one coefficients are replaced by classical probabilities ψ_0 for state $|0\rangle = \binom{1}{0}$ and ψ_1 for state $|1\rangle = \binom{0}{1}$. We can think of a coin, fair whenever $\psi_0 = \psi_1 = 1/2$, and unfair otherwise. From a mathematical point of view, rbits are convex combinations of the two bit values 0 and 1.

There are computational problems which can be solved efficiently using randomized computation, although no efficient or tractable deterministic algorithms are known. We will not go into the details and the applications of randomized computing. But we show that it is very easy to upgrade the reversible, deterministic circuit model to a random circuit model of computation. We must only replace the extreme bit values 0 and 1 by convex combinations. There are two major differences to reversible computing:

- Randomized computing is *not reversible*.
- *Entangled* registers can occur.

Nevertheless, with the background of reversible computing we can quickly derive the essential distinctions.

A string of two rbits allows the states 00, 01, 10, and 11 with probabilities ξ_{00}, ξ_{01}, ξ_{10}, ξ_{11} yielding a four-dimensional vector

$$\begin{aligned} |\xi\rangle \;&=\; \xi_{00}|00\rangle + \xi_{01}|01\rangle + \xi_{10}|10\rangle + \xi_{11}|11\rangle, \\ &\quad\; \xi_{ij} \geq 0, \quad \sum_{i,j=0}^{1} \xi_{ij} = 1. \end{aligned} \tag{111}$$

The difference to (82) is the replacement of the condition $\xi_{ij} \in \{0, 1\}$ by a convex combination.

For two independent rbits $|\varphi\rangle$ and $|\psi\rangle$, both defined as in (110), the probabilities must be multiplied yielding the tensor product

$$|\xi\rangle = |\psi\rangle \otimes |\varphi\rangle = \begin{pmatrix} \psi_0\varphi_0 \\ \psi_0\varphi_1 \\ \psi_1\varphi_0 \\ \psi_1\varphi_1 \end{pmatrix}. \tag{112}$$

In fact, the coefficients are nonnegative, and

$$\sum_{i,j=0}^{1} \psi_i\varphi_i = (\psi_0 + \psi_1)(\varphi_0 + \varphi_1) = 1. \tag{113}$$

Thus they represent probabilities. These states are called *product states*. We can think of two coins that are thrown independently.

The definition (111) enables states that are not product states, but are correlated. An example is the state

$$|\xi\rangle = \frac{1}{2}|00\rangle + \frac{1}{2}|11\rangle. \tag{114}$$

This is not a product state, since the equations

$$\psi_0\varphi_0 = \frac{1}{2}, \quad \psi_0\varphi_1 = 0, \quad \psi_1\varphi_0 = 0, \quad \psi_1\varphi_1 = \frac{1}{2} \tag{115}$$

are contradictory. States that are not product states are called *correlated* or *entangled*. The entangled state (114) can be physically realized as two welded coins, such that either both coins show Head or both show Tail in each case with probability 1/2.

Product states are exceptional. Typical in our world are entangled or correlated situations that reflect the various randomized interactions between physical objects. So not surprisingly, entanglement is of particular importance in quantum information theory. It is useful to keep in mind that entanglement does not occur in deterministic reversible or irreversible computing. The reason is that exactly one of the coefficients is one, all others are zero, and thus the related equations (115) are always solvable. Entanglement is a fundamental property of stochastic physical models.

It is simple to generalize registers of n binary bits, described in (91) and (94), to registers of n randomized bits. Let

$$i = (i_0 \dots i_k \dots i_{n-1}) \tag{116}$$

denote a register in form of a multiindex with $i_k \in \{0, 1\}$, and let

$$|i\rangle = |i_0 \dots i_k \dots i_{n-1}\rangle \tag{117}$$

be the related canonical unit vector with 2^n components, as already defined in the deterministic reversible circuit model. We mention again that sometimes m do not distinguish between the register form and its equivalent vector representation. This will cause no confusion.

A *random register* $|\xi\rangle$ is a superposition

$$|\xi\rangle = \sum_i \xi_i |i\rangle, \quad \xi_i \geq 0, \ \sum_i \xi_i = 1, \tag{118}$$

with base states $|i\rangle$. This superposition is interpreted as follows: with probability ξ_i the register is found in state $|i\rangle$.

We have seen that in reversible computation gates are described in (95) by permutation matrices, since canonical unit vectors are transformed onto canonical unit vectors. *Random gates* $\hat{U} = (\hat{U}_{ij})$ must transform random registers onto random registers, that is,

$$|\xi'\rangle = \hat{U}|\xi\rangle, \quad \xi'_i, \xi_i \geq 0, \quad \text{and} \quad \sum_i \xi'_i = \sum_i \xi_i = 1 \tag{119}$$

for all registers $|\xi\rangle$. Hence, it follows that

$$\hat{U}_{ij} \geq 0 \quad \text{for all} \quad i, j, \tag{120}$$

and

$$1 = \sum_i \xi'_i = \sum_i \sum_j \hat{U}_{ij} \xi_j = \sum_j \xi_j \left(\sum_i \hat{U}_{ij} \right). \tag{121}$$

Therefore, a random gate $\hat{U}$ is a nonnegative matrix where the entries in each column add up to one. Such matrices are called *stochastic matrices*, more precisely *left stochastic matrices*. They describe transitions in *Markov processes*.

All permutation matrices are stochastic matrices, and thus all reversible gates can be used in randomized computations. But there are several stochastic matrices that are not permutations:

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}. \tag{122}$$

The first one transforms both base states $|0\rangle$ and $|1\rangle$ onto $|0\rangle$, the second one transforms $|0\rangle$ and $|1\rangle$ on $|1\rangle$, and the third one is a random switch gate that transforms both $|0\rangle$ and $|1\rangle$ onto $(\frac{1}{2}\ \frac{1}{2})^T$. The random switch gate can act as a source creating random bits. Obviously, these randomized gates are not reversible.

From a mathematical point of view, the random registers are convex combinations of the unit vectors. The set of convex combinations

$$S = \{|\xi\rangle = \sum_i \xi_i |i\rangle : \xi_i \geq 0, \quad \sum_i \xi_i = 1\}, \tag{123}$$

forms a simplex, a special convex polytope with vertices corresponding to the canonical unit vectors. A vertex of S is a vector that cannot be represented as a convex combination of two other vectors in S. In the two-dimensional case the simplex S corresponds to a line segment, and in the three-dimensional case S is a triangle, see Figure 54.

Any left stochastic matrix $\hat{U}$ must map S into itself. If $\hat{U}$ is reversible, thus forming a bijection from S to S, then it is easy to prove that vertices must be mapped to vertices in a bijective way. Since the vertices are just the canonical unit vectors, it follows that each left stochastic reversible matrix where the inverse is also left stochastic, is a permutation matrix.

We can compose stochastic gates in the same way as reversible gates. In series we have to multiply the stochastic matrices, in parallel we have to apply the tensor product of matrices, and for controlled stochastic matrices we use the direct sum. It is easy to convince oneself that in all cases these operations lead to left stochastic matrices. Notice that the compositions of gates in random computation are based on Feynman's addition and multiplication rule but with classical probabilities. We have discussed this issue already for compositions of reversible gates.

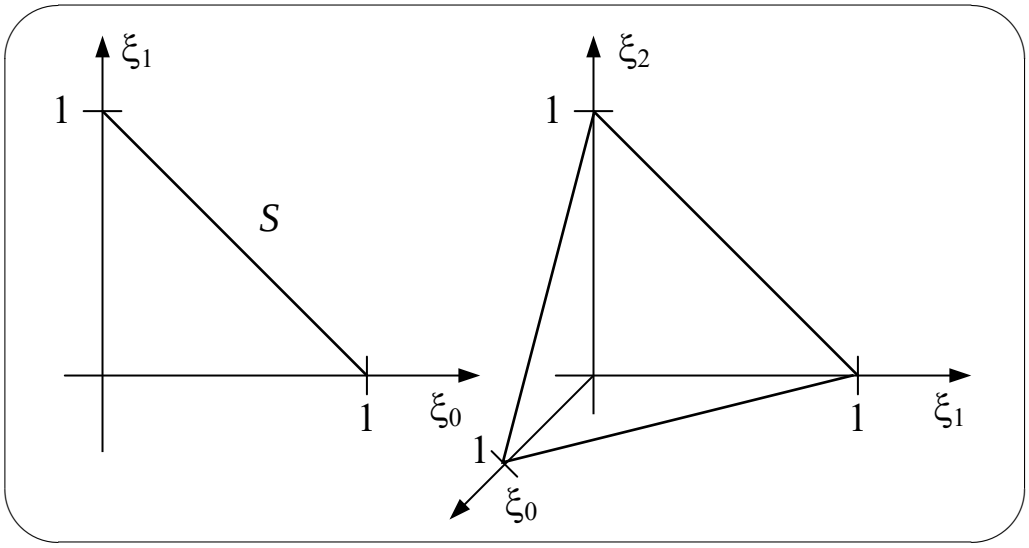


Figure 54: Simplices in $\mathbb{R}^2$ and $\mathbb{R}^3$.

Let for example

$$\hat{R} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \tag{124}$$

then we obtain for the series

$$\hat{R} \cdot \mathrm{NOT} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = \hat{R}, \tag{125}$$

in parallel

$$\hat{R} \otimes \mathrm{NOT} = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}, \tag{126}$$

and conditionally

$$\hat{R} \oplus \mathrm{NOT} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{127}$$

All matrices are left stochastic.

Let us apply the CNOT gate to the random register consisting of the random product state

$$|\psi\rangle \otimes |0\rangle = \begin{pmatrix} \psi_0 \\ 0 \\ \psi_1 \\ 0 \end{pmatrix}. \tag{128}$$

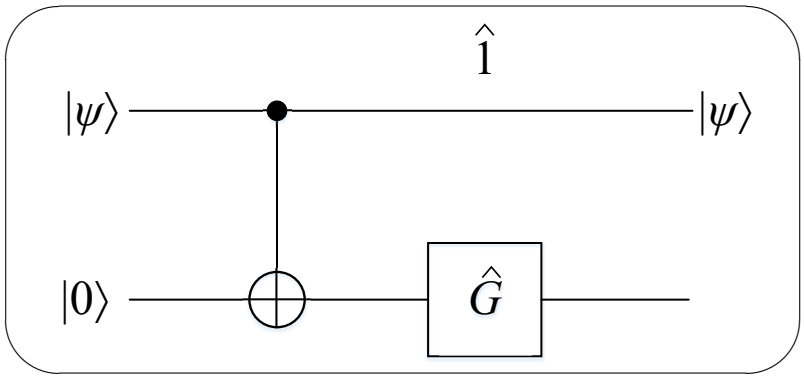


Figure 55: At first CNOT is applied to the random register $|\psi\rangle \otimes |0\rangle$, and then $\hat{G}$ is applied to the second rbit.

If follows that

$$\mathrm{CNOT}(|\psi\rangle \otimes |0\rangle) = \psi_0|00\rangle + \psi_1|11\rangle, \tag{129}$$

since

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \psi_0 \\ 0 \\ \psi_1 \\ 0 \end{pmatrix} = \begin{pmatrix} \psi_0 \\ 0 \\ 0 \\ \psi_1 \end{pmatrix} = \psi_0 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \psi_1 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{130}$$

Therefore, the CNOT gate entangles the rbits .

As a further example we consider the random gate

$$\hat{G} = \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 1 \end{pmatrix}. \tag{131}$$

This is a reversible left stochastic matrix. But the inverse cannot be left stochastic, since otherwise $\hat{G}$ would be a permutation matrix. This gate transforms as

$$\hat{G}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, \quad \hat{G}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{132}$$

We look at the series, where at first CNOT is applied, and then $\hat{G}$ to the second rbit, see Figure 55.

We have to evaluate

$$(\hat{1} \otimes \hat{G}) \cdot \mathrm{CNOT} \cdot (|\psi\rangle \otimes |0\rangle) =$$

$$\begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \psi_0 \\ 0 \\ \psi_1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\psi_0 \\ \frac{1}{2}\psi_0 \\ 0 \\ \psi_1 \end{pmatrix} \tag{133}$$

$$= \tfrac{1}{2}\psi_0|00\rangle + \tfrac{1}{2}\psi_0|01\rangle + \psi_1|11\rangle.$$

For

$$|\psi\rangle = \frac{1}{2}|0\rangle + \frac{1}{2}|1\rangle \tag{134}$$

we get from (133) the result

$$\frac{1}{4}|00\rangle + \frac{1}{4}|01\rangle + \frac{1}{2}|11\rangle. \tag{135}$$

This is what we expected. CNOT entangles the random register to

$$\frac{1}{2}|00\rangle + \frac{1}{2}|01\rangle, \tag{136}$$

and $\hat{G}$ applied to the second rbit yields with (131) formula (135).

Finally, some remarks to the superposition of random registers (118). In most textbooks the random superposed state of registers at some intermediate time is interpreted as follows: at each time the register has always definite zero-one values. The probabilistic uncertainty is due to the observer's ignorance about these definite values. If we observe rbits, then the register collapses to definite values that reflects the information we have obtained. For example, if we look at the second rbit in (135), then we observe with probability $\frac{1}{4} + \frac{1}{2} = \frac{3}{4}$ the state $|1\rangle$, and the register collapses to

$$\frac{1/4}{3/4}|01\rangle + \frac{1/2}{3/4}|11\rangle = \frac{1}{3}|01\rangle + \frac{2}{3}|11\rangle. \tag{137}$$

Then, if we look at the first rbit, we observe with probability $\frac{1}{3}$ the state $|0\rangle$, and thus the register collapses to the base state $|01\rangle$. This process is called *measurement* in classical stochastic computing.

3.6 Quantum Computation

Quantum computation refers to the capability to operate with quantum bits, the short form being *qubits*:

$$|\psi\rangle = \psi_0|0\rangle + \psi_1|1\rangle = \begin{pmatrix}\psi_0\\ \psi_1\end{pmatrix} \in \mathbb{C}^2, \qquad \psi_0, \psi_1 \in \mathbb{C}, \quad |\psi_0|^2 + |\psi_1|^2 = 1. \tag{138}$$

The essential difference to random computation is the replacement of classical non-negative probabilistic coefficients ψ_0 and ψ_1 by complex *quantum amplitudes* satisfying the law of Born and Malus. Almost all things that we have learned in random computation apply to quantum computation, and thus we quickly go through the quantum constructs. There is one aesthetic, fundamental exception compared with randomized computation:

- Quantum computation is reversible.

This is the basic property common with deterministic reversible computing. All information in a quantum circuit is preserved, nothing is lost.

A string of two qubits allows the states 00, 01, 10, and 11 with *probability amplitudes* $\xi_{00}, \xi_{01}, \xi_{10}, \xi_{11}$ yielding a four-dimensional vector

$$|\xi\rangle = \xi_{00}|00\rangle + \xi_{01}|01\rangle + \xi_{10}|10\rangle + \xi_{11}|11\rangle \qquad \xi_{ij} \in \mathbb{C}, \quad \sum_{i,j=0}^{1} |\xi_{ij}|^2 = 1 \tag{139}$$

The difference to (111) is the replacement of convex combinations by complex amplitudes, where the squared amplitude $|\xi_{ij}|^2$ is the probability of the base state $|ij\rangle$. As in the case of random computation (112) we define the *tensor product state* of two qubits $|\varphi\rangle$ and $|\psi\rangle$, both defined as in (138) via the tensor product

$$|\xi\rangle = |\psi\rangle \otimes |\varphi\rangle = \begin{pmatrix}\psi_0\varphi_0\\ \psi_0\varphi_1\\ \psi_1\varphi_0\\ \psi_1\varphi_1\end{pmatrix} \in \mathbb{C}^4. \tag{140}$$

The products $\psi_i\varphi_j \in \mathbb{C}$ are probability amplitudes that define a normalized vector, since

$$\sum_{i,j=0}^{1} |\psi_i\varphi_j|^2 = \big(|\psi_0|^2 + |\psi_1|^2\big)\big(|\varphi_0|^2 + |\varphi_1|^2\big) = 1. \tag{141}$$

As in random computation entangled states which cannot be represented as tensor products of qubits are typical.

A register of n qubits, called *quantum register*, is a *superposition*

$$|\xi\rangle = \sum_i \xi_i|i\rangle, \quad \xi_i \in \mathbb{C}, \quad \sum_i |\xi_i|^2 = 1, \tag{142}$$

where the base states

$$|i\rangle = |i_0 \dots i_k \dots i_{n-1}\rangle \tag{143}$$

denote the register of n bits with values $i_k \in \{0, 1\}$. They correspond to canonical unit vectors in the 2^n dimensional complex Hilbert space $\mathbb{C}^{2^n}$. This definition is the same as in (118) with the exception that the superposition is interpreted according to the rule of Born and Malus: with probability $|\xi_i|^2$ the register is in state $|i\rangle$ when measured.

Quantum gates must transform quantum register onto quantum register linearly, that is,

$$|\xi'\rangle = \hat{U}|\xi\rangle. \tag{144}$$

Hence,

$$1 = \sum_i |\xi'_i|^2 = \sum_i \xi'^*_i \xi'_i = \langle \hat{U}\xi | \hat{U}\xi \rangle, \tag{145}$$

and

$$1 = \sum_i \xi^*_i \xi_i = \langle \xi | \xi \rangle. \tag{146}$$

Therefore, the matrix $\hat{U}$ leaves the inner product

$$\langle \hat{U}\xi | \hat{U}\xi \rangle = \langle \xi | \xi \rangle \tag{147}$$

invariant. Such a matrix is called *unitary*.

Unitary matrices have many nice properties, among them:

- $\hat{U}^\dagger$ is unitary, and $\hat{U}^\dagger \hat{U} = \hat{U}\hat{U}^\dagger = \hat{1}$;
- the inverse $\hat{U}^{-1} = \hat{U}^\dagger$;
- the columns and the rows of $\hat{U}$ form an orthonormal set of vectors, respectively;
- the eigenvectors of $\hat{U}$ are orthogonal with eigenvalues lying on the complex unit circle;
- $\hat{U}$ has a decomposition of the form

  $$\hat{U} = \hat{V}\hat{D}\hat{V}^T, \tag{148}$$

  where $\hat{V}$ is unitary, and $\hat{D}$ is diagonal and unitary;
- each unitary matrix can be written as

  $$\hat{U} = e^{i\hat{H}} = \hat{1} + i\hat{H} - \frac{1}{2}\hat{H}^2 + \dots \tag{149}$$

  where $\hat{H}$ is a *Hermitian matrix* satisfying

  $$\hat{H}^\dagger = \hat{H}. \tag{150}$$

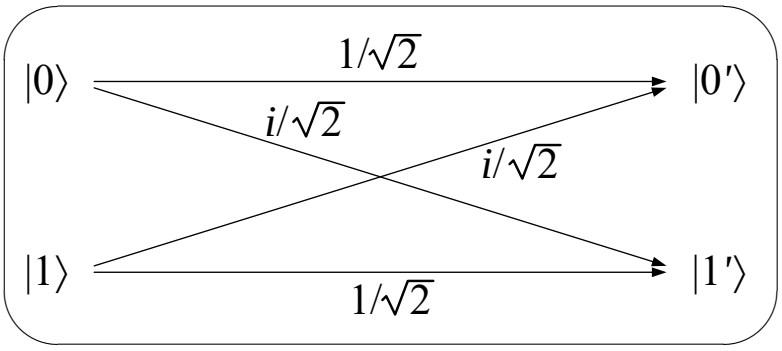


Figure 56: Transition amplitudes of a beamsplitter.

All permutation matrices and *orthogonal matrices* (real matrices $\hat{O}$ satisfying $\hat{O}^T = \hat{O}^{-1}$) are unitary, as can easily be seen. Stochastic matrices are in general not unitary, since they can be irreversible. The essential distinction between random computation and quantum computation is the reversibility of the quantum circuit model. Both models work with probabilities.

In quantum computation there are various further gates. A basic important one is the $\sqrt{\text{NOT}}$ gate which can be realized physically by a half-silvered mirror. In Section 2.9, we have considered a *Mach-Zehnder interferrometer*, see Figure 32. The beamsplitter in the lower part has two input paths $|0\rangle$ and $|1\rangle$, and two exit paths $|0'\rangle$ and $|1'\rangle$. Using the same amplitudes as before, Feynman's addition rule yields

$$|0'\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{i}{\sqrt{2}}|1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\ i\end{pmatrix}, \tag{151}$$

$$|1'\rangle = \frac{i}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}i\\ 1\end{pmatrix}, \tag{152}$$

which can be displayed graphically as in Figure 56.

Both equations can be written as the matrix

$$\sqrt{\text{NOT}} = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & i\\ i & 1\end{pmatrix}. \tag{153}$$

Then

$$\sqrt{\text{NOT}}\begin{pmatrix}1\\ 0\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\ i\end{pmatrix}, \quad \text{and} \quad \sqrt{\text{NOT}}\begin{pmatrix}0\\ 1\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}i\\ 1\end{pmatrix}. \tag{154}$$

If we compose two beamsplitter in series, we must multiply the matrices

$$\sqrt{\text{NOT}} \cdot \sqrt{\text{NOT}} = \frac{1}{2}\begin{pmatrix}1 & i\\ i & 1\end{pmatrix}\begin{pmatrix}1 & i\\ i & 1\end{pmatrix} = i\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}. \tag{155}$$

The global phase factor $i = e^{i\pi/2}$ can be ignored. Physically, we can remove it by putting two $-\pi/4$ phase shifters at both exit ports. Then we obtain the surprising result

$$\sqrt{\text{NOT}}\sqrt{\text{NOT}} = \text{NOT}. \tag{156}$$

This result is unexpected since transition probabilities of $\sqrt{\text{NOT}}$ are obtained by squaring the magnitudes of the amplitudes. Hence, the resulting random matrix has all coefficients equal to 1/2 due to (153). But putting them in series the randomness disappears since the result is the deterministic reversible NOT gate. This contradicts clearly the postulate of additivity in classical probability theory.

One of the most frequently used single qubit gates is the *Hadamard gate*

$$\hat{H} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{157}$$

It can be physically realized as a beamsplitter with two $-\pi/2$ phase shifters as shown in Figure 57.

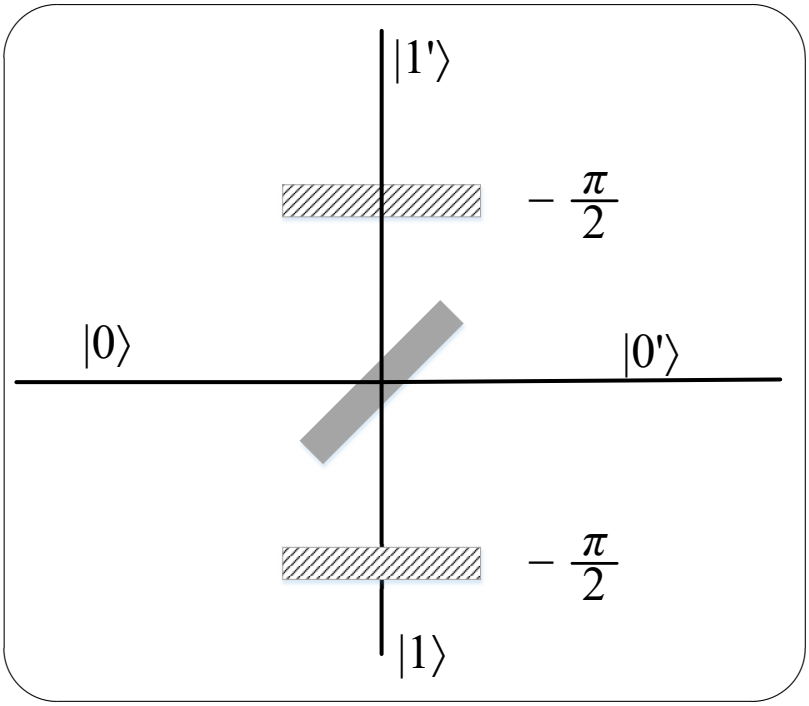


Figure 57: The Hadamard gate realized optically by a beamsplitter and two $-\pi/2$ phase shifters.

From Figure 57 it follows immediately with multiply and add that

$$|0'\rangle = \frac{1}{\sqrt{2}}|0\rangle + e^{-i\pi/2}\frac{i}{\sqrt{2}}|1\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle, \tag{158}$$

$$|1'\rangle = \frac{i}{\sqrt{2}}e^{-i\pi/2}|0\rangle + e^{-i\pi/2}\frac{1}{\sqrt{2}}e^{-i\pi/2}|1\rangle = \frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle. \tag{159}$$

Both equations imply the Hadamard gate (157) such that

$$\hat{H}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \hat{H}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}. \tag{160}$$

Two Hadamard gates composed in series

$$\hat{H} \cdot \hat{H} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{161}$$

give the identity, demonstrating interference once more.

The Hadamard gate is sometimes called the king of quantum computing, since it produces a superposition of all possible n bit registers, and is thus responsible for the high degree of parallelism in quantum computing. If we compose n Hadamard gates in parallel, and apply the gates to n qubits that are all prepared in base state $|0\rangle$, then we obtain a superposition of all 2^n binary registers, see Figure 58.

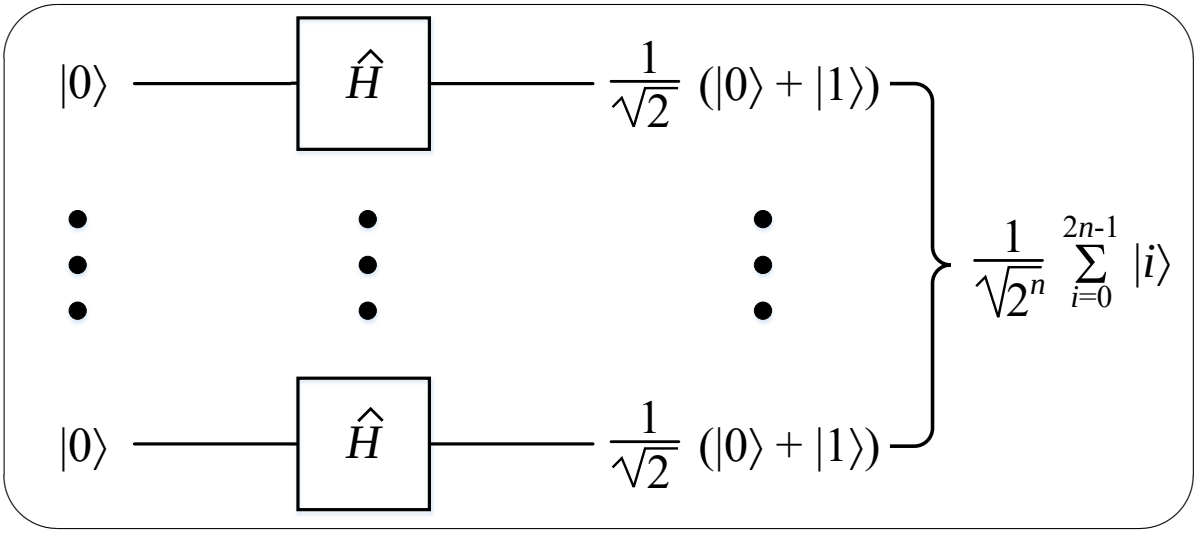


Figure 58: n Hadamard gates H in parallel are applied to n qubits each prepared in state $|0\rangle$.

It is easy to see that

$$\begin{aligned} &\hat{H}|0\rangle \otimes \hat{H}|0\rangle \otimes \ldots \otimes \hat{H}|0\rangle = \\ &\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes \ldots \otimes \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) = \\ &\frac{1}{\sqrt{2^2}}(|00\rangle + |01\rangle + |10\rangle + |11\rangle) \otimes \ldots \otimes \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) = \\ &\frac{1}{\sqrt{2^n}} \sum_{i=0}^{2^n-1} |i\rangle, \end{aligned} \tag{162}$$

where $|i\rangle$ is the base state (143)

$$|i\rangle = |i_0 \ldots i_k \ldots i_{n-1}\rangle. \tag{163}$$

In other words, n parallel Hadamard gates prepare a superposition of all integers from 0 to $2^n - 1$. In a 6 qubit register, for instance, the state $|010010\rangle$

corresponds to $18 = 1 \cdot 2^4 + 1 \cdot 2^1$. This important capability to obtain a simultaneous superposition of an exponential number of registers is used frequently in quantum algorithms.

As already seen in Figure 56 a simple beamsplitter and phase shifters suffice to implement the Hadamard gate. It can be proved that the Hadamard gate $\hat{H}$, the two phase shifters

$$\hat{S} = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}, \qquad \hat{T} = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}, \tag{164}$$

and the CNOT gate form a universal set of gates for quantum computation, that is, any n-qubit unitary gate can be constructed with a finite number of these gates. This is not the only universal set of gates. For instance, $\{\text{CCNOT}, \hat{H}\}$ is another universal set of gates. Obviously, quantum computation is at least as powerful as classical deterministic computation which can be performed only with CCNOT. The two phase gates $\hat{S}$ and $\hat{T}$ can be implemented by phase returners like wave plates. Hence is there a simple construction to implement CNOT for photons? Yes, there is a nice way to implement the 2-qubit gate CNOT by using a beamsplitter, and an optical element that rotates the polarization qubit by 90 degrees, like a simple sugar solution, see Figure 59. The location and polarization of the photon are the control and the target qubit, respectively.[50]

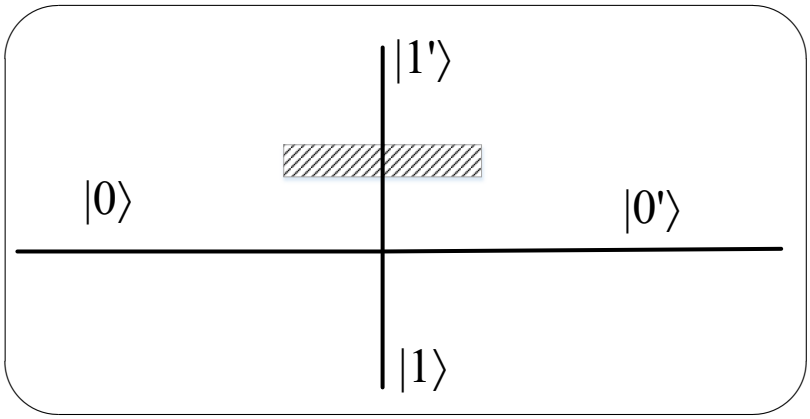


Figure 59: A CNOT gate implemented by a photon direction and its polarization.

For illustration, we show the simple quantum circuit that multiplies a single bit by two, that is, 0 is mapped to 0 and 1 is mapped to $2 \triangleq 10$ in binary code. A quantum circuit is reversible, and has the same number of input and output wires.

The binary code suggests immediately to use a CNOT gate where the control bit is the bit which is multiplied and the target bit is a carry bit which is initialized in state $|0\rangle$. Before the application of CNOT we prepare the input bit as a superposition of both values 0 and 1 of the control bit using the Hadamard gate, see Figure 60.

[50] Cerf, Adami, Quiat???

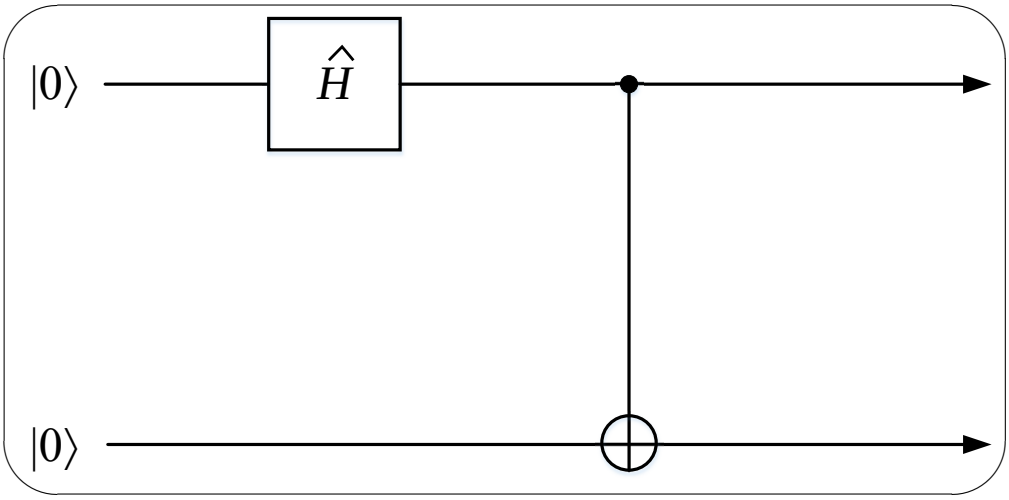


Figure 60: Gate that multiplies any bit by 2.

After the Hadamard gate the state of both qubits is

$$\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|10\rangle, \tag{165}$$

since the second qubit is unchanged. After CNOT, the state is

$$\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle, \tag{166}$$

since the second bit flips only if the control bit is one. This state contains the result of the multiplication with 2. If the control bit has value 0, the carry bit has the value 0 yielding the state $|00\rangle$. If the control bit has value 1, the carry bit must be one. This is expressed in state $|11\rangle$. Hence, the superposition (166) represents the multiplication with 2 on both values of the control bit simultaneously. In other words, we perform two calculations in parallel in one step. But we should have in mind that reading out the result is classically random as a coin toss, either $|00\rangle$ or $|11\rangle$ in each case with probability 1/2.

The state (166) is a famous entangled state, also called *Bell state*. There are four well-known Bell states that form a basis in $\mathbb{C}^4$, see Figure 61.

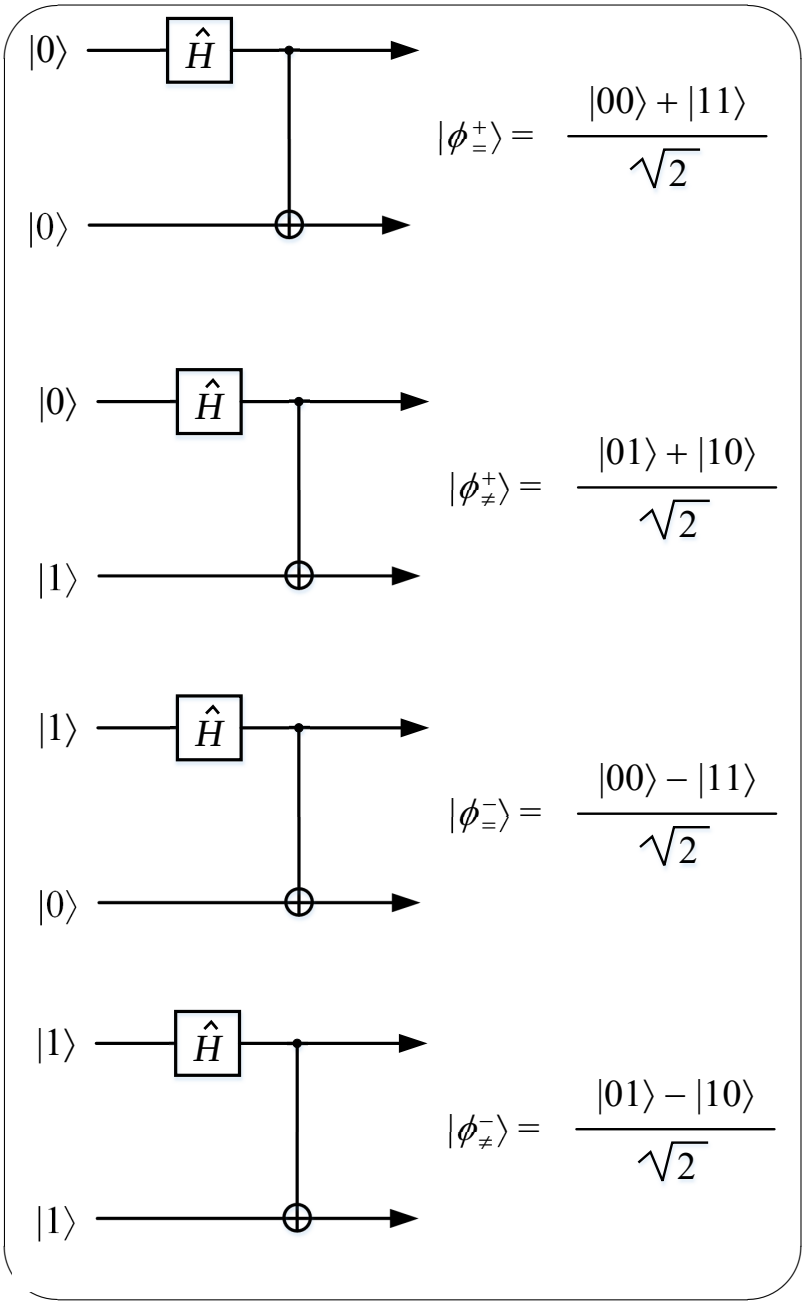


Figure 61: The quantum circuits for the four Bell states.

As a further example, we consider the extended quantum circuit by appending a second Hadamard gate, as displayed in Figure 62.

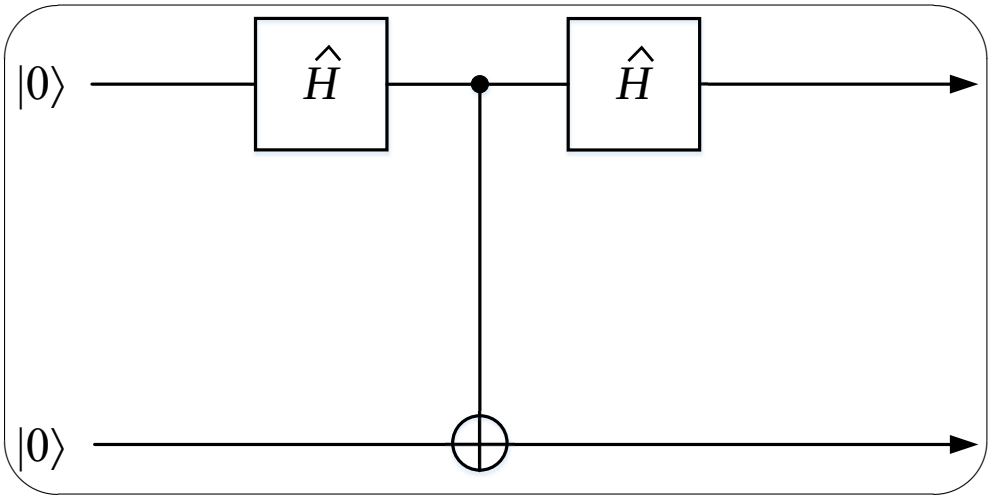


Figure 62: Appending at the gate for the Bell state $\frac{1}{\sqrt{2}}(|00\rangle+|11\rangle)$ a Hadamard gate yields the state $\frac{1}{2}(|00\rangle+|01\rangle+|10\rangle-|11\rangle)$, such that each of the four states $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$ can be observed with probability 1/4.

Then the control bit is split once more, and we obtain

$$\begin{aligned} & \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}|00\rangle+\frac{1}{\sqrt{2}}|10\rangle\right)+\frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}|01\rangle-\frac{1}{\sqrt{2}}|11\rangle\right) \\ = \; & \frac{1}{2}\left(|00\rangle+|10\rangle+|01\rangle-|11\rangle\right). \end{aligned} \tag{167}$$

Hence, each of the four base states has a probability amplitude 1/2, yielding the squared probability 1/4.

3.7 Quantum Parallelism

The aim of this section is to describe the tremendous potential of quantum computing and the advantage over classical computing. Let $x = (x_1 \cdots x_n) \in \{0,1\}^n$ be an n-bit binary register that represents any of the 2^n possible configurations. Let $f(x)$ denote a Boolean function that returns a single bit with value 0 or 1 for each register x.

In Section 3.2 we have seen how a classical irreversible Boolean function f can be described equivalently as a reversible function $\widetilde{f}$. This reversible function has the form

$$\widetilde{f} : \{0,1\}^{n+1} \to \{0,1\}^{n+1},\ (x,y) \to (x, y \oplus f(x)), \tag{168}$$

where x has n components and y has one component. The binary operation $y \oplus f(x)$ represents the exclusive XOR. If we set $y = 0$, then we get the value $f(x)$ on the last bit. Hence, the $(n+1)$-bit register $(x,0)$ is transformed as

$$(x,0) \to (x, f(x)). \tag{169}$$

We know that each reversible Boolean function can be realized by an appropriate composition of reversible logical gates. Actually, we need only CCNOT since this gate is universal in reversible computation. Mathematically, $\tilde{f}$ can be described as a permutation matrix, and thus is unitary. We denote this matrix representing $\widetilde{f}$ by $\hat{U}_f$ and use Dirac's notation for the register such that (169) is written as

$$\hat{U}_f(|x,0\rangle) = |x, f(x)\rangle. \tag{170}$$

In reversible deterministic computing we can only evaluate the function value for a given register. Since quantum mechanics allows the simultaneous superposition of states, we can ask: is it possible to realize the superposition of all function values simultaneously? In other words, can we realize the superposed state

$$|\psi\rangle := \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} |x, f(x)\rangle \tag{171}$$

by one quantum circuit for each Boolean function f? The answer is yes, and this fundamental feature is called *quantum parallelism*, referring to the property to evaluate a Boolean function $f(x)$ for all possible registers x simultaneously.

We have already seen that n parallel Hadamard gates, see Figure 57, suffice to prepare a superposition of all possible registers:

$$\begin{aligned} \hat{H}^{\otimes n}|0\rangle^{\otimes n} &= (\hat{H}|0\rangle) \otimes \ldots \otimes (\hat{H}|0\rangle) \\ &= \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes \ldots \otimes \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \\ &= \frac{1}{\sqrt{2^n}} \sum_{x_1,\ldots,x_n=0}^{1} |x_1 \cdots x_n\rangle, \end{aligned} \tag{172}$$

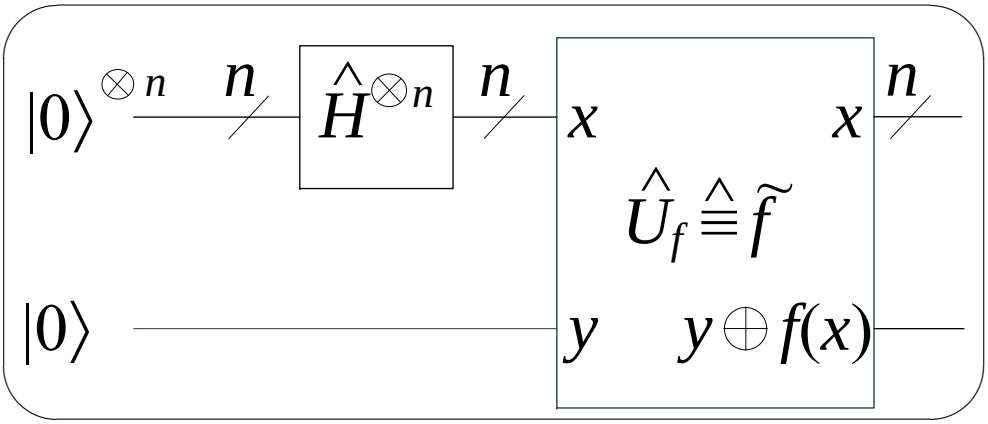


Figure 63: The quantum circuit that implements the simultaneous superposition of all function values $f(x)$. The vertical bar and the number n over the wire symbolize n parallel qubits.

or shortly

$$\hat{H}^{\otimes n}|0\rangle^{\otimes n} = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} |x\rangle. \tag{173}$$

We have only considered the case where each bit in the register $x = (x_1 \cdots x_n)$ has value 0. We can prove with mathematical induction the more general case, where arbitrary registers are allowed. It is

$$\hat{H}^{\otimes n}|x\rangle = \frac{1}{\sqrt{2^n}} \sum_{z\in\{0,1\}^n} (-1)^{\langle x,z\rangle}|z\rangle, \tag{174}$$

where

$$\langle x, z\rangle = x_1 z_1 + \ldots + x_n z_n. \tag{175}$$

This is perhaps not so easy to see. Let us consider in detail the case $n = 2$:

$$\begin{aligned}
&(\hat{H}\otimes\hat{H})|x_1x_2\rangle = \hat{H}|x_1\rangle \otimes \hat{H}|x_2\rangle \\
&= \tfrac{1}{\sqrt{2}}(|0\rangle + (-1)^{x_1}|1\rangle) \otimes \tfrac{1}{\sqrt{2}}(|0\rangle + (-1)^{x_2}|1\rangle) \\
&= \tfrac{1}{\sqrt{2^2}}(|00\rangle + (-1)^{x_2}|01\rangle + (-1)^{x_1}|10\rangle + (-1)^{x_1+x_2}|11\rangle) \\
&= \tfrac{1}{\sqrt{2^2}} \sum_{z_1=0}^{1} \sum_{z_2=0}^{1} (-1)^{x_1z_1+x_2z_2}|z_1z_2\rangle.
\end{aligned} \tag{176}$$

In the same manner the mathematical induction can be easily performed for arbitrary n.

For a realization of all function values $f(x)$ simultaneously, it seems natural to prepare the $(n+1)$-bit register in state $|0\rangle^{\otimes n} \otimes |0\rangle$, then apply the Hadamard gate to the first n qubits, and then the gate $\hat{U}_f$ to all qubits, see Figure 63.

It is now easy to go through the following sequence that coincides with the ordering in Figure 63. We obtain with (173) and (170) the simultaneous superposition of function values:

$$\begin{aligned} |0\rangle^{\otimes n} \otimes |0\rangle &\overset{\hat{H}^{\otimes n}\otimes \hat{1}}{\longmapsto} \left(\sum_{x\in\{0,1\}^n} \tfrac{|x\rangle}{\sqrt{2^n}} \right) \otimes |0\rangle \\ &\overset{\hat{U}_f}{\longmapsto} \tfrac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} \hat{U}_f|x,0\rangle = \tfrac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} |x, f(x)\rangle = |\psi\rangle. \end{aligned} \tag{177}$$

Keep in mind: Quantum parallelism is the ability to implement a simultaneous superposition of all function values of a Boolean function. It provides a potential exponential speed up in computing.

This simultaneous superposition of all function values is really amazing and should allow an *exponential speed up* in quantum computing. But unfortunately, when measuring the superposed state in (171) with Born's rule, each function value occurs with probability $\frac{1}{2^n}$, and the superposition and its parallelism is destroyed. Therefore, the utilization of quantum parallelism is a rather deep phenomenon and requires a way of thinking that is different to a sequential rationality. We shall discuss this issue in the next section. *Measurement* means that any machine, such as polarizing filters, filters out one of the base states.

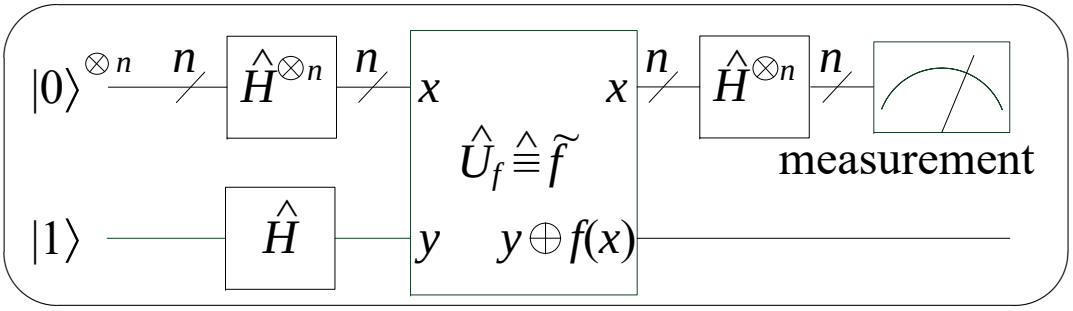


Figure 64: The quantum circuit implementing the Deutsch-Josza algorithm. At first the simultaneous superposition of all function values is implemented, similar as in Figure 63. Then the tensor product of n Hadamard gates is applied to the top n qubits which are measured afterwards.

3.8 Deutsch-Jozsa Algorithm

In this section we show the power of quantum parallelism for a problem solved by Deutsch and Jozsa. They considered Boolean functions $f(x)$, defined on the set of n-bit binary registers x with single binary output values 0 and 1. Such a function is *constant* if it returns the same value $f(x)$ for all binary input registers x, and is *balanced* if for half of the registers the value $f(x) = 1$, and for the other half of the registers $f(x) = 0$. Then they looked at the class of functions that are either balanced or constant, and asked for the algorithm with the fewest function calls that decides whether a function of this class is balanced or constant.

Of course, this problem seems to be rather artificial and not very practical. But it is useful, since many key steps of quantum computation can be explained in a clear and understandable manner.

Thinking classically, there are 2^n n-bit registers, and we need to check the function values for the half of them plus one in the worst case. Thus, $2^{n-1} + 1$ function values must be checked to see in each case with certainty whether the function is constant or balanced. Now we will see that an algorithm, called the *Deutsch-Josza algorithm*, can be implemented as a single quantum circuit, see Figure 64. One call of this circuit suffice to solve our decision problem, whereas an exponential number of calls may be necessary in the classical case.

In the first step we initialize n qubits in state $|0\rangle$, and the $(n+1)$th qubit in state $|1\rangle$. Then we apply the tensor product of $n+1$ Hadamard gates to the initialized qubits. Using (173) it follows that

$$(\hat{H}^{\otimes n} \otimes \hat{H})(|0\rangle^{\otimes n} \otimes |1\rangle) = \frac{1}{\sqrt{2^n}} \left(\sum_{x \in \{0,1\}^n} |x\rangle \right) \otimes \frac{1}{\sqrt{2}} (|0\rangle - |1\rangle). \tag{178}$$

Secondly, the operator $\hat{U}_f$ is applied. Using binary addition it follows that $y \oplus f(x) = y$ if $f(x) = 0$, and $y \oplus f(x) = \mathrm{NOT}(y)$ if $f(x) = 1$. Since the last

qubit is the superposition $|y\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$, we get

$$|y \oplus f(x)\rangle = \left\{ \begin{array}{l} \frac{|0\rangle - |1\rangle}{\sqrt{2}} \text{ for } f(x) = 0 \\ \frac{|1\rangle - |0\rangle}{\sqrt{2}} \text{ for } f(x) = 1 \end{array} \right\} = (-1)^{f(x)} \frac{|0\rangle - |1\rangle}{\sqrt{2}}. \tag{179}$$

Applying $\hat{U}_f$ to the result in (178) gives

$$\stackrel{\hat{U}_f}{\longmapsto} \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} (-1)^{f(x)} |x\rangle \otimes \left(\frac{|0\rangle - |1\rangle}{\sqrt{2}} \right). \tag{180}$$

Thirdly, using (174), the tensor product operator $\hat{H}^{\otimes n} \otimes \hat{1}$ acts as

$$\begin{aligned} &\stackrel{H^{\otimes n} \otimes \hat{1}}{\longmapsto} \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} (-1)^{f(x)} \left(\frac{1}{\sqrt{2^n}} \sum_{z \in \{0,1\}^n} (-1)^{\langle x,z \rangle} |z\rangle \right) \otimes \left(\frac{|0\rangle - |1\rangle}{\sqrt{2}} \right) \\ &= \left(\sum_{x,z \in \{0,1\}^n} \frac{(-1)^{f(x) + \langle x,z \rangle}}{2^n} |z\rangle \right) \otimes \left(\frac{|0\rangle - |1\rangle}{\sqrt{2}} \right) = |\psi\rangle \end{aligned} \tag{181}$$

Finally, the top n qubits are measured.

Now we argue that f is balanced if and only if at least one of the first n qubits is in state $|1\rangle$. Otherwise, f is constant.

To justify this statement, assume that the first n qubits in (181) are measured in state $|0\rangle$, that is, $|z\rangle = |0\,0 \cdots 0\rangle$. In this case

$$\langle x, z \rangle = x_1 z_1 + \ldots + x_n z_n = 0, \tag{182}$$

and (181) yields the amplitude for this event

$$\psi_{z=0} = \sum_{x \in \{0,1\}^n} \frac{(-1)^{f(x)}}{2^n}. \tag{183}$$

With Born's rule we obtain the related probability

$$\text{Prob}_{z=0} = \left| \sum_{x \in \{0,1\}^n} \frac{(-1)^{f(x)}}{2^n} \right|^2. \tag{184}$$

If f is constant, then $f(x)$ is either 0 or 1 regardless of the input x. Hence, the probability for measuring the first n qubits in state $|0\rangle$ is

$$\text{Prob}_{z=0} = \left| 2^n \frac{\pm 1}{2^n} \right|^2 = 1. \tag{185}$$

On the other hand, if f is balanced half of the terms in (184) are positive and the other ones are negative. They cancel out and yield zero probability to measure $|z\rangle = |0\,0 \cdots 0\rangle$. Hence, one of the first n qubits must be in state $|1\rangle$, proving our statement.

The Deutsch-Jozsa quantum circuit is exponentially faster than the best deterministic algorithm on a classical computer. There are several other quantum algorithms that solve certain problems faster than classical algorithms. Well-known algorithms are *Shor's algorithm* for factoring and the *quantum Fourier transformation*. Shor's quantum algorithm to factor an integer n takes about $\mathcal{O}((\log n)^3)$ quantum gate operations which is substantially faster than the best-known classical factoring algorithm which requires about $\mathcal{O}(e^{1.9(\log n)^{1/3} \cdot (\log \log n)^{2/3}})$ operations.

3.9 No-Cloning Theorem

In Section 3.2 we have seen that FANOUT can be put in a reversible form by using CNOT. Therefore, classical bits and classical information can easily be copied. At a first glance it seems that the reversible classical circuit representing FANOUT can be used to copy quantum states $|\psi\rangle$, similarly as displayed in Figure 65.

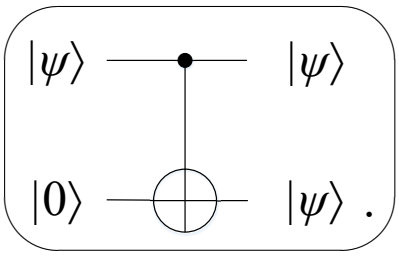


Figure 65: Copying states $|\psi\rangle$ with FANOUT is in general not possible.

The input before CNOT is $|\psi\rangle \otimes |0\rangle$, and the output after CNOT is $|\psi\rangle \otimes |\psi\rangle$. Let $|\psi\rangle = \psi_0|0\rangle + \psi_1|1\rangle$, then we have

$$|\psi\rangle \otimes |0\rangle = \psi_0|00\rangle + \psi_1|10\rangle. \tag{186}$$

Applying CNOT yields

$$\mathrm{CNOT}(|\psi\rangle \otimes |0\rangle) = \psi_0|00\rangle + \psi_1|11\rangle. \tag{187}$$

Unfortunately, this is not the desired state $|\psi\rangle \otimes |\psi\rangle$, since

$$|\psi\rangle \otimes |\psi\rangle = \psi_0^2|00\rangle + \psi_0\psi_1|01\rangle + \psi_1\psi_0|10\rangle + \psi_1^2|11\rangle. \tag{188}$$

Only if $\psi_0 = 0$ and $\psi_1 = 1$ or $\psi_0 = 1$ and $\psi_1 = 0$ both states coincide. Hence, only classical bits are copied appropriately. For all superposed states the above circuit fails.

But is there any other quantum circuit which copies arbitrary qubits? Each quantum circuit can be represented as a unitary matrix $\hat{U}$. Let us assume that $\hat{U}$ acts on n qubits as displayed in Figure 66.

The first qubit $|\psi\rangle$ should be copied onto the second qubit which is initialized with $|0\rangle$. We allow an arbitrary number of ancillary bits, say $n-2$, that might be necessary. A quantum copy machine $\hat{U}$ producing clones of quantum states $|\psi\rangle$ must satisfy the equation

$$\hat{U}(|\psi\rangle \otimes |0\rangle \otimes |0\rangle^{\otimes(n-2)}) = |\psi\rangle \otimes |\psi\rangle \otimes |\gamma\rangle \tag{189}$$

for all qubits $|\psi\rangle$, where $|\gamma\rangle$ is any superposed state of $n-2$ qubits. Therefore, if $|\psi\rangle = |0\rangle$ then (189) implies

$$\hat{U}(|00\rangle \otimes |0\rangle^{\otimes(n-2)}) = |00\rangle \otimes |\gamma^0\rangle. \tag{190}$$

If $|\psi\rangle = |1\rangle$, then it follows that

$$\hat{U}(|10\rangle \otimes |0\rangle^{\otimes(n-2)}) = |11\rangle \otimes |\gamma^1\rangle, \tag{191}$$

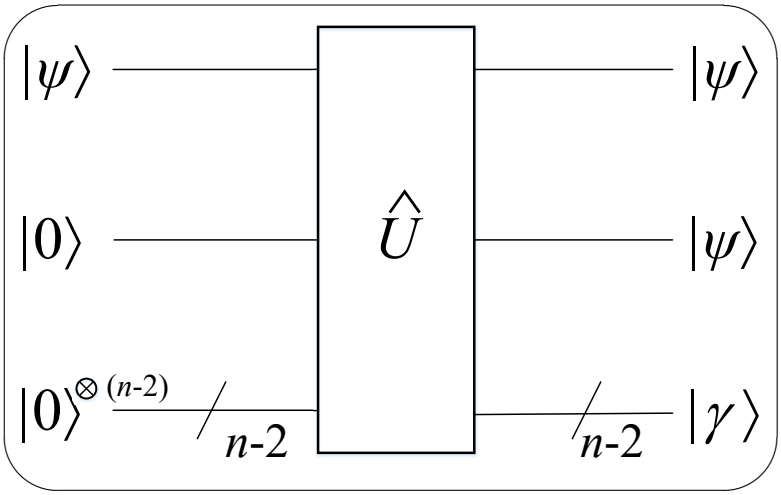


Figure 66: A quantum circuit represented by a matrix $\hat{U}$ which copies the state $|\psi\rangle$ onto the second qubit initialized with $|0\rangle$. An arbitrary number of qubits, namely $n-2$ ancillary qubits, can be used for copying.

and if $|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ then we get

$$\begin{aligned}
&\hat{U}\left(\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes |0\rangle \otimes |0\rangle^{\otimes(n-2)}\right) = \\
&\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes |\gamma^{01}\rangle = \\
&\frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle) \otimes |\gamma^{01}\rangle.
\end{aligned} \tag{192}$$

So far we have derived three equations that describe the action of the copy machine $\hat{U}$ for three states. We have assumed that $\hat{U}$ is a matrix and thus describes a linear transformation. Hence, the equations

$$\begin{aligned}
&\hat{U}\left(\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes |0\rangle \otimes |0\rangle^{\otimes(n-2)}\right) = \\
&\hat{U}\left(\frac{1}{\sqrt{2}}(|00\rangle + |10\rangle) \otimes |0\rangle^{\otimes(n-2)}\right) = \\
&\frac{1}{\sqrt{2}}\left(\hat{U}(|00\rangle \otimes |0\rangle^{\otimes(n-2)}) + (\hat{U}|10\rangle \otimes |0\rangle^{\otimes(n-2)})\right) = \\
&\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \otimes |\gamma\rangle
\end{aligned} \tag{193}$$

must be satisfied, where we have used (190) and (191) for the last identity. Comparing (193) with (192) we see that the base states $|01\rangle$ and $|10\rangle$ in formula (192) do not occur in (193). Hence, both equations do not coincide. There exists no quantum copy machine $\hat{U}$ that is linear and can copy at least three

states $|0\rangle$, $|1\rangle$ and $\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)$. This fact is known as the *No-Cloning Theorem*. No linear procedure exists by which superposed quantum states can be copied or cloned exactly. This result is due to Wootters and Zurek[51].

We have seen that deterministic states can be cloned, quantum states not. The reasons are the linearity of the machine $\hat{U}$ and the superposed states. Going through the derivation of the No-Cloning Theorem, once more it follows that this theorem is also true for classical random states. More generally, we can say that probabilistic states, random or quantum ones, cannot be cloned by linear machines.

Keep in mind: Probabilistic states, classically random or complex quantum ones, cannot be cloned.

[51] W.K. Wootters, W.H. Zurek: A single quantum cannot be cloned. Nature 299(5886): 802–803, 1983.

3.10 Quantum Teleportation

In this section we return to the puzzles that are related to entanglement, EPR effects, the apparent conflict with the theory of relativity, and spooky actions at distance. The distant communication of entangled pairs of photons seems to violate Einstein's demand that signals cannot travel faster than light.

One puzzle is *quantum teleportation*, which can be viewed as a strange effect of quantum entanglement. It shows a so-called *non-local interaction* in zero time. A local interaction employs either a direct contact or a contact via some medium. Friction, for instance, is a local interaction between two bodies. Even gravitation can be viewed as a local interaction when accepting the existence of *gravitons* that are the mediators between gravitating objects. Local interactions are limited: they cannot propagate faster than the speed of light. That is a postulate of the theory of relativity.

In some field theories it is shown that all known forces — the electromagnetic, the gravitational, the weak and the strong force — are local. Hence, reality should be local. There seem to be, however, exceptions. Beside entangled pairs of photons, the so-called "collapse of the state vector", also known as the "measurement problem" (see Principle 3 in Section 1), is a non-local effect at least in the Copenhagen interpretation of quantum mechanics. The reason is that nothing explains, determines or mediates the collapse of the state vector to any base vector of an observable. Non-local interactions, if they exist, are really magic and seem to be absurd, since they are not mediated by particles or fields, and they are not limited by any speed. But do they exist?

The basic task of *teleportation* is to send physical objects or information from one place in space to another. Of course, there is no problem to transport any classical things in an ordinary way. We experience daily the transmission of classical information. This is clear since classical information, that is sequences of classical bits, can be copied and transported. But this is not true for qubits, and thus for quantum information, as we have learned before. If teleportation of qubits is possible, then during the teleportation operation the original quantum state must be destroyed. Otherwise, teleportation would produce a perfect copy, violating the No-Cloning Theorem.

Let us investigate the following quantum circuit displayed in Figure 67. The idea is that Alice and Bob create an entangled Bell pair $(|01\rangle - |10\rangle)/\sqrt{2}$, and Bob leaves Alice with the second qubit of the entangled pair. Some times later, Alice wants to send an unknown qubit state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ to Bob who is far away. She uses the quantum circuit visualized in Figure 67.

After the creation of the entangled Bell state, there are three parallel qubits, defining a three qubit register. This register is mathematically described by the tensor product:

$$\begin{aligned} |\xi_0\rangle &= (\alpha|0\rangle + \beta|1\rangle) \otimes \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \\ &= \frac{\alpha}{\sqrt{2}}(|001\rangle - |010\rangle) + \frac{\beta}{\sqrt{2}}(|101\rangle - |110\rangle). \end{aligned} \tag{194}$$

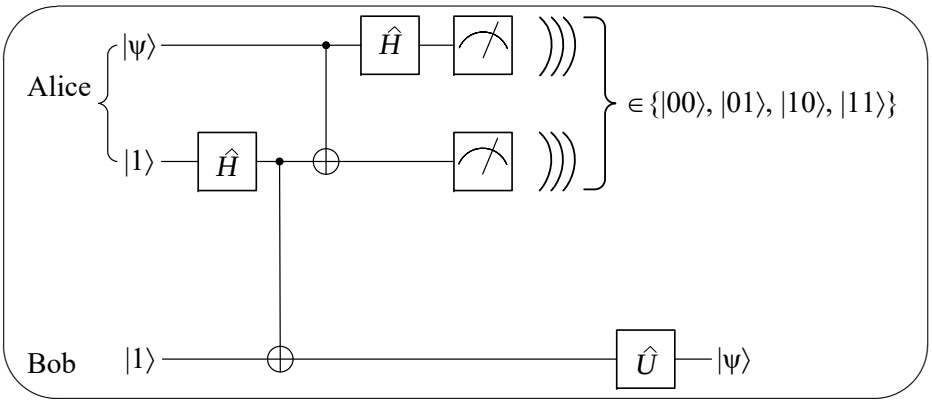


Figure 67: Quantum circuit for teleporting a quantum state $|\psi\rangle$ from Alice to Bob. The first and the second qubit is held by Alice, and the third qubit is given to Bob, who is somewhere far far away. This experimental set-up starts with creating an entangled Bell pair of states $\frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$. Bob has taken the second entangled qubit, before he left Alice. Then Alice performs a so-called Bell measurement between the first two qubits and obtains two classical bit values which are passed via any classical channel to Bob. With this information Bob can restore the original state $|\psi\rangle$ when applying an appropriate unitary transformation to his qubit.

Now, Alice applies a CNOT gate to the first and second qubit, and obtains

$$|\xi_1\rangle = \frac{\alpha}{\sqrt{2}}(|001\rangle - |010\rangle) + \frac{\beta}{\sqrt{2}}(|111\rangle - |100\rangle). \tag{195}$$

Next, performing the Hadamard gate yields

$$\begin{aligned} |\xi_2\rangle &= \frac{\alpha}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)\right) \otimes (|01\rangle - |10\rangle) \\ &\quad + \frac{\beta}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\right) \otimes (|11\rangle - |00\rangle) \\ &= \frac{\alpha}{2}(|001\rangle - |010\rangle + |101\rangle - |110\rangle) \\ &\quad + \frac{\beta}{2}(|011\rangle - |000\rangle - |111\rangle + |100\rangle). \end{aligned} \tag{196}$$

We can rearrange this state:

$$\begin{aligned} |\xi_2\rangle &= \frac{1}{2}\left(|00\rangle \otimes (\alpha|1\rangle - \beta|0\rangle) + |01\rangle \otimes (\beta|1\rangle - \alpha|0\rangle)\right. \\ &\quad \left. + |10\rangle \otimes (\alpha|1\rangle + \beta|0\rangle) + |11\rangle \otimes (-\beta|1\rangle - \alpha|0\rangle)\right). \end{aligned} \tag{197}$$

This is really surprising. Instantaneously, with the last application of the Hadamard gate, before measuring the first two qubits, on Bob's qubit the state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ reappears in this superposition. This is expressed in Table 3, where we see that the infinite sequence of bits required for α and β are instantaneously teleported. Only two bits of information are missing.

Alice' measurement	Bob's qubit	Bob's action
$\|00\rangle$	$\alpha\|1\rangle - \beta\|0\rangle$	$\hat{U} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$
$\|01\rangle$	$\beta\|1\rangle - \alpha\|0\rangle$	$\hat{U} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$
$\|10\rangle$	$\alpha\|1\rangle + \beta\|0\rangle$	$\hat{U} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$
$\|11\rangle$	$\alpha\|0\rangle + \beta\|1\rangle$	$\hat{U} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Table 3: Instantaneous teleportation after measurement. Bob's qubit has one of four states that can be transformed to $|\psi\rangle$ when using the unitary gates $\hat{U}$ above.

If Alice measures her two qubits and sends to Bob her result via a classical channel, then Bob can restore the original state $|\psi\rangle$ when using the unitary gates given in Table 3. For example, if the measurement result is $|00\rangle$, Bob's qubit is $(-\beta, \alpha)^T$, and applying $\hat{U}$ yields the desired state $(\alpha, \beta)^T$. The original state $|\psi\rangle$ is destroyed, which is in agreement with the No-Cloning Theorem.

So far, so good, and in almost all text books (at least that I have read) the interpretation is that there are fortunately no conflicts and no paradoxes with the theory of relativity, because we have a complete teleportation only when the two final bits from the measurement are send from Alice via a classical channel to Bob. This can be done only with a velocity not larger than the speed of light. OK, then it seems that it doesn't matter that in a quarter of all cases an instantaneous teleportation has happened without sending two bits from Alice to Bob.

The statement that teleportation does not violate the theory of relativity, however, is not always shared. One exception can be found in the well-known book of Penrose[52] in Chapter 23.9. His scenario was that he had to send the state $|\psi\rangle$ to his colleague who stays on Titan 5 light years away. He writes:

> *Yet, we shall find this direction leading into a territory that many people would, no doubt, be most reluctant to enter- and with reason as we shall see. [...] What is particularly striking about quantum teleportation is that, by sending my colleague 2 bits of classical information (one of the numbers 0,1,2,3,4, which could have been coded as 00, 01, 10, 11, respectively), I have conveyed the 'information' of a point on the entire Riemann sphere [...] the information contained in the unrestricted choice of a point in a continuum: strictly $\aleph_0$ bits, for perfect accuracy! At this point I should mention that real experiments have been performed which confirm the expectations of quantum mechanical exportation (over distances of the order of metres, not Earth-Saturn spans, of course) so we must take these things seriously. Not only that, but the blossoming subject of quantum cryptography depends upon things of this general nature; so do many ideas of quantum computation.* Penrose 2004

Penrose displayed a spacetime diagram, see Figure 68. It describes a link between the state $|\psi\rangle$ before teleportation and after teleportation. But this path goes 5 years back into the past. We shall come back to this apparent conflict later.

[52] Penrose, Roger. The Road to Reality: A Complete Guide to the Physical Universe. Jonathan Cape, 2004.

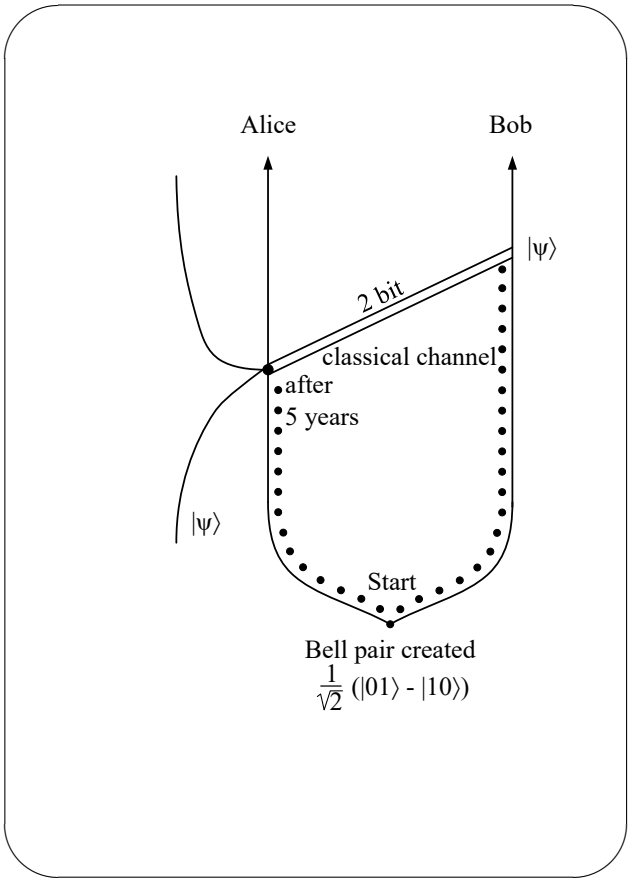


Figure 68: Penrose's spacetime diagram for quantum teleportation shows the acausal propagation of quantum information $|\psi\rangle$. The unknown superposed quantum state $|\psi\rangle$ with its $\aleph_0$ bits is with certainty teleported by simply sending 2 bits of classical information. In a quarter of all cases, however, an instantaneous teleportation occurs without sending classical information. There seems to be something completely wrong, in quantum mechanics or in relativity theory or in both. Nevertheless there is a link in the spacetime diagram, the dotted path. Unfortunately, this path goes 5 years back into the past leading to an acausal propagation of quantum information.

4 Unification via Semimodules

Based on our previous description of reversible, random, and quantum computation, we present a unification of physical concepts such as states, observables and dynamics. In the literature these notions are defined in different ways depending on the physical theory. This unification, independent of the physical model, makes use of semimodules. A semimodule is a linear space where the field of numbers is replaced by a number system without requiring subtraction and division. It turns out that the weird quantum principles presented in Section 1 actually are very natural and coincide with what we feel and observe. In particular, we describe quantum mechanics as a probabilistic theory of possibilities that characterize the future. This theory applies just as well to macroscopic objects. **The central goal of this section is to show various consequences when we replace the external parameter time t by a trinity, namely the partitioning future, present, and past.** Clearly, then spacetime vanishes. However, among many other consequences, we show an alternative approach to the Lorentz transform, the key to the theory of relativity. Additionally, we discuss the dimensions of the underlying spaces in different physical theories. From the point of view of information theory we investigate in which dimensions a physical theory could be reasonable.

4.1 Preliminaries

In the introduction of these lecture notes we displayed the strange principles of quantum physics. If we look in more detail into physics, it turns out that even basic concepts become very vague. When asking what a state in physics is, one cannot find a simple answer. For example, on the page "Physics Stack Exchange" in December 2015, the following question occured:

> *What is a state in physics? While reading physics, I have heard many a times a "system is in ... state" but the definition of a state was never provided (and googling brings me totally unrelated topic of solid state physics), but was loosely told that it has every information of the system you desire to know. On reading further, I have found people talking of Thermodynamic state, Lagrangian, Hamiltonian, wave-function etc etc which I think are different from one another. So in general I want to know what do we mean by state in physics and is there a unique way to describe it?* Manish Kumar Singh[53]

The last question is really important and seemingly difficult to answer. Such questions show on the one hand a great interest in science, but on the other hand they exhibit insufficient explanations of basic physical concepts in many textbooks.

It is common sense that each physical theory must work out three fundamental tasks:

1. *States* must be defined.

2. All possible outcomes or *observable* values of an experiment must be specified precisely.

3. The *dynamics*, which allows to predict a future state given the current state, must be presented. In other words, we need a rule about how states change with time.

In physics the concept of a state and the notion of an observable is sometimes presented ambiguously. Most physicists say that the precise definition of a *state* depends on the underlying physical theory. We look at some examples:

- In classical mechanics, the space of states can be the six-dimensional phase space, the four-dimensional spacetime, or discrete subsets due to some *coarse graining*. For a system of n particles the phase space is $\mathbb{R}^{6n}$, three spatial coordinates and three coordinates for the momentum of each particle. In all cases states are points in a real space.

- In quantum mechanics, a state is an element of a complex Hilbert space. Each state is a superposition of base states. The base states form an orthonormal basis of the underlying Hilbert space. For example, the

[53] see Physics Stack Exchange, December 2015, http://physics.stackexchange.com/questions/223564/what-is-a-state-in-physics

base states of a polarized photon are described by its polarization angle α and two possibilities, namely vertical and horizontal polarization. The base states of an electron are described by its momentum and its spin.

- In classical probability theory the set of states is the *sample space*, which is defined as the set of all possible *outcomes* of a random experiment. The outcomes are also called *elementary events*. Events are specific subsets of the sample space.

- The formalism of thermodynamics and statistical physics can be developed classically, but also in terms of quantum mechanics. There, systems of a very large number of particles are considered. In the absence of interaction, each particle can be in one of its possible states, a classical or a quantum one. The realization where each particle is in a certain state i is called a *microstate* of the system. The information given by microstates is excessive. A compressed meaningful information is the number of particles that are in state i, say N_i. These numbers specify the *macrostates* in thermodynamics.

- More subtle is the definition of states in quantum field theories.

Summarizing, we have a colorful diversity of concepts for the definition of a state. Moreover, the states of classical mechanics and of quantum mechanics seem to differ significantly. For instance, in the book of Susskind and Friedman[54] we find:

> *In Volume I, it took a little more than a page to explain what a state is in classical mechanics. The quantum version has taken three lectures, three mathematical interludes, and according to my rough count, about 17,000 words to get to the same place. But I think the worst is over. We now know what a state is.* Susskind and Hrabovsky 2013

There seems to be little hope to find a short, precise and all-encompassing definition of a state, independent of the underlying physical theory. Closely related is the problem to find a universal valid definition of an observable. In the following we will elaborate on these concepts in detail.

At least as difficult seems to be the interpretation of the mathematical formalisms that describes a physical theory. In this context Peres says that in a laboratory an experimental set up consist of various machines that produce outputs in terms of color, current, voltage, tension and so on. These measured outputs are described by numbers, and he writes:

> *Some authors state that the last stage in this chain of measurements involves "consciousness," or the "intellectual inner life" of the observer, by virtue of the "principle of psycho-physical parallelism." Other authors introduce a wave function for the entire*

[54] Susskind [2014], page 93

> *universe. In this book, I shall refrain from using concepts that I do not understand.* Peres[55] 1995.

About the difference between conclusions, such as many mind interpretations derived from the mathematical formalism, and the reality he says:

> *Quantum phenomena do not occur in a Hilbert space. They occur in a laboratory.* Peres[56] 1995

Like Peres we try to avoid any weird concepts which are difficult to understand, if at all. This includes even the widely used concept of an *observer*.

[55] Peres [1995], pages 25-26
[56] Peres [1995], page 373

4.2 Base States

All physical theories are based on experimental set-ups or machines that are characterized by a fundamental property, namely the existence of mutually exclusive and empirically decidable *alternatives*. We postulate this property, since a theory without alternatives is hardly imaginable. We call these alternatives in the following *base states*. Let us look at some examples.

In classical mechanics *dynamical variables*, like position, momentum, or energy provide alternatives for describing the freedom of a particle or of a physical system. For example, a particle can be in this or that position, but not simultaneously in several positions. In the same way the particle can occupy one of many different geometrical objects. For instance, a harmonic oscillator may be on one of the orbits in phase space. In a slit experiment the base states are the different positions at the wall of detectors.

In classical probability theory *outcomes*, also called *elementary events*, are the base states. They are the alternatives that define the *sample space*. For a coin the base states are Head and Tail. For a dice the base states are 1,2,3,4,5, and 6. For a slit experiment the base states are the positions at the wall of detectors. When detectors are added at the slits, then the base states are pairs of detectors. The first element of this pair contains the detected position, and the second element contains the located slit. The base states are the same as in classical mechanics, only their names differ.

The same definition of base states applies to quantum mechanics. As before, the positions define the base states in a slit experiment. In optics, a photon passing a calcite polarizer has the possibility to be in one of two beams with perpendicular polarization, yielding two base states, namely horizontal and vertical polarization.

In thermodynamics, the base states are the microstates or the macrostates, depending on whether we consider the microscopic or the macroscopic model, respectively.

In some field theories, the base states are the classical fields, namely the solutions of classical field equations.

> **Keep in mind:** For any physical theory we postulate the existence of mutually exclusive alternatives, the *base states*. A theory without base states is hardly imaginable.

Now we have a precise definition of base states as alternatives, that is, mutually exclusive distinguishable situations. This definition applies to all well-known physical theories. But how can we represent base states? Looking at the previous examples it follows that base states can be represented by numbers. For a dice or a coin toss the base states are natural numbers. Similarly, the base states of other physical models can be represented, at least by grouped numbers. In summary, any experimental set-up can be characterized by a collection of numbers that have a specific meaning in the experimental context. We call this the *number representation* of base states. It is the most commonly used representation in physics.

Additionally, there are two other mathematically equivalent representations. Any number can be represented as a string of bits, called register. Commonly, a *bit* is defined as a quantity that can only take one of two values 0 and 1. It is the basic unit in information theory. The two values are interpreted as logical values TRUE or FALSE. Thus, in addition to the number representation we have an equivalent *register representation* of base states. The register representation is the most commonly used representation in information theory.

In the following, we consider only systems with a finite number N of base states, or at most a countable set of base states. The N registers representing all base states can be expressed equivalently as the orthonormal standard unit vectors in $\mathbb{C}^N$. We call this the *vector representation* of base states, and the transition from the number representation or the register representation to the vector representation is called *vectorization*. The vector representation is widely used in quantum theory, when working with bits and qubits.

Frequently, in these notes our interpretation of a bit is as follows:

- A bit is a question at a system that has exactly two possible answers, say YES for 0 and NO for 1.

For certain systems the base state can be represented by a register consisting of a sequence of binary questions. In such situations we obtain a better insight into certain physical relationships. Then a register of n bits can be represented equivalently as a standard unit vector in $\mathbb{C}^{2^n}$ via the tensor product of the bits in the register. This vector representation is widely used in quantum theory.

The four numbers $0, 1, 2, 3$ have the register and vector representations

$$|00\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, |01\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, |10\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, |11\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{198}$$

We use as before Dirac's bracket notation for representing states.

Summarizing, we have three mathematical equivalent representations of base states: the number, register, and vector representation. This has some important consequences, as we will see later. Among them are consistent definitions of states and observables for various physical models, a vectorized classical mechanics, or a unified treatment of classical mechanics, classical probabilistic mechanics, and quantum theory via semimodules. In the literature quantum mechanics is presented as a linear theory in contrast to classical mechanics. But it turns out that a vectorized classical mechanics is a linear theory as well, working with permutation matrices, a special class of unitary matrices.

Keep in mind: The mutually exclusive alternatives, the *base states*, can be described in terms of three mathematical equivalent representations. The *number representation* that is well-known from classical theories. The *register representation* that is frequently used in information theory. For certain systems this representation shows the divisibility of physical knowledge into decidable YES / NO questions. Finally, the *vector representation* that is used mainly in quantum theory. In particular, classical mechanics has also a vector representation, and thus can be viewed as a linear theory. In the same way quantum mechanics has a number representation, for instance, the componentwise Schrödinger equation. We make use of all three representations. A major reason that physical theories seem to be incompatible or inconsistent sometimes goes back to the ignorance of these equivalent representations.

A first important consequence is the distinguishability of base states. In classical mechanics two base states, such as position or momentum, with different number representations are called *distinguishable*. In quantum mechanics we have to learn the seemingly strange definition that only orthogonal states are distinguishable. It follows that the notion distinguishability seems to depend on the physical theory. Actually, it does not depend on the physical model, but on its representation. We have the obvious natural equivalences:

- Two base states i and j in their number representation are distinguishable if their numbers are different. $\Leftrightarrow$
- Two base states in their register representation are distinguishable if their registers $i = (i_0, ..., i_n)$ and $(j = j_0, ..., j_n)$ are different. $\Leftrightarrow$
- Two base states in their vector representation are distinguishable if their corresponding standard unit vectors $|i\rangle$ and $|j\rangle$ are orthogonal, that is,

$$\langle i|j\rangle = \sum_{k=0}^{n-1} i_k j_k = 0. \tag{199}$$

The latter property follows because different register representations imply different orthogonal standard unit vectors. Orthogonality and distinguishability of base states are exactly the same, only expressed in different representations.

In order to get a feeling for the above notions, we consider the example displayed in Figure 69. It has only six base states. This is sufficient for our purposes, since most of the conceptual issues have nothing to do with an infinite number of base states.

In this example a circuit is chopped into a collection of six cells that are consecutively represented in terms of numbers, namely their corresponding half-open intervals of angles:

$$|[0, \frac{\pi}{3})\rangle, |[\frac{\pi}{3}, \frac{2\pi}{3})\rangle, |[\frac{2\pi}{3}, \pi)\rangle, |[\pi, \frac{4\pi}{3})\rangle, |[\frac{4\pi}{3}, \frac{5\pi}{3})\rangle, |[\frac{5\pi}{3}, 0)\rangle. \tag{200}$$

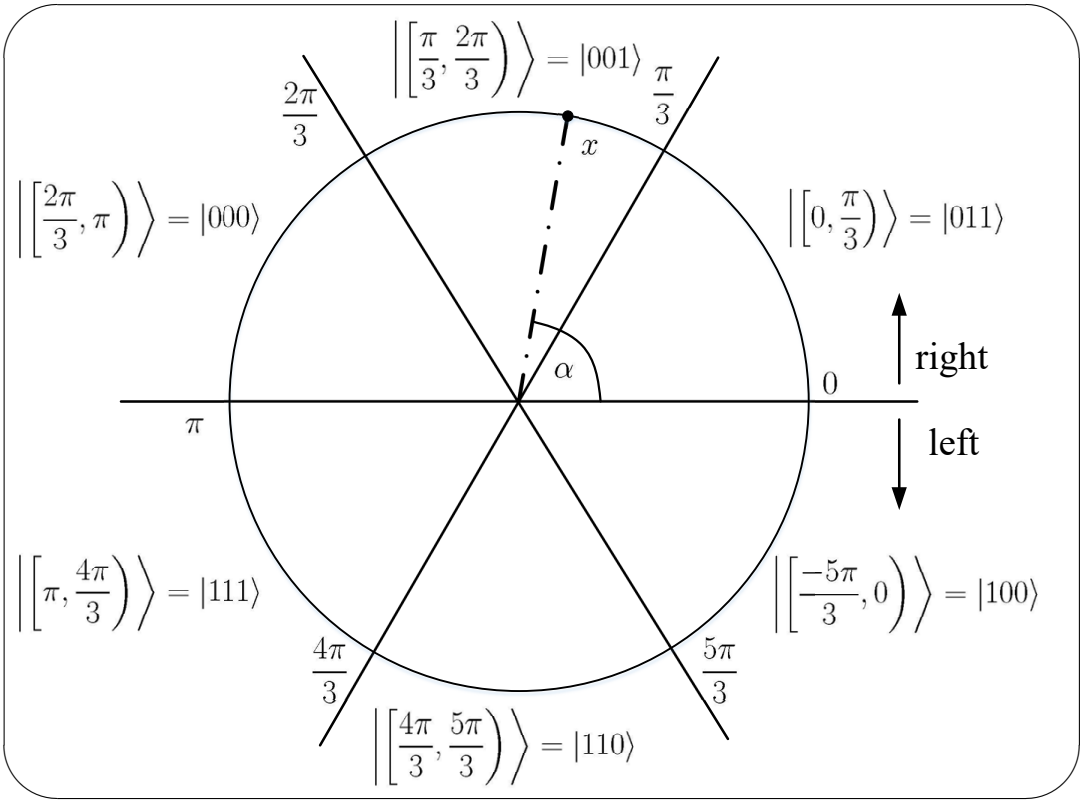


Figure 69: A circuit is chopped into a collection of 6 cells by bisection three times the circuit. A point x is positioned on the circumference. For each line bisecting the circuit we can ask the question whether point x is on the right or on the left side of the line yielding the value 0 or 1, respectively. Hence, each cell with the number representation $|[\frac{k\pi}{3},\frac{(k+1)\pi}{3})\rangle$ has an equivalent register representation $|i_0,i_1,i_2\rangle$ with $i_0,i_1,i_2 \in \{0,1\}$, thus an equivalent vector representation. They represent the 6 cells. This example requires 3 binary alternatives yielding 6 base states where a specific point x may be positioned.

Of course, we could make the cells arbitrarily small, but leave the number of base states finite. Then nothing would change the following considerations.

In many cases, we use for each representation sometimes Dirac brackets in order to keep in mind the fundamental vector representation. This will cause no ambiguities.

We obtain the register representation as follows: we start with the line through 0 and π, and ask whether a fixed specific point x is on the right side or on the left side of this line, when looking from 0 to π. If x is on the right side we write a 0, and otherwise we write a 1. Then we take the line through $\pi/3$ and $4\pi/3$ and ask the same question. Finally, we take the line through $2\pi/3$ and $5\pi/3$ and ask this question once more. Since we have only six base states representing the mutually exclusive cells and $6 \leq 2^3$, we need only three questions that yield a 3-bit register representation.

Obviously, the point x displayed in Figure 69 is in the cell $|[\frac{\pi}{3},\frac{2\pi}{3})\rangle$, which

has the number, register and vector representation

$$|[\frac{\pi}{3}, \frac{2\pi}{3})\rangle = |001\rangle = (01000000)^T \in \mathbb{C}^8, \tag{201}$$

respectively. Here, we have represented a register of 3 bits (questions) as a standard unit vector in $\mathbb{C}^8$ via the tensor product of the bits in the register. The representations of the other cells are displayed in the same figure. If one point is positioned on the circuit, it is in exactly in one of these six non-overlapping cells. Mathematically, these alternatives form an orthonormal set of canonical unit vectors in $\mathbb{C}^8$, when using the vector representation. This justifies the name *base states*.

In our example, the registers $|101\rangle$ and $|010\rangle$ do not occur, since both are impossible when interpreting the related bits as questions. We can avoid such impossible registers when we switch to $\mathbb{C}^6$, which can be viewed as a subspace of $\mathbb{C}^8$. It is interesting that there are examples where impossible registers do not occur. If we bisect the circuit only two times, then we obtain $4 = 2^2$ registers that only represent possible base states. Impossible ones do not occur. Another example is a sphere in the three dimensional real space where an axis passes through the center of the sphere. The plane through the center orthogonal to this axis divides the sphere into two halves, and we can ask the question on which half a point may be positioned. This gives the first bit. Then we rotate the axis. This gives a new plane, and we can ask the same question yielding a second bit. Then we rotate once more and obtain a third bit. The 2^3 corresponding 3-bit registers represent base states where a point possibly may be positioned on the sphere. Taking more than three planes leads to impossible base states. In general, we can always avoid impossible base states by passing to an appropriate subspace.

Another aspect is that these representations are ambiguous, as usual in physics. In the number representation, for instance, we could simply write the numbers 1 to 6. In the register representation, we can interchange 0 and 1 for each bit.

This example can be used in various physical models and situations, some of them are described below. Although it has nothing to do with small particles, we shall see later that it generates the mathematical formalism of quantum mechanics. Let us think of a point as a soccer ball, and think of the circuit as a large hula hoop. Then the cells are an appropriate partitioning of the hula hoop, and we have a simple macroscopic example.

(A) We can number the six base states from 1 to 6 as above yielding another number representation. This set of numbers can also be viewed as the sample space of a dice. The states are called *outcomes* in *classical probability theory*. In particular, it follows that these outcomes have a natural register representation as well as a vector representation. Usually, both representations are not exerted in classical probability theory.

(B) In terms of classical mechanics this example may describe one particle on some position x on the circuit. The classical states of this particle are

the possible positions given by the cells. The cells are the base states that have a number, register, and vector representation. More general, the theory of classical mechanics can be described equivalently in one of these three representations, but only the number representation is widely used. Sometimes it is advantageous to describe classical mechanics in its vector representation.

(C) Classical fields can be described as well. Suppose we want to measure the temperature in a room with a circuit wall. In each cell we have placed an air thermometer at the wall. The temperature in each cell can be written approximately as a finite register, say with 5 bits. One realization of a classical field could be

$$\begin{array}{ll} \text{cell } 011 & \text{temperature } 10100 \\ \text{cell } 001 & \text{temperature } 10011 \\ \text{cell } 000 & \text{temperature } 10010 \\ \text{cell } 111 & \text{temperature } 10001 \\ \text{cell } 110 & \text{temperature } 10011 \\ \text{cell } 100 & \text{temperature } 10010 \end{array} \tag{202}$$

When we merge consecutively the bits of this field, we obtain its register representation

$$|i\rangle = |011\ 10100\ 001\ 10011...\rangle. \tag{203}$$

In a field theory the possible fields $|i\rangle$ of the form (203) define the base states.

(D) The idea of thermodynamical microstates and macrostates can be explored in our simple example as follows. Suppose we have 6 chips, black on one side and red on the other. We toss them and place them in the 6 cells. A *microstate* is a list of 6 elements with a color on each side. The microstates form a set of base states that can binary be represented with 0 for black and 1 for red. This gives their register representation. In our case there are 64 microstates that are equally likely with probability 1/64 for each one.

A *macrostate* is characterized by the total number of black coins, since the others are red. If this number is 1, for example, then 6 microstates have this property and the probability of this macrostate is 6/64. The macrostates form a set of base states that may be labeled as

$|0\rangle, |1\rangle, |2\rangle, |3\rangle, |4\rangle, |5\rangle, |6\rangle.$

(E) We can use this model for describing the dynamics of particles. Fermions can be viewed as soccer balls such that at most one of them is in each cell. If a particle is in a certain cell we write a 1, otherwise we write a zero. Hence, a base state is a 6-bit register, for instance $|100110\rangle$. In its vector

representation it is a standard unit vector in $\mathbb{C}^{2^6}$. This toy model allows us to work with a variable number of particles. It is closely related to a *Fock space* in *quantum field theory*. Particles may be annihilated. For example, the base state $|100110\rangle$ may change to $|000110\rangle$. The vacuum is defined as the space without any particles, that is, the base state $|000000\rangle$. The vacuum is mathematically represented as the standard unit vector $|000000\rangle$. This vector has in the first component a 1 and otherwise zeros. It represents our hula hoop without soccer balls. The annihilation operator maps the vacuum onto the zero vector, since there is no particle that can be annihilated. In reverse, we can create particles in specific cells that contain no particle. If we want to create a particle in a cell which is already occupied, the creation operator maps this base state onto the zero vector since each cell can contain at most one fermion. In particular, a creation operator maps the base state $|111111\rangle$ always onto the zero vector.

Bosons have the property that any integer number of particles can be in one cell. Also in this case we can create and annihilate particles. But now the creation operator does never map a state onto the zero vector. Notice that the creation and annihilation operators act on base states.

Keep in mind: Given a machine or any experimental set-up. Its *base states* are the mutually exclusive alternatives. Two base states are different and distinguishable, if and only if

(i) in the number representation the corresponding numbers are not equal,

(ii) in the register representation the corresponding registers are not equal, and

(iii) in the vector representation the vectors are orthogonal.

4.3 Irreversibility and Trinity of Time

A major goal of these lecture notes is to describe how physics might work without an external time parameter t and a (3+1)-spacetime. This is a radically different entrance to physics. We completely avoid to use a real time variable t. At a first glance, it seems that most physical theories must vanish: no classical theories like electromagnetism, no special or general relativity theory, no quantum field theories, and no particle physics. Actually, the mathematical formalisms of these theories can be recovered, but the interpretations change drastically. It turns out that it is possible to formulate physics as a timeless language without apparent paradoxes, but with unified definitions of states, observables, and the change of states.

Firstly, we need to consider some aspects of the nature of time, perhaps the deepest of all philosophical puzzles. Newton believed that absolute time exists independently of other influences. Mathematically, it means that time appears as an external parameter in the equations of motion characterizing the temporal development of a physical system.

Many physicists and philosophers thought about time and came to diverse definitions. One of the most radical consequences in the theory of relativity is that any moving clock slows down relative to a stationary observer. This phenomenon is called *time dilation*. Moreover, the relativity of simultaneity, that is, whether two spatially separated events occur at the same time, depends on the observer's reference frame, thus is not absolute.

Although quantum theory is unified with special relativity in relativistic quantum theory, the meaning of time becomes even more strange in general relativity. This results in the well-known problem of *quantum gravity* when trying to unify quantum theory with general relativity. One way to deal with this problem is a well-known standard recipe with which one can generate a suitable quantum theory from a Hamiltonian formulation. Applying this recipe to general relativity leads to the *Wheeler-de Witt equation* (we cannot go into details), formally written as $\hat{H}\Psi = 0$, where Ψ is the wave function of the universe, which depends on the geometry and the matter content of the entire space-time. The crucial point is that this equation does not involve any time. It is static and thus has generated much controversial discussions. This is not very surprising, however, since even general relativity theory does not contain a separate time parameter, but describes the entire spacetime as a variable.

Is a quantity t with such strange properties a time or an illusion? However, it seems to be hardly possible to define fundamental physical terms absolute exactly. Feynman wrote in his famous Feynman Lectures Volume I, 8-1:

> *We can't define anything precisely. If we attempt to, we get into that paralysis of thought that comes to philosophers ... one saying to the other:"you don't know what you are talking about". The second one says: "what do you mean by talking? What do you mean by you? What do you mean by know?"*

We try, nevertheless, to define several physical terms as accurately as possible, and in accordance with language, experience, and experiments. If we

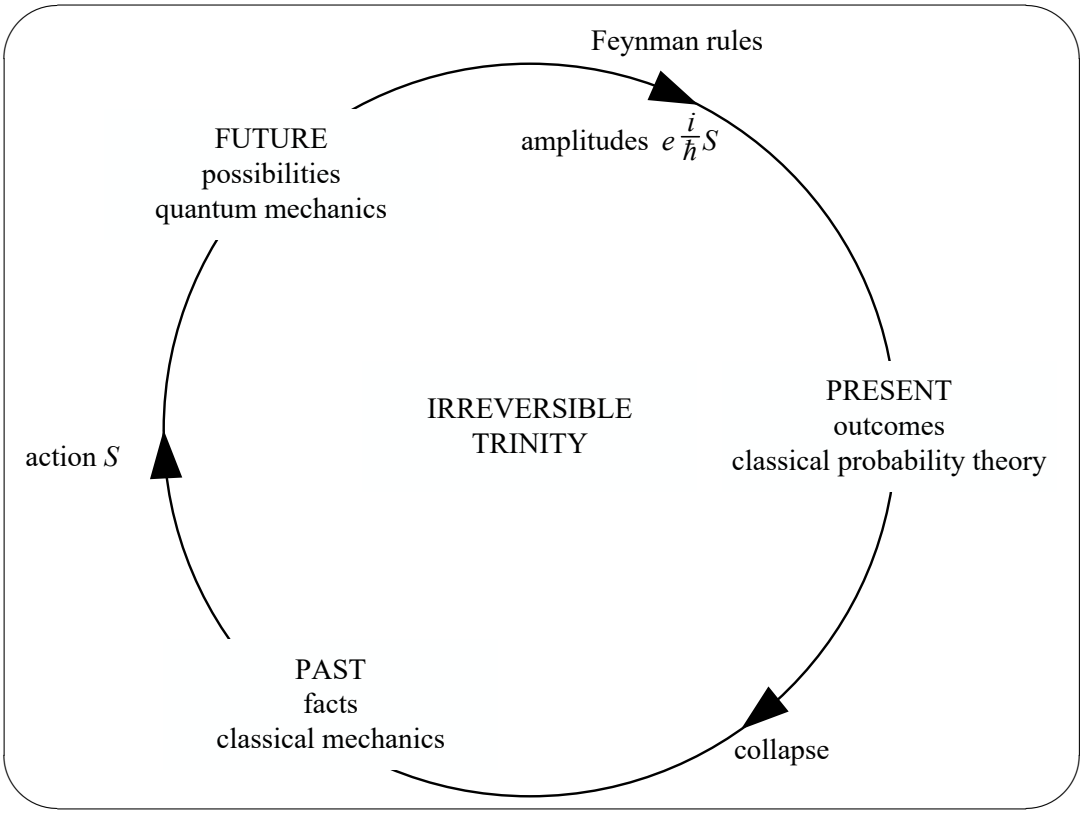


Figure 70: Trinity of time. Go through the slit experiments with this picture in mind.

have an experience about time at all, then of course not about an external time parameter. But it is obvious to everyone that our experiences show that all phenomena behave irreversible, and moreover we observe a trinity of time " future, present, and past". These experiences are unmistakable and not disproved by any experiments.

The *irreversibility* in our world exhibits for instance in the observation that "broken cups do not fit together by themselves". Another funny demonstration is to run a film backwards.

The *past* is fixed. It is characterized by the things that happened, that we do remember, thus by *facts* that exist, that are unique, that are determined, that perhaps we know or that are measured by any machine. The *future* is open, is characterized by all things that might happen, that is, by *possibilities*. Between future and past is the *present*. The present is neither an instant of time nor a time interval. It describes what actually happens, that is, it selects from a set of possibilities exactly one outcome which then in turn becomes a fact in the past.

This trinity is deeply rooted in our behavior, our thinking, our language, our genetics. Perhaps, even Neanderthalers, or even animals, have this experience. This trinity, which may be visualized as a tripartite circle as displayed in Figure 70, forms a basis of this lecture notes.

Perhaps a very very naive but helpful picture is a pinball machine. If this machine is built, there are a lot of possible games and results. These are the

possibilities one has in the future. When you play a game you are in the present. When the game is over the result is a fact in the past. The past can influence the future. Perhaps you have won some additional games.

Keep in mind: Our experience of future, present, past, and irreversibility is deeply rooted in our behavior, our thinking, our language, our genetics. Physics can be formulated on the basis of this experience as a timeless language without apparent paradoxes, but with unified definitions of states, observables, and the change of states. A major goal of this lecture notes is to describe physics without an external time parameter in a simple manner from the very beginning, and thus supporting the *Wheeler-de Witt equation* derived formally from a Hamiltonian formulation of general relativity.

4.4 States

Above, we have replaced time by the trinity future, present, and past. Moreover, we have introduced a common unique notion of base states as mutually exclusive and empirically decidable *alternatives* which are suitable for describing several different physical models. But what is a state? At the beginning of this chapter we have seen that in the literature states are defined in different manners depending on the physical theory. Now, using our definition of base states we can unify the idea of a state.

Given a machine or an experimental set-up which is characterized by a set of base states $\{|i\rangle\}$. A *state* is defined as a (probabilistic) *superposition* of base states

$$|\xi\rangle = \sum_i \xi_i |i\rangle, \qquad \xi_i \in \mathbb{S}, \quad \sum_i |\xi_i|^2 = 1, \tag{204}$$

where $\mathbb{S}$ is a number system, see Section 2.2. In the following we classify states according to the three number systems: the non-negative integers $\mathbb{N}$, the non-negative real numbers $\mathbb{R}_+$, and the complex numbers $\mathbb{C}$.

Firstly, let us consider *two-state systems* that is, we look at binary alternatives. These systems correspond to single bits, that is, YES/NO questions.

$$\text{A} \left\{ \begin{array}{l} \textit{classical} \\ \textit{random} \\ \textit{quantum} \end{array} \right\} \textit{state} \text{ , short form is } \left\{ \begin{array}{l} \textit{bit} \\ \textit{rbit} \\ \textit{qubit} \end{array} \right\}, \tag{205}$$

is represented in the form

$$|\xi\rangle = \xi_0|0\rangle + \xi_1|1\rangle = \begin{pmatrix} \xi_0 \\ \xi_1 \end{pmatrix} \in \left\{ \begin{array}{l} \mathbb{N}^2 \\ \mathbb{R}_+^2 \\ \mathbb{C}^2 \end{array} \right\}, \tag{206}$$

where

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{207}$$

and the coefficients satisfy the normalization condition

$$|\xi_0|^2 + |\xi_1|^2 = 1. \tag{208}$$

The meaning of the coefficients is related to the number system. In a certain sense, however, they have a probabilistic meaning.

For integer coefficients we obtain the *classical states*. The normalization condition implies that either $\xi_0 = 1, \xi_1 = 0$ or $\xi_0 = 0, \xi_1 = 1$. Hence, the only states are the base states, either zero or one. This definition coincides with the definition of a classical bit in Section 3.3. Bits represent *facts*, things that have happened. Facts describe the *past*. When a coin is thrown, for example, it shows either Head or Tail. When a particle has passed an interferometer,

it interacts exactly with one detector. The past is deterministic. Nevertheless, the coefficients ξ_0 and ξ_1 of a classical bit can be viewed as "extreme probabilities", where exactly one has the value 1 and the other is 0.

For *random states* the coefficients ξ_0 and ξ_1 are nonnegative, and their squares sum up to one. These squares $|\xi_0|^2$ and $|\xi_1|^2$ represent the corresponding probabilities for the two outcomes of the random bit. Here we have defined random states in a way such that the coefficients must be squared in order to obtain the classical probabilities. At a first glance, it would be better to define random states as usual in the form

$$|\xi\rangle = \xi_0|0\rangle + \xi_1|1\rangle,\ \xi_0 + \xi_1 = 1,\ \xi_0, \xi_1 \geq 0. \tag{209}$$

But then we must change the normalization condition. Because there is unique correspondence between positive numbers and their squares, it is more appropriate to maintain the normalization condition and to use Born's rule for the square roots of classical probabilities. Then the definition of squared random bits coincides with the definition of a random bit in Section 3.5.

Random states describe experimental outcomes that happen at this moment, that is, in this moment the two-state system collapses to one base state $|0\rangle$ or $|1\rangle$, respectively. Random states represent the *present*. We call the transition from the present to the past and from classical probabilistic outcomes to facts the *collapse*. In a coin toss either Head or Tail emerges. In experiments with an interferometer exactly one detector clicks. In a slit experiment exactly one of the position detectors at the wall clicks. In this interpretation the collapse is not a second dynamics in quantum theory as pointed out in the literature. Instead it happens when going from the present (classical probability) to the past (classical mechanics). This transition is irreversible. We cannot go reversely from the past (unique facts) to the present (several outcomes).

For *quantum states* the coefficients ξ_0 and ξ_1 are complex numbers, the so called *probability amplitudes*. Squaring their magnitudes gives the classical probabilities for outcomes. The definition of quantum bits coincides with the definition of a quantum bit in Section 3.6. The obvious difference to classical probabilities are that complex amplitudes are numbers that allow cancellation or *interference*, in contrast to non-negative numbers.

Quantum states represent the *future* in terms of *"as well as"* distinguishable possibilities. The transition from the future to the present and from quantum probabilistic outcomes to classical probabilities is given by *Born's rule*, namely to square the magnitudes of complex amplitudes. In this interpretation Born's rule is applied when going from the future (quantum amplitudes) to the present (classical probability). This transition is likewise irreversible. Due to the squaring procedure, we cannot go reversely from the present to the future. But the other way around works. The facts of the past affect the possibilities of the future. Changed facts offer new possibilities. A cup must be produced at first. If this is done it belongs to the past. Only then one has the mutually exclusive possibilities to fill the cup with water, tee, or coffee. In this sense we get an irreversible circle of future, present, and past, as displayed in Figure 70.

In physics the *arrow of time* expresses our observed one-way property of time, although almost all equations are reversible with respect to time. The

British astronomer Eddington introduced this concept 1928, and, using several arguments, he concluded that the time's arrow is a thermodynamical property of entropy alone. Physical processes at the microscopic level, however, are believed to be reversible in time. More precisely, if we reverse the direction of time all mathematical statements remain true. Our approach is fundamentally different. Firstly, we have no time t, hence, we cannot reverse its direction. Secondly, we have a circle representing the trinity, not an arrow. In particular; the circle is closed such that our experience "facts influence the future" is incorporated. That physical quantum processes at the microscopic level are reversible can be simply resolved in our treatment of the trinity of time. We describe quantum theory in terms of as future actions or possibilities and reversible unitary transformations that map base states onto base states.

The same mathematical formalism applies not only to two-state systems, but also to *multi-state systems*. These systems can be represented as registers

$$i = (i_0 \dots i_{n-1}), \quad i_j \in \{0, 1\} \tag{210}$$

with vector representation $|i\rangle$. A

$$\left\{ \begin{array}{l} \text{classical} \\ \text{random} \\ \text{quantum} \end{array} \right\} \textit{ state} \text{ or } \textit{register}$$

is a superposition $|\xi\rangle$ of base states $|i\rangle$:

$$|\xi\rangle = \sum_i \xi_i |i\rangle = \begin{pmatrix} \xi_0 \\ \vdots \\ \xi_{N-1} \end{pmatrix} \in \left\{ \begin{array}{c} \mathbb{N}^N \\ \mathbb{R}_+^N \\ \mathbb{C}^N \end{array} \right\}. \tag{211}$$

with vectorized base states

$$|i\rangle = |i_0, ..., i_{n-1}\rangle, \tag{212}$$

satisfying the normalization condition

$$\sum_i |\xi_i|^2 = 1.$$

The meaning of the coefficients is exactly the same as for two-state systems.

Keep in mind: Base states are non-overlapping, distinguishable alternatives that define the facts in the past, the outcomes or elementary events in the present, and the distinguishable possibilities of the future. They do not change when going from the future to the present and then to the past. This is important and is in contrast to states. They are defined as superpositions of base states and differ with respect to the trinity of time: in the past we have classical states, in the present we have classical probabilistic states, and in the future we have quantum states.

4.5 Some Experiments Revisited

In this section we explain the notions base states, states and possibilities in terms of some examples.

Let us first consider a fair die. Its base states are the numbers $1, 2, 3, 4, 5, 6$. They define the outcomes before we toss the die, and exactly one of these numbers becomes a fact if the die is tossed. In this simple example we can also say that the base states provide six distinguishable possibilities. In the future, when the die is not yet tossed, we can assign probability amplitudes $\pm 1/\sqrt{6}$ to the base states. They are not uniquely defined. When we toss the coin in the present, then each base state has the probability 1/6. This follows from Born's rule. The present is best described by the classical probability rules of Kolmogorov. In this theory probabilities are non-negative numbers such that *interference* cannot occur. Probabilities of mutually exclusive events are added and probabilities for independent events are multiplied, the same rules as for probability amplitudes. Finally, when the die is tossed one of the six numbers becomes a fact.

Next, we consider *Renninger's* negative-result experiment, where around a light source two spherical scintillation screens with different radii are placed, see Figure 34. The outer radius is very large compared to the inner radius. The light source emits in sufficiently large "time intervals" a photon that can move in all directions. It was designed to show that knowledge is gained, although not any detection has happened.

In our approach, given the experimental set-up without any emitted photon, we calculate the probability amplitudes for the two future actions: the photon is detected at the inner or at the outer screen. These are all base states. The probability amplitudes are not uniquely defined, but their squared magnitudes must correspond to the areas of both screens. The experimental set-up determines completely the probabilities, the only quantities that can be measured. When a photon is emitted in the present, then each base state has via Born's rule a probability related to the area. Finally, when the photon is detected, one of the base states becomes a fact. That settles the experiment. Since we have no time parameter, we have no time-dependent wave-function, and we cannot gain any knowledge. Nothing strange happens. It is a short exercise to explain in the same way *Schrödinger's cat*.

Some important aspects occur when we look at slit experiments. Suppose that the experimental set-up is given, but no particle is in the experiment. Hence, we ask what might happen in the future. There are several possibilities where a particle can go from the source s through some hole i in the first wall, then through some hole j in the second wall, and finally to a detector x at the screen. The corresponding amplitude related to this possibility or path is

$$\langle x|j\rangle\langle j|i\rangle\langle i|s\rangle. \tag{213}$$

We know already that these probability amplitudes depend on the geometry of the experimental set up, and can be calculated easily. Our next question is: what are the base states? It seems to be natural to identify the base states with the possible paths. But this would not work, since the paths are not

distinguishable, provided we have no detectors at the slits. The particle can move in a future action on this path *as well as* on any other path from source s to detector x. What are the mutually exclusive alternatives? These are the positions. At the screen exactly one detector would observe a particle. Thus, the positions form distinguishable, non-overlapping outcomes. The amplitudes for these *outcomes* $\langle x|s\rangle$ are calculated as follows: the amplitude to arrive from source s at a specific position x is a sum of all amplitudes:

$$\langle x|s\rangle = \sum_{\substack{i=1,\ldots,N \\ j=1,\ldots M}} \langle x|j\rangle\langle j|i\rangle\langle i|s\rangle. \tag{214}$$

The single amplitudes (213) in this superposition correspond to the various non-distinguishable, but mutually exclusive *possibilities*. In particular, possibilities have not the property that *either* this *or* this may happen, as is true for outcomes. Possibilities are in general non-distinguishable elements of the future, where nothing happens. Thus, one possibility *as well as* other possibilities lead to the same outcome (base state) described by formula (214). In other words, possibilities form a set of decisions or paths that may be not distinguishable, in general. Base states are the distinguishable possibilities. However, in some models as in the previous examples the possibilities may coincide with the base states.

In order to obtain the correct probabilities in experiments we have to define the base states very carefully. Moreover, we should have in mind that a base state represents a possibility, but not reversely. When we have additionally detectors at all slits, then the paths (possibilities) through the various slits are distinguishable and become base states. Thus we obtain a completely new and larger set of base states, leading to another probability distribution. In particular, *interference* vanishes. We have discussed these aspects extensively in Section 2.6, and we have seen that the slit experiment is not strange or paradox.

Actually, our unified definition of base states and states has lead us to a deeper insight into slit experiments. Paradoxes vanish. So far we have described how all probability amplitudes can be calculated, and this calculation depends on the geometry of the experimental set-up. In the future no particle is in the experiment. Everything else follows as in the examples before.

Via Born's rule we obtain the probabilities of the present. Finally, the collapse, the transition between present and past, provides the facts in the past. In general, the new facts change the past and thus provide new possibilities in the future.

Feynman's rules describe the transition from possibilities (future) to outcomes (present), from quantum mechanics to classical probability theory. The probability amplitudes are calculated as follows: if a transition occurs in terms of "as well as" possibilities, the probability amplitude for this transition is the sum of all complex probability amplitudes of these possibilities. For transitions, that occur in a series of steps which happen independently, the probability amplitudes are multiplied for each of these steps. This shows the meaning and the content of quantum superposition, and exhibits quantum theory as the mathematical theory of the future that allows *interference*.

We can formulate Feynman's calculations mathematically equivalent in terms of matrix multiplications with unitary matrices. In Section 3.4 we have seen that in the matrix product

$$|\psi'\rangle = \hat{V}(\hat{U}|\psi\rangle) = (\hat{V}\cdot\hat{U})|\psi\rangle. \tag{215}$$

of unitary matrices Feynman's rules are hidden. The input ψ_i evolves to the output ψ_i' via the intermediate values $V_{jk}U_{kj}$. Therefore, we can calculate the probability amplitudes of the slit experiment in terms of the multiplication of three unitary matrices:

$$|\psi'\rangle = \hat{W}(\hat{V}(\hat{U}|\psi\rangle)) = (\hat{W}\cdot\hat{V}\cdot\hat{U})|\psi\rangle. \tag{216}$$

The vector $|\psi\rangle$ describes the state (quantum register) in the space of positions where the source s is placed. An appropriate choice of the probability amplitudes would be the value $\psi_s = 1$ and zero otherwise. Thus, only particles come out of source s. The first matrix $\hat{U}$ describes the transition of this state to the state at the first wall. The second matrix $\hat{V}$ describes the transition of this state to the state at the second wall. Finally, the third matrix $\hat{W}$ describes the transition of this state to the state at the wall of detectors.

Keep in mind: The *superposition* principle in classical probability theory means that *either* this outcome *or* another outcome happens with corresponding non-negative probabilities. The *superposition* principle in quantum theory means that this possibility *as well as* another possibility could happen in the future with corresponding complex probability amplitudes. The reason is that outcomes are distinguishable, but not possibilities, in general.

Another example is a calcite crystal with a polarization axis along the z-axis. This is a two-state system with the two base states, horizontally polarized photons and vertically polarized photons, say $|0\rangle$ and $|\pi/2\rangle$, respectively. Now we take a second calcite with optical axis along an angle α with respect to the z-axis. This apparatus is characterized by the two base states $|\alpha\rangle$ and $|\alpha+\pi/2\rangle$. The law of Malus (11) and (12) implies that the base states of one apparatus can be expressed as an "as well as" superposition of the base states of the other apparatus:

$$|\alpha\rangle = \cos\alpha\,|0\rangle + \sin\alpha\,|\pi/2\rangle. \tag{217}$$

The geometrical coefficients $\cos^2\alpha$ and $\sin^2\alpha$ obtained with Born's rule denote the related probabilities. Note that on the left hand side of this equation is the base state $|\alpha\rangle$ of a calcite. On the right hand side this base state is expressed as a superposition of base states of the other calcite, but these base states must be interpreted as possibilities. This mathematical identity says that the base state $|\alpha\rangle$ can be expressed as a superposition of the possibilities of another calcite in the sense: if in a future action a photon would pass the beam $|\alpha\rangle$ it would pass the beams $|0\rangle$ and $|\pi/2\rangle$ with probabilities $\cos^2\alpha$ and $\sin^2\alpha$, respectively. This is the contents of quantum *superposition* for this example.

Let us now look again at the EPR paradox and Bell's inequality. We derived this inequality by defining eight mutually exclusive base states, each consisting of the three symbols which are either "+" or "−". These eight elements define the sample space. Then we used classical probability theory and considered non-elementary events such as $(+\,+\,\pm) = \{(+\,+\,+), (+\,+\,-)\}$ which consist of two elements. But when we have shown that Bell's inequality is violated, we have considered in the experiment with entangled photons only possibilities with two elements. For example, $(+\,+\,\pm)$ is the possibility where $(+\,+\,+)$ as well as $(+\,+\,-)$ may happen in a future action. Therefore, the classical sample space used in Bell's inequality consisting of eight outcomes differs from the set consisting of four base states when using the quantum rules for the pair of entangled photons. Not surprisingly, different stochastic models imply different statistics.

It is frequently stated that quantum theory is the most fundamental branch in physics that replaces and improves Newtonian mechanics and classical electromagnetism at the atomic level. From our point of view these theories do not contradict each other. They differ because they are part of different time sections, future and past, and use different representations.

The future is timeless. Thus, quantum mechanics, the theory of future actions, can be viewed as a timeless theory. This is also expressed, for instance, in the well-known *Wheeler-de Witt equation*.

The past is timeless as well. Everything has already happened. Classical mechanics is a deterministic theory describing the past. The base states are *facts*, things that have happened. For a harmonic oscillator its orbit in the phase space can be represented in an implicit and an explicit form. The parameter t in the explicit form has only a geometrical meaning. It cannot be a time. Clocks do not work in the past.

Delayed choice experiments can easily be explained when accepting our trinity of time. Since quantum theory is a timeless theory of future actions, the probability amplitudes correspond to possibilities. Particles are not in the experiment. Only until a particle is put in, we switch to the present and apply Born's rule. In our approach there can be no inversion of the normal order of time.

4.6 Semimodules

Physical theories deal with systems of numbers, which share some basic indispensable properties. *Numbers* can be **added** and **multiplied**, and both operations are **associative** and **commutative**. Moreover, they have a **neutral element** 0 for addition, and a **neutral element** 1 for multiplication. Such a number system is called a *commutative semiring* $\mathbb{S}$. In other words, semirings are just "fields without postulating subtraction and division".

There are a lot of semirings that describe well-known number systems used in science. For example, the sets of natural numbers $\mathbb{N}$, integers $\mathbb{Z}$, rational numbers $\mathbb{Q}$, real numbers $\mathbb{R}$, complex numbers $\mathbb{C}$, and the corresponding positive cones $\mathbb{Q}_+$, and $\mathbb{R}_+$. We use $|\mathbb{S}|$ to denote the *cardinality* for any set $\mathbb{S}$.

Semirings first appeared in a work of Dedekind 1894 about the algebra of ideals of a commutative ring. Later very significant applications were found in automata theory, optimization theory, algebraic theory of communicating processes, fuzzy computation, Baysian Networks, and several others. Detailed properties together with references and proofs can be found in the literature[57].

A basic generalization of a linear space over a field, is a *semimodule*. The only difference is the replacement of the field by a semiring.

More precisely, let $\mathbb{S}$ be a semiring. A *semimodule* is a set $\mathbb{M}$ of objects $|\varphi\rangle, |\psi\rangle, \ldots$, the vectors, equipped with an addition $|\psi\rangle + |\varphi\rangle$ and a *scalar multiplication* $r|\psi\rangle$, which satisfies the following conditions for all $r, s \in \mathbb{S}$ and $|\varphi\rangle, |\psi\rangle \in \mathbb{M}$:

1. The addition is associative, commutative, and there exists a unique neutral element, usually denoted by $0 = 0_{\mathbb{M}}$[58], such that $|\psi\rangle = |\psi\rangle + 0$.

2. Both operations satisfy $(rs)|\psi\rangle = r(s|\psi\rangle)$, $r(|\psi\rangle + |\varphi\rangle) = r|\psi\rangle + r|\varphi\rangle$, and $(r + s)|\psi\rangle = r|\psi\rangle + s|\psi\rangle$.

3. The neutral elements satisfy $1|\psi\rangle = |\psi\rangle$ and $r0_{\mathbb{M}} = 0_{\mathbb{M}} = 0|\psi\rangle$.

If the semimodule is the field of real numbers, we obtain the real linear space, the fundamental space of classical physics. If the semimodule is the field of complex numbers, we obtain the complex Hilbert space, the fundamental space of quantum mechanics. The set of non-negative real numbers lead to the cone of non-negative vectors, the semimodule used in classical probability theory.

Keep in mind: The simplest way to remember a semimodule, is to take the rules of a linear space and to replace the word "field" by a "field without subtraction and division". This suggests that many properties of linear spaces and linear operators transfer without any difficulty to semimodules.

[57] Tan [2014], Tan [2014]

[58] We don't use the symbol $|0\rangle$ for the neutral element 0 of addition in order to avoid confusion with the standard unit vector $|0\rangle$ for the register 0=(0,...,0).

The most obvious and important example of a semimodule is defined as follows: given a finite or at least a countable index set $M = \{1, 2, ..., m, ...\}$, we define a *vector* as a vertical list of numbers of a given semiring $\mathbb{S}$:

$$|\psi\rangle = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_m \\ \vdots \end{pmatrix}, \tag{218}$$

where the number $\psi_m \in \mathbb{S}$ is assigned to the element $m \in M$. With $\mathbb{S}^M$ we denote the set of all such vectors with index set M. Defining the vector addition

$$|\psi\rangle + |\varphi\rangle = (\psi_1 + \varphi_1, \psi_2 + \varphi_2, \ldots, \psi_m + \varphi_m, \ldots)^T \tag{219}$$

and the scalar multiplication

$$r|\psi\rangle = (r\psi_1, r\psi_2, \ldots, r\psi_m, \ldots)^T, \tag{220}$$

it follows that $\mathbb{S}^M$ is a semimodule.

The well-known concepts from linear spaces over fields can be generalized immediately to semimodules. We can transpose column vectors to row vectors, and we can define their conjugate complex. The conjugate complex operation $*$ is well-defined, since we consider only subsets of the complex numbers. We define the *standard inner product* as usual

$$\langle\phi|\psi\rangle = |\phi\rangle^\dagger|\phi\rangle = \phi_1^*\psi_1 + \phi_2^*\psi_2 + \ldots + \phi_m^*\psi_m + \ldots. \tag{221}$$

Two vectors $|\psi\rangle, |\phi\rangle \in \mathbb{M}$ are called *orthogonal*, denoted by $|\psi\rangle \perp |\phi\rangle$, if $\langle\phi, \psi\rangle = 0$.

The canonical unit vectors $\{e_1, e_2, \ldots, e_m, \ldots\}$ in the semimodule $\mathbb{S}^M$ are orthogonal for the standard inner product. For the semimodules with semirings $\mathbb{S} = \mathbb{N}$ and $\mathbb{S} = \mathbb{R}_+$ the product of two numbers is always nonnegative. Hence, two vectors $|\psi\rangle, |\phi\rangle \in \mathbb{S}^n$ are orthogonal iff $\phi_m\psi_m = 0$ for all $m \in M$. In other words, both vectors are orthogonal, if their supports are disjoint. The support of a vector is the set of indices where the corresponding coefficients are nonzero.

Subsemimodules, *linear independence* of vectors, *orthogonal basis*, the addition and multiplication of matrices, *invertible* matrices, eigenvalues and eigenvectors, and many other issues are rather similar to those in linear algebra.

Semimodules allow to formulate mathematical characteristics of various physical theories in the same manner. In Section 4.4 we have established a unified concept of states for classical mechanics, statistical mechanics, and quantum mechanics. States are defined as superpositions of base states. The latter describe mutually disjoint alternatives that are obtained by asking binary questions to the system under consideration. This approach leads in a natural and easy understandable way to the concept of a state space.

Principle 1 (state space): To every object, process or experimental set-up a semimodule over a semiring $\mathbb{S}$ is associated, called the *state space*. The normalized vectors $|\psi\rangle$ in this space are called *state vectors*, or shortly states.

For $\mathbb{S} \in \{\mathbb{N}, \mathbb{R}_+, \mathbb{C}\}$ the states are called classical, random, or quantum states, respectively.

Notice that at the beginning of this section we started with an imprecise notion of states, depending on concrete physical models. Now we arrived at a reasonable, unified, and uncomplicated principle defining the state space of several physical models.

The frequent statements that points are the states in classical mechanics and vectors or rays in a Hilbert space are the states in quantum mechanics is misleading. In contrast, vectors are the states in both theories, after vectorization of the theory. The difference is that the coefficients depend on the chosen semiring.

Now the state-space Principle 1 in Section 1 turns out to be a special case of our Principle 1. The latter occurs as a simple consequence of asking questions and of the vectorization of registers.

4.7 Change of States

So far we have been dealing with a unified definition of states as superpositions of base states. The base states can be described as registers which correspond to canonical unit vectors. In this section we investigate how states may change. The essential condition of motion in physical theories is that states are transformed to states. Let's work through different theories.

In classical mechanics, states and base states coincide as we have seen. Hence, states are transformed to states if and only if standard unit vectors are transformed to standard unit vectors. Therefore, only the components of the vectors must be permuted. In other words, the transformations of states are permutation matrices. They have only coefficients $0, 1 \in \mathbb{S} = \mathbb{N}$. In classical reversible computation these matrices are called gates.

In probability theory, the states

$$|\xi\rangle = \sum_i \xi_i |i\rangle, \quad \xi_i \geq 0, \quad \sum_i |\xi_i|^2 = 1 \tag{222}$$

correspond in a unique manner to the squared states

$$|\xi\rangle^2 = \sum_i |\xi_i|^2 |i\rangle, \quad \xi_i \geq 0, \quad \sum_i |\xi_i|^2 = 1, \tag{223}$$

since the coefficients ξ_i are assumed to be non-negative. Hence, states are transformed to states if and only if the squared states are transformed to squared states. This can be done by *left stochastic matrices*, as we have already seen in Section 3.5.

Here we see one reason why random computation is irreversible in contrast to classical reversible computation and quantum computation: the square root of a non-negative number is uniquely defined in $\mathbb{S} = \mathbb{R}_+$, but not in $\mathbb{C}$. In other words, in $\mathbb{C}$ the superposition (223) has many solutions (222).

In quantum mechanics, states are linear combinations of base states with complex coefficients. States are transformed by reversible unitary matrices with coefficients in $\mathbb{C}$ as we already know.

In summary, we obtain the following unified description of the change of states:

Principle 2 (change of states): The change of state vectors is $\mathbb{S}$-linear, that is, it is described by matrices with coefficients in $\mathbb{S}$.

We have proved this principle for $\mathbb{S} \in \{\mathbb{N}, \mathbb{R}_+, \mathbb{C}\}$, but it holds true also for several other semirings. It generalizes Principle 5 in Section 1 to classical mechanics and classical probability theory.

4.8 Composition Rules

In Section 3.4 we have considered in detail the composition of gates. They are matrices that transform between states. Using semimodules, we can summarize compositions in the following unified form:

Keep in mind:

$$\left\{ \begin{array}{l} \text{Classical} \\ \text{Random} \\ \text{Quantum} \end{array} \right\} \text{ gates}$$

or transformations change states. They correspond to

$$\left\{ \begin{array}{l} \text{permutation} \\ \text{stochastic} \\ \text{unitary} \end{array} \right\} \text{ matrices}$$

with coefficients in the semiring

$$\left\{ \begin{array}{l} \mathbb{N} \\ \mathbb{R}_+ \\ \mathbb{C}, \end{array} \right\}$$

respectively. They are composed in

$$\left\{ \begin{array}{l} \text{series} \\ \text{parallel} \\ \text{controlled} \end{array} \right\}$$

by using the

$$\left\{ \begin{array}{l} \text{matrix product} \\ \text{tensor product} \\ \text{direct sum} \end{array} \right\},$$

respectively.

This generalizes and unifies Principle 6 in Section 1 to classical mechanics, classical probability theory, and quantum theory.

4.9 Alternative Bases

We have described registers in their vector representation form. It was shown that n-bit registers are represented as standard unit vectors in a semimodule $\mathbb{S}^N$. These vectors form an orthonormal basis in this semimodule. Now we ask whether we can find alternative orthonormal bases?

For the semiring $\mathbb{S} = \mathbb{N}$ the normalization condition implies that each normalized vector must have zero components with exception of one component that is equal to one. Hence, each normalized vector is a standard unit vector, and thus the basis of standard unit vectors is the only one in classical mechanics.

In classical probability theory, where $\mathbb{S} = \mathbb{R}_+$, all coefficients are non-negative, and two states

$$|\xi\rangle = \sum_i \xi_i |i\rangle, \quad |\varphi\rangle = \sum_i \varphi_i |i\rangle \tag{224}$$

are orthogonal, thus distinguishable, if

$$\langle \xi | \varphi \rangle = \sum_i \xi_i^* \varphi_i = 0. \tag{225}$$

Therefore,

$$\xi_i \varphi_i = 0 \quad \text{for all} \quad |i\rangle \in \mathbb{C}^N. \tag{226}$$

Hence, the intersection of the supports of both states must be empty. Since each orthonormal basis in $\mathbb{C}^N$ has N vectors, it follows that in classical probability theory there exists exactly one orthonormal basis, as in classical mechanics.

In quantum physics the situation changes fundamentally. Actually, many other orthonormal bases are possible. These include rotations of an orthonormal basis, but also seemingly unusual choices such as entangled states, like the Bell basis described in Section 3.6.

Another example is light. A photon is described as an electric field which oscillates with respect to some given axis horizontally or vertically. Both base states are denoted by $|0\rangle$ and $|1\rangle$. Light can also be polarized with respect to some other axis. If we rotate the axis at an angle $\pi/4$, then we obtain two other base states, namely the orthonormal ones $|+\pi/4\rangle$ and $|-\pi/4\rangle$ that can be expressed as superpositions of the orthonormal base states $|0\rangle$ and $|1\rangle$:

$$|+\pi/4\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad |-\pi/4\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle), \tag{227}$$

For a better understanding, let us consider the example in Section 4.2 once more, but in its simplest form as a two-state system, see Figure 71.

The standard unit vectors

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{228}$$

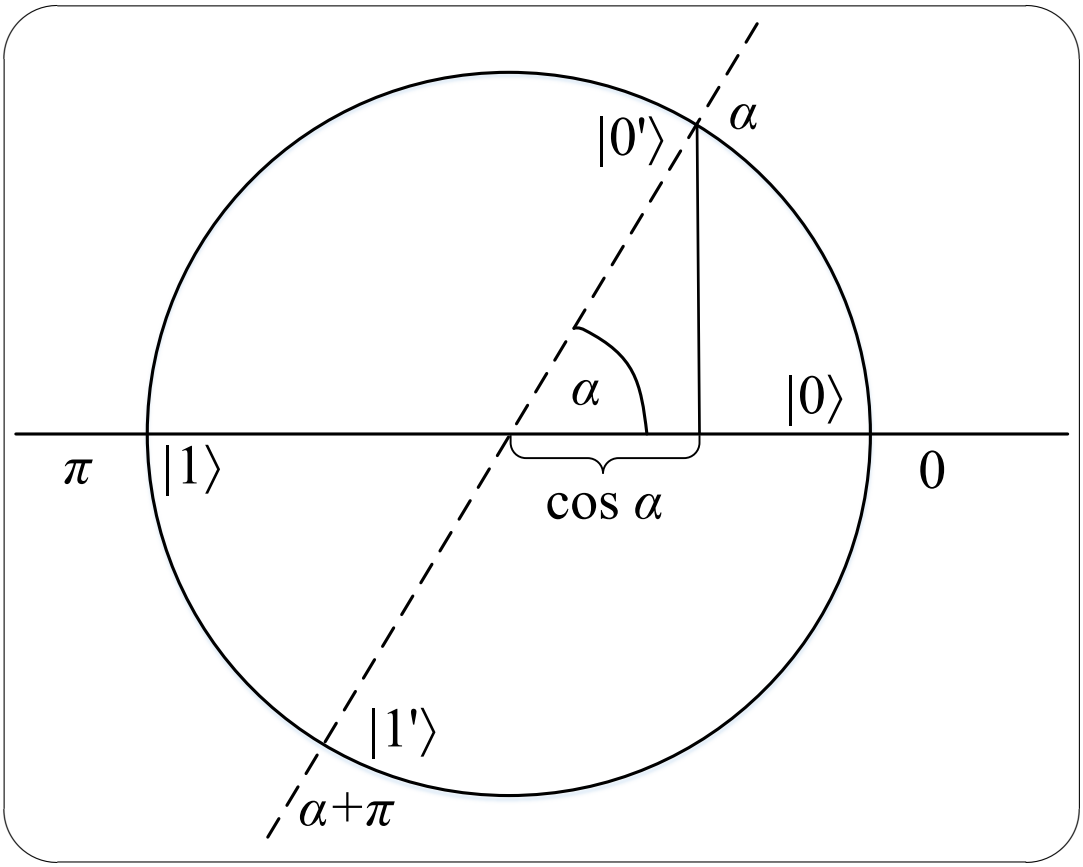


Figure 71: The solid line divides the circuit into two parts, the right and the left side of the line through 0 and π, denoted by $|0\rangle$ and $|1\rangle$. The dashed line has an angle α to the solid line and divides the circuit into the parts $|0'\rangle$ and $|1'\rangle$.

describe on which side of the line through 0 and π a point on the circuit can be positioned. These vectors represent this bit, and they form an orthonormal basis in $\mathbb{C}^2$.

From Figure 71 it follows that for $\alpha = 0$ we obtain $|0'\rangle = |0\rangle$, $|1'\rangle = |1\rangle$. For $\alpha = \pi$ we get $|0'\rangle = -|1\rangle$, $|1'\rangle = |0\rangle$.

For general α we just rotate with angle $\alpha/2$:

$$\begin{aligned} |0'\rangle &= \cos\frac{\alpha}{2}|0\rangle - \sin\frac{\alpha}{2}|1\rangle, \\ |1'\rangle &= \sin\frac{\alpha}{2}|0\rangle + \cos\frac{\alpha}{2}|1\rangle. \end{aligned} \tag{229}$$

This coincides with the formulas for $\alpha = 0$ and $\alpha = \pi$. In the case $\alpha = \pi/2$, it follows

$$\begin{aligned} |0'\rangle &= \frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}, \\ |1'\rangle &= \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}. \end{aligned} \tag{230}$$

Geometrically, the semi-circle $|0'\rangle$ consists half of the semi-circle $|0\rangle$ and half of the semi-circle $|1\rangle$. The same holds true for the semi-circle $|1'\rangle$. Since $1/2 = (1/\sqrt{2})^2$, the squared magnitudes of the coefficients in (230) are the related lengths of the circle segments. This holds true in the general case (229). Thus, we have derived the rule of Born and Malus from a purely macroscopic geometrical example, namely on which side of a line a point on the circuit may be placed. It has nothing to do with polarization, spin, or small particles. However, this geometrical point of view is not really surprising. The complex probability amplitudes $e^{iS/\hbar}$ in slit experiments use the action S, which is a fundamental geometrical concept related to curves.

Mathematically, we need to know how the eigenvectors in two orthonormal bases, say $\{|i\rangle\}$ and $\{|i'\rangle\}$, are related. Both bases have the same number of vectors, and we assume an order where each $|i\rangle$ corresponds uniquely to a primed $|i'\rangle$.

We ask, how a particular state $|\xi\rangle$, represented in the $\{|i\rangle\}$-basis

$$|\xi\rangle = \sum_i \xi_i |i\rangle, \qquad \xi_i \in \mathbb{C}, \quad \sum_i |\xi_i|^2 = 1, \tag{231}$$

can be represented in the $\{|i'\rangle\}$-basis, that is,

$$|\xi\rangle = \sum_{i'} \xi'_{i'} |i'\rangle, \qquad \xi'_{i'} \in \mathbb{C}, \quad \sum_{i'} |\xi'_i|^2 = 1. \tag{232}$$

Both bases are orthonormal. Hence, calculating the inner products of state $|\xi\rangle$ with the basis vectors yields the amplitudes

$$\xi_i = \langle i|\xi\rangle \text{ and } \xi'_{i'} = \langle i'|\xi\rangle. \tag{233}$$

Now, we need to know how both orthonormal bases are related. It is easy to see that the operator

$$\hat{U} = \sum_i |i'\rangle\langle i|, \tag{234}$$

where $\langle i|$ is the corresponding conjugate transposed vector, is unitary. Both bases are orthonormal, yielding

$$|j'\rangle = \hat{U}|j\rangle. \tag{235}$$

Thus, the unitary operator $\hat{U}$ maps the unprimed basis onto the primed basis, and vice versa the inverse unitary operator $\hat{U}^\dagger$ maps the primed basis onto the unprimed basis.

It is easy to express the amplitudes $\xi'_{j'}$ as follows:

$$\xi'_{j'} = \langle j'|\xi\rangle = \langle j'|(\sum_i |i\rangle\langle i|)\xi\rangle = \sum_i \langle j'|i\rangle\langle i|\xi\rangle, \tag{236}$$

where we have inserted the identity $\sum_i |i\rangle\langle i| = \hat{1}$. From equation (235) we obtain

$$\xi'_{j'} = \sum_j \langle j|\hat{U}^\dagger|i\rangle\langle i|\xi\rangle = \sum_i (\hat{U}^\dagger)_{ji}\xi_i. \tag{237}$$

Thus, the column vector of amplitudes $(\xi'_{j'})$ representing the state $|\xi\rangle$ in the primed basis is obtained from the amplitudes (ξ_i) by matrix-vector multiplication with the unitary matrix defined as $(\hat{U}^\dagger)_{ji} = \langle j|\hat{U}^\dagger|i\rangle$.

Summarizing, these formulas show how a given state can be represented in different orthonormal bases. The proper way to think of orthonormal bases is to think in terms of observables. This is discussed in the next section.

4.10 Observables

In physics observables are associated to experiments. They describe things that can be measured like position, momentum, energy, spin, temperature, or an electric field. Just as states are defined in the literature with respect to a certain physical model, the concept of observables depends on the model.

In classical mechanics, an *observable* is a *dynamical variable*, that is, a real-valued function defined on the phase space, like position, momentum, or energy.

In probability theory observables are usually called random variables. A *random variable* is a real-valued function defined on a sample space of outcomes. This function assigns a real number, thus a label, to each outcome or elementary event. However, notice that a function itself is neither random or a variable.

Let a random variable be defined on a finite sample space I, say $A : I \to \mathbb{R}$ with values A_i where $i \in I$. Let us assume that probabilities $\mathrm{Prob}(i)$ are given for the outcomes $i \in I$. This collection of probabilities is called a discrete *probability distribution*. For the random variable A the *expectation value* is defined as

$$\langle A \rangle = \sum_i A_i \, \mathrm{Prob}(i), \tag{238}$$

and the *variance* takes the form

$$(\Delta A)^2 = \langle (A - \langle A \rangle)^2 \rangle = \langle A^2 \rangle - \langle A \rangle^2. \tag{239}$$

In quantum mechanics states are nonzero vectors $|\xi\rangle$ in a Hilbert space. Observables are defined as Hermitian operators $\hat{A}$ that act on states. Observables can be measured with the non-intuitive property that the system's state collapses non-deterministic and irreversible to an eigenvector of the operator. This is well-known under the name *collapse postulate*. Notice that in our approach, using the trinity of time, the collapse is not part of quantum theory, but describes the transition from the present to the past. The *expectation value* of a quantum observable is approximately the average value that we get from a large number of measurements. It is defined with respect to a normalized state $|\xi\rangle$ as

$$\langle \hat{A} \rangle_\xi = \langle \xi | \hat{A} | \xi \rangle, \tag{240}$$

and the *variance* is

$$(\Delta \hat{A})^2 = \langle \hat{A}^2 \rangle_\xi - \langle \hat{A} \rangle_\xi^2. \tag{241}$$

Our aim is to give a unified definition of observables that is independent of the physical model, similar as the definition of a state.

The unified definition is as follows: given an orthonormal basis, say

$$|I\rangle = \{ |i\rangle \in \mathbb{C}^{2^n} : i = i_0 ... i_n \in I \}, \tag{242}$$

where the base state $|i\rangle$ is the vector representation of an equivalent register or number representation i. An *observable* A is a map

$$A : I \to \mathbb{R},\ A_i = A(i),\ i \in I. \tag{243}$$

Obviously, this definition coincides with the concepts of dynamical variables and random variables. It is defined independently of any probability distribution $|\xi\rangle$.

But does this definition fit into quantum mechanics? Yes, when we switch to the equivalent vector representation, a register i is transformed to a standard unit vector $|i\rangle$. Consequently the function A becomes a diagonal matrix

$$\hat{A} = \begin{pmatrix} A_0 & & & & & \\ & \ddots & & & & \\ & 0 & \ddots & 0 & & \\ & & & \ddots & & \\ & & & & A_{2^n-1} \end{pmatrix} \tag{244}$$

such that the eigenvalue equation

$$\hat{A}|i\rangle = A_i|i\rangle \tag{245}$$

is fulfilled. In particular, it follows that

$$\langle i|\hat{A}|i\rangle = A_i,\ \langle i|\hat{A}|j\rangle = 0 \text{ for } i \neq j, \tag{246}$$

and

$$\hat{A} = \sum_i A_i |i\rangle\langle i|. \tag{247}$$

Hence, the transition from the register representation to the vector representation corresponds to a transition from a real valued function to a Hermitian diagonal matrix.

Observables are defined on a set of registers as functions, or equivalently on a set of base state vectors as matrices. Hence, they do not act on a space of states, or transform states like stochastic or unitary matrices. But what is their interplay with states? Let us look at a superposed state

$$|\xi\rangle = \sum_i \xi_i |i\rangle, \quad \sum_i |\xi_i|^2 = 1, \tag{248}$$

where the coefficients ξ_i are elements of a semiring $\mathbb{S} \in \{\mathbb{N}, \mathbb{R}_+, \mathbb{C}\}$. The squared magnitude $|\xi_i|^2$ defines the probability $\text{Prob}(i)$ for a system to be in base state $|i\rangle$. Then the classical formulas (238) and (239) define the expectation value and the variance. We obtain

$$\langle \hat{A} \rangle = \langle \hat{A} \rangle_\xi = \sum_i A_i |\xi_i|^2 = \langle \xi | \hat{A} | \xi \rangle, \tag{249}$$

and

$$(\Delta A)^2 = \langle \hat{A}^2 \rangle_\xi - \langle \hat{A} \rangle_\xi^2. \tag{250}$$

This follows immediately, since

$$\langle\xi|\hat{A}|\xi\rangle = \sum_{i,j} \xi_j^* \xi_i \langle i|A_j|j\rangle\langle j|i\rangle = \sum_i A_i \xi_i^* \xi_i = \sum_i A_i \text{ Prob}(i). \tag{251}$$

In summary, we have obtained a unified mathematical framework for observables, expectation value, and variance in classical mechanics, classical probability theory, and quantum mechanics. To go from one theory to the other we have only to choose the appropriate semiring $\mathbb{S} \in \{\mathbb{N}, \mathbb{R}_+, \mathbb{C}\}$.

At a first glance, this result seems to be surprising, perhaps wrong. In the classical theories we have only diagonal matrices. But in quantum mechanics also non-diagonal Hermitian matrices are allowed. What is the reason for this discrepancies? We have seen in Section 4.9 that in the classical cases with semirings $\mathbb{S} \in \{\mathbb{N}, \mathbb{R}_+\}$ there exists only one orthonormal basis, due to the inner product in the related semimodule. Contrary, in quantum mechanics where $\mathbb{S} = \mathbb{C}$ various other orthonormal bases exist, since orthogonality does not imply that the supports of base states must be disjoint. Cancellation can occur.

Keep in mind: The strange looking Principle 2 in the introduction, telling that "to every observable of a quantum system a Hermitian operator acting on a quantum state space is associated" turns out to be an old acquaintance from classical theories. Now we have a unified definition. What seems to be strange, can easily be explained with the transition from the number representation to the equivalent vector representation where functions become matrices, and the existence of various orthonormal bases in quantum mechanics, the theory of future events.

In quantum mechanics there exist different orthonormal bases. The representation of a state depends on the used bases. In Section 4.9 we have seen how a particular state represented in one basis can be represented in another one. In the same way we can view a matrix in different bases. Let us consider a matrix $\hat{A}$, where its coefficients $A_{ji} = \langle j|\hat{A}|i\rangle$ are defined with respect to the unprimed basis $|i\rangle$. Now we calculate its matrix elements $\langle j'|\hat{A}|i'\rangle$ for the primed basis $|i'\rangle$. We insert the identity matrix $\sum_i |i\rangle\langle i|$ twice such that

$$\langle j'|\hat{A}|i'\rangle = \langle j'|(\sum_k |k\rangle\langle k|)\ \hat{A}\ (\sum_l |l\rangle\langle l|)|i'\rangle, \tag{252}$$

yielding

$$\langle j'|\hat{A}|i'\rangle = \sum_k \sum_l \langle j'|k\rangle\langle k|\hat{A}|l\rangle\langle l|i'\rangle. \tag{253}$$

Using equation (235), it follows that

$$\langle j'|\hat{A}|i'\rangle = \sum_k \sum_l \langle j|\hat{U}^\dagger|k\rangle\langle k|\hat{A}|l\rangle\langle l|\hat{U}|i\rangle = \sum_k \sum_l (\hat{U}^\dagger)_{jk} \hat{A}_{kl} \hat{U}_{li}. \tag{254}$$

Thus, in matrix form we get the matrix product

$$\hat{A}' = \hat{U}^\dagger\ \hat{A}\ \hat{U}. \tag{255}$$

This is a *similarity transformation*, an equation encountered routinely in linear algebra.

It is a simple exercise and a well-known fact that similarity transformations do not change the eigenvalues, the determinant, the expectation value, and the variance of an observable. Thus, the characteristic quantities of an observable are invariant under a change of the orthonormal bases, fortunately. Otherwise the concept of an observable would be meaningless.

Keep in mind: The column vector $|\xi'\rangle$ of amplitudes representing some quantum state in the primed basis is obtained from the amplitudes of this state in the unprimed basis by matrix-vector multiplication with a unitary matrix, that is, $|\xi'\rangle = \hat{U}|\xi\rangle$. For a matrix a change of the basis is expressed as a similarity transformation, that is, $\hat{A}' = \hat{U}^\dagger\ \hat{A}\ \hat{U}$. The eigenvalues, the expectation value and the invariance of an observable $\hat{A}$ are invariant under a change of the orthonormal basis.

Let us consider again the example described in Section 4.9, where a line divides a circuit into two parts. The orthonormal basis $\{|0\rangle, |1\rangle\}$ describe the bit related to the question whether a point is on the right or on the left side of the line through 0 and π. The map

$$\begin{aligned} &\sigma_z: \quad \left\{|0\rangle = \begin{pmatrix}1\\0\end{pmatrix}, |1\rangle = \begin{pmatrix}0\\1\end{pmatrix}\right\} \quad \to \quad \{-1, +1\}, \\ &\sigma_z|0\rangle = +1, \quad \sigma_z|1\rangle = -1 \end{aligned} \tag{256}$$

is an observable with values 1 and (-1). With vectorization using (247) the observable becomes a matrix $\hat{\sigma}_z$. Its matrix representation is

$$\hat{\sigma}_z = 1 \cdot \begin{pmatrix}1\\0\end{pmatrix}\begin{pmatrix}1 & 0\end{pmatrix} + (-1)\begin{pmatrix}0\\1\end{pmatrix}\begin{pmatrix}0 & 1\end{pmatrix} = \begin{pmatrix}1 & 0\\0 & -1\end{pmatrix}. \tag{257}$$

Now we rotate the line with angle $\pi/2$, and obtain from (230) the orthonormal basis $|0'\rangle$ and $|1'\rangle$.

The observable

$$\begin{aligned} &\sigma_x: \quad \left\{|0'\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}, |1'\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}\right\} \quad \to \quad \{-1, +1\}, \\ &\sigma_x|0'\rangle = -1, \quad \sigma_x|1'\rangle = 1 \end{aligned} \tag{258}$$

has a matrix representation (247) with eigenvalues -1 and $+1$:

$$\begin{aligned} \hat{\sigma}_x &= -1\frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix}1 & -1\end{pmatrix} + 1\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\end{pmatrix} \\ &= -\frac{1}{2}\begin{pmatrix}1 & -1\\-1 & 1\end{pmatrix} + \frac{1}{2}\begin{pmatrix}1 & 1\\1 & 1\end{pmatrix} = \begin{pmatrix}0 & 1\\1 & 0\end{pmatrix}. \end{aligned} \tag{259}$$

Our simple geometric two-state example leads to the Pauli matrices $\hat{\sigma}_z$ and $\hat{\sigma}_x$. Notice that this example doesn't require the imagination of small and

invisible particles. This mathematical framework appears for a macroscopic problem, namely that a soccer ball may be placed on a large circle, and we ask on which side of a line the ball may be.

4.11 Geometry, Polarization, and Spin

Quantum theory, including quantum field theories, is known to be the fundamental branch of physics concerned with very small particles like atoms, electrons, or photons. There, the rather difficult to understand concepts such as polarization, spin or chirality are mathematically described as quantum two-state systems. There arises at least two questions. To what extent have these concepts to do with small particles? Is this mathematical framework only limited to objects outside our direct experience, or does it occur in our daily life also? It will pay off to read this section carefully.

We look at a large sphere where an axis, called the z-axis, passes through the center of the sphere. The plane through the center orthogonal to the z-axis is called the xy-plane. It partitions the sphere into two parts, see Figure 72.

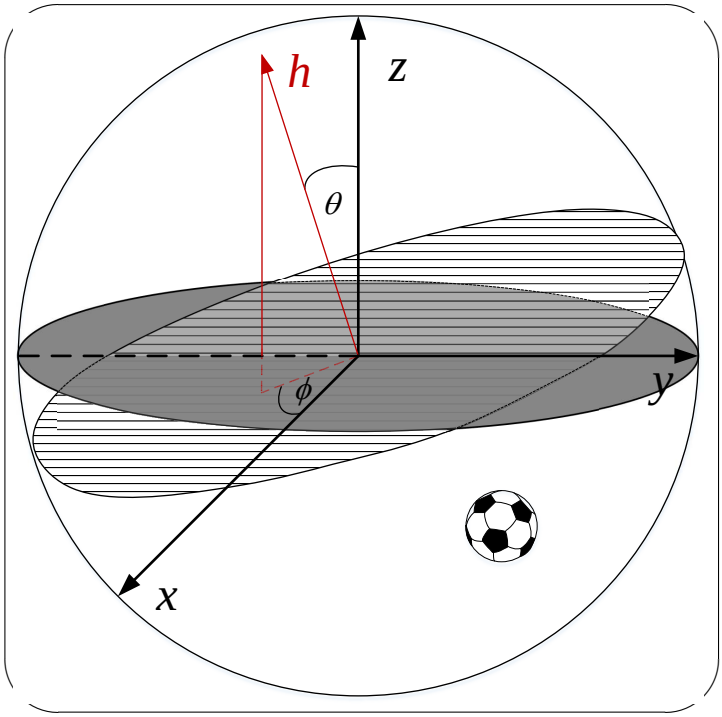


Figure 72: A socker ball on a large spherical surface. The questions "on which side of the xy-, yz-, and xz-plane is the soccer ball?" yield immediately to the quantum probabilistic framework that describes also spin and polarization. The probabilities correspond to cut-out areas, as will be described below. The three coordinates of a normal vector correspond to two cut-out areas described by complex numbers. This visualizes the well-known mathematical isomorphism between the algebras so(3) and su(2), see the appendix for the related definitions.

The two parts of the sphere represent two mutually exclusive possibilities to position, say a soccer ball, on the spherical surface. Now we ask the simple question: on which half of the sphere is the soccer ball? Our goal is to build a mathematical framework that encompasses everything about a ball on the

spherical surface, at a first glance a seemingly trivial macroscopic problem.

There are two mutually exclusive answers, namely the ball is on the upper half or on the lower half. This question represents a bit with two base states $|0z\rangle$ and $|1z\rangle$, respectively. For its vectorized form we set

$$|0z\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1z\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{260}$$

If the ball is already positioned on the spherical surface, the action to position the ball belongs to the past, and the ball is *either* in the half $|0z\rangle$ *or* in the half $|1z\rangle$. This is the point of view of classical mechanics, the theory of *facts* that can be described with the semiring $\mathbb{S} = \mathbb{N}$.

The action "position the ball" happens in the present. There can be only two outcomes: position the ball *either* in the half $|0z\rangle$ *or* in the half $|1z\rangle$ with a specific probability. It is reasonable to assume that the ball will be uniformly distributed on the surface. Since the surface area of one half of the sphere is 1/2, the uniformly distributed probabilities are assumed toe be equal to 1/2. This is the point of view of classical probability theory and statistical mechanics, the theory of *outcomes* that can be described with the semiring $\mathbb{S} = \mathbb{R}_+$. The sample space is $\{|0z\rangle, |1z\rangle\}$.

If the ball is yet not positioned on the spherical surface, in may be positioned in a future action in one of the two halves, that is, in $|0z\rangle$ *as well as* in $|1z\rangle$. When the future becomes the present we square the magnitudes of the amplitudes according to *Born's rule*. To be consistent with the probability 1/2, a reasonable choice of both amplitudes is $1/\sqrt{2}$.

Now we consider a second plane through the center of the spherical surface, say with normal vector $|n\rangle$. This plane can be described by the polar angle θ and the azimuthal angle ϕ. It divides the surface into two halves, $|0n\rangle$ and $|1n\rangle$. These are mutually exclusive orthogonal base states. The plane cuts out parts of the surface areas $|0z\rangle$ and $|1z\rangle$, see Figure 72. We can express the relationship of these surface areas by the simple superposition

$$\begin{aligned} |0n\rangle &= \xi_{00}|0z\rangle + \xi_{01}|1z\rangle, \\ |1n\rangle &= \xi_{10}|0z\rangle + \xi_{11}|1z\rangle, \end{aligned} \tag{261}$$

where the squared magnitudes $|\xi_{ij}|^2$ of the complex coefficients for $i, j = 0, 1$ are the corresponding cut-out areas. For example, $|\xi_{00}|^2$ is the surface area of $|0z\rangle$ intersected with that of $|0n\rangle$. In this geometric language a *superposition* is expressed very naturally in terms of concrete surface areas.

It is important to notice the close relationship between the three-dimensional real space that contain the normalized real vectors such as $|n\rangle$ and the two-dimensional normalized complex vectors $|0n\rangle$ and $|1n\rangle$ expressed in terms of the basis vectors $|0z\rangle$ and $|0z\rangle$. This relationship is not unique because of the ambiguities due to the squared magnitudes. But when fixing these quantities like global phases this correspondence becomes for small angles at least locally unique.

Let us look at the special case where the second plane through the center of the surface is perpendicular to the z-axis. Its normal vector $|x\rangle$ is contained

in the xy-plane and defines the x-axis. The plane orthogonal to the x-axis divides the surface into two halves, the two base states that we denote by $|0x\rangle$ and $|1x\rangle$. They represent two mutually exclusive possibilities to position a ball. They are distinguishable. Hence, represented as vectors, they must be orthonormal, that is,

$$\langle 0x|0x\rangle = \langle 1x|1x\rangle = 1, \quad \langle 0x|1x\rangle = \langle 1x|0x\rangle = 0. \tag{262}$$

If the ball would be in $|0x\rangle$, it could be in $|0z\rangle$ or in $|1z\rangle$. Analogously, if the ball would be in $|1x\rangle$, it could be in $|0z\rangle$ or in $|1z\rangle$. Thus, we can express the base states $|0x\rangle$ and $|1x\rangle$ in terms of the other base states as in (261):

$$\begin{aligned} |0x\rangle &= \psi_{00}|0z\rangle + \psi_{01}|1z\rangle, \\ |1x\rangle &= \psi_{10}|0z\rangle + \psi_{11}|1z\rangle. \end{aligned} \tag{263}$$

The orthonormality yield the coefficients

$$\psi_{ij} = \langle iz|jx\rangle, \quad i,j \in \{0,1\}. \tag{264}$$

The plane orthogonal to the x-axis bisects the halves $|0z\rangle$ and $|1z\rangle$, suggesting the formula

$$|0x\rangle = \frac{1}{\sqrt{2}}|0z\rangle + \frac{1}{\sqrt{2}}|1z\rangle. \tag{265}$$

Then the squared amplitudes are 1/2. This is equal to the related cut-out surface areas.

There is some ambiguity in the signs of the amplitudes, but this reflects the ambiguity when we choose the directions for the x and y axes.

Next, we consider the base state $|1x\rangle$. As above the geometry implies the cut-out areas 1/2, and the orthonormality (262) fixes this base state:

$$|1x\rangle = \frac{1}{\sqrt{2}}|0z\rangle - \frac{1}{\sqrt{2}}|1z\rangle. \tag{266}$$

Now we choose the y-axis, the axis that is orthogonal to the x- and z-axis. We denote the two halves of the surface, obtained by bisecting the surface with the xz-plane being orthogonal to the y axis, by $|0y\rangle$ and $|1y\rangle$. These two base states can be expressed in terms of the orthonormal basis $\{|0z\rangle, |1z\rangle\}$ as well as in terms of $\{|0x\rangle, |1x\rangle\}$. There are some obvious necessary conditions. They are mutually exclusive, hence must be orthonormal:

$$\langle 0y|0y\rangle = \langle 1y|1y\rangle = 1, \quad \langle 0y|1y\rangle = \langle 1y|0y\rangle = 0. \tag{267}$$

They can be expressed in terms of cut-out areas with respect to $\{|0z\rangle, |1z\rangle\}$ as

$$\begin{aligned} |0y\rangle &= \varphi_{00}|0z\rangle + \varphi_{01}|1z\rangle, \\ |1y\rangle &= \varphi_{10}|0z\rangle + \varphi_{11}|1z\rangle. \end{aligned} \tag{268}$$

With respect to the basis $\{|0x\rangle, |1x\rangle\}$ it is

$$\begin{aligned} |0y\rangle &= \delta_{00}|0x\rangle + \delta_{01}|1x\rangle, \\ |1y\rangle &= \delta_{10}|0x\rangle + \delta_{11}|1x\rangle. \end{aligned} \tag{269}$$

Both formulas have phase ambiguity, that is, if we multiply all amplitudes with a global phase factor $e^{i\alpha}$ then orthonormality and the cut-off areas remain invariant. Thus, global phases can be ignored.

From (268) we obtain the coefficients

$$\varphi_{ij} = \langle iz|iy\rangle \quad \text{for} \quad i,j \in \{0,1\}. \tag{270}$$

Because the surfaces areas of both halves are equal, it follows that

$$|\varphi_{ij}|^2 = \langle iz|jy\rangle^* \langle iz|jy\rangle = \frac{1}{2} \quad \text{for} \quad i,j \in \{0,1\}. \tag{271}$$

Similarly,

$$\delta_{ij} = \langle ix|jy\rangle \quad \text{for} \quad i,j \in \{0,1\}, \tag{272}$$

and

$$|\delta_{ij}|^2 = \langle ix|jy\rangle^* \langle ix|jy\rangle = \frac{1}{2} \quad \text{for} \quad i,j \in \{0,1\}. \tag{273}$$

Using (265) and (270), we get

$$\begin{aligned} \langle 0x|0y\rangle &= \frac{1}{\sqrt{2}}\langle 0z|0y\rangle + \frac{1}{\sqrt{2}}\langle 1z|0y\rangle \\ &= \frac{1}{\sqrt{2}}\varphi_{00} + \frac{1}{\sqrt{2}}\varphi_{01}. \end{aligned} \tag{274}$$

From (271) it follows that

$$\begin{aligned} \frac{1}{2} &= \langle 0x|0y\rangle^* \langle 0x|0y\rangle \\ &= \frac{1}{2}(\varphi_{00}^*\varphi_{00} + \varphi_{00}^*\varphi_{01} + \varphi_{01}^*\varphi_{00} + \varphi_{01}^*\varphi_{01}) \\ &= \frac{1}{2}\left(\frac{1}{2} + \varphi_{00}^*\varphi_{01} + \varphi_{01}^*\varphi_{00} + \frac{1}{2}\right) \\ &= \frac{1}{2} + \varphi_{00}^*\varphi_{01} + \varphi_{01}^*\varphi_{00}. \end{aligned} \tag{275}$$

Hence,

$$\varphi_{00}^*\varphi_{01} + \varphi_{01}^*\varphi_{00} = 0. \tag{276}$$

Similarly, (266) and (270) imply

$$\langle 1x|1y\rangle = \frac{1}{\sqrt{2}}\varphi_{10} + \frac{1}{\sqrt{2}}\varphi_{11}, \tag{277}$$

and

$$\frac{1}{2} = \langle 1x|1y\rangle^* \langle 1x|1y\rangle = \frac{1}{2} + \varphi_{10}^* \varphi_{11} + \varphi_{11}^* \varphi_{10}. \tag{278}$$

Thus,

$$\varphi_{10}^* \varphi_{11} + \varphi_{11}^* \varphi_{10} = 0. \tag{279}$$

Since all coefficients φ_{ij} are nonzero, the equations (276) and (279) can only be solved in the field of complex numbers. Obviously, $\varphi_{00}^* \varphi_{01}$ and $\varphi_{01}^* \varphi_{00}$ must be purely imaginary. But if the product of two numbers is purely imaginary, the numbers cannot both be real.

If we solve the above equations in $\mathbb{C}$, then we obtain the orthonormal solutions

$$|0y\rangle \;=\; \frac{1}{\sqrt{2}}|0z\rangle + \frac{i}{\sqrt{2}}|1z\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\ i\end{pmatrix}, \tag{280}$$

and

$$|1y\rangle \;=\; \frac{1}{\sqrt{2}}|0z\rangle - \frac{i}{\sqrt{2}}|1z\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\ -i\end{pmatrix}. \tag{281}$$

At the beginning of this lecture notes we have argued that the field of complex numbers is the maximal field of numbers. Now, we have seen that we require complex numbers, even for solving a simple macroscopic geometric problem that has nothing to do with small particles.

Keep in mind: The need of complex numbers is not only a necessity in quantum mechanics, it is a general feature when we try to describe geometrical properties of macroscopic objects.

In Section 4.10 we have defined an observable as a map from an orthonormal basis to the real numbers. In the vector representation observables are matrices satisfying the eigenvalue equation (245). For our geometric two-state model we have derived in (257) the Pauli-z matrix

$$\hat{\sigma}_z = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \tag{282}$$

and in (259) the Pauli-x matrix

$$\hat{\sigma}_x = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}. \tag{283}$$

For the orthonormal basis (280) and (281) we obtain in the same manner

$$\hat{\sigma}_y = 1 \cdot |0y\rangle\langle 0y| + (-1)|1y\rangle\langle 1y = \begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}, \tag{284}$$

the Pauli-y matrix as observable in its vector representation.

In summary, three binary questions "On which side of a plane is a ball?" lead us via vectorization immediately to three orthonormal bases as well as to the famous Pauli matrices. This shows that the well-known mathematical framework, used for properties of small invisible particles like polarization or spin, applies also to macroscopic objects. In particular, two-state systems are not restricted to tiny physical objects as frequently stated. It is a natural framework that applies to binary questions, in other words to bits.

There is another aspect. In all textbooks known to me this mathematical framework including the Pauli matrices is presented as a part of quantum mechanics, and is motivated by spin or polarization. But this enforces students to believe in the properties of invisible small particles, thus are educated to faith and magics although it is not necessary. Hopefully, students now think that this framework is easy to understand and as natural as the concept of numbers.

In fact, the ideas which we have about small particles are dubious, and are also discussed controversial. In this context an interview (given in German and translated here) of the well-known physicist Dürr may be of interest[59].

> *Dürr: Let us take an electron. As I understand, such a physical particle does not exist. Basically, there is something much bigger.*
>
> *Interviewer: Do you really think an electron does not exist at all?*
>
> *Dürr: Not at all in the form of a conventional particle. In my language, I call it a "work (Wirks)" or "happening (Passierchen)". It is a tiny articulation of reality, something that works, that happens, something that triggers something.*

Now we consider more general an arbitrary line through the center of a spherical surface with normalized direction

$$|n\rangle = (n_x, n_y, n_z) = (\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta)), \tag{285}$$

where θ is the polar angle and ϕ is the azimuthal angle. The plane orthogonal to $|n\rangle$ through the center partitions the sphere into two halves that correspond to the base states (261). We make the Ansatz that the observable $\hat{\sigma}_n$, corresponding to this orthonormal basis, is composed linearly of the Pauli matrices in the form

$$\hat{\sigma}_n = n_x\hat{\sigma}_x + n_y\hat{\sigma}_y + n_z\hat{\sigma}_z, \tag{286}$$

that is, the components of $|n\rangle$ are multiplied with the related Pauli matrices, and the resulting terms are added. We expect that the eigenvectors of $\hat{\sigma}_n$ are $|0n\rangle$ and $|1n\rangle$.

[59] P.M. Magazin 05/2007

We write this matrix explicitly:

$$\begin{aligned}\hat{\sigma}_n &= n_x\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} + n_y\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix} + n_z\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\\ &= \begin{pmatrix}n_z & n_x - in_y\\ n_x + in_y & -n_z\end{pmatrix} \qquad (287)\\ &= \begin{pmatrix}\cos\theta & \sin\theta e^{-i\phi}\\ \sin\theta e^{i\phi} & -\cos\theta\end{pmatrix}.\end{aligned}$$

After some work we get its eigenvectors

$$\begin{aligned}|0n\rangle &= \cos\frac{\theta}{2}|0z\rangle + \sin\frac{\theta}{2}e^{i\phi}|1z\rangle,\\ |1n\rangle &= -\sin\frac{\theta}{2}e^{-i\phi}|0z\rangle + \cos\frac{\theta}{2}|1z\rangle.\end{aligned} \qquad (288)$$

They are orthonormal and reduce to the orthonormal bases along the z, x and y axes. For example, let $n = (0,0,1)$ then (285) yields $\theta = 0, \phi = \frac{\pi}{2}$, and

$$|0n\rangle = |0z\rangle, \quad |1n\rangle = |1z\rangle. \qquad (289)$$

For $n = (1,0,0)$ we get $\theta = \frac{\pi}{2}$, $\phi = 0$, and

$$|0n\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) = |0x\rangle, \quad |1n\rangle = (-|0\rangle + |1\rangle) = -|1x\rangle. \qquad (290)$$

For $n = (0,1,0)$ we have $\theta = \frac{\pi}{2}, \phi = \frac{\pi}{2}$ and

$$|0n\rangle = |0y\rangle, \quad |1n\rangle = i|1y\rangle. \qquad (291)$$

Notice that in (290) the second state $|1x\rangle$ is multiplied by the global phase factor -1, and in (291) the state $|1y\rangle$ is multiplied by the global phase factor i. But we know that the squared magnitudes of the amplitudes correspond to cut-out areas of the surface and are invariant when multiplied by a global phase factor.

Finally, let us look at the special case

$$|n\rangle = (\sin\theta, 0, \cos\theta). \qquad (292)$$

Then we obtain from (287) the observable

$$\hat{\sigma}_n = \begin{pmatrix}\cos\theta & \sin\theta\\ \sin\theta & -\cos\theta\end{pmatrix} \qquad (293)$$

with eigenvectors (288)

$$\begin{aligned}|0n\rangle &= \cos\frac{\theta}{2}|0z\rangle + \sin\frac{\theta}{2}|1z\rangle,\\ |1n\rangle &= -\sin\frac{\theta}{2}|0z\rangle + \cos\frac{\theta}{2}|1z\rangle.\end{aligned} \qquad (294)$$

to eigenvalues +1 and −1, respectively.

We suppose that the area of the spherical surface is normalized to one. What is the area that the observable $\hat{\sigma}_n$ takes the eigenvalue +1? Noticing that $\sin^2\theta/2 + \cos^2\theta/2 = 1$, the answer is

$$\text{Area } (+1) = |\langle 0n|0z\rangle|^2 = \cos^2\frac{\theta}{2}. \tag{295}$$

Similarly, we get

$$\text{Area } (-1) = |\langle 1n|0z\rangle|^2 = \sin^2\frac{\theta}{2}. \tag{296}$$

This is what we expect from the geometrical point of view.

Until now we have investigated a simple geometric problem, and we have obtained the quantum mechanical framework of polarization and spin. Quantum mechanics is a probabilistic theory. So, where occur probabilities? Obviously, if we assume that the ball on the surface is uniformly distributed, then the areas correspond to probabilities.

With this interpretation of areas as probabilities the expectation value of the observable $\hat{\sigma}_n$ can be written as

$$\langle\hat{\sigma}_n\rangle = (+1)\cos^2\frac{\theta}{2} + (-1)\sin^2\frac{\theta}{2} = \cos\theta, \tag{297}$$

a well-known formula from the spin formalism.

Keep in mind: The mathematical framework describing polarization and spin is not restricted to small particles. It is a consequence when describing physics in terms of binary questions, namely bits. From this point of view we can say that a photon or an electron behaves like a big ball on a spherical surface, at least from the mathematical point of view.

In Section 4.17 we show that the uncertainty principle has a simple geometric background. We remember that the Pauli observables $\hat{\sigma}_z$ and $\hat{\sigma}_x$ label the two halves that result from bisecting the sphere with the xy-plane and the yz-plane, respectively. The geometric contents of the uncertainty principle turns out to be as follows: if in a future action we would position a ball in the half $|0y\rangle$, then it is undetermined if the ball would be above or below the xy-plane or above or below the yz-plane. This is simple, not weird or paradox, clear from the geometry, and supports our view that quantum mechanics is the theory of the future.

4.12 Creation and Annihilation Operators

We have worked through several different classes of operators or matrices. Firstly, there are operators $\hat{U}$ that act on states $|\xi\rangle$. Their physical meaning is that they change states. The operators $\hat{U}$ are special $\mathbb{S}$-*linear matrices*: permutation matrices ($\mathbb{S} = \mathbb{N}$), stochastic matrices ($\mathbb{S} = \mathbb{R}_+$), and unitary matrices ($\mathbb{S} = \mathbb{C}$).

Secondly, there are operators that are related primary to base states,namely the *observables* A that map base states i to real numbers. Actually, this is only a labeling procedure. In other words, observables label base states. Via vectorization they become matrices. Matrices $\hat{A}$ are linear, and thus they can act also on states $|\xi\rangle$. For an observable the result $\hat{A}|\xi\rangle$ has no physical meaning in general, but it can be used to define some important quantities, in particular, the expectation value $\langle\xi|\hat{A}|\xi\rangle$ and the variance.

Thirdly, there is, and for the sake of completeness should be, a type of operators that act on base states and changes them. They do not preserve possibilities and don't act on states. But since they are linear matrices they can be applied formally to states just like observables. They are not mentioned in the postulates of quantum mechanics, where only unitary and Hermitian operators are considered. One consequence of this failure is that the number of particles do not change in quantum mechanics. This restriction in quantum mechanics is not necessary. However, a change of the number of particles belongs to quantum field theories and Fock spaces.

A well-known type of operators that act on base states and change them are creation and annihilation operators. In its simplest form these operators act on two-state systems. The *annihilation operator* converts the base state $|1\rangle$ to base state $|0\rangle$, but for the base state $|0\rangle$ this operator produces the number 0. Obviously, its matrix representation is

$$\hat{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \tag{298}$$

A frequently used interpretation is to imagine one particle moving on a line. This represents the base state $|1\rangle$. The annihilation operator removes this particle. The line without particle is represented as $|0\rangle$. A second application of the annihilation operator cannot produce a base state, since there is nothing that can be annihilated. This is represented by the number 0.

Its conjugate complex matrix is

$$\hat{A}^\dagger = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \tag{299}$$

This matrix converts the base state $|0\rangle$ to $|1\rangle$, thus generates a particle on a line. When operating on the base state $|1\rangle$ there is no further state above that can be created, thus the matrix converts it to the number 0. This operator is called *creation operator*.

In Section 4.4 we emphasized in the "Keep in mind" that the distinguishable base states do not change when going from the future to the present and

then to the past. They define the facts in the past, the outcomes or elementary events in the present, and possibilities in the future.

Keep in mind: The phrase "one particle moves on a line" describes a fact in the past, an outcome in the present, or a possibility in the future. It will become clear whether we talk about future (quantum mechanics), present (classical probability theory), or past (classical mechanics). Therefore, we do not change a phrase like "one particle moves on a line" to "if, in a future action one particle would move on a line", although the latter is correct when talking about quantum mechanics.

An important aspect is that each 2×2-matrix can be constructed with these two matrices only. If we compose both matrices in series, we obtain the product matrices

$$\hat{A}\hat{A}^\dagger = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{300}$$

and

$$\hat{A}^\dagger\hat{A} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{301}$$

When multiplying these four matrices with scalars and adding them, then we obtain every 2×2-matrix. Via the tensor product we obtain all important gates of quantum computing. For example, the *controlled-NOT* can be written as

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \hat{A}^\dagger\hat{A} \otimes (\hat{B} + \hat{B}^\dagger) + (\hat{A}\hat{A}^\dagger \otimes \hat{1}), \tag{302}$$

where $\hat{B}$ and $\hat{B}^\dagger$ are the annihilation and creation operators for the second bit. Feynman[60]1982 was the first who considered how computers can be built using the laws of quantum theory. He started to build up all gates using creation and annihilation operators. Notice that these operators are neither unitary nor observables.

In quantum physics these operators are used in many-particle physics and quantum field theories. Please look in those books for an extensive discussion of this subject. See also Section 5.

Keep in mind: In physics there are three important types of operators: operators that transform states (permutation matrices, stochastic and unitary matrices), operators that label base states (observables, Hermitian operators), and operators that change base states (creation and annihilation operators).

[60]Feynman [1982], Feynman [1986]

Finally, we want to mention that there is a set of matrices that belong to all three classes, namely the *Pauli matrices*, which we have introduced in Section 4.11. Simple computations show:

- The Pauli matrices are unitary, thus transform states.
- The Pauli matrices are Hermitian, thus they are observables.
- The Pauli matrices transform base states, since they are unitary.

4.13 Space and Entanglement

It is often mentioned in the literature that there are two fundamental problems with quantum theory, namely the meaning of **superposition** and **entanglement**.

We have explained superposition by experimental set-up's like the slit experiments. In order to avoid paradoxes we have formulated quantum mechanics as a theory describing the *future* in terms of "as well as" possibilities. The future is timeless, there is no " future clock". Actually, in the future happens nothing. Perhaps, future might be best described by the phrase " What could happen, when nothing happens? "

Future itself is timeless, of course. But are there hints that quantum theory is timeless such that we can classify quantum theory as the theory describing the future? Yes, the well-known *Wheeler-de Witt equation* describes quantum mechanics without any time-dependence. This equation has generated much controversy, but is out of the scope of these notes. Nevertheless, when we describe physics in terms of our trinity "future-present-past" we must prove that the fundamental equations of physics, in particular those of gravitation and electromagnetism, must be fulfilled, at least approximately.

Indeed, if quantum mechanics is timeless, that is the theory has no external time t, then we run into big problems. Spacetime vanishes, and quantum mechanics and the theory of relativity can hardly work together. Relativistic quantum mechanics and quantum field theories should be impossible. The Lorentz transformation and its invariant quantities seem to vanish. But could there be any other space and an appropriate theory that can replace the four-dimensional spacetime and the theory of relativity?

Related to these questions is the problem of entanglement. There are easy understandable forms of entanglement, for instance, two welded coins or when the cat is entangled with a box. It is, however, far from being simple when two photons are entangled although they are separated by a large distance. Einstein referred to it with the phrase "spooky action at a distance". Distance, however, is a property of the underlying space. Could it be that changing the space leads to a better understanding of this phenomenon?

When we change the underlying space and additionally avoid time t, then it seems to be not possible to recover our fundamental equations. Let us summarize some well-known facts about spaces used in physics[61].

- Physicists love symmetry. But just when starting with the fundamental concept of space, the non-symmetric (3+1)-spacetime with three spatial coordinates and one time coordinate is used. An alternative is to start with the (3+3)-phase space. If we don't involve mass, then the phase

[61] Our notation when using coordinates should be mentioned. The Newtonian equation of motion is a second order differential equation usually formulated in *Cartesian coordinates* x,v and p. This equation is not invariant for transformations to other coordinate systems like cylindrical or polar coordinates. A big advantage of Hamilton's equations is their invariance when transforming coordinate systems. Then we speak of *generalized coordinates*, namely generalized position q and generalized momentum p. In these notes this distinction does not matter, and we will use both notation for the coordinates.

space consists of symmetric pairs $(x, v) \in X \times V$ in a position-velocity space, where X denotes a set of points and V denotes a set of velocities. The set of points allow to define arrows in V with starting and terminal points. Vice versa the arrows depict to points. Without points no arrows, and conversely without arrows no points. Positions and their change require velocities, and velocities generate starting and terminal points. This remembers at Esher's "drawing hands", where out of a sheet of paper two hands, seemingly paradoxical, draw one another into existence. The hands are both, the object and the machine of creation. Neither hand seems to have an origin, a beautiful symmetry of (3+3)-space.

- The "drawing hands" and their symmetry is in some sense reflected in the *Hamilton equations*

$$\dot{p}_k = -\frac{\partial H}{\partial x_k}, \quad \dot{x}_k = \frac{\partial H}{\partial p_k}, \tag{303}$$

the fundamental equations of classical mechanics. These equations have six independent variables, three position variables x and three momentum variables p. The interpretation in physics is: given any Hamilton function $H(x, q)$ and the values of the position and momenta coordinates at some time, the equations (303) give the coordinates of the resulting path at an infinitesimal time later. The complete trajectory in the phase space is obtained by successively updating the coordinates.

This is the usual interpretation of Hamilton's equations. But in our interpretation these equations describe facts of the past, and we know that the past is timeless. What means t? Well it's just a geometrical parameter that allows to describe the solution of the equations in explicit form. The solution is a geometrical object, an ellipse or a more complicated curve. Hamilton's equation are part of the deterministic past.

- The phase space has dimension (3+3). Are there other spaces with this dimension? Yes, of course. The fundamental theory of electromagnetism has the same dimension. Maxwell's equations work with a three-dimensional electric field E and a three-dimensional magnetic field B.

- But what does this mean for the theory of general relativity? Does this theory vanish? At least not completely. *Einstein's field equations* describe how matter and radiation determine the curvature of spacetime, and vice versa. These equations are nonlinear. But it is fascinating that the linearized field equations turned out to be of the type of Maxwell's equations in a $3 + 3$-space. Moreover, all of the experimental results like the deflection of light, the gravitational time dilation, the gravitational radiation, or the correction formula for GPS, emerge in the linear approximation, see for example Ohanian[62]. In addition, since Maxwell's equations can easily be quantized (see Section 5.12), the linearized field

[62] Ohanian [1994]

equations of Einstein can easily be quantized. In this sense gravitation can be easily quantized. Hence, our most fundamental theories can be described in a (3+3)-space.

- But what about *Heisenberg's uncertainty principles*, when we change spacetime and use a (3+3)-space? These principles are known to assert a fundamental limit to the precision we can know about certain pairs of complementary variables, like position and momentum or time and energy. The position-momentum uncertainty is in accordance with the (3+3)-space. There exists, however, no position-time uncertainty or momentum-energy uncertainty. Uncertainty principles don't like non-symmetric (3+1)-spaces. They prefer a symmetric dimension.

- **Keep in mind:** The (3+3)-position-velocity space $X \times V$is distinguished by a wonderful symmetry which reminds of Esher's "drawing hands". Moreover, the "spooky action at a distance" of two entangled photons in (3+1)-spacetime vanishes. The two photons are welded in the velocity space V. In other words, they can be connected via a velocity $v \in V$. Notice that the notion of distance depends on the space.

We have found some basic arguments that support a (3+3)-position-velocity space. In particular, the fundamental physical theories such as classical mechanics, gravitation and electromagnetism can be formulated using dimension (3+3). Of great advantage is that this space allows to use our preferred language of machines.

Remember that we looked at birefringent plates as machines. They are characterized by a polarization axis, and by two possibilities, namely their two polarization directions. In the same way we define a *position machine*. This machine is characterized by a velocity $v \in V$, and by the set X that represent all possible positions. The difference to the calcite is that the angle of polarization is replaced by the velocity, and that we have now a large number of possibilities, namely all positions.

Think of a position machine as a "train" moving at constant speed. Don't think that this machine is merely a reference frame. This would be true in (3+1)-spacetime. Contrary, this is a real existing machine which is consistent with our experience. If you are inside a train moving at constant speed and you toss a ball into the air, then the ball will go straight up and then comes straight down. The ball has the same inertia as the train. Much more beautiful examples are provided by astronauts in a gravitational free space. This interpretation of trains in a position-velocity space explains the famous riddle of *inertia*. Compare the comments in Section 1. Recognize that this explanation does not hold valid in the (3+1)-spacetime, thus giving one more argument for the position-velocity space.

The whole space $X \times V$ consists of many position machines. As in the case of polarization of light, where only the difference between the angles of

two polarization axes is relevant, symmetry requires that only the difference between the velocities of two position machines is useful. Frequently, we denote by $(X, 0)$ one machine and by (X', v) the other machine, where v denotes this difference.

The positions in X represent the base states, that is, all possible positions that may be occupied in a future experiment. Now we can apply our mathematical formalism developed above. The *number representation* is given by the numbers $x \in X$. Asking binary questions in which half a point is, we obtain the *register representation* of positions. Finally, vectorization leads to the base states $|x\rangle$.

The *states* of a position machine are *superpositions*

$$|\xi\rangle = \sum_x \xi_x |x\rangle, \qquad \xi_x \in \mathbb{S}, \quad \sum_x |\xi_x|^2 = 1, \tag{304}$$

which are classical, random, or quantum depending on whether $\mathbb{S} = \mathbb{N}, \mathbb{S} = \mathbb{R}_+$, or $\mathbb{S} = \mathbb{C}$, respectively. Using our unified definition of observables, we arrive in a natural and simple manner to observables described as Hermitian operators.

In quantum mechanics the transition from classical observable values to Hermitian operators is usually postulated. Here, it turns out to be a consequence of carefully distinguishing between states and base states and our simple process of vectorization.

The (3+3)-position-velocity space contains also *velocity machines* or "dual trains". This machine is characterized by a distance $x \in X$, and by the set V representing all possible velocities. Given two velocity machines we denote by $(0, V)$ one machine and by (x, V') the other machine, where x denotes the distance between both.

We have visualized the concept of "trains" (X', v) and "dual trains" (x, V') in Figure 73. One might imagine motion of physical objects as a change from (x, v) to $(x + \Delta x, v + \Delta v)$ between trains and dual trains. Since quantum theory is probabilistic this change must be probabilistic one, and must be in agreement with Feynman's rules.

Keep in mind: There are good arguments to describe physical processes on the basis of a $3+3$-space. In particular, this space can be viewed as composed of position and velocity machines.

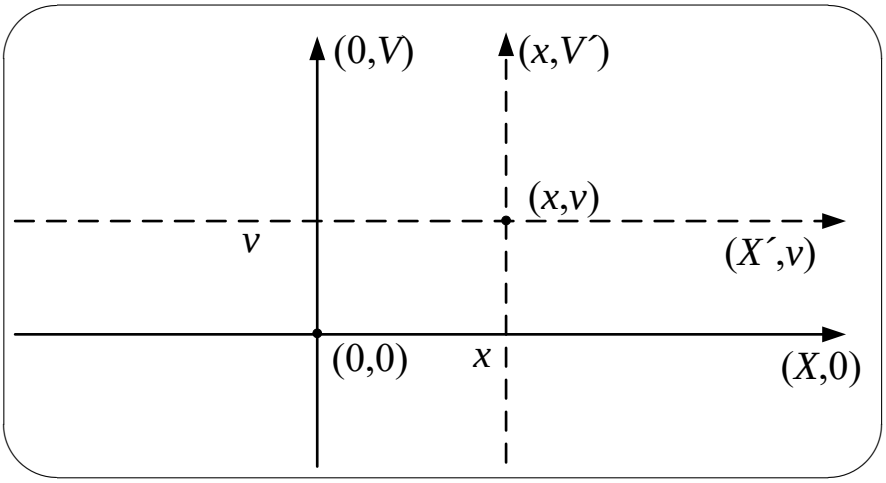


Figure 73: The $(3+3)$ position-velocity space (X, V) can be viewed as a collection of machines: the trains (X', v) and the dual trains (x, V'). These are displayed for one position and one velocity coordinate. In this space in a future experiment a particle can change from $(0, 0)$ to (x, v). Feynman's (never violated) rules are applied to obtain the related probabilities.

4.14 Clocks and the Lorentz Transform

Above we have given several arguments to replace the (3+1)-spacetime by a (3+3)-position-velocity space. A major reason was that we view quantum mechanics as a theory describing the future, and the future is timeless. But if there is no spacetime, then there is no theory of relativity. Nevertheless, we ask whether we can reproduce the mathematical formalism of special relativity. In particular, we are interested in the key of this theory, namely the *Lorentz transform*. A second question is: From which principles can we derive the framework of relativity?

Let us consider what is possibly the deepest of all philosophical puzzles, namely the nature of time. Clocks are used to compare the speed of different objects, say a group of runners on a certain distance. When looking at his clock each runner observes a value, called the time for the given distance. Instead of runners, let us look at two trains $(X, 0)$ and (X', v), or position machines if you like, see Figure 74. Each train is equipped with a clock at places $0 \in X$ and $0' \in X'$. Both trains differ with respect to the relative velocity v. How can we measure the velocity v.

We assume that at the beginning both clocks are synchronized, that is, the positions 0 and $0'$ as well as the clockhands are coincident. The length $x_{clock} > 0$ denotes the distance traveled by the tip of the clockhand in one round, and $|v_{clock}|$ denotes the clockhand's constant magnitude of velocity. Hence, the ratio $x_{clock}/|v_{clock}|$ characterizes the clock itself, and we can choose the clock unit as

$$\frac{x_{clock}}{|v_{clock}|} = 1\ [sec = \text{meter/mps}], \tag{305}$$

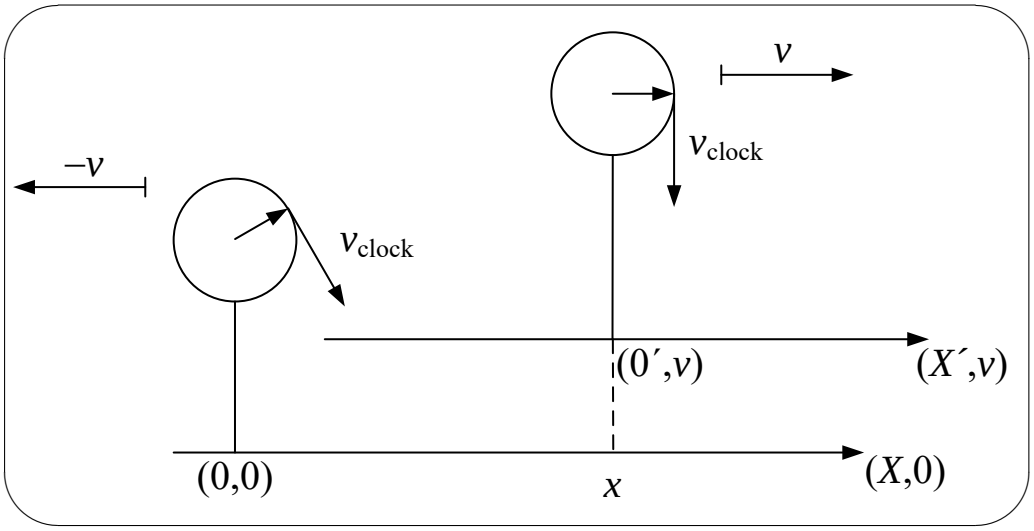


Figure 74: Two trains $(X, 0)$ and (X', v). Both trains are equipped with a clock at $(0, 0)$ and $(0', v)$. At the beginning, we assume that both places, 0 and $0'$, and the position of the clockhands coincide. Then the clocks are synchronized.

where we have defined the time-independent velocity unit as mps, a shortcut for meter per second. Notice that now time is a derived quantity in the position-velocity space.

Let us assume that the second train has covered the distance x in X. Moreover, we assume that the tip of the clockhand of the first clock has covered a distance x_c. After 1 round the clockhand has covered the distance $x_c = x_{clock}$. After 2 rounds the clockhand has covered the distance $x_c = 2x_{clock}$. After 2.5 rounds the clockhand has covered the distance $x_c = 2.5x_{clock}$, and so on. Hence, it is natural to define the time t of the first clock via the number of rounds covered by the clockhand:

$$t = \frac{x_c}{x_{clock}}. \tag{306}$$

The parameter t denotes the clock's measurable or observable value which defines a quantity called time, derived in $X \times V$. Now t is not an external fundamental parameter, but a parameter that is related to a unique specific clock.

We assume uniform motion, that is, the position-independent quantity v is constant. Then the ratio of the distances must be equal to the ratio of the velocities, that is,

$$\frac{x}{x_c} = \frac{v}{|v_{clock}|}. \tag{307}$$

Then

$$x = v\,\frac{x_c}{|v_{clock}|} = v\,\frac{x_c}{x_{clock}}\,\frac{x_{clock}}{|v_{clock}|}, \tag{308}$$

From these formulas we obtain the fundamental kinematic equation

$$x = vt, \tag{309}$$

where time t is the derived value for the first clock in our (3+3)-space, depending on distance and velocity only. This coincides with Einstein's advice:

Time is what clocks measure.

But in contrast to Einstein we do not postulate the (3+1)-spacetime, and do not assume a maximal speed of light.

There are various clocks like pendulum clocks, spring-driven clocks, quartz clocks where a quartz crystal vibrates, and atomic clocks based on the vibration of electrons in atoms. Clocks are physical systems that change positions and velocities. They are machines inside our universe, and thus can be described as observables, that is, matrices.

In most cases clocks perform periodic processes of motion, which occur again and again in the same way, for example, the reciprocation of a pendulum. We assume that they always take approximately the same period, so that they can provide us with a uniform time standard. The clock can be described as a machine in position-velocity space such that the observed time $t = t(x, v)$ depends on position and velocity.

Let (X, t) and (X', t') denote the coordinates of trains and times of the first and second clock, respectively. Neither clock is preferred. From the reversal point of view, the first clock ist uniformly moving with velocity $-v$ with respect to the second clock. Hence, we obtain as before the fundamental kinematic equation

$$x' = -v\, t'. \tag{310}$$

It is always possible to transform between the coordinates of both trains. The question is whether there exists a non-singular linear transformation. Let us make the Ansatz that there is a linear transformation

$$\begin{pmatrix} x' \\ t' \end{pmatrix} = \Lambda \begin{pmatrix} x \\ t \end{pmatrix} = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix}. \tag{311}$$

This is the most simple model. Obviously, the matrix $\Lambda = \Lambda(v)$ depends on the scale $v \neq 0$. Moreover, we can assume that the covered distance x, and thus the time t of the first clock, are non-zero.

This equation must hold true for all possible coordinates. If we set $x' = 0$, then the first line of this equation implies $\Lambda_{11} \neq 0$, since otherwise Λ_{12} must be zero, yielding a singular transformation matrix. Dividing the first equation by Λ_{11} gives

$$-\frac{\Lambda_{12}}{\Lambda_{11}} = \frac{x}{t} = v. \tag{312}$$

Hence, $\Lambda_{12} = -\Lambda_{11} \cdot v$ and (311) implies

$$x' = \Lambda_{11} \cdot (x - vt). \tag{313}$$

Since $x' = -v \cdot t'$, the equations (311) and (313) yield

$$t' = -\frac{1}{v}x' = -\frac{\Lambda_{11}}{v}x + \Lambda_{11}t = \Lambda_{21}x + \Lambda_{22}t. \tag{314}$$

Hence, $\Lambda_{11} = \Lambda_{22}$ and

$$t' = \Lambda_{22}\left(\frac{\Lambda_{21}}{\Lambda_{22}}x + t\right). \tag{315}$$

Let us switch to the more appropriate notation

$$\gamma = \gamma(v) = \Lambda_{11} = \Lambda_{22} \text{ and } \beta = \beta(v) = \Lambda_{21}/\Lambda_{22}. \tag{316}$$

Then the equations (313) and (315) can be written as

$$x' = \gamma(x - vt), \quad t' = \gamma(\beta x + t), \tag{317}$$

yielding the transformation matrix

$$\Lambda = \gamma\begin{pmatrix} 1 & -v \\ \beta & 1 \end{pmatrix}. \tag{318}$$

This is almost the Lorentz transformation. We have only to determine the constants β and γ.

Let us look at three trains. If we can transform coordinates between (X, t) and (X', t') with respect to velocity v, then we can transform coordinates between (X', t') and (X'', t'') with respect to a velocity u. Of course, we can transform coordinates between (X'', t'') and (X, t) by consecutive composition. In matrix form it follows that

$$\begin{pmatrix} x' \\ t' \end{pmatrix} = \gamma(v)\begin{pmatrix} 1 & -v \\ \beta(v) & 1 \end{pmatrix}\begin{pmatrix} x \\ t \end{pmatrix}, \quad \begin{pmatrix} x'' \\ t'' \end{pmatrix} = \gamma(u)\begin{pmatrix} 1 & -u \\ \beta(u) & 1 \end{pmatrix}\begin{pmatrix} x' \\ t' \end{pmatrix}, \tag{319}$$

and

$$\begin{pmatrix} x'' \\ t'' \end{pmatrix} = \gamma(v)\gamma(u)\begin{pmatrix} 1 - \beta(u)v & -u - v \\ \beta(v) + \beta(u) & 1 - \beta(v)u \end{pmatrix}\begin{pmatrix} x \\ t \end{pmatrix}. \tag{320}$$

Since the transformation matrix must be of type (318), the diagonal elements are equal to one, yielding for every u and v:

$$1 - \beta(u)v = 1 - \beta(v)u, \text{ or equivalently } \frac{u}{\beta(u)} = \frac{v}{\beta(v)}. \tag{321}$$

Clearly, the ratio $v/\beta(v) = \varrho$ must be a nonzero constant. In other words, it is invariant for any scale $v \neq 0$, and therefore

$$\Lambda(v) = \gamma(v)\begin{pmatrix} 1 & -v \\ v/\varrho & 1 \end{pmatrix}. \tag{322}$$

Now we can transform from (X', t') to (X, t), and then backwards yielding

$$\begin{pmatrix} x' \\ t' \end{pmatrix} = \gamma(v)\begin{pmatrix} 1 & -v \\ v/\varrho & 1 \end{pmatrix}\begin{pmatrix} x \\ t \end{pmatrix}, \quad \begin{pmatrix} x \\ t \end{pmatrix} = \gamma(-v)\begin{pmatrix} 1 & v \\ -v/\varrho & 1 \end{pmatrix}\begin{pmatrix} x' \\ t' \end{pmatrix}, \tag{323}$$

and

$$\begin{aligned} \begin{pmatrix} x \\ t \end{pmatrix} &= \gamma(-v)\gamma(v)\begin{pmatrix} 1 & v \\ -v/\varrho & 1 \end{pmatrix}\begin{pmatrix} 1 & -v \\ v/\varrho & 1 \end{pmatrix}\begin{pmatrix} x \\ t \end{pmatrix} \\ &= \gamma(-v)\gamma(v)\begin{pmatrix} 1+v^2/\varrho & 0 \\ 0 & 1+v^2/\varrho \end{pmatrix}\begin{pmatrix} x \\ t \end{pmatrix}. \end{aligned} \tag{324}$$

This equation must hold for every x and t, and we obtain

$$\gamma(-v)\gamma(v)\cdot(1+\frac{v^2}{\varrho}) = 1. \tag{325}$$

Since both systems (X, t) and (X', t') are on an equal footing, and $x' = \gamma(v)(x - vt)$ translates symmetrical to $x = \gamma(-v)(x' - (-v)t')$, the nonzero value $\gamma(v)$ does not depend on the direction of v. Therefore, $\gamma(v) = \gamma(-v)$ implying

$$\gamma(v)^2(1+\frac{v^2}{\varrho}) = 1. \tag{326}$$

Taking the square root and substituting into equation (314), we get the well-known two-dimensional *Lorentz transformation*

$$x' = \gamma(x - vt),\ t' = \gamma(\frac{v}{\varrho}\cdot x + t),\ \gamma(v) = \pm 1/\sqrt{1+\frac{v^2}{\varrho}}. \tag{327}$$

A simple computation shows the invariant quantity

$$-\varrho(t')^2 - (x')^2 = -\varrho t^2 - (x)^2. \tag{328}$$

The fundamental constant ϱ is nonzero, and thus can be positive or negative. Since γ is dimensionless, ϱ has the units of a squared velocity. Firstly, let us consider the case $\varrho < 0$, and let $c = \sqrt{-\varrho}$. Then

$$\gamma(v) = \pm 1/\sqrt{1-\frac{v^2}{c^2}}, \tag{329}$$

and the invariant (328) takes the form

$$(ct')^2 - (x')^2 = (ct)^2 - (x)^2. \tag{330}$$

This is the famous *Minkowsky distance* which is invariant with respect to Lorentz transformations. Therefore, every relative speed distance $|v|$ between two clocks has to satisfy the condition $|v|^2 < c^2$, and thus must be smaller than c; otherwise γ would be imaginary, illicitly.

In physics the constant c is identified with the maximal speed of light. Hence, the maximal speed of light is a strictly upper bound for relative velocity distances between clocks. The identity

$$x = vt = \kappa(ct) = \kappa x^0,\ 0 < \kappa = v/c < 1 \tag{331}$$

shows that we have obtained for all clocks a common length ruler $x^0 = ct$. Obviously, the clocks differ in their characteristic ratios. All clocks have the same ruler ct, and the invariant (328) has the value $(ct)^2(1-\kappa^2)$.

At the beginning the coordinates of the relative velocity v could have unsubstantially numerical values, and in some sense there was some arbitrariness when choosing $x = vt$. In other words the labeling was not reasonable. But now we have a bounded characteristic number κ for each train, and the common ruler ct for all trains. Summarizing, we have obtained physical reasonable coordinates in the form of state-coordinates and the same scale in all possible position machines.

Additionally, in this case we get a nice interpretation. Since $\varrho = (ic)^2$, formula (328) implies the invariant

$$(x)^2 + (ict)^2 = x'^2 + (ict')^2 = c^2\bar{t}^2. \tag{332}$$

Hence, for negative ϱ, Lorentz transformations can be viewed as rotations about the origin with a real position coordinate x and an imaginary coordinate, the imaginary velocity ic. This suggest a position-velocity space $X \times iV$ with imaginary velocities.

Finally, let us consider the case $\varrho > 0$, and let $c = \sqrt{\varrho}$. Then the transformation describes simply an orthogonal rotation of the coordinates (ct, x). The coordinates of v cannot be further specified, the set of possible values κ is unbounded, and the equation (328) makes not much sense, because we have the same arbitrariness as before. Therefore, this case is physically uninteresting.

Keep in mind: The position-velocity space allows to define clocks as machines that provide a derived quantity, the time. A few simple arguments yield the mathematical framework of special relativity. This derivation does not require the usual postulates of special relativity, namely that (i) the laws of physics are the same in all inertial frames of reference, and that (ii) the speed of light in free space has the same value c in all inertial frames of reference.

In summary, the position-velocity space allows to describe Hamilton's classical mechanics, the theory of special relativity, and a reasonable explanation of entanglement. To emphasize the central difference, we remember what Herman Minkowski wrote:

> *Henceforth, space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality.* (80th Assembly of German Natural Scientists and Physcists at Cologne 1908)

Our approach is rather radical when viewing t as a variable depending on (x, v), and describing reality as a kind of union of position and velocity. The derived parameter t has no relation to motion and should be not mixed up with the time that is used in physics for describing the dynamics.

In almost all books about relativity the word *observer* is in use. Observing is based on human beings. But the laws of physics should not depend on the existence of human beings, whatever they are. Only physical objects of reality can exist. Mathematical entities such as frames of reference do not exist in reality, and in no manner they can move. We arrived at the Lorentz transformation via machines, called clocks, in the position-velocity space. The fact, however, that the times t and t' differ seems to be strange. It comes out solely of simple transformations, but the mathematics explains nothing. Why differ these times? Let us look at an example in the one-dimensional position space. Assume you have a pencil, and the pencil point is 1 meter away from your eye. You have measured this with a tape measure. What does this value mean? Actually nothing. Another person, a so called observer, measuring from another position obtains another value. Each value is possible. The observers disagree. But there is another important quantity, the length of the pencil, that is, the difference between the pencil tip and its end. Independent of the position, all observers measure the same length. The length is independent of the observer. That is the key to special relativity. Clocks can provide different values. This is not surprising, and these values are unimportant. Important are the values that are invariant with respect to the Lorentz transform, such as the distance (330). The well-known *Hafele-Keating experiment* has shown different times between clocks that are moving with different velocities. In the four-dimensional spacetime this experimental result is surprising and perhaps weird, since velocity is a derived quantity. In a position-velocity space the velocity is an independent quantity leading to reasonable results.

Keep in mind: We have seen that the mathematical framework of quantum mechanics applies also to large objects. Now we have seen that the theory of relativity can be derived from simple transformation rules that are necessary from a mathematical point of view. It is not necessary, however, to postulate the existence of light itself or a maximal speed of light.
Time dilation and length contraction are not strange, they are simple consequences of our viewpoint that clocks are derived machines in a position- velocity space without external time parameter.

4.15 Dimension of Space

There are some deep fundamental questions in physics. What can we say about the dimension of physical reality? Why is the underlying space in most physical models the (3+1)-dimensional spacetime? Are there more useful spaces with other dimensions? Can mathematics perhaps dictate dimensions?

Unfortunately, we can only describe superficially whether and how these fundamental questions might be answered. It requires knowledge about the difficult theory of Lie groups and Lie algebras, which we cannot treat extensively within the context of these lecture notes. However, a rough introduction is presented in the Appendix 7. Hopefully, it might help to understand the essential aspects.

There are several physical theories that are based on various dimensions. For example, we have considered a (3+3)-dimensional position-velocity space, and we defined therein a dependent variable time. This time is what a clock measures. The *Kaluza-Klein theory* is a field theory that unifies gravitation and electromagnetism in a 5-dimensional space which is beyond the usual (3+1)-dimensional spacetime. *String theories* require extra dimensions of space. In bosonic string theory, the space is 26-dimensional, in superstring theory it is 10-dimensional, and in M-theory it is 11-dimensional. *Quantum chromodynamics*, the theory of strong interactions between quarks and gluons, is a quantum field theory that is based on the 3-dimensional complex space. At a first glance each dimension might be useful, say 37, 543, or any other number.

What is a reason why the previous questions about the dimensions are important and even fundamental? There is an old well-known story. During a meeting in the year 1953 Fermi criticized a model developed by Dyson. He quoted John von Neumann who said:

> *With four parameters I can fit an elephant, and with five I can make him wiggle his trunk.*

He pointed out that nobody should be impressed when complicated, high-dimensional models fit all data sets. With enough parameters, you can fit anything. Obviously, the same argument applies also to the number of variables in a physical theory defining its dimension; a fixed value of a variable is a parameter. Hence, a major goal is to develop useful physical theories with lowest possible dimensions and with few parameters. It is amusing that Mayer et al.[63] gave a five parameter model that encodes the elephant including its wiggling trunk, see Figure 75.

When thinking about spaces, we can imagine the most diverse geometries: non-smooth geometries, twisted geometries like a pretzel, and many others. Of course, without any reasonable assumptions we cannot answer our questions above. It seems to be natural, however, to assume that in a plausible space with reasonable geometry we can define simple geometric objects, as is possible in an inner-product space like $\mathbb{R}^n$ or $\mathbb{C}^m$. In these inner-product spaces we can define normal vectors, planes, rotations, unitary operators, and we have a great understanding and experience with such spaces.

[63] Mayer, Khairy, Howart [2010]

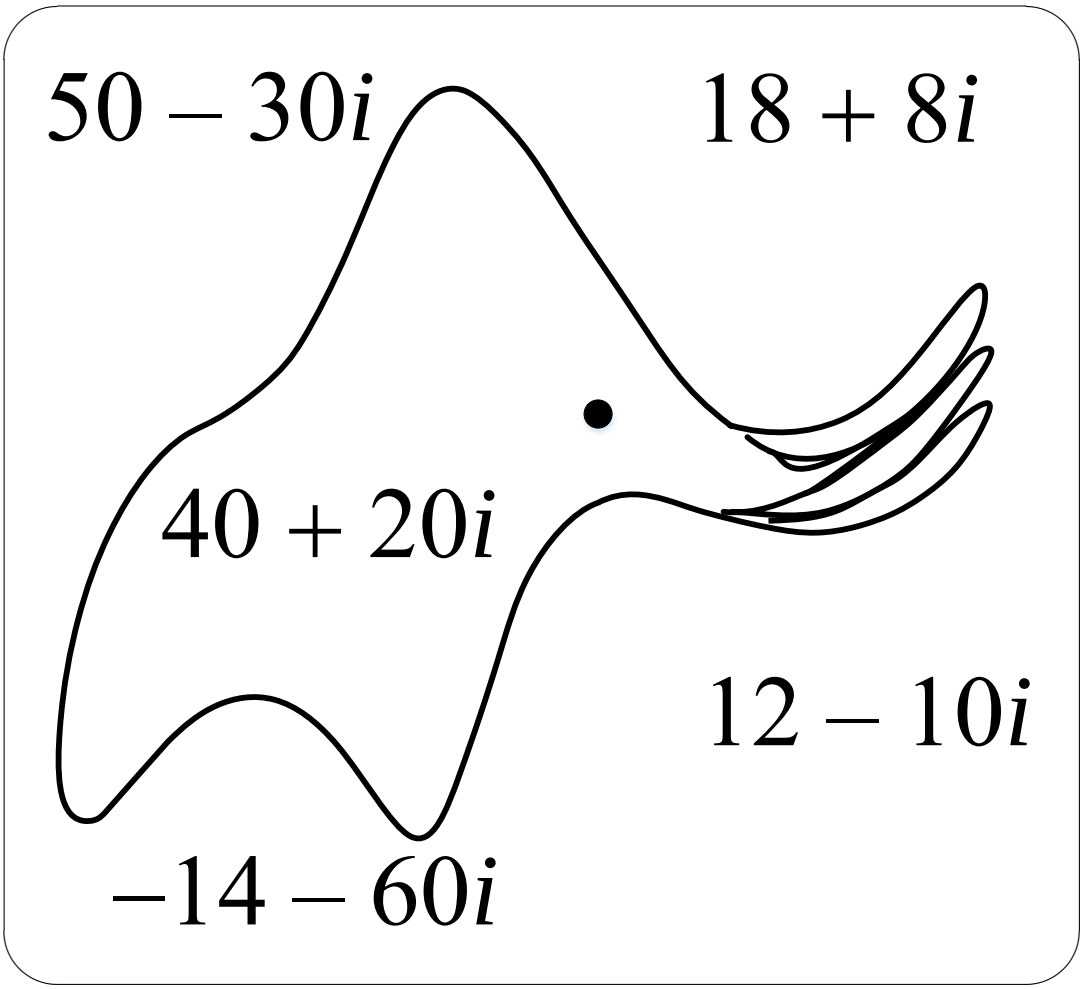


Figure 75: Three snapshots of the wiggling trunk with the five complex parameters that encode the elephant.

In Section 4.11 we have developed the quantum mechanics of two-state systems in terms of real rotations acting on real three-dimensional vectors, and on unitary matrices acting on two-dimensional complex vectors. Actually, we have seen that for each real rotation there must correspond a unitary matrix and vice versa. Consequently, both spaces, the three-dimensional real space $\mathbb{R}^3$ and the two-dimensional complex space $\mathbb{C}^2$, are compatible in this sense; not $\mathbb{R}^4$ and $\mathbb{C}^2$ as one would expect since each complex number is described by two real numbers.

This example is so simple and fundamental that it should apply to all reasonable spaces that are used in physics. Hence, the physical space can be thought of in two different ways. There is the real space that allows to rotate real vectors, and then there is the complex space where unitary matrices act on complex vectors. We accept only spaces that are compatible in this sense. Hence we try to answer the following question: For which pairs (m, n) is the n-dimensional real space $\mathbb{R}^n$ compatible with some m-dimensional complex space $\mathbb{C}^m$ with respect to rotations?

In other words, for given pairs (m, n) we ask whether there is a unique correspondence between real n-dimensional rotations and complex unitary m-dimensional operators. All other dimensions (m, n) that don't satisfy this basic, simple and reasonable requirement would be ruled out. We call this requirement the *space-compatibility postulate*. Notice that spin and polarization satisfy this postulate, since two state-systems can be described in terms of the real space $\mathbb{R}^3$ and the complex space $\mathbb{C}^2$.

But if we consider only the spaces $\mathbb{R}^n$ and $\mathbb{C}^m$, then we would rule out the (3+1)-dimensional spacetime, since the spacetime inner product is not the same as in the four-dimensional real Euclidean space $\mathbb{R}^4$. In Section 4.6 we have defined the inner product (221) of two vectors. In the theory of relativity the concept of pseudo-Euclidean spaces is used. In the following we allow also pseudo-Euclidean spaces in our considerations.

Let p, q be two integers with $p+q = n$. Then the *pseudo-inner product* of two vectors $|\phi\rangle, |\psi\rangle \in \mathbb{C}^n$ is defined as the sum of the first p terms $\phi_i^* \psi_i$, $i = 1, \ldots, p$ minus the sum of the remaining q terms $\phi_j^* \psi_j$, $j = p+1, \ldots, n$, that is,

$$\langle \phi | \psi \rangle = \phi_1^* \psi_1 + \ldots \phi_p^* \psi_p - \phi_{p+1}^* \psi_{p+1} - \ldots - \phi_n^* \psi_n. \tag{333}$$

We denote the complex space equipped with an pseudo-inner product by $\mathbb{C}^{p,q}$, and the real space equipped with an pseudo-inner product by $\mathbb{R}^{p,q}$.

In our geometrical example where a ball is placed on a spherical surface we have considered real rotations and unitary transformations. Both leave the inner product invariant, namely the length of the real normal vectors and the areas of the spherical surface. Such transformations are called symmetry operations.

Now comes our central question. Which pairs of real and complex groups are compatible and admit a unique correspondence, at least between their infinitesimal rotations?

This is an old classical question which has long ago been answered by Barut and Raczka[64] 1965. They provide a table of all possible compatible

[64] Barut et al. [1965]

correspondences. We have displayed these correspondences in Table 5. This

Table 4: Compatible groups and spaces

Real	compatible	Complex
$so(2)$	$\cong$	$u(1)$
$so(1,2)$	$\cong$	$su(1,1)$
$so(3)$	$\cong$	$su(2)$
$so(1,3)$	$\cong$	$sl(2,\mathbb{C})$
$so(4)$	$\cong$	$su(2) \otimes su(2)$
$so(2,4)$	$\cong$	$su(2,2)$
$so(6)$	$\cong$	$su(4)$

is in fact a very surprising result. Normally, one would expect that a compatible complex space exists for every real space. This is obviously not the case. Only 7 pairs exist.

The numbers in both columns denote the dimensions of the underlying spaces, and the symbols "so" or "su' denote the infinitesimal rotations on these spaces. For example, $so(2)$, $so(1,3)$, and $su(2)$ denote the infinitesimal real rotations of the Euclidean space $\mathbb{R}^2$ with determinant equal one, the infinitesimal rotations on the pseudo-Euclidean space $\mathbb{R}^{1,3}$ with determinant equal one, and the infinitesimal unitary rotations on the complex space $\mathbb{C}^2$ with determinant equal one, respectively. The basic definitions as well as the idea of the proof are sketched in appendix 7. Readers without any knowledge of Lie theory should read this appendix first. With capital letters $SO(2)$, $SO(1,3)$, and $SU(2)$ we denote the corresponding groups.

The next surprising observation is that the *Kaluza-Klein theory*, all *string theories*, and *quantum chromodynamics* are ruled out; the real space dimensions $n = 5, 10, 11, 26$ and the complex space dimension $m = 3$ do not occur. The Standard Model of particle physics, a gauge quantum field theory containing the internal symmetries of the product group $SU(3) \otimes SU(2) \otimes U(1)$ does not fit into this table. All these theories do not satisfy our *compatibility postulate*. Hence, the simple example, where we derived everything important about *two-state systems* such as spin and polarization, does not work in these theories. Hence, the simple question "on which side of a plane a ball may be positioned on a spherical surface" cannot not be discussed there. But many physicists, like von Weizsäcker with his *Ur-theory*, think that physical reality can be built up with two-state systems. Remember that two-state systems are the basic building blocks in quantum information theory.

Let us now look more precisely on Table 4. The first compatible pair contains the simplest complex group $U(1)$. Geometrically, it describes the rotational symmetry of a circle using complex (1×1)-matrices. This symmetry is well-known as the symmetry group for electromagnetic interactions.

The second pair is rarely used in physics. However, there are some relations with symplectic algebra and canonical transformations. In the paper of Marcel

Novaes[65] a nice application is presented how to obtain the energy spectrum of the hydrogen without solving the Schrödinger equation.

The third pair shows the compatibility of the three-dimensional real Euclidean space with the two-dimensional complex space. This relationship has various applications, among them the three-dimensional spaces of position and velocity, or electric and magnetic fields. Polarization and spin can be described with two-dimensional complex vectors. The group $SU(2)$ is also known to model weak nuclear interactions.

The fourth pair is well-known. It contains the widely used (3+1)-dimensional spacetime. It satisfies our space-compatibility postulate. But the pseudo-orthogonal Lorentz rotations are related to two-dimensional complex matrices that are not necessarily unitary. Thus they do not preserve the inner product. With regard to our example, this has the consequence that the total area of the spherical surface is not preserved. This fact is another disadvantage in addition to the asymmetry of spacetime. Moreover, it is a non-Euclidean space. These arguments may be hints that the spacetime is helpful, but seemingly not the fundamental space of a physical theory. Consequently, quantum field theories that are based on spacetime would be ruled out, although they might be useful in many situations.

The fifth pair is very promising. The space is Euclidean. The group of real four-dimensional rotations $SO(4)$ contains the group of all real three-dimensional rotations which is compatible with the group $SU(2)$ of complex two-dimensional rotations as shown above. This shows the close relationship between the groups $SO(4)$, $SO(3) \otimes SO(3)$, and $SU(2) \otimes SU(2)$. Both latter groups imply dual (complementary) variables that lead naturally to Heisenberg's uncertainty relations. Moreover, they explain why in electromagnetism we have an electric field and a magnetic field. The same applies to position and velocity or momentum. Finally, the corresponding space explains entanglement and thus avoids distant spooky effects. All these properties coincides with our observations in Section 4.13.

The sixth pair is substantial in twistor theory which is out of scope of this lecture notes.

The last pair is perhaps the most promising space. The group of real six-dimensional rotations $SO(6)$ contains $SO(4)$ as a subgroup, and thus contains all its nice properties. Moreover, it is a well-understood Euclidean space.

Mirman investigated in a series of papers how symmetry principles and group theory can be used in order to determine the dimension of space. A nice presentation, proofs, many references and interesting aspects can be found in his excellently written book "Our Almost Impossible Universe: Why the Laws of Nature Make the Existence of Humans Extraordinarily Unlikely"[66]. He concluded, however, that physics and our universe is only possible in the $3 + 1$-dimensional spacetime. Using further arguments, with which I do not agree, he ruled out any other dimension.

[65] Noveas [2004]
[66] Mirman [2001, p. 197]

4.16 Teleportation: the Experiment

Zeilinger[67] writes in his articles that "the science-fiction dream of 'beaming' objects from place to place is now reality - at least for particles of light". Moreover, he describes rather detailed his experimental set-up. We have already discussed teleportation in Section 3.7 with the association that Alice and Bob share an entangled pair of photons, and Alice wants to send an unknown photon state to Bob who is far far away with his photon. Since it is not possible to transport a specific photon, not surprisingly in the Innsbruck experiment Alice and Bob or any other observers are absent. Especially, Bob does not carry a photon. Thus, let us now consider to what extend the described association with Alice and Bob is realized in the Innsbruck experimental set-up. This experiment is briefly described in Figure 76.

In this experiment a brief pulse of ultraviolet laser light passes from left to the right through a beta barium borate crystal BBB and creates an entangled pair of photons $|AB\rangle$. Photon $|A\rangle$ is reflected at mirror M3 and sent to the semireflecting mirror SM that serves as a polarizing beam splitter. The other one $|B\rangle$ moves to the calcite crystal and is detected either in D3 or D4. The remaining part of the pulse is reflected back at mirror M1, passes through the crystal once more, and creates a second entangled pair of photons $|CD\rangle$. Photon $|C\rangle$ interacts with detector D5, and thus certifies that its colleague $|D\rangle$ is ready for teleportation. Photon $|D\rangle$ passes a calcite crystal which has a certain given polarization axis. In one of both exit beams of the calcite this photon comes out with a new polarization state, say $|\psi\rangle$. This new prepared state is known, because the calcite's polarization axis is known. The interaction of $|\psi\rangle$ with $|A\rangle$ at the SM implies four possibilities: both photons are reflected, both photons pass, $|\psi\rangle$ is reflected and $|A\rangle$ passes, and vice versa. It turns out, and we show it below that quantum interference implies that photon state $|B\rangle$ coincides with state $|\psi\rangle$ whenever both detectors D1 and D2 register a photon. Then it is said that teleportation is testified. Subsequently, we describe the essential parts more detailed.

The first experimental task is the creation of entangled pairs of photons. One way to do this is called *parametric down-conversion*: a laser emits photons of high frequency ω_P and momentum k_P, say ultraviolet light, and the photon passes through a BBB which creates two photons of lower energy, say red ones. The two photons of lower energy must satisfy the conservation of energy ($\omega_P = \omega_A + \omega_B$) and the conservation of momentum ($k_P = k_A + k_B$). These conservation conditions imply that photon pairs are emitted which have no definite polarization, but are entangled. Actually, using a second BBB and a half wave plate it is possible to create every Bell state. Nowadays, producing entangled pairs of photons is experimentally an easy task.

The second task is to entangle two independent photons, photon 1 in state $|\psi\rangle$ and photon 2 in state $|A\rangle$. This is the most difficult experimental part of teleportation. It uses a polarizing beam splitter SM which, roughly spoken, consists of a birefringent material and splits light into beams of different polarization.

[67] **?**Zeilinger [2000]

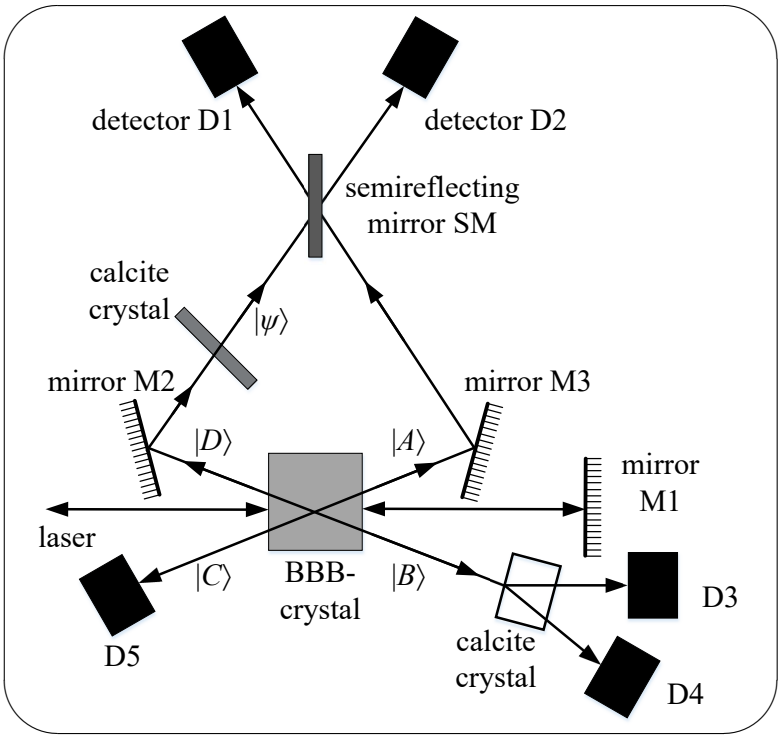


Figure 76: In the Innsbruck experiment two short pulses of laser light pass from left to right through a beta barium borate crystal BBB. Let the pulse produce the entangled pair of photons $|AB\rangle$. Let the second pulse, reflected back at mirror M1, create a further entangled pair $|CD\rangle$ when passing the crystal. The photon $|D\rangle$ is reflected at mirror M2 and passes a calcite that prepares photon $|D\rangle$ in the specific state $|\psi\rangle$. If a photon $|C\rangle$ is detected in detector D5, it follows that in fact $|\psi\rangle$ is sent to the polarizing beam splitter SM that (assuming an idealized experimental set-up) superposes there with photon $|A\rangle$ yielding a Bell state. This superposition destroys $|\psi\rangle$. It turns out that the photon state $|B\rangle$ coincides with $|\psi\rangle$ provided the detectors D1 and D2 have registered two photons. Hence, teleportation is claimed.

Let the photon to be teleported have the polarization

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \tag{334}$$

and let

$$|AB\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \tag{335}$$

be the entangled Bell state of the pair of photons $|A\rangle$ and $|B\rangle$. The current state of all three photons is

$$\begin{aligned} |\psi AB\rangle &= (\alpha|0\rangle + \beta|1\rangle) \otimes \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \\ &= \frac{1}{\sqrt{2}}(\alpha|001\rangle + \beta|101\rangle - \alpha|010\rangle - \beta|110\rangle). \end{aligned} \tag{336}$$

We have already considered the four maximally entangled Bell states

$$\begin{aligned} |\phi^-_{\neq}\rangle &= \frac{1}{\sqrt{2}}(|10\rangle - |01\rangle), \\ |\phi^+_{\neq}\rangle &= \frac{1}{\sqrt{2}}(|10\rangle + |01\rangle), \\ |\phi^-_{=}\rangle &= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), \\ |\phi^+_{=}\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle). \end{aligned} \tag{337}$$

The first one is antisymmetric, that is, it changes the sign if we change 0 to 1 and 1 to 0. The remaining three Bell states are invariant when interchanging 0 and 1. Hence, they are symmetric.

Using the coordinate representation of these states, it follows immediately that the four Bell states form an orthonormal base in $\mathbb{C}^4$. Hence, we can express the state $|\psi AB\rangle$ in terms of the Bell states. A short computation shows

$$\begin{aligned} |\psi AB\rangle = \frac{1}{\sqrt{2}} \{ &|\phi^-_{\neq}\rangle \otimes (-\alpha|0\rangle - \beta|1\rangle) \\ &+|\phi^+_{\neq}\rangle \otimes (-\alpha|0\rangle + \beta|1\rangle) \\ &+|\phi^-_{=}\rangle \otimes (\alpha|1\rangle - \beta|0\rangle) \\ &+ |\phi^+_{=}\rangle \otimes (\alpha|1\rangle + \beta|0\rangle)\} . \end{aligned} \tag{338}$$

Notice that this is simply a mathematical manipulation that expresses $|\psi AB\rangle$ with respect to the basis of Bell states; nothing physical has happened.

Therefore, the photons $|\psi\rangle$ and $|A\rangle$ interact at the polarizing beam splitter yielding one of the four Bell states, instantaneously Bob's third photon is in one of the four states

$$\begin{aligned} |B\rangle \in \{ \quad & (-\alpha|0\rangle - \beta|1\rangle), \quad (-\alpha|0\rangle + \beta|1\rangle), \\ & (\alpha|1\rangle + \beta|0\rangle), \quad (\alpha|1\rangle - \beta|0\rangle) \quad \}, \end{aligned} \tag{339}$$

in each case with probability 1/4 according to Born's rule.

There are three possibilities that can happen: both particles are detected in D1, or both particles are detected in D2, or both detectors register one particle. Let us consider the latter case. A deeper analysis shows that in the case where both detectors register a photon, the third photon $|B\rangle$ has the original state $|\psi\rangle$. Thus, instantaneous teleportation is claimed.

Actually, the situation is more complicated. Teleportation means that a complete object is teleported. But the complete object or unity is the entangled pair, not a fragment of the pair. Since we underlay the space $X \times V$, the entangled pairs $|AB\rangle$ and $|CD\rangle$ are welded in V. Spooky distant effects don't occur in this space. This entanglement is also expressed in the conservation of energy and momentum. The complete pair $|CD\rangle$ is not teleported, only its fragment $|D\rangle$ which changes to $|\psi\rangle$ when passing the calcite crystal. Thus, it is doubtful to speak of teleportation. Actually, an interaction between $|A\rangle$ and $|\psi\rangle$ takes place at SM, which can be described as a superposition of the four Bell states. This, of course, must have a related influence on $|B\rangle$, due to the entanglement between $|A\rangle$ and $|B\rangle$.

An important aspect is that the conflict with the theory of relativity has completely disappeared, since there are no more distant effects in the position-velocity space. The pair is a unit. The usual arguments about the conflict between quantum teleportation and relativity theory, as discussed in Section 3.10, vanish.

4.17 Uncertainty Principle

Some years ago a doctor in a hospital asked me whether the uncertainty in the medical diagnosis and the medical treatment is a manifestation of Heisenberg's uncertainty principle, that "nothing in nature is precisely defined". Actually, I said "No". My consideration was that a person consists of about 10^{28} interaction atoms that average to classical values. A person walking one body length performs an action value of about $10^{36}\hbar$. From the point of view of medicine human beings are best described by average values and probability theory.

Frequently, *uncertainty* reflects situations that involve imperfect or unknown *information*. But it arises in several other different ways. For instance, if one observable is measured, or if predictions about future events are required. It appears in many areas of research, including physics, engineering, economics, sociology or information science.

Probability theory is widely used to describe uncertainty, in many cases for one observable or random variable. But in quantum theory, the famous *Heisenberg uncertainty principle* makes a statement about **two observables**. This principle is known to put limits on the accuracy about two complementary quantities like position and momentum or different spin directions. Historically, Heisenberg (1927) derived the position-momentum uncertainty in two ways. Firstly, by localizing a particle through scattering with high-energy photons, and secondly by investigating the Fourier relationships of free wave packets. In both cases he showed that the product of position and momentum uncertainty is of the order of Planck's constant $\hbar$. Heisenberg's notation of uncertainty cannot be explained just by ignorance, but it is believed by most physicists that this principle is a fundamental property of nature. Exemplary, we quote from the beautiful written textbook of Zetteli[68] about quantum mechanics:

> On the one hand, the Davisson-Germes and the double-slit experiments have shown that microscopic material particles give rise to interference pattern. To account for the interference patterns, we have seen that it is imperative to describe microscopic particles by means of waves. Waves are not localized in space. As a result we have to give up on accuracy to describe microscopic particles, for waves give at best a probabilistic account. .... The classical concepts of exact position, exact momentum, and unique path of a particle therefore make no sense at the microscopic scale. This is the essence of Heisenberg's uncertainty principle.

In the following we shall argue that the uncertain principle is not restricted to microscopic problems, but applies also to macroscopic ones. Moreover, we give another interpretation which results from looking at quantum mechanics as a theory of the future. Let us understand what we mean by *uncertainty* from the mathematical point of view. Given a state

$$|\psi\rangle = \sum_{i\in I} \xi_i |i\rangle, \quad \sum_i |\xi_i|^2 = 1, \quad \xi_i \in \mathbb{S} \tag{340}$$

[68] Zetteli [2009]

we derived in Section 4.10 the *expectation value* of an observable

$$\langle \hat{A} \rangle_\psi = \langle \psi | \hat{A} | \psi \rangle, \tag{341}$$

and its *variance*

$$(\Delta \hat{A})^2 = \langle \hat{A}^2 \rangle_\psi - \langle \hat{A} \rangle_\psi^2 = \langle \psi | (\hat{A} - \langle \hat{A} \rangle_\psi \hat{1})^2 | \psi \rangle \tag{342}$$

the latter also called *uncertainty*. Since $\hat{A}$ is Hermitian, equation (342) implies the identity

$$\begin{aligned} (\Delta A)^2 &= \langle \psi | (\hat{A} - \langle \hat{A} \rangle_\psi \hat{1})^\dagger (\hat{A} - \langle \hat{A} \rangle_\psi \hat{1}) | \psi \rangle \\ &= \| (\hat{A} - \langle \hat{A} \rangle_\psi \hat{1}) | \psi \rangle \|^2. \end{aligned} \tag{343}$$

Hence, the variance vanishes if and only if

$$(\hat{A} - \langle \hat{A} \rangle_\psi \hat{1}) | \psi \rangle = 0, \tag{344}$$

that is, state $|\psi\rangle$ must be an eigenvector of $\hat{A}$. Equation (245) tells us that $|\psi\rangle$ must be one of the base states $|i\rangle$. We notice that uncertainty of an observable is not present if state $|\psi\rangle$ is one of its base states. In classical mechanics $\mathbb{S} = \mathbb{N}$, each state is a base state, and there can be no uncertainty.

In classical probability theory and in quantum mechanics the set of base states and the set of states differ such that uncertainty of a random variable or observable is present. But this uncertainty has nothing to do with Heisenberg's uncertainty principle in quantum mechanics.

A fundamental question in quantum theory is:

- Under which conditions are two observables $\hat{A}$ and $\hat{B}$ certain in conjunction, that is, their variances $(\Delta \hat{A})^2$ and $(\Delta \hat{B})^2$ vanish always simultaneously?

Equations (343) and (344) tell us that both observables are certain simultaneously, if and only if they have the same eigenvectors, that is,

$$\hat{A} | i \rangle = A_i | i \rangle \quad \text{and} \quad \hat{B} | i \rangle = B_i | i \rangle \tag{345}$$

for all base states $|i\rangle$ with $i \in I$. In particular, both observables must be defined on the same set of base states. Only their labeling, namely their eigenvalues, can be different. In classical mechanics and classical probability theory we have seen that there is only one bases, and each number of observables are certain simultaneously.

In quantum theory, where $\mathbb{S} = \mathbb{C}$, this nice property does not hold true, since their exist different bases, as we have seen. Let us look in greater detail at the two observables

$$\hat{C} := \hat{A} - \langle \hat{A} \rangle_\psi \hat{1}, \tag{346}$$

and

$$\hat{D} := \hat{B} - \langle \hat{B} \rangle_\psi \hat{1}. \tag{347}$$

Both observables are Hermitian, since $\hat{A}$ and $\hat{B}$ are Hermitian. Some simple matrix multiplication yield the equations

$$\begin{aligned} (\hat{C}\hat{D})^\dagger &= \hat{D}^\dagger\hat{C}^\dagger = \hat{D}\hat{C}, && (348)\\ \langle\hat{C}\hat{D}\rangle_\psi &= \langle\psi|\hat{C}\hat{D}|\psi\rangle = x + iy, && (349)\end{aligned}$$

and

$$\langle\hat{D}\hat{C}\rangle_\psi = \langle\psi|(\hat{C}\hat{D})^\dagger|\psi\rangle = x - iy.$$

Moreover, the commutator

$$[\hat{C},\hat{D}] := \hat{C}\hat{D} - \hat{D}\hat{C}, \tag{350}$$

satisfies

$$[\hat{C},\hat{D}] = [\hat{A},\hat{B}], \tag{351}$$

and

$$\begin{aligned} \langle[\hat{C},\hat{D}]\rangle_\psi &= \langle\psi|\hat{C}\hat{D}|\psi\rangle - \langle\psi|\hat{D}\hat{C}|\psi\rangle = 2iy \\ &= 2i\mathrm{im}(\langle\hat{C}\hat{D}\rangle_\psi). \end{aligned} \tag{352}$$

Finally, it follows that

$$\begin{aligned} \langle\psi|\hat{C}^2|\psi\rangle &= \langle\psi|\hat{A}^2|\psi\rangle - 2\langle\hat{A}\rangle_\psi\langle\psi|\hat{A}|\psi\rangle + \langle\hat{A}\rangle_\psi^2 \\ &= \langle\hat{A}^2\rangle_\psi - \langle\hat{A}\rangle_\psi^2 \\ &= (\Delta\hat{A})^2, \end{aligned} \tag{353}$$

and analogously

$$\langle\psi|\hat{D}|\psi\rangle = (\Delta\hat{B})^2. \tag{354}$$

We make use of (351), (352), the Cauchy-Schwarz inequality, (353) and (354) yielding

$$\begin{aligned} |\langle\psi|[\hat{A},\hat{B}]|\psi\rangle|^2 &= |\langle\psi|[\hat{C},\hat{D}]|\psi\rangle|^2 \\ &= 4|\langle\psi|[\hat{C}\hat{D}]|\psi\rangle|^2 \\ &= 4|(\langle\psi|\hat{C})\cdot(\hat{D}|\psi\rangle)|^2 \\ &\leq 4\|\langle\psi|\hat{C}\|^2\cdot\|\hat{D}|\psi\rangle\|^2 \\ &= 4\langle\psi|\hat{C}^2|\psi\rangle\cdot\langle\psi|\hat{D}^2|\psi\rangle \\ &= 4(\Delta\hat{A})^2(\Delta\hat{B})^2. \end{aligned} \tag{355}$$

We can rewrite this inequality as

$$\Delta\hat{A}\cdot\Delta\hat{B} \geq \frac{1}{2}|\langle\psi|[\hat{A},\hat{B}]|\psi\rangle|. \tag{356}$$

This mathematical inequality, derived by Robertson 1929, is called Heisenberg's general *uncertainty principle*.

Since variances are always nonnegative, this inequality is meaningless and has no content if the commutator $[\hat{A}, \hat{B}] = 0$. A well-known mathematical theorem says that this commutator vanishes if and only if the sets of eigenvectors of $\hat{A}$ and $\hat{B}$ are equal.

If on the other hand both observables do not commute, that is, their commutator $[\hat{A}, \hat{B}]$ does not vanish, one cannot know the value of both observables at once exactly. Or expressed in another way: knowing the value of one observable precisely, makes the other observable completely vague. This can only happen in quantum theory, where various orthonormal bases are allowed.

Now we apply this uncertainty principle to the Pauli matrices. Remember that we have derived these matrices by considering a macroscopic problem, namely to position a soccer ball on a spherical surface. It is a straightforward matrix-matrix multiplication to compute the commutator for the *Pauli matrices* (282), (283) and (284):

$$\hat{\sigma}_x, \hat{\sigma}_y = [\hat{\sigma}_x \hat{\sigma}_y] - \hat{\sigma}_y \hat{\sigma}_x = 2i\hat{\sigma}_z, \quad [\hat{\sigma}_y, \hat{\sigma}_z] = 2i\hat{\sigma}_x, \quad [\hat{\sigma}_z, \hat{\sigma}_x] = 2i\hat{\sigma}_y. \tag{357}$$

Let

$$|\psi\rangle = \psi_0|0z\rangle + \psi_1|1z\rangle, \quad \psi_0, \psi_1 \in \mathbb{C}, \quad |\psi_0|^2 + |\psi_1|^2 = 1 \tag{358}$$

be a general state of the two-state system. Then

$$\begin{aligned} \langle\psi|[\hat{\sigma}_z, \hat{\sigma}_x]|\psi\rangle &= \langle\psi|2i\hat{\sigma}_y|\psi\rangle \\ &= 2i \quad \{(\psi_0^*)^2\langle 0z|\hat{\sigma}_y|0z\rangle + \psi_0^*\psi_1\langle 0z|\hat{\sigma}_y|1z\rangle \\ &\qquad +\psi_1^*\psi_0\langle 1z|\hat{\sigma}_y|0z\rangle + (\psi_1^*)^2\langle 1z|\hat{\sigma}_y|1z\rangle\}\,. \end{aligned} \tag{359}$$

Since

$$\hat{\sigma}_y|0z\rangle = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ i \end{pmatrix}, \tag{360}$$

and

$$\hat{\sigma}_y|1z\rangle = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -i \\ 0 \end{pmatrix}, \tag{361}$$

it follows

$$\begin{aligned} \langle\psi|[\hat{\sigma}_z, \hat{\sigma}_x]|\psi\rangle &= 2i\,\{(\psi_0^*)^2 \cdot 0 + \psi_0^*\psi_1(-i) + \psi_0\psi_1^* i + (\psi_1^*)^2 0\} \\ &= 4\ \mathrm{im}(\psi_0^*\psi_1) \end{aligned} \tag{362}$$

From (355) we obtain the uncertainty principle

$$\Delta\hat{\sigma}_z \cdot \Delta\hat{\sigma}_x \geq\ \mathrm{im}(\psi_0^*\psi_1). \tag{363}$$

When ψ_0 or ψ_1 vanishes, the right hand side is zero, and this inequality is irrelevant. But otherwise we obtain a lower limit for the product of the uncertainties of $\hat{\sigma}_z$ and $\hat{\sigma}_x$. In the case $\psi_0 = 1/\sqrt{2}$ and $\psi_1 = i/\sqrt{2}$ we get

$$\Delta\hat{\sigma}_z \cdot \Delta\hat{\sigma}_x \geq \frac{1}{2}. \tag{364}$$

Let us investigate the meaning of Heisenberg's uncertainty principle for our macroscopic problem. From equation (280) it follows that $|\psi\rangle = |0y\rangle$. This denotes one half that results from bisecting the sphere with the xz-plane which is perpendicular to the y-axis. We remember that the observables $\hat{\sigma}_z$ and $\hat{\sigma}_x$ label the two halves that result from bisecting the sphere with the xy-plane and the yz-plane, respectively. The area of the halve $|0y\rangle$ contains parts of all areas $|0z\rangle$, $|1z\rangle$, $|0x\rangle$, and $|1x\rangle$. Hence, the knowledge that a ball is positioned on the surface of $|0y\rangle$ does not determine whether the ball is above or below the xy-plane or above or below the yz-plane. In other words, if in a future action we would position a ball in the half $|0y\rangle$, then it is uncertain if the ball would be above or below the xy-plane or above or below the yz-plane. This is the content of the uncertainty principle in this case. It is simple, not weird or paradox, clear from the geometry, supports our view that quantum mechanics is the theory of the future, and applies also to macroscopic problems.

4.18 Position-Momentum Commutation Relation

For understanding the position-momentum commutation relation it is perhaps the best to go back to an idea of the electrical engineer Heaviside (1850-1925). He observed that differential operators often can be processed as numbers. This fact can be used to solve certain types of differential equations via simple algebraic equations.

If we want to solve the differential equation

$$2y(x) + \frac{d}{dx}y(x) = h(x), \tag{365}$$

we can treat the differential operator $\frac{d}{dx}$ as an ordinary number and write the equation in the form

$$\left(1 + \frac{1}{2}\frac{d}{dx}\right)y(x) = \frac{1}{2}h(x). \tag{366}$$

Formally, we can divide through $(1 + \frac{1}{2}\frac{d}{dx})$ and apply the geometric series corresponding to the term $\frac{1}{2}\frac{d}{dx}$. Then we get

$$y(x) = \frac{1}{2}\left(1 - \frac{1}{2}\frac{d}{dx} + \frac{1}{4}\frac{d^2}{dx^2} - \frac{1}{8}\frac{d^3}{dx^3} + \ldots\right)h(x). \tag{367}$$

If we set for example $h(x) = 2x^3$, then we find the solution

$$y(x) = x^3 - \frac{3}{2}x^2 + \frac{3}{2}x - \frac{3}{4}. \tag{368}$$

This formal procedure can be used in various applications. In particular, it can be viewed as a key procedure in quantum mechanics.

Heaviside was confronted with declination and opposition. Of course, when treating operators as numbers one must be very careful, because numbers commute, but in general operators don't commute. If we think of x and $\frac{d}{dx}$ as operators, namely multiplication and differentiation operators that act on functions $\psi(x)$, then

$$\left(\frac{d}{dx}x - x\frac{d}{dx}\right)\psi(x) = \frac{d}{dx}(x\psi(x)) - x\frac{d}{dx}\psi(x) = \psi(x). \tag{369}$$

Hence, both operators don't commute and satisfy the *commutation relation*

$$\frac{d}{dx}x - x\frac{d}{dx} = \hat{1}, \tag{370}$$

where $\hat{1}$ denotes the identity operator. In first order for small d

$$\psi(x + d) = \psi(x) + d\frac{d}{dx}\psi(x). \tag{371}$$

Thus the differential operator $\frac{d}{dx}$ generates an infinitesimal translation $x \to x+d$. If ψ describes the state function of any object, then there are two cases to consider. Either the object is transported through a distance d, then we speak

of an *active transformation*, or the object is at rest and the origin of coordinate x is displaced by $-d$. The latter case is called a *passive transformation*.

Infinitesimal translations of a system in an n-dimensional configuration space are closely related to what we mean by *momentum*, and thus not surprisingly in quantum mechanics the differential operators in generalized coordinates

$$\hat{P}_j = -i\hbar\frac{\partial}{\partial q_j} \tag{372}$$

describe the linear momentum. The operator $\hat{P}_j$ is called the *momentum operator* and acts on state functions that are elements in an appropriate Hilbert space. The momentum operator coincides with the same interpretation as in classical mechanics, where momentum can be viewed as the generator of translations. *Planck's constant* $\hbar$ has the dimension of the product of position and momentum. Since we differentiate with respect to $\frac{\partial}{\partial q_j}$, thus dividing by position, we should multiply with a product of position and momentum in order to obtain the dimensions of momentum. This is the reason for multiplying the differentiation operator with Planck's constant. The complex factor i is introduced to ensure that the momentum operator is Hermitian, thus implying real measurable values.

If we are interested in the position of a particle or a system, then in quantum mechanics position states $|q_k\rangle$ are introduced that describe the kth position coordinate as a Dirac delta function, that is,

$$|q_k\rangle = \delta(q_k' - q_k). \tag{373}$$

The position operator $\hat{Q}_k$ is defined as the Hermitian operator[69] which has the position eigenvalues q_k, that is,

$$\hat{Q}_k|q_k\rangle = q_k|q_k\rangle, \tag{374}$$

where the position coordinate q_k runs continuously from $-\infty$ to ∞.

Exactly as deriving (370), we find immediately the *commutation relations* for the momentum operators $\hat{P}_j$ and the position operators $\hat{Q}_k$, that multiplies a function ψ with the position q_k:

$$[\hat{P}_j, \hat{P}_k] = [\hat{Q}_j, \hat{Q}_k] = 0, \quad [\hat{P}_j, \hat{Q}_k] = -i\hbar\delta_{jk}\hat{1}. \tag{375}$$

The commutation relation is one of the most fundamental equations in physics. For example, the typical values of an electron in a hydrogen atom are 5×10^{-11} m for the position coordinate describing the radius, and 2×10^{-24} kg $\times$ m/s for the momentum. Planck's constant has the value

$$\hbar = 1.054\times10^{-34}\text{kg m}^2/s, \tag{376}$$

that is close to the product of the radius and the momentum. These are very small quantities, and the commutation relation $[\hat{P}_j, \hat{Q}_j] = -i\hbar 1$ shows that

[69] For a derivation and more details about position and momentum operators see Susskind [2014, Chapter 8], Penrose [2005, Chapter 21], or Schiff [1968, page 175].

there is a small difference between $\hat{P}_j\hat{Q}_j$ and $\hat{Q}_j\hat{P}_j$. However, what is small? Smallness has not an absolute meaning, but makes sense only when comparing objects. Our units kg, m, and s are natural to us, because in daily life we are surrounded by big objects consisting of more than 10^{26} atoms. But why are we so big? Well, to obtain living and functioning objects, such as planets or animals, requires a lot of atoms, moving slowly. But when looking at small objects like atoms, Planck's constant becomes really large.

Now we can apply the uncertainty principle (356) using the commutation relation (375). Then we obtain

$$\Delta\hat{Q}_j\Delta\hat{P}_j \geq \frac{1}{2}\hbar \tag{377}$$

In particular, we see that for the position-momentum pair the uncertainty principle turns out to be state independent. The interpretation is similar as above. If we would know in the future the exact position of a particle, then its momentum would be completely undetermined. The same holds true to the momentum. Formulated in terms of machines: if the momentum is known exactly, then all positions of the position machines are possibilities.

Finally, it should be mentioned that there is a very interesting, beautiful written, and extensive physic book by Schiller [70] where he investigated in part IV "The Quantum of Change". This book is highly recommendable for students in engineering. His starting thesis is:

- the action values $S(t)$ and $S(t+\Delta t)$ between two successive events of a quantum system, a time Δt apart, cannot vanish. Roughly spoken, action is energy E times time t. More precisely, they satisfy the inequality

$$|S(t+\Delta t) - S(t)| \geq \frac{\hbar}{2}. \tag{378}$$

This *minimum action principle* is in complete contrast to classical physics and has never failed a single test, as pointed out in his book. He writes that it would perhaps be better named the *quantum of change*.

He derives from this action principle in a modest manner several spectacular statements using only simple mathematics. Formula (378) takes the form

$$|(E \pm \Delta E)(t+\Delta t) - Et| = |E\Delta t \pm \Delta E t \pm \Delta E\Delta t| \geq \frac{\hbar}{2}. \tag{379}$$

If we choose a suitable coordinate system, we can set E and t equals to zero, yielding

$$\Delta E\Delta t \geq \frac{\hbar}{2}, \tag{380}$$

when ΔE is the change of the energy of the system, and Δt is the time between two successive events. This inequality is called the *time-energy uncertainty principle*, which is derived from the *minimum action principle* (378).

[70] Schiller [2016]

For a free system with momentum p and mass 1, we can choose a coordinate system such that $E = 0$ but $\Delta E = \Delta p^2$. Since for Cartesian coordinates $x = q$ it is $\Delta x = \Delta p \Delta t$, the inequality (380) yield

$$\Delta p \Delta x \geq \frac{\hbar}{2}, \tag{381}$$

which is the *position-momentum uncertainty principle*. This is really a very simple derivation of the uncertainty principle.

Based on the quantum of change, Schiller deduced several consequences that cannot be found in (all?) other textbooks:

- In nature there is no *rest*.
- In nature there is no perfectly straight or perfectly *uniform motion*.
- Perfect *clocks* do not exist.
- Motion *backwards in time* is possible over microscopic times and distances.
- The *vacuum* is not empty.
- Photons have no position and cannot be localized.
- Microscopic systems behave randomly.
- Light can move faster than the speed of light c.

So what is the truth in physics?

5 Quantization and Fields

A *field* is a physical quantity that has a value for each point in space and time. This value may be a number, a vector, or an operator. *Quantum Field Theory (QFT)* is a mathematical framework describing the extension of quantum mechanics, dealing with a finite number of particles, to fields. Teller writes in the preface of his well-known book [71]:

> *QUANTUM FIELD THEORY is a notoriously hard theory to learn. The best physics students do well with it, but many able students flounder, eventually resigning themselves to going through the motions with their problem sets to make it through the course. Among philosophers of physics, I have heard many colleagues express interest, only to learn a year or two later that they had somehow gotten involved in other things. I too have found the subject extremely difficult, and after much effort, I have managed to understand only some of the basics.*

In fact, QFT is a really difficult subject. I hope that this section might help students who have an interest in its conceptual structure and want to get a quick view of some basic ideas. An interesting survey can be found in the "Stanford Encyclopedia of Philosophy".

The major goal of this section is to show that our imagination of machines, characterized by possibilities, and some results from the section about unification via semimodules leads in a natural way to quantization and QFT. Moreover, we have to show that our physical model without an external time parameter t is suitable for QFT.

QFT has grown successively in a very complex way, and its interpretation is sometimes notably weird. In contrast to many other physical theories QFT is not well-defined. Instead one can formulate several totally different models that all have their merits but also limits. The most important applications are in particle physics. In particular, QFT describes scattering processes where particles are created while others are destroyed.

We start with the quantum harmonic oscillator. Its energy values are discrete in contrast to the classical harmonic oscillator possessing continuous energy values. Moreover, the mathematical derivation of the quantum harmonic oscillator from the classical one is a perfect example of quantization. A harmonic oscillator can be viewed as a quantum field theory in one parameter t, but without any space dimensions. We shall develop the quantization of fields on the background of the simplest continuum theory, namely vibrations of an atomic chain. This provides an entrance to quantum electrodynamics and other quantum field theories.

[71] Teller [1997]

5.1 Quantization Rules

The word quantum comes from the Latin word quantus, which means "how great". In physics, this word is referred to the property that certain quantities can take on only discrete values. *Quantization* is known as the process of constraining continuous quantities, such as real or complex numbers, to a discrete set, such as the integers. Hence, it is a little bit similar to discretization as we shall see in Section 5.4.

Are there different processes of constraining quantities to discrete sets? What do these processes look like? Is there a common reason for quantization?

The first quantization rule is due to Planck 1900 who postulated the quantization of energy $E = n\hbar\omega$ for his successful explanation of blackbody radiation. In order to understand certain properties of atoms Bohr 1913 postulated that the electron's orbital angular momentum $L = n\hbar$ is quantized. Three years later Wilson and Sommerfeld offered a scheme that quantized the action of a system. They derived Planck's and Bohr's postulates as special cases. Remember, that we have mentioned this scheme and consequences already in Section 4.18

Today, it is common to quantize a physical system by replacing the classical dynamical variables by means of Hermitian operators. It is a recipe that takes us from the Hamiltonian formalism of classical mechanics to quantum theory. This scheme, called *canonical quantization*, comprises the previous quantization rules. At a first glance this recipe sounds really strange as Penrose writes:

> *In fact, it seems to be mathematically completely crazy! ... What are we to do with this crazy-looking operator/momentum? The role of this 'quantum-mechanical momentum', $i\hbar\frac{\partial}{\partial x^a}$, is that it is to be slotted into the classical Hamiltonian function $H(p_1, ..., p_N; x^1, ..., x^N)$, just where the old classical momentum p_a used to be. This is the key to the procedure known as (canonical) quantization.* Penrose 2004, Chapter 21.2

From the point of view of Section 4, however, this replacement turns out to be rather natural and not strange. There we started with the basic question

> What are the base states of a physical model?

When answering this question, it was easy to define states and observables.

Our initial situation are the classical dynamical variables: position q, momentum p, and functions $F(p, q)$, like energy or angular momentum. For example, the one-dimensional harmonic oscillator has the energy function $H(p, q) = p^2/2m + m\omega^2 q^2/2$. Position and momentum can be represented as registers, that is, a sequence of binary questions. Each real value can be approximated sufficiently accurate by a finite register. Both, position and momentum, are observables that map the registers to values p and q, respectively. Now we go to the equivalent vector representation. Then the registers become base states $|p\rangle$ and $|q\rangle$, and the observables become diagonal matrices $\hat{P}$ and

$\hat{Q}$ with respect to their base states. The values q and p are the eigenvalues of these observables[72].

In classical mechanics we have shown that there exists exactly one orthonormal basis. Hence, both observables $\hat{P}$ and $\hat{Q}$ are diagonal matrices with values p and q on the diagonal. They contain all information about position and momentum. Diagonal matrices commute. The obvious way to get all values $F(p,q)$ is to replace the values p and q by the diagonal matrices $\hat{P}$ and $\hat{Q}$, such that we obtain the observable $\hat{F} = \hat{F}(\hat{P},\hat{Q})$. This is the diagonal matrix with all classical values $F(p,q)$ on the diagonal, thus containing all information.

In quantum mechanics the probabilistic theory of the future, the orthonormal basis of position vectors and the basis of momentum vectors are different, as we have seen. From (375) we obtain a non-zero commutator $[\hat{P},\hat{Q}] = i\hbar$. Hence, if we use the same idea and take the observable $\hat{F} = \hat{F}(\hat{P},\hat{Q})$ we cannot expect to obtain the eigenvalues $F(p,q)$. We must carefully incorporate the commutator. Then some simple manipulations give us the eigenvalues and eigenvectors of $\hat{F}$. It turns out that the eigenvalues are quantized and do not coincide with the classical values. Quantization is the result of some inevitable mathematical calculations using the commutation relations, but not of certain ad hoc postulates.

However, due to the fact of non-commuting operators, the replacement of dynamical variables by Hermitian operators is not obvious. Already the simple classical product term qp may be written as pq or $0.5(qp+pq)$. When replacing these expressions by non-commuting operators they lead to different results. This problem is called the *factor-ordering problem*. One helpful way to find the correct operator expressions is to use *Poisson brackets*; the interested reader is referred to the rich literature.

Keep in mind: Quantization is the process of constraining continuous quantities, such as real or complex numbers, to a discrete set. In canonical quantization the classical dynamical variables position and momentum in expressions are replaced by the Hermitian position and momentum operators. Due to the fact that operators don't commute, dynamical values are quantized. This replacement must be done carefully, since the *factor-ordering problem* emerges. Canonical quantization comprises the well-known quantization rules of Planck, Bohr, Sommerfeld, and Wilson. The underlying reason of quantization is the fact that quantum mechanics has various orthonormal bases in contrast to classical mechanics and classical probabilistic mechanics.

[72] If all possible values for q would be discrete, the eigenvalue problem can be written down in the usual form. But if q is a continuous real variable, then in order to be mathematical rigorous, we require distribution theory. However, there is a way to avoid this theory. The rational numbers, dense in the set of real numbers, are countable, that is, there exists an injective map from the set of rational numbers to the natural numbers. Hence, if we assume that the position values q are rational, then we can replace q by an integer $j(q)$, and can write the eigenvalue problem for $\hat{Q}$ in the form $\hat{Q}|j(q)\rangle = q|j(q)\rangle$. The same holds true for the momentum. An advantage is that we can work with vectors indexed by natural numbers and can avoid integrals.

So far states, superpositions, and probabilities, thus many basic ingredients of quantum mechanics, are not employed in the process of canonical quantization. Quantization works in a rather general way. For example, it is well-known that the quantized classical equation of motion becomes the Heisenberg equation describing motion in quantum mechanics.

If we have a finite number of separate physical systems with position observables $\hat{Q}_j$ and momentum observables $\hat{P}_j$ for $j = 1, ..., N$, then we use the commutation relations

$$\left[\hat{P}_j, \hat{Q}_l\right] = i\hbar\delta_{jl}. \tag{382}$$

for quantization.

More general, we can replace the index j by rational or real index, say x and t. Then we obtain the observables $\hat{Q}(x,t)$ and $\hat{P}(x,t)$. We use for fixed t the commutation relations

$$\left[\hat{Q}(x,t), \hat{P}(y,t)\right] = i\hbar\delta(x-y) \text{ for all } x, y. \tag{383}$$

Now we have a rough impression of quantization. In the following we describe the details for several concrete systems.

5.2 Classical Harmonic Oscillator

A fundamental model in physics, perhaps the most important one, is the harmonic oscillator. This model is not a special physical pattern like that of an electron or a fullerene molecule. Instead it is a mathematical model that applies to a large number of physical phenomena.

The harmonic oscillator describes particles moving under a restoring force which pulls a particle back to the equilibrium position. Think of a spring, or a ball moving inside a basin. Many physical systems can be described, at least approximately, by a quadratic energy function. When the system is disturbed out of the equilibrium it will oscillate. The atoms in crystals are an example of oscillations. A wave is a disturbance or oscillation that travels through matter or space, and transfers energy and momentum. Mechanical waves propagate, due to a disturbance, through a medium by a deformation of the corresponding substance. The deformation leads to certain restoring forces. For instance, sound waves propagate via air molecules that collide with their neighbors and oscillate. However, electromagnetic waves, consisting of periodic oscillations of electrical and magnetic fields generated by charged particles, are assumed to travel through a vacuum without any substance. Mathematically, it turns out that the energy function of the electromagnetic field can be written as the sum of infinitely many harmonic oscillators. The quantization of a field can be viewed as the quantization of its harmonic oscillators.

Let's work through the harmonic oscillator. The total energy of a non-damped, simple harmonic oscillators is described by the *Hamiltonian*

$$H(p,q) = \frac{p^2}{2m} + \frac{1}{2}\kappa_s q^2, \tag{384}$$

that is, the total energy is the sum of kinetic and potential energy. The variables q and p are the position and conjugate momenta coordinates, respectively, and κ_s denotes the spring constant.

The classical trajectory of motion in the phase space with coordinates (q_k, p_k) is described by two coupled differential equations of first order, called the *Hamilton equations*

$$\dot{p}_k = -\frac{\partial H}{\partial q_k}, \quad \dot{q}_k = \frac{\partial H}{\partial p_k}. \tag{385}$$

Hence, given any Hamilton function and the values of the position and momenta coordinates at t, the equations (385) give the coordinates an infinitesimal time later. The complete trajectory in the phase space is obtained by successively updating the coordinates. Remember that in our interpretation t is a geometrical parameter describing the oscillator's trajectory in explicit form.

For the Hamiltonian (384) we obtain from (385) the equations

$$\dot{q} = \frac{p}{m}, \quad \dot{p} = -\kappa_s q, \tag{386}$$

yielding the equation

$$m\ddot{q} + \kappa_s q = 0. \tag{387}$$

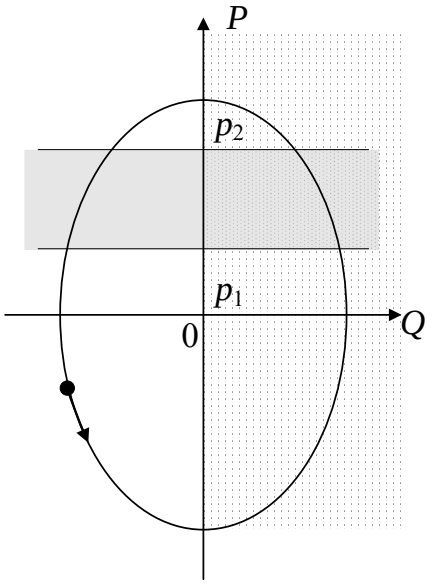


Figure 77: The ellipse describes a certain orbit of a harmonic oscillator in a two-dimensional phase space $Q \times P$, consisting of the Cartesian product of a one-dimensional position space Q and a one-dimensional momentum space P. The grey area corresponds to the momentum interval $p_1 \leq p \leq p_2$ in the phase space, and the pointed area is the position set $q \geq 0$.

Frequently, the spring constant is written in the form

$$\kappa_s = m\omega^2, \tag{388}$$

and the Hamiltonian becomes

$$H(q,p) = \frac{p^2}{2m} + \frac{m\omega^2}{2}q^2. \tag{389}$$

Then the general solution is given by the functions

$$q = A\cos(\omega t - \varphi) \quad \text{and} \quad p = m\dot{q} = -mA\omega\sin(\omega t - \varphi) \tag{390}$$

with *amplitude* A, *angular frequency* ω, *phase* φ, and *mass* m. In Figure 77 the motion of an harmonic oscillator is described in phase space. This geometric form is similar to Kepler's laws for planetary motion which are also described by functions, namely ellipses.

If we change the position variable $q' = (\kappa_s m)^{1/2}q$, then the Hamiltonian takes the form

$$H = \frac{1}{2m}(p^2 + (q')^2). \tag{391}$$

Hence, the Hamiltonian has a circle symmetry between the coordinates p and q. In the phase-space with q and p coordinates the harmonic oscillator circles around the origin. This can be described by the *symmetric group* of rotations. Notice the almost complete symmetry of Hamilton's equations (385).

The minus sign in the second equation is the only asymmetry in Hamilton's formulation.

For later purposes it is useful to write the Hamiltonian in product form. Let

$$\alpha = \sqrt{\frac{m\omega}{2\hbar}}q + i\frac{1}{\sqrt{2\hbar m\omega}}p, \tag{392}$$

$$\alpha^\dagger = \sqrt{\frac{m\omega}{2\hbar}}q - i\frac{1}{\sqrt{2\hbar m\omega}}p, \tag{393}$$

then we get the real quantity

$$\alpha^\dagger\alpha = \frac{m\omega}{2\hbar}q^2 + \frac{1}{2\hbar m\omega}p^2 - i\frac{1}{2\hbar}qp + i\frac{1}{2\hbar}pq. \tag{394}$$

Since the numbers qp commute, the Hamiltonian can be written as a product of complex coefficients:

$$H = \hbar\omega\alpha^\dagger\alpha. \tag{395}$$

It is useful to view the solutions (390) in the complex plane with real position q-axis and imaginary momentum p-axis. From (390) and (393) it follows that the complex amplitude evolves as

$$\alpha^\dagger(t) = \sqrt{\frac{m\omega}{2\hbar}}A\cos(\omega t-\varphi) - i\frac{1}{\sqrt{2\hbar m\omega}}mA\omega\sin(\omega t-\varphi) = (A\sqrt{\frac{m\omega}{2\hbar}}e^{i\varphi})\ e^{-i\omega t}. \tag{396}$$

In this plane the solutions can be visualized as a clockhand rotating clockwise with frequency ω.

5.3 Quantum Harmonic Oscillator

It is natural to view the oscillator as a machine described by its mutually exclusive alternatives, the base states, and corresponding numbers that describe the positions and momenta. This is the concept of an observable. Now, what are the base states? We apply *canonical quantization*.

We insert the operators $\hat{P}$ and $\hat{Q}$ into the classical Hamiltonian (389) and look what happens. For the harmonic oscillator we obtain the so-called *Hamilton operator*

$$\hat{H} = \frac{1}{2m}\hat{P}^2 + \frac{1}{2}m\omega^2\hat{Q}^2. \tag{397}$$

This operator is Hermitian. Because formula (397) contains the necessary information about all positions and momenta, it should contain all energy eigenvalues and energy base states of the oscillator.

Quantization is a purely algebraic way solely based on the *commutation relation* for the position and momentum operator discussed in Section 4.18:

$$[\hat{P}, \hat{Q}] = \hat{P}\hat{Q} - \hat{Q}\hat{P} = -i\hbar \tag{398}$$

No further information is required to obtain all energy eigenvalues and eigenstates of the Hamilton operator.

We try to rewrite (397) as a product of two operators. Finding the eigenvalues of a product of operators is much easier than of their sum. For numbers we have factorized a sum in (394), (395) using complex numbers. But in contrast to complex numbers, operators don't commute, and we have to take account of the commutator (398). Using our rule of canonical quantization, according to (393) we define the two operators

$$\hat{A} = \sqrt{\frac{m\omega}{2\hbar}}\hat{Q} + \frac{1}{\sqrt{2\hbar m\omega}}i\hat{P}, \tag{399}$$

$$\hat{A}^\dagger = \sqrt{\frac{m\omega}{2\hbar}}\hat{Q} - \frac{1}{\sqrt{2\hbar m\omega}}i\hat{P}. \tag{400}$$

The operator $\hat{A}$ is called the *annihilation operator*, and $\hat{A}^\dagger$ is called the *creation operator*. Their names will become clear below. Reversely, we obtain

$$\hat{Q} = \sqrt{\frac{\hbar}{2m\omega}}(\hat{A}^\dagger + \hat{A}), \quad \hat{P} = i\sqrt{\frac{m\hbar\omega}{2}}(\hat{A}^\dagger - \hat{A}), \tag{401}$$

Some simple calculations using the commutation relations (398) yield

$$[\hat{A}, \hat{A}^\dagger] = \hat{1}, \quad [\hat{A}, \hat{A}] = 0, \quad [\hat{A}^\dagger, \hat{A}^\dagger] = 0. \tag{402}$$

Multiplying out we get

$$\hat{H} = \frac{1}{2}\hbar\omega(\hat{A}\hat{A}^\dagger + \hat{A}^\dagger\hat{A}) = \hbar\omega\hat{N} + \frac{1}{2}\hbar\omega\hat{1}, \tag{403}$$

where the operator $\hat{N} = \hat{A}^\dagger \hat{A}$ is called *number operator*. This factorization is only possible when using complex numbers. Real numbers are not sufficient. This is a further argument for the need of complex numbers.

Obviously, the number operator satisfies the commutation relations

$$[\hat{N}, \hat{A}^\dagger] = \hat{A}^\dagger, \quad [\hat{N}, \hat{A}] = -\hat{A}. \tag{404}$$

Moreover, the commutation relations

$$[\hat{H}, \hat{A}^\dagger] = \hbar\omega \hat{A}^\dagger, \quad [\hat{H}, \hat{A}] = -\hbar\omega \hat{A}, \quad [\hat{H}, \hat{N}] = 0. \tag{405}$$

follow immediately. The equations (402) and (405) show that the set of operators $\hat{1}$, $\hat{A}$, $\hat{A}^\dagger$ and $\hat{H}$ is closed under the commutation operation, and such sets are called *commutator algebras*. Now we show the power of the commutator algebra: it solves the eigenvalue problem of the Hamilton operator (397). This operator is Hermitian, and thus has an orthonormal basis of eigenvectors $|n\rangle$ with observable energy eigenvalues E_n. Since the Hamiltonian $\hat{H}$ and the number operator $\hat{N}$ commute, they have the joint set of eigenvectors $\{|n\rangle\}$.

Firstly, let $E_n, |n\rangle$ be an arbitrary eigenpair, that is,

$$\hat{H}|n\rangle = E_n|n\rangle. \tag{406}$$

Then using the previous commutation relations, we obtain

$$\hat{H}\hat{A}^\dagger|n\rangle = (\hat{A}^\dagger\hat{H} + \hbar\omega\hat{A}^\dagger)|n\rangle = (E_n + \hbar\omega)\hat{A}^\dagger|n\rangle, \tag{407}$$

$$\hat{H}\hat{A}|n\rangle = (\hat{A}\hat{H} - \hbar\omega\hat{A})|n\rangle = (E_n - \hbar\omega)\hat{A}|n\rangle. \tag{408}$$

Therefore, we see that for each eigenvector $|n\rangle$ the vectors $\hat{A}^\dagger|n\rangle$ and $\hat{A}|n\rangle$ are (unnormalized) eigenvectors of $\hat{H}$ and $\hat{N}$ with energy eigenvalues $E_n + \hbar\omega$ and $E_n - \hbar\omega$, respectively.

Secondly, if we repeat this process, we obtain a sequence of eigenpairs with eigenvalues

$$\ldots, E_n - 3\hbar\omega, E_n - 2\hbar\omega, E_n - \hbar\omega, E_n, E_n + \hbar\omega, E_n + 2\hbar\omega, E_n + 3\hbar\omega, \ldots \tag{409}$$

and unnormalized eigenvectors. Since the Hamiltonian is the sum of two squared Hermitian operators, all eigenvalues E_n are nonnegative. Because of (403), the previous process stops, if the energy value $E_0 = \hbar\omega/2$ occurs. Therefore, in contrast to the classical harmonic oscillator, we obtain a lowest positive energy.

Thirdly, since the lowest energy value satisfies $E_0 = \hbar\omega/2$, the expression in (409) vanishes if and only if $E_n = E_0$. The value E_0 is the ground state energy corresponding to the *ground state* $|0\rangle$. For the ground state the equation

$$\hat{A}|0\rangle = 0. \tag{410}$$

must be satisfied, since otherwise (408) would imply negative eigenvalues of the Hamiltonian. All eigenvalues of the Hamiltonian operator are

$$E_n = \hbar\omega(n + \frac{1}{2}), \quad n = 0, 1, 2, \ldots \tag{411}$$

If there would be another eigenvalue, we apply the previous steps of creation and annihilation to this eigenpair, yielding an eigenvector with eigenvalue smaller than E_0, what is impossible. Summarizing, we have obtained with the commutation relations the complete energy spectrum of the harmonic oscillator, which is discrete with a positive lowest energy. This seems to be a weird property from the point of view of classical mechanics, where the energy spectrum is continuous with lowest value zero.

Finally, using (403) and (411), it follows that

$$\hat{N}|n\rangle = \hat{A}^\dagger A|n\rangle = \left(\frac{1}{\hbar\omega}\hat{H} - \frac{1}{2}\right)|n\rangle = \left(\frac{1}{\hbar\omega}E_n - \frac{1}{2}\right)|n\rangle, \tag{412}$$

and hence

$$\hat{N}|n\rangle = n|n\rangle. \tag{413}$$

This is the reason why the operator $\hat{N}$ is called *number operator*, see Figure 78.

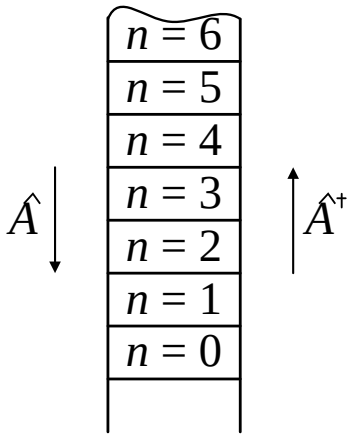


Figure 78: The energy levels of the harmonic oscillator are discrete and can be visualized as a ladder. The operators $A^\dagger$ and $\hat{A}$ raise and lower the energy level, respectively, and thus explains their names. The number operator $\hat{N}$ has the lower bound zero corresponding to the ground state, but no upper bound.

Since $\hat{A}|n\rangle$ is an eigenvector of the number operator and

$$\hat{N}\hat{A}|n\rangle = \hat{A}\hat{N}|n\rangle + [\hat{N}, \hat{A}]|n\rangle = \hat{A}n|n\rangle - \hat{A}|n\rangle = (n-1)\hat{A}|n\rangle, \tag{414}$$

it follows that $\hat{A}|n\rangle$ is an eigenvector of $\hat{N}$ corresponding to the eigenvalue $n-1$. Thus it must be (up to a normalization constant) the eigenvector $|n-1\rangle$. We set

$$\hat{A}|n\rangle = \sqrt{n}|n-1\rangle. \tag{415}$$

Analogously, we obtain

$$\hat{A}^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle. \tag{416}$$

With the square root factors the equation (413) is fulfilled In particular, we can express all base states by acting repeatedly on the ground state $|0\rangle$ with the creation operator:

$$|1\rangle = \hat{A}^\dagger|0\rangle,\ |2\rangle = \frac{1}{\sqrt{2}}\hat{A}^\dagger|1\rangle = \frac{1}{\sqrt{2}}(\hat{A}^\dagger)^2|0\rangle, ... \tag{417}$$

and more general

$$|n\rangle = \frac{1}{\sqrt{n!}}(\hat{A}^\dagger)^n|0\rangle. \tag{418}$$

We have found all eigenvalues and all orthonormal eigenvectors of the Hamilton operator. The eigenvectors (418) form a set of distinguishable base states. **The approach to work with non-commutative operators instead of commutative numbers has lead us to** *quantization***: the spectrum of the Hamiltonian is infinite, but discrete. To obtain this result we have used the operators $\hat{H}, \hat{N}, \hat{A}, \hat{A}^\dagger$ that act on base states together with their commutation relations. Unitary transformations acting on superposed states are not involved.**

We can represent these operators as matrices related to the basis of orthonormal energy eigenvectors $\{|n\rangle\}$. Since an operator $\hat{M}$ can be described by a matrix with coefficients

$$\hat{M}_{n'n} = \langle n'|\hat{M}|n\rangle, \tag{419}$$

it follows that the Hamilton operator is represented by an infinite diagonal matrix with the energy eigenvalues (411) on the diagonal.

Formula (416) implies immediately that the creation and annihilation operators have the non-Hermitian matrix representations

$$\hat{A}^\dagger_{n'n} = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots \\ \sqrt{1} & 0 & 0 & 0 & \dots \\ 0 & \sqrt{2} & 0 & 0 & \dots \\ 0 & 0 & \sqrt{3} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \cdots \end{pmatrix}, \quad \hat{A}_{n'n} = \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & \dots \\ 0 & 0 & \sqrt{2} & 0 & \dots \\ 0 & 0 & 0 & \sqrt{3} & \dots \\ 0 & 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \cdots \end{pmatrix} \tag{420}$$

All these matrices are infinite dimensional. However, since energy is never infinite, we can view these as large finite-dimensional matrices.

Since the harmonic oscillator is perhaps the most fundamental physical model, it is interesting for understanding quantum mechanics to investigate the Hermitian matrix representations of the most important observables. The number operator (413) has the representation

$$\hat{N} = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 2 & 0 & \dots \\ 0 & 0 & 0 & 3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \cdots \end{pmatrix} \tag{421}$$

From (401) we obtain the matrix representations of the position and momentum operator

$$\hat{Q} = \sqrt{\frac{\hbar}{2m\omega}} \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & \dots \\ \sqrt{1} & 0 & \sqrt{2} & 0 & \dots \\ 0 & \sqrt{2} & 0 & \sqrt{3} & \dots \\ 0 & 0 & \sqrt{3} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \cdots \end{pmatrix}, \quad \hat{P} = i\sqrt{\frac{m\hbar\omega}{2}} \begin{pmatrix} 0 & -\sqrt{1} & 0 & 0 & \dots \\ \sqrt{1} & 0 & -\sqrt{2} & 0 & \dots \\ 0 & \sqrt{2} & 0 & -\sqrt{3} & \dots \\ 0 & 0 & \sqrt{3} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \cdots \end{pmatrix}. \tag{422}$$

These matrices should satisfy the commutation relations (398), and thus the position-momentum uncertainty principle. In fact, some simple computations show

$$\hat{Q}\hat{P} = i\frac{\hbar}{2} \begin{pmatrix} 1 & 0 & -\sqrt{2} & \dots \\ 0 & 1 & 0 & \dots \\ \sqrt{2} & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \cdots \end{pmatrix}, \quad \hat{P}\hat{Q} = i\frac{\hbar}{2} \begin{pmatrix} -1 & 0 & -\sqrt{2} & \dots \\ 0 & -1 & 0 & \dots \\ \sqrt{2} & 0 & -1 & \dots \\ \vdots & \vdots & \vdots & \cdots \end{pmatrix}, \tag{423}$$

implying the desired result $\hat{P}\hat{Q} - \hat{Q}\hat{P} = -i\hbar$.

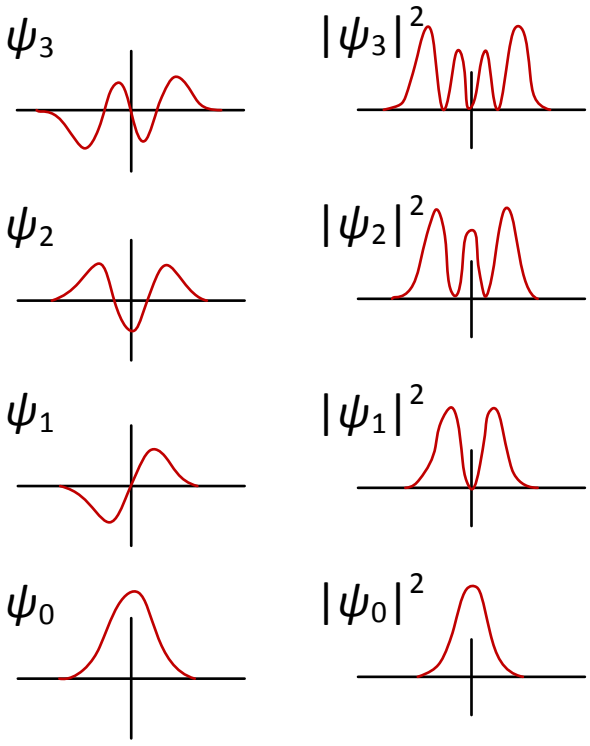


Figure 79: The first four wave functions ψ_i with corresponding probability densities $|\psi_i^2|$.

So far we have described algebraic consequences of the commutation relations and the quadratic form of the Hamilton operator. In particular, we

have found all important algebraic facts about this operator. It's eigenvalues are quantized. What has all this to do with quantum mechanics, known as a stochastic process? Nothing. Until now this theory is deterministic since only distinguishable base states are used. Hence, it belongs more to classical deterministic mechanics, not to the stochastic part of quantum theory. Quantization is the process of finding base states, the fundamental building blocks of the future, present, and past.

The Hamilton operator is defined via the position and momentum operator. Hence, when inserting both operators in their position representation, it should be possible to obtain the eigenfunctions $|\psi_n(x)\rangle = |n\rangle$ corresponding to the energy eigenvalues. Here, we use Cartesian coordinates x as usual. Then for the lowest energy level the equation (410) in position representation has the form

$$\left(\sqrt{\frac{m\omega}{2\hbar}}x + i\sqrt{\frac{1}{2m\hbar\omega}}\frac{\hbar}{i}\frac{\partial}{\partial x}\right)\psi_0(x) = 0. \tag{424}$$

It is easy to see that

$$\psi_0(x) = \sqrt[4]{\frac{m\omega}{\pi\hbar}}e^{-\frac{m\omega}{2\hbar}x^2} \tag{425}$$

is a solution. Moreover it is a normalized solution, because the integral over space $\int_{-\infty}^{\infty}\psi_0^*(x)\psi_0(x)dx = 1$. With (418) all other eigenfunctions can be obtained:

$$\psi_n(x) = \frac{1}{\sqrt{n!}}\left(\sqrt{\frac{m\omega}{2\hbar}} + i\sqrt{\frac{1}{2m\hbar\omega}}\frac{\hbar}{i}\frac{\partial}{\partial x}\right)^n\psi_0(x). \tag{426}$$

Some computations that can be found in most textbooks about quantum mechanics yield the normalized wave functions

$$\psi_n(x) = \sqrt[4]{\frac{m\omega}{\hbar\pi}}\frac{1}{\sqrt{2^n n!}}e^{-\frac{m\omega}{\hbar}\frac{x^2}{2}}H_n\left(\sqrt{\frac{m\omega}{\hbar}}x\right) \tag{427}$$

where H_n are the Hermitian polynomials, see Figure 79 for the first four wave functions. The corresponding formulas are:

$$\begin{aligned}
\psi_0 &= \left(\frac{a}{\pi}\right)^{\frac{1}{4}}e^{-y^2/2},\\
\psi_1 &= \left(\frac{a}{\pi}\right)^{\frac{1}{4}}\sqrt{2}ye^{-y^2/2},\\
\psi_2 &= \left(\frac{a}{\pi}\right)^{\frac{1}{4}}\frac{1}{\sqrt{2}}(2y^2-1)e^{-y^2/2},\\
\psi_3 &= \left(\frac{a}{\pi}\right)^{\frac{1}{4}}\frac{1}{\sqrt{3}}(2y^3-3y)e^{-y^2/2},
\end{aligned} \tag{428}$$

where $a = \dfrac{m\omega}{\hbar}$ and $y = \sqrt{a}x$.

One can prove that these normalized functions are orthogonal, that is,

$$\int_{-\infty}^{\infty} \psi_n^*(x)\psi_m(x)dx = 0 \quad \text{for} \quad n \neq m. \tag{429}$$

Thus they form a set of distinguishable base states.

In summary, we have shown that a purely algebraic method, solely based on base states together with observables acting on base states, lead to the quantization of the energy values. Moreover, for the position representation we have obtained the corresponding orthogonal eigenfunctions $\psi_n(x)$. These are purely geometric objects and represent mutually exclusive alternatives. It is most helpful to keep in mind that these function do not depend on time, thus they don't require the (3+1)-spacetime.

But what can be the meaning of these timeless functions? What have these functions to do with an oscillating object? Where does quantum probability come into play? These functions are related to the position x of the oscillator. It is a good choice to interpret these functions probabilistic in terms of probability amplitudes defined on the position base states. The latter are the mutually exclusive alternatives where the oscillator might be. As usual we apply *Born's rule* for obtaining the probabilities. This rule says that the probability of finding the oscillator at any given position x is the square of the magnitude of the function $|\psi_n(x)|^2$. From formula (427) it follows that each eigenfunction is an exponential function in x multiplied by a polynomial. Hence, as $x \to \pm\infty$ all eigenfunctions go to zero, and thus the probability that the mass moves far away from the equilibrium point goes to zero for each energy value. It can be shown that each eigenfunction has one more real zero, if the energy is increased by one unit $\hbar\omega$. Thus for increasing energies the eigenfunctions oscillate more rapidly. Moreover, the eigenfunctions spread out for increasing energies.

If we think in terms of an oscillating ball, we can conclude that the zeros of the eigenfunctions are points where the oscillating ball will never be found. For the lower energy states the most probable values of position are different from the classical harmonic oscillator. Classically, the oscillator spends more time close at the turning points of its motion, since there its velocity is almost zero. This looks rather strange. But for increasing energies, the chance of being far away from the equilibrium point increases. Then the probability density becomes concentrated at the turning points of the classical oscillator, where the particle spends most of its time, see Figure 79. In this context time means: when performing many experiments in most cases the oscillating particle is found close to the turning points. In contrast to classical mechanics, it follows that there is a small positive probability to find the ball outside the basin. This effect is called *quantum tunneling*, an important physical phenomenon with many applications, such as the tunnel diodes, scanning tunneling microscope or quantum computing.

Thus, we have obtained a completely different picture of the quantum harmonic oscillator compared with the classical case. What is the reason for that? If the position and the momentum operator would be diagonal matrices, their commutator would be zero, and nothing interesting would happen. Hence, the

commutation relations are the key to quantization. But what is the mathematical reason for the commutation relations? Obviously, this is caused by the fact that the underlying complex linear space allows different bases such that the operators cannot be simultaneously diagonalized.

Keep in mind: Replacing in the classical Hamiltonian the classical variables position and momentum by their corresponding Hermitian observables leads to the energy Hamilton operator. A purely algebraic method, based only on the commutation relations, shows that the energy values are discrete, and thus justifies the name *quantization*. Hence, whenever you see operators, figure out their commutators. The related base state functions do not depend on an external time parameter t, but have a probabilistic interpretation.

5.4 Matrix Mechanics in Position Representation

In Section 5.3 we have considered the matrix theory of the harmonic oscillator in the energy representation. Moreover, we have presented position, momentum, energy, annihilation, and creation operators as matrices with respect to the orthonormal energy eigenstates. It is a good exercise to investigate these operators with respect to other basis representations. In this section we consider the *position representation*. This provides a deeper understanding of matrix mechanics.

The classical dynamical variable position is the most natural quantity in physics. Although this variable is continuous in classical mechanics, we consider a discrete one-dimensional position space, which may be defined via a discrete coarse graining such that

$$q = n \cdot \varepsilon, \quad -N \leq n \leq N, \tag{430}$$

with grid spacing $\varepsilon > 0$, and integer N which may also be infinite. The generalization of the following investigation to higher dimensional infinite or continuous position spaces is straightforward. Let the set of vectors

$$|\psi\rangle := \sum_{n=-N}^{N} \psi_n |n\rangle, \quad \psi_n \in \mathbb{C} \tag{431}$$

denote the position state space, where $|n\rangle$ is the normalized basis vector corresponding to the position $q = n\varepsilon$. Then the coefficients $\psi(n) = \psi_n$ are the amplitudes generating the corresponding wave function $|\psi\rangle$. The squared magnitude of the amplitude is the probability of finding the particle in state $|n\rangle$. The basis states $|n\rangle$ are the orthonormal eigenvectors of the position operator $\hat{Q}$, that is

$$\hat{Q}|n\rangle = n\varepsilon|n\rangle, \quad \langle n|n'\rangle = \delta_{nn'}, \tag{432}$$

and $-N \leq n, n' \leq N$. Therefore, the matrix position representation of $\hat{Q}$ takes the form

$$\hat{Q}_{n'n} = \langle n'|\hat{Q}|n\rangle = n\varepsilon\delta_{n'n}, \tag{433}$$

and thus yields the diagonal matrix

$$\hat{Q} = \varepsilon \begin{pmatrix} -N & & & & & & & \\ & \ddots & & & & & & \\ & & -2 & & & & 0 & \\ & & & -1 & & & & \\ & & & & 0 & & & \\ & & & & & 1 & & \\ & & 0 & & & & 2 & \\ & & & & & & & \ddots \\ & & & & & & & & N \end{pmatrix}. \tag{434}$$

Its expectation value is

$$\langle \hat{Q} \rangle_\psi = \langle \psi|\hat{Q}|\psi\rangle = \sum_{n=-N}^{N} n\varepsilon|\psi_n|^2. \tag{435}$$

The momentum operator $\hat{P} = -i\hbar\frac{\partial}{\partial q}$ in position presentation can be obtained via discretization by using centered differences, that is, we replace the differential operator $\frac{\partial}{\partial q}$ by the difference operator $\frac{\Delta}{\varepsilon}$ such that

$$\frac{\Delta}{\varepsilon}\psi(n) := \frac{1}{2}(\psi(n+1) - \psi(n-1)). \tag{436}$$

Then

$$\hat{P}\psi(n) = -i\hbar\frac{\Delta}{\varepsilon}\psi(n) \tag{437}$$

is an approximation of $\hat{P}\psi(q) = -i\hbar\frac{\partial\psi(q)}{\partial q}$, where $q = n\varepsilon$. Assuming periodic boundary conditions $\psi(-N) = \psi(N)$ such that the states $|n\rangle$ may be viewed as discrete points on a circle, we obtain the momentum matrix

$$\hat{P} = -\frac{i\hbar}{2\varepsilon}\begin{pmatrix} 0 & 1 & 0 & 0 & \dots & -1 \\ -1 & 0 & 1 & 0 & \dots & 0 \\ 0 & -1 & 0 & 1 & \dots & 0 \\ 0 & 0 & -1 & 0 & \dots & 0 \\ \vdots & & & & & \\ 1 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}. \tag{438}$$

As expected, the discretization yields a matrix with 1's above and -1's below the diagonal. Observe that both matrices $\hat{Q}$ and $\hat{P}$ are Hermitian matrices in the energy representation (422) as well as in the position representation (434) and (438).

A simple computation yields the commutator

$$[\hat{Q}, \hat{P}] = \frac{i\hbar}{2}\begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & 1 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}. \tag{439}$$

Thus the matrix has 1's above and below the diagonal. As the discretization becomes finer, the nonzero entries seem to be more and more concentrated at the diagonal. But it is not an identity matrix as we expect from the commutation relation and (423). But the equation

$$[\hat{Q}, \hat{P}]\begin{pmatrix} \vdots \\ \psi(n) \\ \vdots \end{pmatrix} = i\hbar\begin{pmatrix} \vdots \\ \frac{1}{2}(\psi(n-1) + \psi(n+1)) \\ \vdots \end{pmatrix} \approx i\hbar\begin{pmatrix} \vdots \\ \psi(n) \\ \vdots \end{pmatrix} \tag{440}$$

shows that for a fine discretization this matrix behaves like an identity matrix, provided ψ is sufficiently smooth such that the arithmetic mean $\frac{1}{2}(\psi(n-1) + \psi(n+1))$ is almost equal to $\psi(n)$.

The eigenstates of the position operator in the position representation are the positions $q = n\varepsilon$ of our coarse graining. What are the eigenstates $|\psi_p\rangle$ of

the momentum operator in the position representation? Using the equation (438), the eigenvalue equation

$$\hat{P}|\psi_p\rangle = p|\psi_p\rangle \tag{441}$$

can be written equivalently in the the componentwise form

$$-\frac{i\hbar}{2\varepsilon}(\psi_p(n+1) - \psi_p(n-1)) = p\psi_p(n). \tag{442}$$

Dividing by $i\hbar$ yields the difference equation

$$\frac{\psi_p(n+1) - \psi_p(n-1)}{2\varepsilon} = \frac{i}{\hbar}p\psi_p(n), \tag{443}$$

which is a discretization of the differential equation

$$\frac{\partial\psi_p(q)}{\partial q} = \frac{i}{\hbar}p\psi_p(q). \tag{444}$$

The general solution of this equation is given by

$$\psi_p(q) = c_p p e^{\frac{ipq}{\hbar}}, \tag{445}$$

where c_p is constant.

Summarizing, when taking the position representation, a discrete coarse graining, and centered differences, the quantum observables are simple matrices and the differential equations are approximated by difference equation.

5.5 Several Harmonic Oscillators

In this section we consider the canonical quantization of N uncoupled harmonic oscillators as a preparation for more complicated problems. The oscillators may be positioned at arbitrarily different places without any contact. Thus, their Hamiltonian takes the form

$$\hat{H} = \sum_{k=1}^{N} \hat{H}_k, \quad \hat{H}_k = \frac{\hat{P}_k^2}{2m_k} + \frac{1}{2} m_k \omega_k^2 \hat{Q}_k^2. \tag{446}$$

From Section 5.3 it follows immediately that the kth oscillator can be described by the creation operator $\hat{A}_k^\dagger$ and the annihilation operator $\hat{A}_k$:

$$\begin{aligned} \hat{A}_k^\dagger |n_1, n_2, \ldots, n_k, \ldots\rangle &= \sqrt{n_k + 1} \quad |n_1, n_2, \ldots, n_k + 1, \ldots\rangle, \\ \hat{A}_k |n_1, n_2, \ldots, n_k, \ldots\rangle &= \sqrt{n_k - 1} \quad |n_1, n_2, \ldots, n_k - 1, \ldots\rangle. \end{aligned} \tag{447}$$

Moreover, since the oscillators are separated, these operators satisfy the commutation relations

$$[\hat{A}_k, \hat{A}_{k'}] = 0, \; [\hat{A}_k^\dagger, \hat{A}_{k'}^\dagger] = 0, \; [\hat{A}_k, \hat{A}_{k'}^\dagger] = \delta_{kk'}. \tag{448}$$

Because of (446) and (403), the Hamiltonian in terms of these operators is

$$\hat{H} = \sum_{k=1}^{N} \hbar \omega_k \left(\hat{A}_k^\dagger \hat{A}_k + \frac{1}{2} \right). \tag{449}$$

The vacuum state $|0\rangle = |0, 0, 0, ...\rangle$ is the state where each oscillator is in its ground state such that

$$\hat{A}_k |0\rangle = 0 \tag{450}$$

for all k. From formulas (418) and (447) we obtain the so-called *occupation number representation* of these oscillators

$$|n_1, n_2, \ldots, n_N\rangle = \frac{1}{\sqrt{n_1! n_2! \cdots n_N!}} (\hat{A}_1^\dagger)^{n_1} (\hat{A}_2^\dagger)^{n_2} \cdots (\hat{A}_N^\dagger)^{n_N} |0, 0, \ldots, 0\rangle. \tag{451}$$

The kth oscillator has n_k quanta of energy.

This simple generalization to several independent oscillators shall serve us when considering more complicated problems. The results above can be immediately applied to a three-dimensional harmonic oscillator. It's Hamiltonian is just the sum of three one-dimensional Hamiltonians.

5.6 The Oscillator Chain

We consider a one-dimensional circular chain of radius R consisting of N atoms with identical masses m. The atoms are connected by springs with spring constant κ_s, as shown in Figure 80. For large N the distance between two atoms in their equilibrium positions is $a = 2\pi R/N$, approximately. The displacement of the jth atom from the equilibrium position x_j is denoted by q_j.

The classical Hamiltonian of this chain is the sum of the kinetic and the potential energy:

$$H(p,q) = \sum_{j=1}^{N} \left[\frac{1}{2m} p_j^2 + \frac{\kappa_s}{2} (q_{j+1} - q_j)^2 \right]. \tag{452}$$

The corresponding *Hamilton equations* are:

$$\begin{aligned} \dot{p}_j &= -\frac{\partial H}{\partial q_j} = \kappa_s (q_{j-1} - 2q_j + q_{j+1}), \\ \dot{q}_j &= \frac{\partial H}{\partial p_j} = \frac{p_j}{m}. \end{aligned} \tag{453}$$

Since $\dot{p}_j = m\ddot{q}_j$ we obtain for the jth atom the equation of motion:

$$m\ddot{q}_j - \kappa_s (q_{j-1} - 2q_j + q_{j+1}) = 0, \quad j = 1, \ldots, N. \tag{454}$$

We consider a closed *oscillator chain* with periodic boundary conditions

$$q_{j+N}(t) = q_j(t), \tag{455}$$

see Figure 80.

In contrast to Section 5.5, we are now dealing with a problem where each mass is strongly coupled to its neighbor by the springs. Thus, the atoms cannot be considered independently. Surprisingly, it will turn out that the quantized system behaves as a system of independent, uncoupled harmonic oscillators.

We solve equation (454) with the Ansatz

$$q_j(t) = e^{i(jp - \omega t)}, \quad j = 1, \ldots, N. \tag{456}$$

Substituted into the equations of motion, we obtain

$$\left[-m\omega^2 - \kappa_s (e^{-ip} - 2 + e^{ip}) \right] e^{i(jp - \omega t)} = 0. \tag{457}$$

The boundary condition (455) requires for $j = 0$

$$e^{i(Np - \omega t)} = e^{-i\omega t}. \tag{458}$$

This can be satisfied if and only if p takes the discrete values

$$p_k = \frac{2\pi k}{N}, \quad k = 0, \pm 1, \ldots, \pm \frac{N}{2}, \tag{459}$$

where we have assumed that N is even. Since

$$e^{-ip} - 2 + e^{ip} = -4 \sin^2 \frac{p}{2}, \tag{460}$$

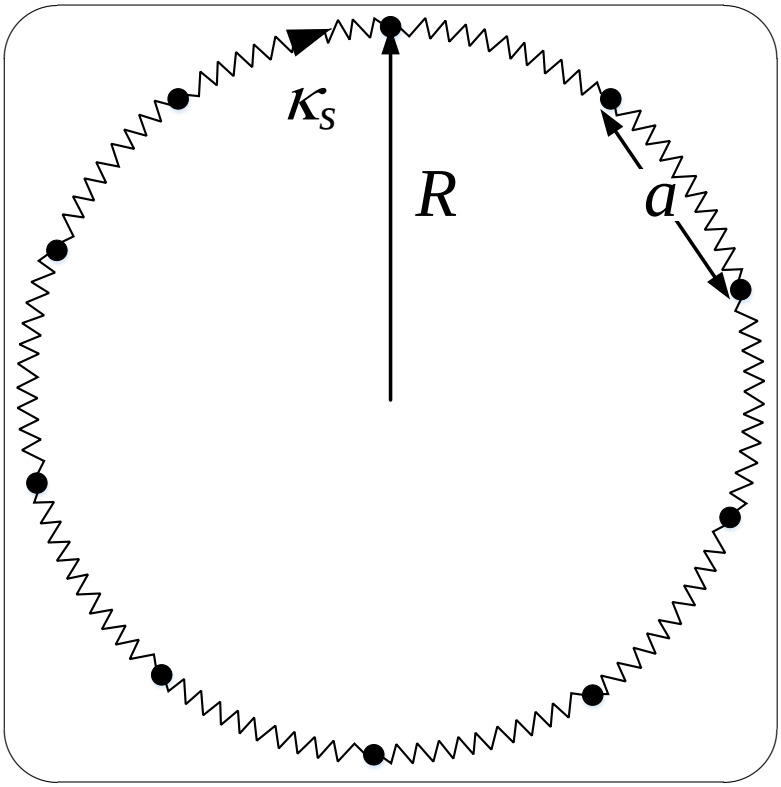


Figure 80: A closed oscillator chain in rest position with spring constant κ_s and total lenght $L = 2\pi R = Na$.

equation (457) implies

$$-m\omega^2 + \kappa_s 4 \cdot \sin^2 \frac{p}{2} = 0. \tag{461}$$

If we set $\kappa_s = m\omega_0^2/4$, then with (459) we obtain the dispersion relation

$$\omega_k^2 = \omega_0^2 \sin^2\left(\frac{\pi k}{N}\right), \quad k = \pm 1, \ldots, \pm\frac{N}{2}. \tag{462}$$

We have excluded $k = 0$ since this case corresponds to a trivial constant displacement. At small momenta p_k, where k is small compared to N, the linear *dispersion law*

$$\omega_k = 2\sqrt{\frac{\kappa_s}{m}}|k| \tag{463}$$

is fulfilled.

In summary, we have obtained for the jth atom the solution

$$q_j(t) = e^{i\left(\frac{2\pi kj}{N} - \omega_k t\right)}, \quad j = 1, \ldots, N, \tag{464}$$

where the frequency ω_k satisfies the dispersion relation (462). The solutions q_j describe an elementary vibrational motion in which a chain of N atoms uniformly oscillates at a fixed common frequency ω_k, see Figure 81. There are

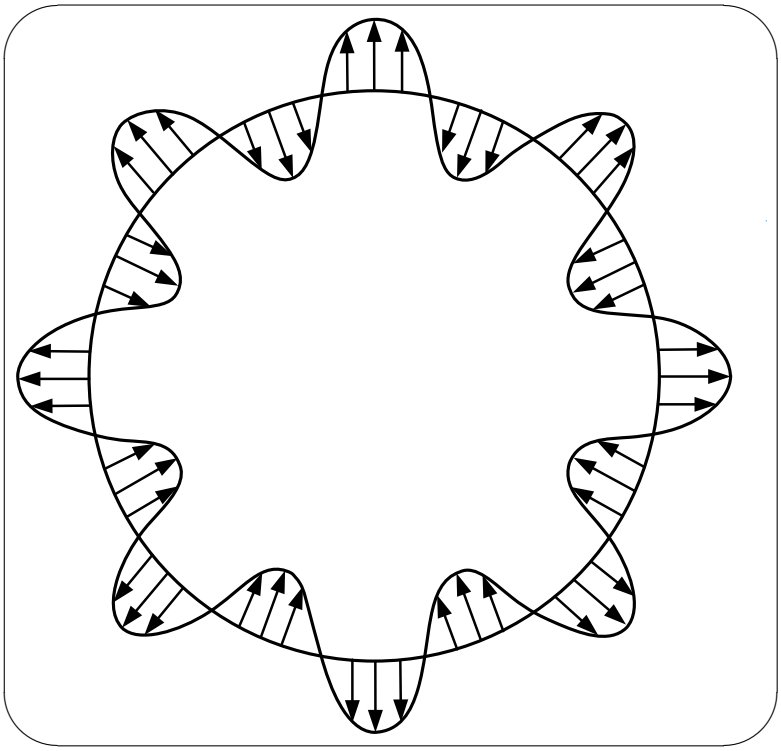


Figure 81: A phonon.

N vibrational motions corresponding to the frequencies ω_k. Each such motion is called a *normal mode*. The modes can be viewed as waves or excitations running along the chain. A *phonon* is its quantum mechanical description, see the next section. Phonons play a major role, for instance in condensed matter physics. They are collective excitations in elastic arrangements of particles.

The general real solution for the jth atom, satisfying the equation of motion (454), can be written as the superposition of normal modes

$$q_j(t) = \frac{1}{\sqrt{N}} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}} c_k e^{i\left(\frac{2\pi kj}{N} - \omega_k t\right)} + c_k^* e^{-i\left(\frac{2\pi kj}{N} - \omega_k t\right)}, \quad j = 1, \ldots, N. \tag{465}$$

This is a Fourier expansion. Combining modes in this way leads to traveling waves, see Figure 82. In other words, motion can be analyzed by assuming that it is the sum of different modes. A nice and very detailed description of this phenomenon is presented in the Feynman Lectures[73].

[73] http://www.feynmanlectures.caltech.edu/I_49.html

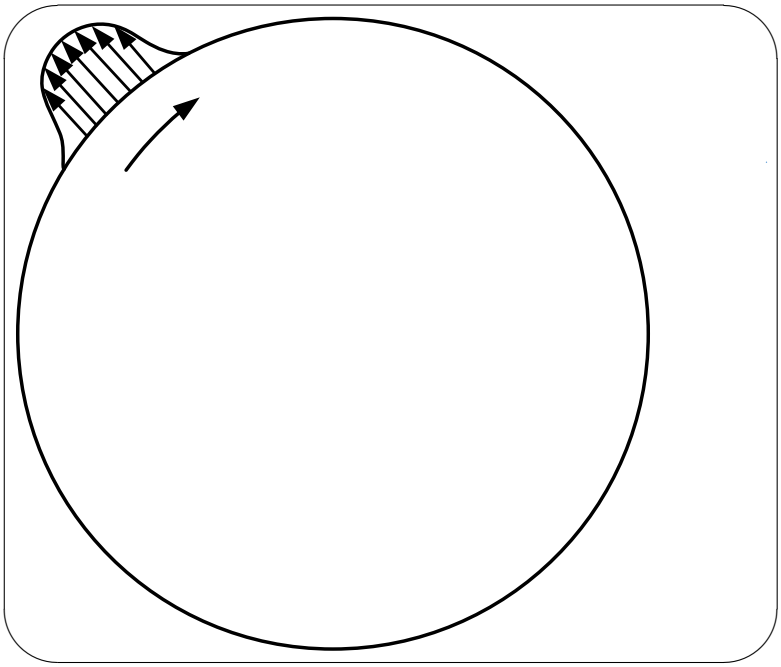


Figure 82: Motion as a superposition of normal modes.

5.7 Quantization of the Chain

We quantize the chain in the same manner as for the harmonic oscillator. We replace in (452) positions q_j and momentum p_j by the observables $\hat{Q}_j$ and $\hat{P}_j$, respectively. Obviously, these Hermitian operators satisfy the commutation relations

$$\left[\hat{P}_j, \hat{Q}_l\right] = i\hbar\delta_{jl}. \tag{466}$$

Not surprisingly, several computations[74] show that the Hamiltonian becomes

$$\hat{H} = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}} \hbar\omega_k \left(\hat{A}_k^\dagger \hat{A}_k + \frac{1}{2}\right) \tag{467}$$

where $\hat{A}_k$ and $\hat{A}_k^\dagger$ are the creation and annihilation operators, respectively. They satisfy the commutation relations

$$\left[\hat{A}_k, \hat{A}_{k'}\right] = \left[\hat{A}_k^\dagger, \hat{A}_{k'}^\dagger\right] = 0, \quad \left[\hat{A}_k, \hat{A}_{k'}^\dagger\right] = \delta_{kk'}. \tag{468}$$

It follows that the closed harmonic chain is expressed as a sum of separate modes k that behave like single harmonic oscillators.

The state $|0\rangle = |0, 0, 0, ...\rangle$ denotes the ground state where

$$\hat{A}_k|0\rangle = 0 \text{ for all } k = -\frac{N}{2}, \ldots, +\frac{N}{2}. \tag{469}$$

[74]For a proof see for example: Lancaster, Tom, and Stephen J. Blundell. Quantum field theory for the gifted amateur. OUP Oxford, 2014. Chapter 2.4. We proof this later for the more difficult continuous oscillator chain.

A general base state

$$|n_1, n_2, \ldots, n_N\rangle = \frac{1}{\sqrt{n_1! n_2! \cdots n_N!}} (\hat{A}_1^\dagger)^{n_1} (\hat{A}_2^\dagger)^{n_2} \cdots (\hat{A}_N^\dagger)^{n_N} |0, 0, \ldots, 0\rangle. \quad (470)$$

contains n_k elementary excitations of the normal mode k. These excitations are obtained by applying n_k times the creation operator $\hat{A}_k^\dagger$ to the ground state. This formalism is called the *occupation number representation* with the occupation numbers n_k, namely the number of phonons that have the same frequency ω_k. As in the case of a single harmonic oscillator, the set of base states $|n_1, n_2, \ldots, n_N\rangle$, that is the excitations, are the eigenstates of the Hamilton operator. They are orthonormal vectors which span a Hilbert space, called *Fock space*. Because of (411), the chain's total energy for the base state $|n_1, n_2, \ldots, n_N\rangle$ is

$$E = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}} \hbar\omega_k \left(n_k + \frac{1}{2}\right). \quad (471)$$

Using the commutation relations (468), it follows for $N = 2$ that

$$|n_1, n_2\rangle = \frac{1}{\sqrt{n_1! n_2!}} (\hat{A}_1^\dagger)^{n_1} (\hat{A}_2^\dagger)^{n_2} |0, 0\rangle = |n_2, n_1\rangle. \quad (472)$$

This means that the base states are symmetric under excitation exchange. In particle physics symmetric states are called *bosons*. It is easy to generalize this symmetry to base states $|n_1, n_2, \ldots, n_N\rangle$. The formalism, using creation and annihilation operators that act on the Fock space, provide a useful tool to study many-body systems.

Is there a difference between phonons, living as an excitation in a lattice like a solid state crystal, and photons, the excitations of the electromagnetic field? Not so much. The *phonon* can be viewed as a quantum of sound, for example in a solid, whereas the *photon* is the quantum of electromagnetic radiation. Both are bosons. They are excitations described by waves. Moreover, both have an energy equal to their frequency times $\hbar$. This is similar as in string theory, where particles are modes of strings.

5.8 The Field as a Continuous Limit

Until now we have considered only discrete systems. If the number of atoms grows such that their spacing becomes very small, we obtain approximately a continuous system, that is, a one-dimensional string of length

$$L = Na = 2\pi R. \tag{473}$$

In the continuous limit we assume a finite mass density ϱ

$$\varrho = m/a, \tag{474}$$

and a finite string tension

$$\sigma = \kappa_s a. \tag{475}$$

Then

$$a, m \to 0,\ N, \kappa_s \to \infty,\ L, \rho, \sigma \text{ are constant},\ x = ja \quad \text{for } j = 1, \ldots, N. \tag{476}$$

Hence, the length L, the density ρ, and the tension σ are fixed. The displacements $q_j(t)$ transform into a continuous displacement field $q(x,t)$. The discrete index is merely replaced by the continuous index x. Then differences become derivatives, and sums are replaced by integrals:

$$q_j(t) \quad \to \quad q(x,t), \tag{477}$$

$$(q_j(t) - q_{j-1}(t))^2 \quad \to \quad a^2 \left(\frac{\partial q(x,t)}{\partial x} \right)^2, \tag{478}$$

$$\sum_j \quad \to \quad \frac{1}{a} \int_0^L dx, \tag{479}$$

and therefore

$$N = \sum_{j=1}^{N} 1 = \frac{L}{a} = \frac{1}{a} \int_0^L 1 dx. \tag{480}$$

From (452) it follows that the Hamiltonian, the sum of kinetic and potential energy, becomes

$$H(p,q) = \int_0^L dx \left[\frac{1}{2\varrho} p^2(x,t) + \frac{\sigma}{2} \left(\frac{\partial q(x,t)}{\partial x} \right)^2 \right], \tag{481}$$

where

$$p(x,t) = \varrho \frac{\partial q(x,t)}{\partial t} \tag{482}$$

is the canonical momentum. According to (454), the equation of motion is:

$$\frac{\partial^2 q(x,t)}{\partial t^2} - c^2 \frac{\partial^2 q(x,t)}{\partial x^2} = 0, \quad c = \sqrt{\frac{\sigma}{\varrho}} = a\sqrt{\frac{\kappa_s}{m}}. \tag{483}$$

This is a linear wave equation with propagation velocity c. The quantity c is a characteristic speed in a specific medium, thus not necessarily the speed of light. The linearity implies that linear combinations of solutions are solutions.

5.9 Renormalization

Until now we have not quantized continuous systems, but it is obvious that there should be problems if the number of modes k_n is infinite. In the continuum limit $N \to \infty$ the ground stage energy

$$E_0 = \sum_{n=-\frac{N}{2}}^{\frac{N}{2}} \frac{1}{2}\hbar\omega_n = \frac{\hbar}{2} \sum_{n=-\infty}^{\infty} \frac{c}{R}|n| \tag{484}$$

is divergent.

Such divergences occur often in QFT. Frequently, divergences appear when going to the continuous limit and if interactions are occur. They require a physical discussion, since calculations with infinities are not so funny, and continuous models are only physically relevant if such divergences are avoided. There are cut-off procedures, so-called *renormalizations*, that replace the infinities by finite values. For instance, such a cut off procedure may be motivated by assuming a physically relevant particle number.

5.10 Canonical Field Quantization

The quantization of fields is very similar to the formalism used for the discrete chain. The major difference is that the finite index j is replaced by the continuous index x. The positions and momenta $q_j(t)$ and $p_j(t)$ of the atoms in (453) are replaced by the fields $q(x,t)$ and $p(x,t)$ as defined in Section 5.8. We quantize the fields by replacing the Hermitian operators $\hat{Q}_j$ and $\hat{P}_j$ by the operator fields $\hat{Q}(x,t)$ and $\hat{P}(x,t)$, respectively. We impose for these operator fields the same commutation relation as in (466), but with continuous index x:

$$\left[\hat{Q}(x,t), \hat{P}(y,t)\right] = i\hbar\delta(x-y) \text{ for all } x, y. \tag{485}$$

This relation is often called an equal-time commutation relation. Obviously, time is not relevant. According to (481) the Hamilton operator is

$$\hat{H} = \int_0^L dx \left[\frac{1}{2\varrho}\hat{P}^2(x,t) + \frac{\sigma}{2}\left(\frac{\partial\hat{Q}(x,t)}{\partial x}\right)^2\right]. \tag{486}$$

The degree of freedom is infinite, and the field values at points (x,t) are operators, that is, machines comprising all possible positions and momenta.

In this form it is not clear how the eigenvalues and eigenstates of the Hamilton operator look like. The idea is to represent the Hamiltonian as a sum over independent oscillators by using a Fourier expansion. Since $q(x,t)$ and $p(x,t)$ are real and periodic for fixed t, the operators $\hat{Q}(x)$ and $\hat{P}(x)$ are Hermitian. Then we can digonalize these operators and perform a Fourier expansion:

$$\hat{Q}(x) = \frac{1}{\sqrt{L}}\sum_k e^{ikx}\hat{Q}_k, \tag{487}$$

$$\hat{P}(x) = \frac{1}{\sqrt{L}}\sum_k e^{ikx}\hat{P}_k, \tag{488}$$

where we set the wave-vectors $k = 2\pi n/L$ and n integer.

The inverse operations are

$$\begin{aligned} \hat{Q}_k &= \frac{1}{\sqrt{L}}\int_0^L dx\, e^{ikx}\hat{Q}(x), \\ \hat{P}_k &= \frac{1}{\sqrt{L}}\int_0^L dx\, e^{ikx}\hat{P}(x), \end{aligned} \tag{489}$$

with commutation relation

$$[\hat{P}_k, \hat{Q}_{k'}] = -i\hbar\delta_{kk'}. \tag{490}$$

This equation follows from the commutation relation (532). Moreover, it follows immediately from (489) that the operators $\hat{Q}_k$ and $\hat{P}_k$ satisfy

$$\hat{Q}_k^\dagger = \hat{Q}_{-k} \quad \text{and} \quad \hat{P}_k^\dagger = \hat{P}_{-k}, \tag{491}$$

and thus are not Hermitian. Since

$$\begin{aligned}\int_0^L dx \left(\frac{\partial}{\partial x}\hat{Q}(x,t)\right)^2 &= \int_0^L dx \left(\frac{1}{\sqrt{L}}\sum_k ike^{ikx}\hat{Q}_k\right)^2 \\ &= \frac{1}{L}\sum_{k,k'}(ik\hat{Q}_k)\cdot(ik'\hat{Q}_{k'})\int_0^L dx\, e^{i(k+k')x} \\ &= \sum_{k,k'}(ik\hat{Q}_k)(ik'\hat{Q}_{k'})\cdot\delta_{k+k',0} \\ &= \sum_k k^2\hat{Q}_k\hat{Q}_{-k},\end{aligned} \tag{492}$$

and similarly

$$\begin{aligned}\int_0^L dx\hat{P}^2(x,t) &= \int_0^L dx \left(\frac{1}{\sqrt{L}}\sum_k e^{ikx}\hat{P}_k\right)^2 \\ &= \frac{1}{L}\sum_{k,k'}\hat{P}_k\hat{P}_{k'}\int_0^L dx\, e^{i(k+k')x} \\ &= \sum_k \hat{P}_k\hat{P}_{-k},\end{aligned} \tag{493}$$

we obtain the Hamiltonian

$$\hat{H} = \sum_k \frac{1}{2\varrho}\hat{P}_k\hat{P}_{-k} + \frac{\delta}{2}\frac{\omega_k^2}{c^2}\hat{Q}_k\hat{Q}_{-k}, \tag{494}$$

where

$$\omega_k^2 = c^2k^2, \quad k = 2\pi n/L, \quad n \quad \text{integer}. \tag{495}$$

is satisfied. From (491) it follows that

$$\hat{H} = \sum_k \frac{1}{2\varrho}\hat{P}_k\hat{P}_k^\dagger + \frac{\delta}{2}\frac{\omega_k^2}{c^2}\hat{Q}_k\hat{Q}_k^\dagger. \tag{496}$$

Inspired by the latter formalism of creation and annihilation for a single harmonic oscillator, we set

$$\hat{A}_k = \sqrt{\frac{m\omega_k}{2\hbar}}\left(\hat{Q}_k + \frac{i}{m\omega_k}\hat{P}_{-k}\right) \tag{497}$$

and

$$\hat{A}_k^\dagger = \sqrt{\frac{m\omega_k}{2\hbar}}\left(\hat{Q}_{-k} - \frac{i}{m\omega_k}\hat{P}_k\right). \tag{498}$$

Then we obtain the computation relation

$$[\hat{A}_k, \hat{A}_{k'}^\dagger] = \delta_{kk'}, \quad [\hat{A}_k, \hat{A}_{k'}] = [\hat{A}_k^\dagger, \hat{A}_{k'}^\dagger] = 0, \tag{499}$$

and the Hamiltonian takes the form

$$\hat{H} = \sum_k \hbar\omega_k\left(\hat{A}_k^\dagger\hat{A}_k + \frac{1}{2}\hat{1}\right). \tag{500}$$

Thus, the Hamiltonian describes an infinite number of separate quantum harmonic oscillators, each indexed by a wavenumber $k = 2\pi n/L$ and its own frequency ω_k satisfying the dispersion law $\omega_k^2 = c^2k^2$.

As in the analysis of a simple harmonic oscillator it follows that $\hat{A}_k^\dagger$ creates a quanta of the mode k and $\hat{A}_k$ annihilates such a quanta. Moreover, for the vacuum

$$\hat{A}_k|0\rangle = 0 \quad \text{for all} \quad k. \tag{501}$$

Now we consider what happens when we apply the Hamiltonian to the vacuum:

$$\begin{aligned} \hat{H}|0\rangle &= \sum_k \hbar\omega_k \left(\hat{A}_k^\dagger \hat{A}|0\rangle + \frac{1}{2}|0\rangle \right) \\ &= \left(\sum_k \frac{1}{2}\hbar\omega_k \right) |0\rangle. \end{aligned} \tag{502}$$

Because there are infinitely many oscillators this sum diverges and requires *renormalization*. The conventional argument is that only energy differences really matters. Then the ground state $|0\rangle$ has zero energy.

5.11 Quantization Using Fourier Analysis

In the previous sections we started with a classical Hamiltonian $H(p,q)$, we quantized H by replacing the classical dynamical variables p and q by Hermitian operators $\hat{P}$ and $\hat{Q}$, and we obtained an Hermitian operator $\hat{H}(\hat{P},\hat{Q})$. Using the commutation relation, we found the eigenvalues and orthonormal eigenstates of $\hat{H}$.

There is an alternative way how field quantization is often presented. We start with a classical field satisfying any wave equation. With Fourier analysis we find that the components of the field's Fourier expansion satisfy the equation of harmonic oscillators. Finally, we quantize the components in the same way as we quantized the harmonic oscillator. Let us do this for the field representing the one-dimensional string starting with the equation of motion

$$\frac{\partial^2 q(x,t)}{\partial t} - c^2 \frac{\partial^2 q(x,t)}{\partial x^2} = 0, \quad c = \sqrt{\frac{b}{\varrho}}. \tag{503}$$

After one circumference the field $q(x,t)$ comes back, that is,

$$q(x+L) = q(x). \tag{504}$$

In order to find the field's Fourier expansion, we try the functions

$$f_{k_n}(x) = \sqrt{\frac{1}{L}} e^{ik_n x}, \quad k_n = \frac{2\pi n}{L}, \quad n = 0, \pm 1, \ldots \tag{505}$$

They satisfy the orthonormality condition

$$\int_0^L dx\, f^*_{k'}(x) f_k(x) = \delta_{k'k}, \tag{506}$$

and the completeness condition

$$\sum_k f^*_k(x') f_k(x) = \delta(x'-x). \tag{507}$$

In the following, we simply write $k = k_n$ and $k' = k_{n'}$.

The theory of Fourier analysis implies that any periodic field can be Fourier expanded in the form

$$q(x,t) = \sqrt{L} \sum_k q_k(t) f_k(x). \tag{508}$$

Because of (505) we multiply with factor $\sqrt{L}$ so that both sides have the same units. Since $q(x,t)$ is real $q^*(x,t) = q(x,t)$, and therefore

$$q^*_k(t) = q_{-k}(t). \tag{509}$$

Substituting the Fourier expansion into the wave equation (503) yields

$$\sqrt{L} \sum_k (\ddot{q}_k(t) f_k(x) - c^2 q_k(t) (ik)^2 f_k(x)) = 0. \tag{510}$$

The functions $f_k(x)$ are orthonormal, thus linear independent. Hence, this sum is zero only if all coefficients are zero, that is,

$$\ddot{q}_k(t) + c^2k^2q_k(t) = 0 \quad \text{for all} \quad k. \tag{511}$$

This is the equation of motion for harmonic oscillators with frequencies $\omega_k = c|k|$.

The kinetic energy is

$$\begin{aligned} T &= \int_0^L dx \frac{\varrho}{2} \frac{\partial q(x,t)^*}{\partial t} \frac{\partial q(x,t)}{\partial t} \\ &= \sum_{k,k'} \frac{\varrho}{2} L \dot{q}_k^* \dot{q}_{k'} \int_0^L dx f_k^*(x) f_{k'}(x) \\ &= \sum_k \frac{\varrho L}{2} \dot{q}_k \dot{q}_{-k}. \end{aligned} \tag{512}$$

In these equations we have used the orthonormality conditions and equation (509). For the potential energy we obtain when using integration by parts:

$$\begin{aligned} V &= \int_0^L dx \frac{\sigma}{2} \frac{\partial q^*(x,t)}{\partial x} \frac{\partial q(x,t)}{\partial x} \\ &= \sum_{k,k'} \frac{\sigma}{2} L q_k^* q_{k'} \int_0^L dx \left(\frac{-\partial^2 f_k^*}{\partial x^2} \right) f_{k'}. \end{aligned} \tag{513}$$

Because of the orthonormal relation (506) and the completeness condition (507) it follows that

$$V = \sum_{k,k'} \frac{\sigma}{2} L q_{-k} q_{k'} k^2 \delta_{k'k}. \tag{514}$$

Finally, using the dispersion relation $\omega_k^2 = c^2k^2$, we find

$$V = \sum_k \frac{\sigma L}{2c^2} \omega_k^2 q_{-k} q_k. \tag{515}$$

The Hamiltonian, the sum of kinetic and potential energy, becomes

$$H = \sum_k \frac{\varrho L}{2} \dot{q}_k \dot{q}_{-k} + \frac{\varrho L}{2} \omega_k^2 q_{-k} q_k. \tag{516}$$

If we set

$$p_k = \varrho L \dot{q}_k, \tag{517}$$

then the Hamiltonian is

$$H(p,q) = \sum_k \frac{1}{2} \frac{1}{\varrho L} p_k p_{-k} + \frac{1}{2} \varrho L \omega_k^2 q_k q_{-k}. \tag{518}$$

Since ϱ is the mass density, the quantity ϱL is the mass m of the string. Now we define appropriate units leading to dimensionless phase space coordinates.

We replace

$$q_k \quad \to \quad \frac{1}{\sqrt{2\hbar/m\omega_k}} \quad q_k, \quad p_k \quad \to \quad \frac{1}{\sqrt{2\hbar m\omega_k}} \quad p_k. \tag{519}$$

Then the Hamiltonian is

$$H(p,q) = \sum_k \hbar\omega_k (p_k p_{-k} + q_k q_{-q}). \tag{520}$$

It can be written in the form

$$H(p,q) = \sum_k \hbar\omega_k \left(\alpha_k^\dagger \alpha_k + \frac{1}{2}\right), \tag{521}$$

where

$$\alpha_k(t) = q_k(t) + ip_{-k}(t) \tag{522}$$

are complex coefficients. Since

$$q_k = \frac{1}{2}(\alpha_k + \alpha_k^\dagger), \tag{523}$$

we obtain from (508) and (519) the solution

$$q(x,t) = \sum_k \sqrt{L} \frac{1}{\sqrt{2\hbar/m\omega_k}} \frac{\alpha_k + \alpha_k^\dagger}{2} f_k(x). \tag{524}$$

We can write this equation also in the form

$$q(x,t) = \sum_k c_k (\alpha_k(0) e^{-i\omega_k t} f_k(x) + \alpha_k^\dagger(o) e^{i\omega t} f_k^\dagger(t). \tag{525}$$

It is a nice exercise to calculate the constants c_k.

Both, the Hamiltonian H given in formula (521) and the solution (525), hold true also in the case, where the elastic string is a straight line. You should try it as a little exercise.

In particular, we have proved that the Hamiltonian for the field can be interpreted as an infinite collection of simple harmonic oscillators, where each oscillator corresponds to a mode k of the field with frequency ω_k. The modes are viewed as excitations of the string.

In order to quantize these classical field, we replace in (520) the dynamical variables q_k and p_k by the corresponding Hermitian operators $\hat{Q}_k$ and $\hat{P}_k$ for all modes k together with their commutation relations

$$\left[\hat{Q}_k, \hat{P}_{k'}\right] = i\hbar\delta_{k,k'}, \ \left[\hat{Q}_k, \hat{Q}_{k'}\right] = 0, \ \left[\hat{P}_k, \hat{P}_{k'}\right] = 0. \tag{526}$$

Hence, the complex coefficients (522) must be replaced by the non-Hermitian operators

$$\hat{A}_k = \hat{Q}_k + i\hat{P}_{-k}, \ \hat{A}_k^\dagger = \hat{Q}_k - i\hat{P}_k. \tag{527}$$

They satisfy the commutation relations

$$\left[\hat{A}_k, \hat{A}_{k'}^\dagger\right] = \delta_{k,k'},\ \left[\hat{A}_k, \hat{A}_{k'}\right] = 0,\ \left[\hat{A}_k^\dagger, \hat{A}_{k'}^\dagger\right] = 0. \tag{528}$$

Therefore, they are the annihilation and creation operators. Any two operators related to different modes commute. These relations express that each mode k is an independent degree of freedom.

From (521) we obtain the Hamilton operator

$$\hat{H} = \sum_k \hbar\omega_k \left(\hat{A}_k^\dagger \hat{A}_k + \frac{1}{2}\right). \tag{529}$$

Comparing this result with (449), we see that both formulas are identical, except for the number of modes. This shows once more that the chain behaves entirely as infinitely many uncoupled separate harmonic oscillators labeled by k, the latter called *phonon* modes. Each phonon can have only energies that are integer multiples of $\hbar\omega_k$. The Hamiltonian of this chain is a linear superposition over non-interacting modes.

Formula (508) lead us to the quantum field

$$\hat{Q}(x) = \sum_k \sqrt{L}\hat{Q}_k f_k(x). \tag{530}$$

Therefore as expected, canonical quantization transforms a real valued field to an operator-valued field. Associated to this field is the operator-valued canonical momentum field

$$\hat{P}(x) = \sum_k \frac{1}{\sqrt{L}}\hat{P}_k f_k^\dagger(x). \tag{531}$$

Because of (507) we obtain the commutation relation for fields

$$\left[\hat{Q}(x), \hat{P}(y)\right] = \sum_{k,k'} \left[\hat{Q}_k, \hat{P}_{k'}\right] f_k(x) f_{k'}^\dagger(y) = i\hbar\delta(x-y) \text{ for all } x, y. \tag{532}$$

This formula coincides with the commutation relation (532) of the previous section.

As in the analysis of a simple harmonic oscillator it follows that $\hat{A}_k^\dagger$ creates a quanta of the mode k and $\hat{A}_k$ annihilates such a quanta. Moreover, for the vacuum

$$\hat{A}_k|0\rangle = 0 \quad \text{for all} \quad k. \tag{533}$$

Now we consider what happens when we apply the Hamiltonian to the vacuum:

$$\begin{aligned} \hat{H}|0\rangle &= \sum_k \hbar\omega_k \left(\hat{A}_k^\dagger \hat{A}|0\rangle + \frac{1}{2}|0\rangle\right) \\ &= \left(\sum_k \frac{1}{2}\hbar\omega_k\right)|0\rangle. \end{aligned} \tag{534}$$

Because there are infinitely many oscillators this sum diverges and requires *renormalization*. The conventional argument is that only energy differences really matters. Then the ground state $|0\rangle$ has zero energy.

5.12 Quantization of the Electromagnetic Field

The electromagnetic field can be quantized similarly as the continuous chain. We use Fourier analysis for the classical electromagnetic vector potential with periodic boundary conditions. Then we obtain as before an infinite discrete number of independent modes k, r. The index r labels the polarisation of the electromagnetic field. Each mode is described by the equation of an harmonic oscillator, and thus is quantized by applying the quantization rules for harmonic oscillators.

Not surprisingly, the resulting Hamiltonian has the form

$$\hat{H} = \sum_k \sum_r \hbar\omega_k \left(\hat{A}^\dagger_{k,r} \hat{A}_{k,r} + \frac{1}{2} \right). \tag{535}$$

where the operators satisfy the commutation relations

$$[\hat{A}_{k,r}, \hat{A}^\dagger_{k',r'}] = \delta_{kk'}\delta_{rr'}, \quad [\hat{A}_{k,r}, \hat{A}_{k',r'}] = [\hat{A}^\dagger_{k,r}, \hat{A}^\dagger_{k',r'}] = 0, \tag{536}$$

From these relations it follows as above that the operators $\hat{A}^\dagger_{k,r}$ and $\hat{A}_{k,r}$ are the creation and annihilation operators. They define the number operator

$$\hat{N}_{k,r} = \hat{A}^\dagger_{k,r} \hat{A}_{k,r}. \tag{537}$$

This operator commutes with the Hamilton operator, and both operators have the corresponding orthogonal eigenvectors

$$|n_{k,r}\rangle = [\hat{A}^\dagger_{k,r}]^{n_{k,r}} |0\rangle. \tag{538}$$

The interpretation is the same as for the chain, except that *phonons* are replaced by the word *photons*. The operator $\hat{A}^\dagger_{k,r}$ creates a photon with momentum $\hbar k$, energy $\hbar\omega_k$, and a polarization depending on k, r. The operator $\hat{A}_{k,r}$ annihilates a photon. The number operator $\hat{N}_{k,r}$ says that the mode k, r is occupied by $n_{k,r}$ photons, where the latter is called *occupation number*.

5.13 States in QFT

Thus far we have used canonical quantization and obtained orthonormal base states of the Hamiltonian, namely the quantized normal modes. This transition from a classical field theory with fields $q(x,t)$ and $p(x,t)$ to QFT is characterized by the entrance of Hermitian operator-valued fields, the Hamilton operator, position and momentum operators $\hat{Q}(x,t)$ and $\hat{P}(x,t)$ respectively, and non-Hermitian creation and annihilation operators. These are solely operators that act on base states and that do not depend on time. But what is the probabilistic part of a QFT?

Actually, the answer is simple. We proceed as in Sections 4.4 and 4.10. The probabilities in QFT appear in the same way as for quantum registers. The state of a quantum field is the quantum superposition of orthogonal field base states.

Moreover, we can calculate for the observables their expectation values. Then we obtain classical fields of expectation values. We have already mentioned in Section 2.2 that classical mechanics and quantum mechanics can be embedded in the same mathematical framework. Real dynamical variables of classical mechanics are just the quantum mechanical average values of Hermitian operators.

Similarly in optics, polarization is described by Jones vectors and Stokes vectors. The Jones vectors are two-dimensinional complex vectors describing the probability amplitudes of the two polarization states. The Stokes vectors are three-dimensional real vectors describing the average values of partially polarized light. This is in accordance with our point of view where two-dimensinional complex vectors correspond to three-dimensional real vectors, such that there algebras $su(2)$ and $so(3)$ are isomorphic. We shall describe these aspects later in a supplement.

5.14 Conclusions

Quantum mechanics, the theory of microscopic objects, and the theory of relativity, suitable for velocities close to the speed of light, are known to form the two fundamental frameworks in physics. Frequently, it is argued that the seemingly strange consequences and the paradoxes are beyond our imagination, since we are macroscopic objects that experience only small relative velocities.

In these lecture notes we have derived physics in terms of the trinity "future, present, and past". An external time parameter is not assumed.

Quantum mechanics was derived as a theory of the timeless future based on "as well as" possibilities. It is a probabilistic theory which may be described with the phrase " what might soon happen when nothing happens". Many quantum mechanical paradoxes vanish in our approach.

The mathematical part of the theory of relativity, in particular the Lorentz transformation, was derived from clocks, special machines in a (3+3)-position-velocity space. This derivation does not require the spacetime and the usual postulates of special relativity, namely that (i) the laws of physics are the same in all inertial frames of reference, and that (ii) the speed of light in free space has the same value c in all inertial frames of reference. Now the apparent paradoxes of this theory turn out to be rather natural in the (3+3)-position-velocity space.

Classical mechanics is the theory of facts in the timeless past. Its smooth solutions can be explained in terms of average values. Summing up, that is integration, is a smoothing procedure. The parameter t is solely a geometrical parameter allowing an explicit description of the solutions.

6 Appendix A: The Theorem of Hurwitz

Hurwitz observed that for any commutative multiplication of two numbers x_1 and y_1 it is

$$x_1^2 y_1^2 = (x_1 y_1)^2, \tag{539}$$

that is, the product of two squares is the square of a product. This equation holds valid for real numbers, and coincides with equation (1) which describe the rule for the squared length of numbers. This equation can also be written as

$$x_1^2 y_1^2 = (z_1)^2, \tag{540}$$

where $z_1 = x_1 y_1$ is a bilinear function of both numbers.

It is natural to look for similar square identities in higher dimensions. Actually, in 1898 Hurwitz proved his famous $(1, 2, 4, 8)$-Theorem:

Let $\mathbb{F}$ be a field of characteristic[75] not equal to 2. If

$$(x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2) = z_1^2 + \cdots + z_n^2 \tag{541}$$

for all $x_1, \ldots, x_n, y_1, \ldots, y_n$ in $\mathbb{F}$, where each z_k is an $\mathbb{F}$-bilinear function of the x's and the y's, then $n = 1, 2, 4$ or 8.

This theorem says that, except for $n = 1, 2, 4, 8$, further identities of this kind are impossible. Real numbers satisfy this equation, as we have seen above.

For $n = 2$ this identity is fulfilled for complex numbers, since multiplying two complex numbers $x = x_1 + ix_2$ and $y = y_1 + iy_2$ we get

$$(x_1^2 + x_2^2)(y_1^2 + y_2^2) = (x_1 y_1 - x_2 y_2)^2 + (x_1 y_2 + x_2 y_1)^2 = z_1^2 + z_2^2, \tag{542}$$

where $z = z_1 + iz_2 = xy$.

For $n = 4$ this identity can be satisfied in the form

$$\begin{aligned}(x_1^2 + x_2^2 + x_3^2 + x_4^2)(y_1^2 + y_2^2 + y_3^2 + y_4^2) \quad = \quad & (x_1 y_1 - x_2 y_2 - x_3 y_3 - x_4 y_4)^2 + \\ & (x_1 y_2 + x_2 y_1 + x_3 y_4 - x_4 y_3)^2 + \\ & (x_1 y_3 + x_3 y_1 - x_2 y_4 + x_4 y_2)^2 + \\ & (x_1 y_4 + x_4 y_1 + x_2 y_3 - x_3 y_2)^2.\end{aligned} \tag{543}$$

For $n = 8$ the identity looks similar and can be found in the literature.

It turns out that, up to a linear change of variables, only the four classical real division algebras of dimensions $n = 1, 2, 4, 8$, namely the real numbers, the complex numbers, the quaternions, and the octonions satisfy this identity.

Originally, Hurwitz proved his theorem for $\mathbb{F} = \mathbb{C}$, but the field of scalars requires only not to be of characteristic equal to 2 for his proof. The idea of the proof is to write equation (541) as a set of equations in $n \times n$ matrices with coefficients in $\mathbb{F}$. Then it is shown that this matrix equations can be solved only when $n = 1, 2, 4$ and 8, see Conrad[76].

[75] For any element x of a finite field F and any integer n, let nx denote the sum of n copies of x. The smallest integer n such that $n1 = 0$ must exist, and is called the characteristic of the field.

[76] Conrad [2010]

Quaternions can be represented as 2×2 complex matrices with non-commutative multiplication, and when we go from quaternions to octonions we loose even associativity. Thus, octonions cannot be represented as matrices with usual matrix-matrix multiplication. Both fields do not satisfy our rules for numbers. This does not mean, however, that these quantities don't have useful applications. But in our sense they are not numbers, and thus the field of complex numbers is maximal.

Frequently it is said that complex numbers are used only for technical reasons, and physics requires only real numbers. This is, however, doubtful, and we close this section with the quote of Susskind:

> *Of course, we've said all along that the space of states is a complex vector space, but until now we have not had to use complex numbers in our calculations. Are the complex numbers in Eqs. 2.10 a convenience or a necessity? Given our framework for spin states, there is no way around them. It's somewhat tedious to demonstrate this, but the steps are straightforward. The following exercise gives you a road map. The need for complex numbers is a general feature of quantum mechanics, and we'll see more examples as we go.*
> Leonard Susskind and Art Friedman[77]

[77] Susskind [2014], page 42

7 Appendix B: Symmetry and Groups

Symmetries play an important role in physics. In particular, symmetries imply conservation laws in classical mechanics: the conservation of energy, momentum and angular momentum. But more important, they provide the fundamental tools in elementary particle physics. We cannot go deeply into these theories, but we will briefly recall some basic facts and refer the reader for more details and elaborate descriptions for instance to Zeidler (2006), and the literature referenced therein. Basically, we can say that a

- *symmetry* is described by a set of operations which can be performed on an object leaving it *invariant.* The object may be a function, an operator, a square, a ball, a physical law, the path of a particle, whatever it is.

Closely related to symmetries is *Noether's theorem* which states that (frequently) to a symmetry there corresponds a conserved quantity, that is, the symmetry expresses itself by the existence of a conserved quantity. Exceptions, for example, are dissipative systems.

A simple example of a symmetry is a ball, described by the function $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$, which is rotational-invariant, that is, the function value does not change if the coordinate system is rotated. The radius is a conserved property. Notice that this function is not translational-invariant. Another example is the symmetry that is satisfied for odd functions. There, the identity and the operation $f(x) \to -f(-x)$ are symmetry operations. The sum $s(x) = f(x) + f(-x) = 0$ is a conserved quantity. The question for symmetries is important when physical processes should be described by different observers, like in the theory of relativity.

These simple examples of symmetry can be put into a formal framework: A *symmetry* is defined by a group of transformations that leave certain properties or observable values of an object invariant. From the mathematical point of view symmetry is described by group theory.

A *group* is a set of elements equipped with an operation that combines any two elements yielding another element of the group. The operation fulfills three basic rules, namely (i) the operation is associativ, (ii) the group contains an identity, and (iii) each element has an inverse element. A *subgroup* is a subset of a group that forms itself a group. Of great importance are *matrix groups* that consist of specific invertible matrices over some field, where the operation is the matrix multiplication.

At first let us look at the symmetries of a square. We may imagine that the vertices of the square are the points $i^0 = 1, i^1 = i, i^2 = -1$ and $i^3 = -i$ in the complex plane. The square is unchanged or invariant by the rotations $R_1 = 1, R_2 = i, R_3 = -1, R_4 = -i$. For the square these four numbers represent all our rotations and define a *group*: The product of any two of these rotations is again one of them, for instance $i^5 = i$, the multiplication is associative, there is an identity element 1, and there are inverses. In our particular example the group is commutative. This group leaves the square invariant, and hence defines a symmetry. We notice that the square is not completely symmetric such as a circle which is invariant by every rotation around its center.

There are two equivalent viewpoints to consider translations, rotations or other transformations. First, in the case of *active transformation* the physical object, system or state changes, and that makes sense even in the absence of a coordinate system. Secondly, in *passive transformation* merely a change in the coordinate system occurs, without any physical significance. The distinction between active and passive transformations should be kept in mind in applications. In most cases, mathematicians usually mean passive transformations, while physicists could mean either. If we look at an object in different ways, and it appears always the same, then the system has a symmetry respectively to the passive operations defined by different chosen points of view. In the theory of relativity, for instance, the same physical system is observed from different inertial frames, and the point of view is the passive one.

An important example are the symmetries in classical mechanics. The Galilean relativity says that the mechanical laws take the same form in all inertial systems (this is the passive point of view). Therefore, the laws should not alter if the time parameter changes by a constant time. This is called the *homogeneity of time*. It is said that classical physics is the same today as yesterday and as tomorrow. Remember, we have argued that classical physics is a theory of the timeless past, where time is a geometrical parameter. Let us look at Newton's equation for a particle $m\ddot{\mathbf{x}}(t) = -\nabla U(\mathbf{x})$ in a potential field $U(\mathbf{x})$, such that the smooth force $\nabla U(\mathbf{x})$ depends only on the position $\mathbf{x}$. Obviously, Newton's law depends only on time differences, and not on the absolute time. For each solution $\mathbf{x}(t)$ the function $\mathbf{x}(t + t_0)$ is also a solution of Newton's equation for every fixed time t_0. Newton's law is invariant under *time translations*, and this invariance implies that the energy $E(t) = m\dot{\mathbf{x}}(t)^2/2 + U(\mathbf{x}(t))$ is conserved for all times t. This is expressed by the fact that the time derivative of the energy is zero for any solution satisfying Newton's equation. Summarizing, the time translation invariance implies conservation of energy. Notice that for a time-dependent potential $U(\mathbf{x}, t)$ this argument fails, since an additional term $\partial U(\mathbf{x}, t)/\partial t$ would appear in the time derivative of the energy.

A physical process, say $x(t)$, is called *reversible* if it is invariant under the *time reversal operation* $x(-t)$. Classical mechanics is reversible since a solution $x(t)$, satisfying Newton's law, implies that the time reversal is also a solution. There is a corresponding conserved quantity yielding the physical effect that antiparticles moving backwards in time must exist. Contrary, the heat equation and diffusion processes are irreversible. If, for instance, friction occurs and the force in Newton's equation depends on the velocity, that is, $m\ddot{x}(t) = -\kappa\dot{x}(t)$ where $\kappa > 0$, then this problem has the irreversible solution of motion $e^{-\kappa t/m}$. In every sense we observe and feel that our world is irreversible. Human beings can distinguish between past, present and future. One never observes that animals or plants develop back in time or become younger.

Symmetries can be described mathematically by groups. The elements of these groups can be represented as matrices, and thus can be parameterized by finitely many real or complex numbers, namely their coefficients. If the multiplication in the group depends in a smooth manner on these parameters, then it is called a Lie group. More precisely, a *Lie group* is a group which is at the same time a finite-dimensiona smooth manifold, so that the group

operations of multiplication and inversion are smooth maps.

The most general matrix group $GL(n,\mathbb{S})$, where $\mathbb{S}=\mathbb{R}$ or $\mathbb{S}=\mathbb{C}$ consists of all nonsingular $n\times n$-matrices with real or complex coefficients, and is called the *general linear group*. This group contains many subgroups. The *special linear group* $SL(n,\mathbb{C})$ consists of all non-singular complex matrices with determinant equal one. This set forms a group because the determinant of the product of two matrices is equal to the product of both determinants. The group $SL(n,\mathbb{R})$, which is the same group but with real coefficients, is a subgroup of $SL(n,\mathbb{C})$.

The *orthogonal group* $O(n)$ is defined as the set of all real $n\times n$-matrices that leave the Euclidean inner product invariant. Since the matrices are real, the orthogonal group is a subgroup of $GL(n,\mathbb{R})$. It follows that $R^\dagger = R^T$ and $R^T R = 1$. Thus, the matrices are orthogonal and $R^T = R^{-1}$. The subgroup $SO(n)$, the subset of $O(n)$ with determinant equal to +1, is called the *special orthogonal group*. Obviously, $SO(n)$ forms a subgroup since $\det(SR) = \det(S)\det(R)$ and $\det(R^{-1}) = 1/\det(R)$. In contrast, the subset of $O(n)$ with determinant equal to -1 does not contain the unity, and thus is no group.

The set of all unitary $n\times n$ matrices forms the group $U(n)$. As in the real case the subset of unitary matrices U with $\det U = 1$ is a subgroup, denoted by $SU(n)$, and is called the *special unitary group*.

A generalization of the orthogonal group is to replace the Euclidean inner product by the pseudo-Euclidean inner product. Let p,q be two integers with $p+q=n$. Then the *pseudo-inner product* of two vectors $|\phi\rangle, |\psi\rangle \in \mathbb{C}^n$ is defined as the sum of the first p terms $\phi_i^*\psi_i,\ i=1,\ldots,p$ minus the sum of the remaining q terms $\phi_j^*\psi_j,\ j=p+1,\ldots,n$, that is,

$$\langle\phi|\psi\rangle = \phi_1^*\psi_1 + \ldots \phi_p^*\psi_p - \phi_{p+1}^*\psi_{p+1} - \ldots - \phi_n^*\psi_n. \tag{544}$$

We denote the complex space equipped with an pseudo-inner product by $\mathbb{C}^{p,q}$, and the real space equipped with an pseudo-inner product by $\mathbb{R}^{p,q}$.

We call the group of real matrices that preserves the pseudo-Euclidean inner product the *pseudo-orthogonal group* $O(p,q)$. Their elements are called *Lorentz transformations*. The subgroup $SO(p,q)$ has a $SO(p)$ subgroup and a $SO(q)$ subgroup, consisting of rotations of the first p and the remaining q components, separately. Analogously, the groups of complex matrices $U(p,q)$ and $SU(p,q)$ are defined.

For a little deeper understanding of groups let us first look on the uncomplicated example of the unitary group $U(1) = \{z\in\mathbb{C} : |z| = 1\}$. The group multiplication is the complex multiplication, and the smooth manifold is the unit circle in the complex plane. Since the smooth manifold is compact and connected, the Lie group is called compact and connected. It is a usual approach to describe the full global structure of nonlinear functions or manifolds by linearization. If we define the set $u(1) := \{iq : q\in\mathbb{R}\}$, then

$$e^{iq}\in U(1) \text{ for all } iq\in u(1), \tag{545}$$

and

$$e^{iq} = 1 + iq + O(q^2) \text{ for } q\to 0. \tag{546}$$

If we compute for small ε the group product of two infinitesimal group elements

$$(1+\epsilon q)(1+\epsilon r) = 1+\epsilon(q+r)+\epsilon^2 qr = 1+\epsilon(q+r), \tag{547}$$

where we have ignored the very small squared number ϵ^2, then we see that we can replace the group product by something much simpler, namely a sum. Moreover, roughly spoken, all information about the local structure in a neighborhood of the unity of the Lie group $U(1)$ is contained in the linear set $u(1)$ which is called the corresponding *Lie algebra*. Of course, in our simple example this observation seems not to be meaningful. But this changes when looking on more complicated Lie groups.

Lie groups, in general, are non-commutative. Hence, only looking on the sum is not sufficient. The nature of non-commutativity can be characterized by the expression

$$(1+\epsilon q)(1+\epsilon r)(1+\epsilon q)^{-1}(1+\epsilon r)^{-1} = 1+\epsilon^2(qr-rq), \tag{548}$$

where we have regarded only terms of order ϵ and ϵ^2 of the power series expression $(1+\epsilon q)^{-1} = (1-\epsilon q+\epsilon^2 q^2+...)$. It follows that the *Lie bracket* or *Lie product*

$$[q,r] = qr - rq \tag{549}$$

describes the distance to commutativity. The Lie group $U(1)$ is commutative and the Lie product is zero. Since matrix multiplication is not commutative, for $n \geq 2$ all Lie groups mentioned above are not commutative. The Lie products are unequal zero, but the Lie brackets satisfy the following three rules that can be checked (as usual, we switch for representing matrices and operators to upper case letter):

$$\textit{Bilinearity}\text{: } [a\hat{Q}+b\hat{R},\hat{S}] = a[\hat{Q},\hat{S}]+b[\hat{R},\hat{S}], \tag{550}$$

$$\textit{Anticommutativity}\text{: } [\hat{Q},\hat{R}] = -[\hat{R},\hat{Q}], \tag{551}$$

$$\textit{Jacobi identity}\text{: } [\hat{Q},[\hat{R},\hat{S}]]+[\hat{Q},[\hat{R},\hat{S}]]+[\hat{Q},[\hat{R},\hat{S}]] = 0. \tag{552}$$

A Lie algebra generalizes this concept and is generated by the hull of all sums and all Lie products. More precisely, a *Lie algebra* is a linear space over some field $\mathbb{F}$ together with a binary operation $[\hat{R},\hat{S}]$, the so-called *Lie product*, that satisfies bilinearity, anticommutativity, and the Jacobi identity.

The prototype of a Lie algebra is the linearization of a Lie group at the identity. This linear space carries almost all information about the Lie group and its nonlinear smooth manifold. We describe in detail the Lie algebra for the special linear group

$$SO(3) = \{\hat{Q} \in GL(n,\hat{\mathbb{R}}) : \hat{Q}^T\hat{Q} = 1,\ det(\hat{Q}) = 1\}, \tag{553}$$

the group of rotations in the three-dimensional real space. The infinitesimal group element $\hat{Q} = 1 + \epsilon\hat{R}$ must satisfy

$$1 = (1 + \epsilon\hat{R})^T(1 + \epsilon\hat{R}) = 1 + \epsilon(\hat{R}^T + \hat{R}) + \epsilon^2\hat{R}^T\hat{R}. \tag{554}$$

Hence the Lie algebra, denoted by lower case letter, is

$$so(3) = \{\hat{R} \in GL(n, \hat{\mathbb{R}}) : \hat{R}^T = -\hat{R}\}, \tag{555}$$

that is the set of skew symmetric matrices, also called sometimes *infinitesimal rotations*. It can be shown that, like in the case of $U(1)$, each rotation $\hat{Q} \in SO(3)$ can be represented in the form

$$\hat{Q} = e^{\hat{R}} = 1 + \hat{R} + ..., \text{ with } \hat{R} \in so(3), \tag{556}$$

where in the power series of the exponential function the variable is the skew symmetric matrix $\hat{R}$.

The matrices

$$\hat{Q}_1 := \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix}, \quad \hat{Q}_2 := \begin{pmatrix} \cos\varphi & 0 & \sin\varphi \\ 0 & 1 & 0 \\ -\sin\varphi & 0 & \cos\varphi \end{pmatrix}, \quad \hat{Q}_3 := \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{557}$$

represent the *principal rotations* around the x_j axes with rotation angle φ. Because of the expansion $\cos\varphi = 1 + O(\varphi^2)$ and $\sin\varphi = 1 - \varphi + O(\varphi^3)$ for $\varphi \to 0$ we obtain the infinitesimal rotations

$$\hat{Q}_j(\varphi) = 1 + \varphi\hat{R}_j + O(\varphi^2), \quad \varphi \to 0, \tag{558}$$

where the skew-symmetric $\hat{R}_j \in so(3)$ are defined as

$$\hat{R}_1 := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \hat{R}_2 := \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad \hat{R}_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{559}$$

For small rotation angle φ the matrix $1 + \varphi\hat{R}_j$ represents the rotation $\hat{Q}_j(\varphi)$ very well. The matrices $\hat{R}_j$ form a basis of the real Lie algebra $so(3)$ with Lie products

$$[\hat{R}_1, \hat{R}_2] = \hat{R}_3, \quad [\hat{R}_2, \hat{R}_3] = \hat{R}_1, \quad [\hat{R}_3, \hat{R}_1] = \hat{R}_2. \tag{560}$$

They are called the *generators* of the Lie algebra. All rotations $\hat{Q} \in SO(3)$ can be written in the form

$$\hat{Q} = e^{\varphi_1\hat{R}_1 + \varphi_2\hat{R}_2 + \varphi_3\hat{R}_3} \quad \text{for } \varphi_1, \varphi_2, \varphi_3 \in \hat{\mathbb{R}}. \tag{561}$$

In a very similar manner other Lie algebras can be described. For the above mentioned Lie groups it can be proved: the Lie algebra $gl(n, \mathbb{R})$ consists of all real $n \times n$-matrices, $gl(n, \mathbb{C})$ consists of all complex $n \times n$-matrices, $u(n)$

consists of all skew-adjoint $n \times n$-matrices in $gl(n, \mathbb{C})$, $su(n)$ consists of all $n \times n$-matrices $\hat{Q} \in u(n)$ with trace equal to one, $o(n)$ consists of all real skew-symmetric $n \times n$-matrices, $so(n)$ coincides with $o(n)$.

A brief discussion of $su(2)$, which is related to the theory of the spin of particles, is useful. This Lie algebra is the real linear space where the *Pauli matrices*

$$\hat{\sigma}_1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{\sigma}_2 := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \hat{\sigma}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{562}$$

multiplied by i together with the identity form a basis of a Lie algebra with Lie products

$$[\hat{\sigma}_1, \hat{\sigma}_2] = 2i\hat{\sigma}_3, \quad [\hat{\sigma}_2, \hat{\sigma}_3] = 2i\hat{\sigma}_1, \quad [\hat{\sigma}_3, \hat{\sigma}_1] = 2i\hat{\sigma}_2. \tag{563}$$

It can be shown that the mapping $\frac{1}{2i}\hat{\sigma}_j \rightarrowtail R_j$ defines an isomorphism between both Lie algebras $su(2)$ and $so(3)$. A necessary condition for this isomorphism is that the number of generators, here three, is equal for both algebras.

Are there other isomorphisms between complex and real Lie algebras? This is an old classical question which has long ago been answered by Barut and Raczka[78] 1965. They provide a table of all possible unique correspondences. We have displayed these correspondences in Table 5. This is in fact a very

Table 5: Compatible groups and spaces

Real	compatible	Complex
$so(2)$	$\cong$	$u(1)$
$so(1,2)$	$\cong$	$su(1,1)$
$so(3)$	$\cong$	$su(2)$
$so(1,3)$	$\cong$	$sl(2,\mathbb{C})$
$so(4)$	$\cong$	$su(2) \otimes su(2)$
$so(2,4)$	$\cong$	$su(2,2)$
$so(6)$	$\cong$	$su(4)$

surprising result. Normally, one would expect that a compatible complex space exists for every real space. This is obviously not the case. Only 7 pairs exist.

In order to get a rough idea of the proof, let us count the number of rotations and their generators of some Lie groups and their Lie algebras. For $n = 2$ rotations keep fixed a point, and for $n = 3$ they keep fixed one axis, and the points in the plane perpendicular to this axis rotate. In general, rotations keep an $(n-2)$-dimensional subspace fixed, and therefore each rotation is in a plane.The *principle rotations* are defined by $x_i x_j$ planes, where the remaining $n-2$ variables are kept fix. All other rotations are products of the principle rotations. If we perform two rotations in different order, and the result is the same, then we say that they *commute*. Obviously, for $n = 2$ all rotations

[78] Barut et al. [1965]

commute (when keeping fixed the same point). For $n = 3$ the rotations around a fixed axis commute, but the principle rotations do not commute, because for instance the rotations in the xy plane and the yz plane share the common variable y. It is immediately clear that pairs of principle rotations commute if and only if they do not share a variable. Hence, for $n = 4$ we obtain six principle rotations in the planes xy, xz, xw, yz, yw, zw, but only two of them mutually commute, for example xy and zw. In the general case there are $n(n-1)/2$ planes $x_1x_2, ..., x_1x_n, x_2x_3, ...$ Thus the number of principle rotations in $\mathbb{R}^n$ is equal to $n(n-1)/2$. Counting the number of planes that do no share a variable, yields immediately the number of mutually commuting rotations, $n/2$ for n even and $(n-1)/2$ for n odd.

Similar as in the real case we can calculate the number of complex principle rotations in $SU(m)$. For $m = 1$ (see $U(1)$ above) we have one principle rotation that rotates the complex variable z, so varies the phase. For $m = 2$ we can rotate around both z_1 and z_2 varying their phases, and we can rotate in the z_1z_2 plane. The relative phase of the resulting two components can be changed after the rotation by another rotation. Summing up, we get $m^2 = 4$ rotations. In the general case, the unitary group has m principle rotations around the m components, $m(m-1)/2$ rotations in the planes, and the same number for changing the relative phases, yielding m^2 principle rotations. The m complex rotations that change the phase of a complex component commute, since changing the phase of one component does not affect another component. Rotations in planes do not commute with the rotation changing the corresponding relative phase. Hence, the largest commuting set of rotations is that of the m relative phase changes. Because in quantum mechanics the overall phases vanish when squaring the probability amplitudes, this rotation is irrelevant and must be neglected. Summarizing, we obtain for $m = 1$ exactly one phase rotation, and for $m > 1$ we get $m^2 - 1$ principle complex rotations, and a largest set of $m - 1$ commuting ones.

Therefore, for a real space of dimension n a complex space of dimension m is compatible, if $n = 2, m = 1$ and for $m > 1$ the two equations

$$n(n-1)/2 = m^2 - 1, \tag{564}$$

$$n/2 = m - 1 \text{ for } n \text{ even, and } (n-1)/2 = m - 1 \text{ for } n \text{ odd} \tag{565}$$

must be fulfilled. It is easy to check that the pairs $n = 3, m = 2$ and $n = 6, m = 4$ define compatible spaces. Solving these equations for m and substituting it into the first equation shows that no other pairs of compatible spaces exist.

We have given a sketch of the proof for three pairs of spaces. But four pairs in the Table 5 are missing. The reason is that we have not considered pseudo-Euclidean inner products. The complete proof is in fact much more complicated and requires a deep knowledge of group theory. However, our sketch of the proof gives at least some insight.

8 Appendix C: Keep in Mind

Keep in mind: Feynman's path integral theory implies Schrödinger's as well as Heisenberg's formalism. It forms the basis of quantum electrodynamics, our best physical theory.

Keep in mind: The field of complex numbers is the **largest commutative field** possessing the indispensable properties of numbers (like commutativity). This could be viewed as a basic reason that quantum mechanics, the most fundamental physical theory, uses complex numbers.

Keep in mind: Simple experiments with light, already performed by Malus in 1810, demonstrate: when accepting that light is built up of photons, the interaction of photons with optical elements is a stochastic process, and the probabilities of the random outcomes are squared magnitudes of numbers.

Keep in mind: However, it must be emphasized that the wave function [of amplitudes] that satisfies the equation is not like a real wave in space; one cannot picture any kind of reality to this wave as was done to a sound wave.

Feynman, Leighton, and Sands: *The Feynman Lectures on Physics*, Vol. 3, pp. 3–4

Keep in mind: A modern stochastic point of view of what **forces** are, is as follows: a force changes the action and thus phases, that is, the directions of probability amplitudes.

Keep in mind: Feynman's theory, a pure particle formalism without any waves, shows the close relationship between the deterministic classical mechanics, described by Hamilton's principle of least action, and quantum mechanics as a stochastic process with complex probability amplitudes. Both theories have in common that the particle prefers the path for which the action does not vary in the first approximation. The considered paths are not assumed to be smooth, they can exhibit a zig-zag curve as in Wiener processes.

Keep in mind: Nature seems to behave in accordance with the predictions of quantum mechanics, but seems to agree less with the theory of relativity.

Jacques et al.: *Experimental realization of Wheelers delayed-choice*, 2006

Keep in mind: Bell's Theorem excludes specific hidden variable theories and confirm quantum mechanics.

Keep in mind: These four rules (Born's rule, addition and multiplication rule, and action probability amplitudes) are the fundamental cornerstones of quantum electrodynamics, our best known physical theory that describes all physical phenomena, except gravitation and radioactive phenomena.

Feynman: *QED: The Strange Theory of Light and Matter*, Princeton University Press, 1985, page 8.

Keep in mind: Quantum parallelism is the ability to implement a simultaneous superposition of all function values of a Boolean function. It provides a potential exponential speed up in computing.

Keep in mind: Probabilistic states, classically random or complex quantum ones, cannot be cloned.

Keep in mind: For any physical theory we postulate the existence of mutually exclusive alternatives, the *base states*. A theory without base states is hardly imaginable.

Keep in mind: The mutually exclusive alternatives, the *base states*, can be described in terms of three mathematical equivalent representations. The *number representation* that is well-known from classical theories. The *register representation* that is frequently used in information theory. For certain systems this representation shows the divisibility of physical knowledge into decidable YES / NO questions. Finally, the *vector representation* that is used mainly in quantum theory. In particular, classical mechanics has also a vector representation, and thus can be viewed as a linear theory. In the same way quantum mechanics has a number representation, for instance, the componentwise Schrödinger equation. We make use of all three representations. A major reason that physical theories seem to be incompatible or inconsistent sometimes goes back to the ignorance of these equivalent representations.

Keep in mind: Given a machine or any experimental set-up. Its *base states* are the mutually exclusive alternatives. Two base states are different and distinguishable, if and only if

(i) in the number representation the corresponding numbers are not equal,

(ii) in the register representation the corresponding registers are not equal, and

(iii) in the vector representation the vectors are orthogonal.

Keep in mind:
Our experience of future, present, past, and irreversibility is deeply rooted in our behavior, our thinking, our language, our genetics. Physics can be formulated on the basis of this experience as a timeless language without apparent paradoxes, but with unified definitions of states, observables, and the change of states. A major goal of this lecture notes is to describe physics without an external time parameter in a simple manner from the very beginning, and thus supporting the *Wheeler-de Witt equation* derived formally from a Hamiltonian formulation of general relativity.

Keep in mind: Base states are non-overlapping, distinguishable alternatives that define the facts in the past, the outcomes or elementary events in the present, and the distinguishable possibilities of the future. They do not change when going from the future to the present and then to the past. This is important and is in contrast to states. They are defined as superpositions of base states and differ with respect to the trinity of time: in the past we have classical states, in the present we have classical probabilistic states, and in the future we have quantum states.

Keep in mind: The *superposition* principle in classical probability theory means that *either* this outcome *or* or another outcome happens with corresponding non-negative probabilities. The *superposition* principle in quantum theory means that this possibility *as well as* another possibility could happen in the future with corresponding complex probability amplitudes. The reason is that outcomes are distinguishable, but not possibilities, in general.

Keep in mind: The simplest way to remember a semimodule, is to take the rules of a linear space and to replace the word "field" by a "field without subtraction and division". This suggests that many properties of linear spaces and linear operators transfer without any difficulty to semimodules.

Keep in mind:

$$\left\{ \begin{array}{l} \text{Classical} \\ \text{Random} \\ \text{Quantum} \end{array} \right\} \text{ gates}$$

or transformations change states. They correspond to

$$\left\{ \begin{array}{l} \text{permutation} \\ \text{stochastic} \\ \text{unitary} \end{array} \right\} \text{ matrices}$$

with coefficients in the semiring

$$\left\{ \begin{array}{l} \mathbb{N} \\ \mathbb{R}_+ \\ \mathbb{C}, \end{array} \right\}$$

respectively. They are composed in

$$\left\{ \begin{array}{l} \text{series} \\ \text{parallel} \\ \text{controlled} \end{array} \right\}$$

by using the

$$\left\{ \begin{array}{l} \text{matrix product} \\ \text{tensor product} \\ \text{direct sum} \end{array} \right\},$$

respectively.

Keep in mind: The strange looking Principle 2 in the introduction, telling that "to every observable of a quantum system a Hermitian operator acting on a quantum state space is associated" turns out to be an old acquaintance from classical theories. Now we have a unified definition. What seems to be strange, can be easily explained with the transition from the number representation to the equivalent vector representation where functions become matrices, and the existence of various orthonormal bases in quantum mechanics, the theory of future events.

Keep in mind: The column vector $|\xi'\rangle$ of amplitudes representing some quantum state in the primed basis is obtained from the amplitudes of this state in the unprimed basis by matrix-vector multiplication with a unitary matrix, that is, $|\xi'\rangle = \hat{U}|\xi\rangle$. For a matrix a change of the basis is expressed as a similarity transformation, that is, $\hat{A}' = \hat{U}^\dagger \hat{A} \hat{U}$. The eigenvalues, the expectation value and the invariance of an observable $\hat{A}$ are invariant under a change of the orthonormal basis.

Keep in mind: The need of complex numbers is not only a necessity in quantum mechanics, it is a general feature when we try to describe geometrical properties of macroscopic objects.

Keep in mind: The mathematical framework describing polarization and spin is not restricted to small particles. It is a consequence when describing physics in terms of binary questions, namely bits. From this point of view we can say that a photon or an electron behaves like a big ball on a spherical surface, at least from the mathematical point of view.

Keep in mind: The phrase "one particle moves on a line" describes a fact in the past, an outcome in the present, or a possibility in the future. It will become clear whether we talk about future (quantum mechanics), present (classical probability theory), or past (classical mechanics). Therefore, we do not change a phrase like "one particle moves on a line" to "if, in a future action one particle would move on a line", although the latter is correct when talking about quantum mechanics.

Keep in mind: In physics there are three important types of operators: operators that transform states (permutation matrices, stochastic and unitary matrices), operators that label base states (observables), and operators that change base states (creation and annihilation operators).

Keep in mind: The (3+3)-position-velocity space $X \times V$is distinguished by a wonderful symmetry which reminds of Esher's "drawing hands". Moreover, the "spooky action at a distance" of two entangled photons in (3+1)-spacetime vanishes. The two photons are welded in the velocity space V. In other words, they can be connected via a velocity $v \in V$. Notice that the notion of distance depends on the space.

Keep in mind: There are good arguments to describe physical processes on the basis of a $3+3$-space. In particular, this space can be viewed as composed of position and velocity machines.

Keep in mind: The position-velocity space allows to define clocks as machines that provide a derived quantity, the time. A few simple arguments yield the mathematical framework of special relativity. This derivation does not require the usual postulates of special relativity, namely that (i) the laws of physics are the same in all inertial frames of reference, and that (ii) the speed of light in free space has the same value c in all inertial frames of reference.

Keep in mind: We have seen that the mathematical framework of quantum mechanics applies also to large objects. Now we have seen that the theory of relativity can be derived from simple transformation rules that are necessary from a mathematical point of view. It is not necessary, however, to postulate the existence of light itself or a maximal speed of light.
Time dilation and length contraction are not strange, they are simple consequences of our viewpoint that clocks are derived machines in a position- velocity space without external time parameter.

Keep in mind: Quantization is the process of constraining continuous quantities, such as real or complex numbers, to a discrete set. In canonical quantization the classical dynamical variables position and momentum in expressions are replaced by the Hermitian position and momentum operators. Due to the fact that operators don't commute, dynamical values are quantized. This replacement must be done carefully, since the *factor-ordering problem* emerges. Canonical quantization comprises the well-known quantization rules of Planck, Bohr, Sommerfeld and Wilson. The underlying reason of quantization is the fact that quantum mechanics has various orthonormal bases in contrast to classical mechanics and classical probabilistic mechanics.

Keep in mind: Replacing in the classical Hamiltonian the classical variables position and momentum by their corresponding Hermitian observables leads to the energy Hamilton operator. A purely algebraic method, based only on the commutation relations, shows that the energy values are discrete, and thus justifies the name *quantization*. Hence, whenever you see operators, figure out their commutators. The related base state functions do not depend on an external time parameter t, but have a probabilistic interpretation.

Index

References

A. Aspect, P. Grangier, and G. Roger. *Experimental realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: a new violation of Bell's inequalities*, Physical review letters **49**(2):91–94, 1982.

R. Bach, D. Pope, S.-H. Liou, and H. Batelaan. *Controlled double-slit electron diffraction*, New Journal of Physics **15**(3): 033018, 2013.

A.O. Barut, and R. Razka. *On Non-Compact Groups. I. Classification of Non-Compact Real Simple Lie Groups and Groups Containing the Lorentz Group*, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. Vol. 287. No. 1411. The Royal Society, 1965.

H.P. Beck-Bornholdt, and H.H. Dubben. *Der Hund der Eier legt*, Rohwohlt, 2004.

Belzasar. *Results of a double-slit-experiment performed by Dr. Tonomura showing the build-up of an interference pattern of single electrons. Numbers of electrons are 11 (a), 200 (b), 6000 (c), 40000 (d), 140000 (e).* `https://commons.wikimedia.org/wiki/File:Double-slit_experiment_results_Tanamura_2.jpg`, provided with kind permission of Dr. Tonomura, 2012.

T. Chivers. *Neuroscience, free will and determinism: 'I'm just a machine'*, The Telegraph, October 12, 2010.

A.J. Coleman, V.I. Yukalov. *Reduced Density Matrices*, Lecture Notes in Chemistry No. 72, Springer, 2000.

K. Conrad. *The Hurwitz theorem on sums of squares*, 2010. `http://www.math.uconn.edu/~kconrad/blurbs/linmultialg/hurwitzlinear.pdf`.

R.P. Crease. *The most beautiful experiment*, Phys. World **15**(9): 19–20, 2002.

P. Davies. *The Mind of God: The Scientific Basis for a Rational World*, Simon & Schuster, 1993.

Dhatfield. Diagram of Schrödinger's cat thought experiment. (Roughly based on Image: Schroedingerscat3.jpg), Wikimedia Commons, 2008, Permission (Reusing this file) CC-by-SA 3.0, `https://commons.wikimedia.org/wiki/File:Schrodingers_cat.svg`.

F. Dyson. *A meeting with Enrico Fermi*, Nature **427**: 297, 2004.

A. Einstein, B. Podolsky, and N. Rosen. *Can Quantum Mechanical Description of Physical Reality Be Considered Complete?*, Physical Review **47**: 777–780, 1935.

J. Erhart, S. Sponar, G. Sulyok, G. Badurek, M. Ozawa, and Y. Hasegawa. *Experimental demonstration of a universally valid error-disturbance uncertainty relation in spin measurements*, Nature Phys. **8**:185–189, 2012

R.P. Feynman, R.B. Leighton, and M. Sands. *The Feynman Lectures on Physics*, California Institute of Technology, 1963, Addison Wesley; Later Printing edition, 1971, `http://www.feynmanlectures.caltech.edu/`.

R.P. Feynman. *The Character of Physical Law*, 1965.

R.P. Feynman, *Simulating physics with computers*, International journal of theoretical physics **21**(6/7):467–488, 1982.

R.P. Feynman. *QED: The Strange Theory of Light and Matter*, Princeton University Press, 1985.

R.P. Feynman. *Quantum mechanical computers*, Foundations of physics **16**(6): 507–531, 1986.

M. Gell-Mann. *The quark and the jaguar: Adventures in the Simple and the Complex*, Henry Holt and Company, LLC, 1994.

M. Gessen. *Die faszinierende Geschichte des Mathematikers Grigori Perelman*, Suhrkamp, 2013.

J. Gleick. *Genius: The Life and Science of Richard Feynman*, Pantheon Books, NY, 1992.

W. Greiner. *Quantenmechanik*, Verlag Harri Deutsch, 2005.

D. Griffiths. *Introduction to Elementary Particles*, Wiley-VCH, 2004.

N. Herbert. *Quantum reality: Beyond the New Physics*, Doubleday, 1985.

V. Jacques, E. Wu, F. Grosshans, F. Treussart, P. Grangier, A. Aspect, and J.-F. Roch. *Experimental realization of Wheelers delayed-choice*, `http://arxiv.org/PS_cache/quant-ph/pdf/0610/0610241v1.pdf`, 2006.

G. Jaroszkiewicz. *Principles of discrete time mechanics*, Cambridge, 2014.

C. Jönsson. *Electron Diffraction at Multiple Slits*, Am. J. Phys. **42**(4):4–11, 1974.

P. Kaye, R. Laflamme, and M. Mosca. *An introduction to quantum computing*, Oxford University Press, 2007, `http://www.sergeev.fiz.univ.szczecin.pl/Dydaktyka/Wyklady/Kaye.pdf`.

P.S. Laplace. *A philosophical essay on probabilities* (trans. F.W. Truscott, and F.L. Emory), John Wiley & Sons, New York and Chapman and Hall, Ltd., London, 1902.

R.B. Laughlin. *A Different Universe: Reinventing Physics from the Bottom Down*, Basic Books, NY, 2005.

H. Lüth. *Quantum Physics in the Nanoworld*, Springer Int., 2015.

J. Mayer, Khairy, and Howard. *Drawing an elephant with four complex parameters*, American Journal of Physics 78.6 (2010): 648-649, 2010.

R. Mirman. *Quantum Mechanics, Quantum Field Theory: geometry, language, logic*, Nova Science Publishers, Inc., Huntington, NY, 2001.

R. Mirman. *Our Almost Impossible Universe: Why the Laws of Nature Make the Existence of Humans Extraordinarily Unlikely*, iUniverse, Inc., 2006.

M.A. Nielsen, and I.L. Chuang. *Quantum computation and quantum information*, Cambridge University Press, 2010, `http://www.johnboccio.com/research/quantum/notes/QC10th.pdf`.

M. Noveas . *Some basics of su (1, 1)*, Revista Brasileira de Ensino de F´isica 26.4: 351-357, 2004.

H. Ohanian. *Gravitation and Spacetime*, W.W. Norton and Company, 1994.

R. Omnes. *The interpretation of quantum mechanics*, Princeton University Press, 1994.

R. Omnes. *Quantum philosophy — understanding and interpreting contemporary science*, Princeton University Press, 1999.

A. Pais. *Subtle is the Lord: The Science and the Life of Albert Einstein*, Oxford University Press, 2005.

R. Penrose. *The road to reality: A Complete Guide to the Laws of the Universe*, Vintage, 2005.

R. Penrose. *Cycles of time: An extraordinary new view of the universe*, Bodley Head, UK, 2010.

A. Peres. *Quantum theory: concepts and methods*, Springer, 1995.

M. Plenio. *Quantum Mechanics*, Imperial College, 2002. `http://www3.imperial.ac.uk/pls/portallive/docs/1/613904.PDF`.

M. Renninger. *Messungen ohne Störung des Messobjekts (Measurement without disturbance of the measured objects)*, Zeitschrift für Physik **158**(4): 417–421, 1960.

T. Rothmann. *Die Physik — Ein baufälliger Turm von Babel*, `http://www.spektrum.de/news/die-physik-ein-baufaelliger-turm-von-babel/1139123`, 2012.

L. I. Schiff. *Quantum mechanics*, 3rd edition, McGraw-Hill, 1968.

C. Schiller. *Motion Mountain*, www.motionmountain.net, 2016.

Ch. Schirm. Schrödinger's Cat, many worlds interpretation, with universe branching. Visualization of the separation of the universe due to two superposed and entangled quantum mechanical states, Wikimedia Commons, 2011, Permission the Creative Commons CC0 1.0 Universal Public Domain Dedication, `https://commons.wikimedia.org/wiki/File:Schroeingers_cat_film.svg`.

F. Sellerie. *Quantum paradoxes and physical reality*, Springer Science & Business Media, 1990.

R. Sheldrake. *The Science Delusion*. Hodder and Stoughton Ltd., 2013.

L. Smolin. *The trouble with physics: The rise of String Theory, the fall of science and what comes next*, Houghton Mifflin Harcourt, 2006.

L. Smolin. *Time reborn: From the crisis in physics to the future of the universe*, Houghton Mifflin Harcourt, 2013.

F. Strocchi. *Complex Coordinates and Quantum Mechanics*, Rev. Mod. Phys. **38**:36, 1966.

L. Susskind, and G. Hrabovsky. *The Theoretical Minimum: What you Need to Know to Start Doing Physics*, Basic Books, 2013.

L. Susskind, and A. Friedman. *Quantum Mechanics: The Theoretical Minimum*, Basic Books, 2014.

Y. Tan. *Basis in semimodules over commutative rings*, LAA 443, 139-152, 2014.

Y. Tan. *Inner products on semimodules over a commutative semiring*, LAA 460, 152-173, 2014.

M. Tegmark. *Shut up and calculate*, `http://arxiv.org/pdf/0709.4024v1.pdf`, 2007.

M. Teller. *An Interpretative Introduction to Quantum Field Theory*, `PrincetonUniversityPress`, 1997.

A. Unzicker. *Vom urknall zum Durchknall. Die absurde Jagd nach der Weltformel*, Springer Berlin/Heidelberg, 2010.

A. Unzicker. *Auf dem Holzweg durchs Universum. Warum sich die Physik verlaufen hat*, Carl Hanser Verlag, München, 2012.

A. Unzicker. *The Higgs Fake. How particle physicists fooled the Novel committee*, CreateSpace Independent Publishing Platform, 2013.

A. Vazsonyi. *Which door has the Cadillac*, Decision Line **30**(1):17–19, 1999.

S. Weinberg. *Lectures on Quantum Mechanics*, Cambridge University Press, 2013.

C.F. Weizsäcker. *Aufbau der Physik*, Carl Hanser Verlag, Munich, 1988.

C.F. von Weizsäcker. *Zeit und Wissen*, Carl Hanser Verlag, 1992.

C.F. Weizsäcker. *The Structure of Physics (Original 1985)*, T. Grönitz, and H. Lyre (eds.), Springer Netherlands, 2006.

J.A. Wheeler. *The "Past" and the "Delayed-Choice" Double Slit Experiment*, In A.R. Marlow, editor, Mathematical Foundations of Quantum Mechanics, Academic Press, New York, 1978.

J.A. Wheeler. *Law without Law*, In J.A. Wheeler, and W.H. Zurek, editors, Quantum Theory and Measurement, Princeton Series in Physics, pages 182–213, Princeton University Press, 1983.

W.K. Wootters, and W.H. Zurek. *A single quantum cannot be cloned*, Nature **299**(5886): 802–803, 1983.

E. Zaslow. Mirror Symmetry, in *Princeton Companion to Mathematics*, T. Gowers, ed., Princeton University Press, 2008.

E. Zeidler. *Quantum Field Theory I*, Springer, 2006.

A. Zeilinger. *Quanten Teleportation*, Spektrum der Wissenschaft, 2000.

A. Zeilinger. *Quantum Teleportation*, Scientific American, 2000.

N. Zetteli. *Quantum mechanics: Concepts and applications*, Wiley, 2009.