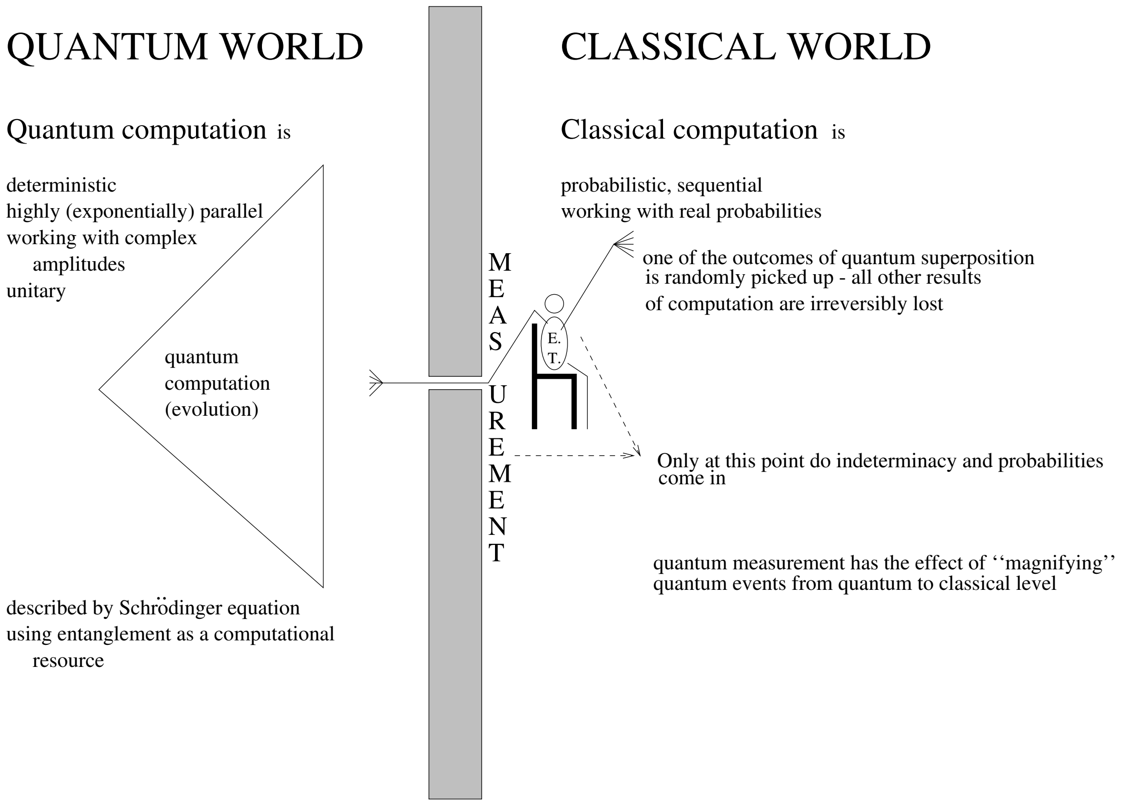
QUANTUM COMPUTING

Jozef Gruska



Appendix

> Our best theories are not only truer than common sense, they make far more sense than common sense does.
>
> David Deutsch, 1997, *The fabric of reality*

The first section of this Appendix is devoted to quantum theory. It contains an informal, often popular, overview and discussion of several basic issues of quantum physics. It is written mainly for those in computing with (almost) no knowledge of the subject. Exposition is therefore necessarily without many details needed if one wants to be (fully) precise. For more the reader is referred, for example, to Peres (1993), Bub (1997), Jammer (1966), Penrose (1990, 1994) and Lindley (1996)—ranging these references, roughly, from more technical to more popular. They mainly influenced presentation of the section.

Section 9.2 presents some basic concepts and results of Hilbert space theory in more detail than in Section 1.4 and it contains additional subjects.

The third part of the Appendix, on the book web pages only, contains in Section 9.3 a survey of the basic concepts, models and results of the complexity theory. This part is oriented mainly towards people outside of computing with (almost) no knowledge of the computation and complexity theory. Section 9.4 contains additional exercises and Section 9.5 additional historical and bibliographical references.

9.1 Quantum Theory

> Physical concepts are free creations of the human mind, and are not, however it may seem, uniquely determined by the external world.
>
> Albert Einstein, *Evolution of physics* (1938)

There are two basic views of the goals a natural science has to meet: (a) to predict—to provide results that allow us to predict behaviours and outcomes of natural processes; (b) to explain—to provide explanations of the corresponding basic phenomena of Nature and to help us to get an understanding of these phenomena.

Quantum mechanics superbly fulfils its prediction role. Concerning its capability to explain phenomena of the quantum world, at least in the sense most of us would like, the situation seems to be quite different.

Quantum theory works perfectly in all practical applications and it describes with unprecedented precision many phenomena of Nature. Predictions made on its basis have enormous value and have been tested to about 14 orders of magnitude. No conflict between its predictions and experiments is known. Quantum theory is at present the best mathematical model to describe the physical world.

Without quantum theory we do not know, for example, to explain the properties of superfluids, the functioning of lasers, the nature of chemistry, the very existence and behaviour of solids, the colours of stars, as well as atomic and nuclear phenomena.

On the other hand, quantum theory is often seen as forcing us to accept, as the best we currently have, philosophically highly unsatisfying views of the world that do not square with our common sense perception of the functioning of the Universe.[6]

The basic reasons for arriving at a quantum theory view of the world were experimental. Quantum theory arose out of the observations of subtle discrepancies between the outcomes of some experiments and predictions classical physics offered.

For many of the issues we discuss there is no unique understanding and it is therefore inevitable that the point of view presented here is not the only one possible.

9.1.1 Pre-history of quantum theory

> Every tool carries with it the spirit by which it has been created.
>
> W. Heisenberg (1958)

Classical physics describes Nature nicely and fully in accordance with our common sense. In classical physics reality exists independently of ourselves. Our bodies and brains are themselves parts of the classical world. An "objective physical reality" seems to correspond to all concepts of classical physics. In particular, particles of classical physics have always position and momentum that can be (statistically) determined.

One of the basic reasons why classical physicss got into difficulties at the end of the last century was the fact that it was not able to cope with the coexistence of two types of physical objects: particles and fields. For a system with both particles and fields to be in equilibrium, all energy gets taken from particles into fields. Since fields have infinitely many degrees of freedom, particles are left without energy.

The question of whether light is a wave process or has a particle character goes back to Newton. The celebrated Thomas Young's two-slit experiment, performed in 1801, demonstrated clearly the wave character and the interference of polarized light and pointed out strongly the particle–wave dichotomy. However, this experiment did not get attention it merited. It was not yet time to handle the problems it brought up.

A famous, historically important, and influential example of the instability of the coexistence of fields and particles was the "black-body radiation problem". Imagine an enclosed empty box at some fixed temperature. Electromagnetic radiation of the object should be in some equilibrium with particles. However, if there were more energy in the walls of the box than in the enclosed radiation, then energy would move from the walls to the interior,

[6] By Penrose: "Quantum theory seems to lead to philosophical standpoints that many find deeply unsatisfying. At best, and taking its descriptions at their most literal, it provides us with a very strange view of the world indeed. At worst, and taking literally the proclamations of some of its most famous protagonists, it provides us with no view of the world at all."

increasing the density of radiation within the box. If there were more energy in the electromagnetic radiation than in the walls, then it would heat up the walls to restore equality. Classical theory predicted that all energy would be sucked up by the field. However, the experiments did not confirm it. At the high frequencies, where classical physics predicted strong discrepancies—a rapid increase of energy—the energy not only did not go up so much, but actually dropped out. To summarize, classical physics could not find out how to define the electromagnetic radiation that would be contained in the box of a constant temperature.

In 1900 Max Planck[7] came with a revolutionary theory that electromagnetic oscillations could carry energy only in "quantas", the energy E of which satisfies the relation

$$E = h\nu$$

to the frequency ν, where h is known today as Planck constant.[8] With this idea the black-body radiation problem could be dealt with.

In spite of Planck's ability to come up with a new theory in accordance with the experimental results, as in the case of the black body radiation problem, his theory did not receive too much attention until the next step occurred. Einstein came up with the theory that electromagnetic field also can exist only in discrete units and on this basis he was able to explain the photoelectric effect.[9] This implied, for example, that light itself must actually be particles, because it was known, due to the results of Maxwell, that light consists of the oscillations of the electromagnetic fields. However, it was also established experimentally that light sometimes behaves as waves. -

The task was then to explain how it can happen that light consists of particles and field oscillations at the same time—some experiments showed light as particles and some as waves.

Niels Bohr[10] made an important use of the Planck relation in 1913. He discovered that the **angular momentum—spin**—of electrons in orbit about the nucleus can occur only in integer multipies of the number $\frac{h}{2\pi}$, for which Dirac introduced the symbol $\hbar$. The only permissible values of the spin of electrons are therefore

$$0, \hbar, 2\hbar, 3\hbar, 4\hbar, \ldots$$

With this approach Bohr was able to put the "solar model" of atoms on a more firm base and to explain many different states energy takes and also special rules for spectral frequencies.

[7]Max Plank (1858–1947), a German physicist. The quantum theory was developed from his hypothesis that atoms emit energy only in discrete bundles (quanta). He received the 1918 Nobel prize for physics—for his work on black-body radiation that depended on his hypothesis.

[8]Planck's ingenious solution was considered controversial for some time. Planck himself did not want to believe that electromagnetic radiation was fundamentally structured in this new way. He hoped to find some overlooked features of classical physics that would be able to explain why waves had to carry energy in discrete quantities. Planck never contributed a genuine physical reality to those "little bundles of electromagnetic energy". He saw them as mathematical constructs hiding some physical principles (LIndley, 1996).

[9]It was formally for this discovery Einstein received the Nobel prize.

[10]Niels Henrik David Bohr (1885–1962), a Danish physicist, one of the best physicists of the twentieth century. He helped to found and then directed the Institute for Theoretical Physics at the University of Copenhagen. He was the main representative of the influential Copenhagen school of quantum physics. Bohr received the 1922 Nobel prize for physics for his work on atomic structures. He combined quantum theory with his new concept of atomic structure. Bohr explained the stability of the nuclear model of atoms by postulating that electrons move on restricted orbits around the atoms nucleus and explaining how atoms emit and absorb energy.

Quantum theory, as known to Bohr's school, is nowadays known as **old quantum theory**. The foundations of **modern quantum theory** were laid down between 1923 and 1927.

In 1923, Louis de Broglie proposed that anything which oscillates with some frequency ν can occur only in discrete units of mass $\frac{h\nu}{c^2}$ (which came from a combination of the Planck identity $E = h\nu$ with Einstein equation $E = mc^2$).[11] This implies that particles sometimes behave as waves and sometimes as fields and that particles and field oscillations are the same things. De Broglie stated the hypothesis that the wave–particle duality is a general property of microscopic objects. This brought an elegant and unified view of Nature but it could not be considered as a new consistent theory yet.

Crucial steps that gave rise to the new quantum theory were done by Heisenberg, Schrödinger and Dirac. In 1925 W. Heisenberg initiated development of so-called "matrix mechanics" in which energy levels were eigenvalues of infinite matrices. In 1926 E. Schrödinger worked out so-called "wave mechanics", suggested by de Broglie, in which energy levels were eigenvalues of differential operators and a matrix was associated with each physical quality. This was a simplification which was shown, by Schrödinger, to be equivalent to the wave mechanics. The new quantum theory became known as **quantum mechanics** and was developed rapidly. Very significant contributions were made: by Dirac—by his generalization of quantum mechanics—and by John von Neumann who also introduced Hilbert space formalism.

Another important discovery was Heisenberg's uncertainty principle, discussed in Section 9.1.2. As a consequence it became clear that the old concept of particles having at each moment a well determined position and momentum was misleading. The naive classical view was replaced by the view that to each particle a quantum state (or a wave function) corresponds and that such properties of the particle as position, momentum and spin are undetermined (do not have a specific value), until they are measured.

However, the main departure from classical physics was the discovery of the intrinsic stochastic nature of quantum processes, which is usually less formally and precisely interpreted as "intrinsic irreproducibility of the experimental results". Determinism, on which classical physics built its philosophy, was shown not to be valid in the quantum world.

Remark 9.1.1 It seems to be not only of historical interest, but also of importance for the future to look into the question why it took so long before quantum computing issues started to be investigated with sufficient vigour, because it is now clear that already 40 years ago it was quite natural to start to do that, both for computer science and for physics.

Computer science, and actually even the most theoretical parts of it, has been developed, and is still developed, basically as a servant of the computer industry and, consequently, with quite restricted and short-term goals. This is why it actually ignored in a sense its most basic scientific goal—to explore fully the potentials and limitations of computing and communication based on the laws of physics[12]. Theoretical computer science, especially complexity theory, found interest and importance to work with computer models far outside the framework of current classical computers, but these models, such as alternating Turing machines, were considered only as tools. Attempts to see the main scientific aims of fundamental computer science (informatics), in a broader context, as in Gruska (1993), have been and still are rare.[13]

[11] In this way Planck's constant, introduced to explain black-body radiation, was found relevant also to the mechanical properties of atoms.

[12] In spite of the fact that Landauer had already, in the early sixties pointed out such a need.

[13] In Gruska (1993), computer science as science is seen as having similar scientific aims as physics and

On the other hand, physics has only very slowly developed an understanding that information is a physically important phenomenon, concept, and resource[14]. In addition, physics practically ignored an important fact that within computer science in general, and in complexity theory in particular, important new concepts, methods, results and insights have been developed that can be useful also for extending our understanding of the physical world. It is also worth noticing that one could hardly expect some deeper developments in quantum computing before complexity theory made significant advances and found ways to classify computational tasks in a deep way, and before such of its modern branches as randomized complexity theory were developed.

9.1.2 Heisenberg's uncertainty principle

> In effect, we have redefined the task of science to be the discovery of laws that will enable us to predict events up to the limits set by the uncertainty principle.
>
> Stephen Hawking: *A brief history of time* (1988)

In 1927 Heisenberg[15] discovered one of the fundamental principles and limitations of quantum mechanics: a bound on the accuracies with which certain pairs of observables are measured. The principle says that measuring the value of one observable more accurately makes the value of another, noncommutative, observable less certain. In addition, there is a certain intrinsic uncertainty with which values of two observables can be measured, and in general once a way of measurement is fixed this uncertainty, in general, increases.

For example, in Stern–Gerlach measurement (see Section 1.5.3), electrons emerging in a definite up-or-down state are in an undetermined half-right, half-left state and vice-versa.

There is a simple form of Heisenberg's principle. The standard deviation of A, ΔA_ψ, when measuring $|\psi\rangle$, is defined by

$$\Delta A_\psi = \sqrt{\langle A^2\rangle_\psi - (\langle A\rangle_\psi)^2}$$

and quantifies the uncertainty with which the values of the observable are given. (In the case of spectral representation of A, $\Delta A_\psi = 0$ if and only if ψ is the eigenvector of A.)

Heisenberg's principle claims that the following lower bound on the uncertainties holds when two observables are measured independently.[16]

$$\Delta A_\psi \cdot \Delta B_\psi \geq \frac{1}{2}|\langle AB - BA\rangle_\psi|$$

similar methodological impacts as mathematics. Its main scientific aims are to study the laws and limitations of the information processing world—whatever it is.

[14]There have been attempts to do so that can be traced back to Maxwell's demon paradox, but these views have not been in the mainstream of physics development.

[15]Werner K. Heisenberg (1901–1976), a German physicist, one of the founders of quantum theory. His main contribution was the development of the matrix mechanics theory and the discovery of the uncertainty principle. In addition, he tried to formulate a theory of elementary particles such that all elementary particles would come as solutions of one field equations. He also contributed to the theory of ferromagnetism, structure of atoms, cosmic rays and field theory. Heisenberg received the 1932 Nobel prize for physics for his work on the development of quantum theory.

[16]For a derivation of Heisenberg's uncertainty principle see, for example, Peres (1993).

Therefore, the only case where two observables A and B can be measured simultaneously (independently) is precisely when they commute.

Many instances of Heisenberg's principle are well known. One of them, and historically the first one, considered already by Heisenberg, concerns the simultaneous measurement of the position and the momentum of particles and it is called Heisenberg's microscope. Classically (see Figure 9.8a), you can measure both, position and momentum. Quantum mechanically, if you measure the position precisely (Figure 9.8b), then for various possible

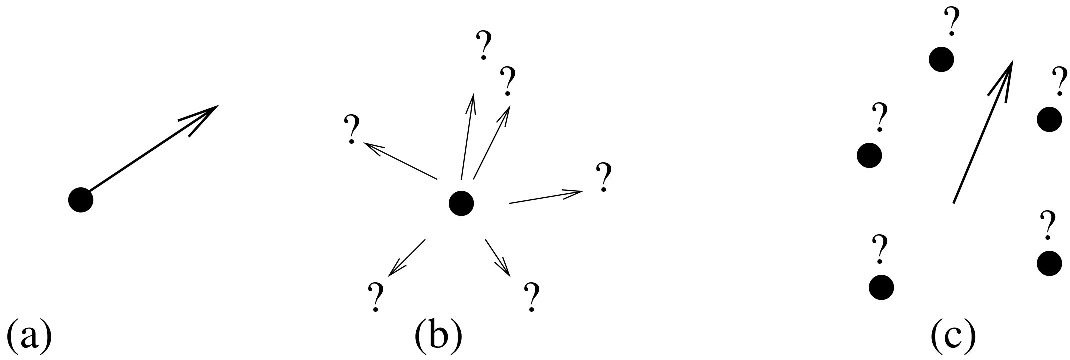


Figure 9.8: Measurement of the position and of the momentum of particles

values of the momentum only probabilities are given. Similarly, if you measure precisely momentum first (Figure 9.8c), then there are many options for positions and for each of them only the probability is known. More exactly, in this case it is assumed that two conjugate observables, the position and momentum, are measured simultaneously, which results in the fact that precision in the measurement of position is obtained at the expense of precision of the measurement of momentum.

For example, one can use a stream of photons to measure an electron's position and momentum. Using an energetic photon, with short wavelength, one can get quite a good idea where the electron is but one then has little idea about its momentum. (The problem is that short-wavelength light implies a large-momentum kick to the electron. On the other hand, a soft collision, with long-wavelength light, provides a poor idea of the electron's position but a good idea of its momentum.)

Another example of uncertainty principle was already demonstrated in connection with the two-slit experiment. Either one can detect through which slit an electron went or one can detect the interference pattern. Detecting through which slit an electron went is a particle measurement; recording the interference pattern is a wave measurement. One can do any of them, but not both in the same experiment.

The following strong form of uncertainty principle (Bennett 1998b) is also of interest for quantum information processing: evolution of a quantum system remains deterministic only if no information about it leaks out into the environment.

The uncertainty principle is a part of Nature and not a consequence of our technological limitations.

In the following example, of importance for quantum key generation, an instance of Heisenberg's uncertainty principle is discussed in detail.

Example 9.1.2 (Uncertainty principle for polarized photons) *Photons*[17] *are elec-*

[17]Photon, or light quantum, is a particle composing light and other forms of electromagnetic radiation.

tromagnetic waves and their electric and magnetic fields are perpendicular to the direction of propagation and also to each other. An important property of photons is polarization—it refers to the bias of the electric field in the electromagnetic field of the photon. If the electric field vector is always parallel to a fixed line we have **linear polarization** *(see Figure 9.9). If the electric field vector rotates about the direction of propagation forming a (right-) left-handed screw, we have (right) left* **eliptic polarization**. *If the rotating electric field vector inscribes a circle, we have (right) left* **circular polarization**.

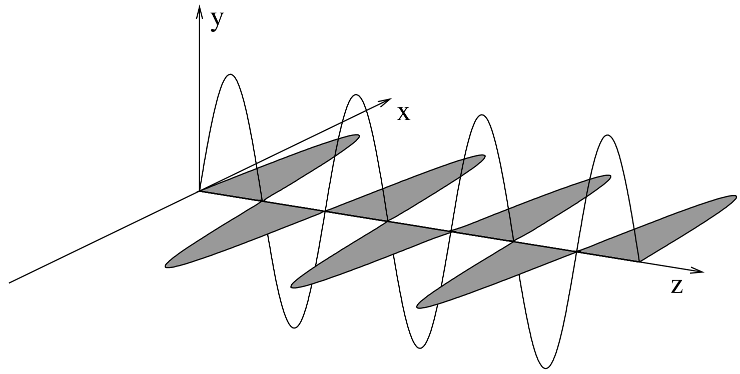


Figure 9.9: Electric and magnetic fields of a linearly polarized photon

Polarization[18] *is a property of photons with which one can demonstrate well what can and cannot be done in the quantum world. In addition, polarized photons are key transmission elements in quantum cryptography.*

Polarization of photons. *There is no way to determine exactly polarization of a single photon. However, for any angle θ there are θ-***polarizers***—"filters"—that always produce θ-polarized photons from an incoming stream of photons. Moreover, they let θ-polarized photons go through and θ_1-polarized photons get through with a probability of* $\cos^2(\theta - \theta_1)$. *In other words, in order to create a photon whose electric field is oscillating in the required plane one can use a polarizer whose polarization axis is set up at the desired angle. More exactly, if the axis of the polarizer makes an angle θ with the plane of the electric field of the photon entering the polarizer, then there is a probability* $\cos^2\theta$ *that the photon will emerge with its polarizer set at the desired angle and the remaining probability,* $\sin^2\theta$, *that it will not be observed.*

Photons whose electronic fields oscillate in a plane at either $0°$ *or* $90°$ *to some reference line are called usually* **rectilinearly polarized** *and those whose electric field oscillates in a plane at* $45°$ *or* $135°$ *as* **diagonally polarized**. *Polarizers that produce only vertically or horizontally polarized photons are depicted in Figure 9.10a,b.*

Generation of orthogonally polarized photons. *For any two orthogonal polarizations (that differ by 90°) there are generators that produce photons of two given orthogonal polarizations. For example, a calcite crystal, properly oriented, can do the job. Figure 9.10c shows a calcite crystal that causes θ-polarized photons to be either horizontally polarized, with probability* $\cos^2\theta$, *or vertically polarized, with probability* $\sin^2\theta$.

[18] Polarization of photons in relation to quantum cryptography is discussed in more detail in book by Williams and Clearwater (1997) and in survey by Lomonaco (1998).

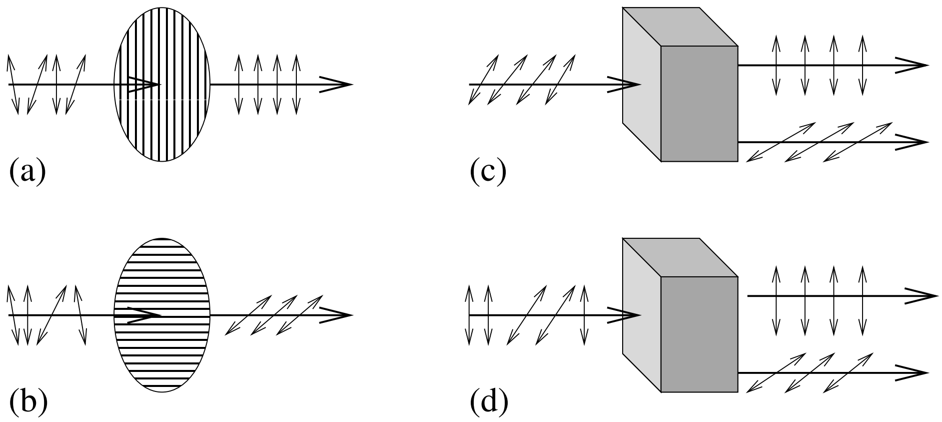


Figure 9.10: Photon polarizers and measuring devices-80%

Exercise 9.1.3 *A stream of randomly polarized photons goes subsequently through a vertical, diagonal and horizontal polarizer. How many of them get through?*

Measuring of polarization. *A calcite crystal, which has the properties of birefringence, can be used to separate horizontally and vertically polarized photons (see Figure 9.10d). Another crystal can be used to distinguish two polarizations for diagonally polarized photons. Can we measure both of them at the same time?*

In the case of the rectilinear basis the observable is the operator represented by the matrix

$$A_r = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{9.9}$$

with two eigenvalues 1 *and* -1 *and two eigenstates,* $|0\rangle$ *and* $|1\rangle$*. The same is true if in the diagonal basis we have the observable* A_d *represented, in that basis, again by the matrix 9.9. (Observe that the matrices* A_r *and* A_d *are the same but they refer to a different basis!) In order to find out whether those two observables commute we have to express both in the same basis. The diagonal basis can be rotated into the rectilinear by the rotation*

$$\begin{pmatrix} \cos -\frac{\pi}{4} & -\sin -\frac{\pi}{4} \\ \sin -\frac{\pi}{4} & \cos -\frac{\pi}{4} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

The observable for diagonal polarization in the rectilinear basis is therefore,

$$A'_d = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Since

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

these two observables do not commute.

9.1.3 Quantum theory versus physical reality

God is subtle, but malicious he is not.

A. Einstein

Quantum theory, like as other theories of Nature, is based on observations and experimental results. Its mathematical abstraction culminated in a theory developed within axiomatically defined Hilbert spaces. It is a theory that aims to provide mathematical abstractions of physical concepts, observations and experimental results, and physical interpretation of mathematical concepts, models, methods and results.

The first part of the aim was achieved well enough. However, from the earliest days of quantum mechanics it was not clear what is the real physical interpretation of quantum mechanical mathematical concepts. The idea of a unique and in the limit infinitely knowable reality of the classical world appeared to be a fiction, but no clear reality for the quantum world has emerged yet. It is also clear that to many natural language concepts which are used to describe classical world phenomena one needs to assign different meanings when using them for describing quantum phenomena. The problem with finding a "real world" interpretation of quantum theory concepts is so severe that the quantum world is seen by some as only an abstract concept.

It is often said that mathematical formalism of quantum theory provides only concepts, results and methods that are superb for developing algorithms to compute probabilities of experimental results. By Peres (1993), "in a strict sense quantum theory is a set of rules allowing the computation of probabilities for the outcomes of tests which follow specific preparations".

For many quantum physicists this seems to be all they need, and they can be quite happy with the current state of the theory. They have not encountered any difficulties with the Copenhagen interpretation (discussed in Section 9.1.7).

On the other hand, it is not clear how important for quantum information processing is the fact that the current state of quantum theory is clearly not fully satisfactory. Could it happen that some other interpretations of quantum theory would significantly change the merit of the current key results concerning quantum information processing, especially concerning the extraordinary computational power of quantum computers (and/or their limitations)?

There are several reasons why current quantum theory is considered by some schools of thought in physics and in the philosophy of science not only as being far from complete but actually as a theory no one fully understands, and a theory that is not able to attach in a fully satisfactory way a definite physical meaning to its mathematical concepts.[19] For example

1. Quantum theory does not seem (to all) to provide a clear notion of what the reality of quantum concepts could be. It provides no fully satisfactory and contradiction-free physical interpretation of theoretical concepts. (However, there can be the following, well-founded objections to this strong statement: What is a reality? Is there a (the)

[19] On the other hand, those standing firmly on the Copenhagen interpretation consider many of these problems as outside quantum mechanics. Moreover, new approaches to such basic and controversial problem as quantum measurement, see Bush et al. (1997), bring new ideas how to deal with some of the open problems of quantum mechanics.

reality?) Mathematical concepts of quantum theory are even considered by some as having no physical meaning at all, and as not being intended to have such a meaning.

2. The existence of non-locality and non-local interactions[20] as pointed out by EPR Gedanken experiment discussed in Section 9.1.5.

Of course, there are attempts to assign physical reality to such concepts as quantum state, quantum systems and quantum measurement. However, they lead to hard-to-accept mysteries and so-called paradoxes.

For that reason a quantum state is seen by some (see Peres, 1993), as having no direct physical meaning and by others (for example von Neumann) as being a complete description of reality.

At the level of the mathematical formalism of Hilbert spaces there are no principal problems with a pragmatic understanding of such concepts as state, observable, entanglement and measurement. However, the attempts to derive these theoretical concepts and principles only from the physical reality and to assign them physical meaning have not worked well. To derive such models as Hilbert space, physical abstractions and reasoning seem to be hardly sufficient. One needs to use principles of abstraction, logic and aesthetic mathematical thinking to derive such models and to utilize by that the richness of mathematical concepts and experiences (for example, to come up with the requirement of completeness for Hilbert spaces).

The existing difficulties with interpretations of quantum theory concepts are also often seen on epistemology versus ontology level. There are views that existing quantum theory is concerned only with our knowledge of reality, or, in other words, that it is directed primarily on epistemology (the studies that focus on questions of how to obtain our knowledge and how to make use of it) and much less on ontology.

9.1.4 Quantum measurements

> No elementary phenomenon is a real phenomenon until it is a measured phenomenon.
>
> *John Wheeler*

The concept of a classical measurement or observation is one of the oldest on which science is based. By a classical measurement, observation, or test, we acquire knowledge about the

[20]The term "non-local interaction" is by itself intriguing. One way to understand it is that a local interaction is one that either exhibits a direct contact, or at least employs an intermediary that is in direct contact. The second part of this interpretation allows local interactions between objects astronomically far apart—see for example gravity, which is considered as a local interaction because it is assumed to be mediated by hypothetical quanta, gravitons, which travel between gravitating objects. It is usually assumed that local interactions meet the following criteria: (1) They are mediated by other entities, such as particles or fields; (2) they do not propagate faster than light; (3) their strength drops off with the distance. It is known that all forces of the universe create local interactions. In such a case it is natural to ask where in the physics we could find something that would allow non-local interactions. It seems that the measurement postulate of quantum mechanics, with a force-free collapse of states provides, such a loophole on which the existence of non-local interactions could be justified. The distaste in a large part of the scientific circle of any theory assuming non-local interactions is based on the assumption that it would contradict Einstein's special theory of relativity, which says that nothing can travel faster than the speed of light.

reality. During the measurement there is an interaction between the measuring devices and the measured objects.

Quantum measurements are, on the other hand, very different and in many respects revolutionary. As already discussed in Section 1.4, quantum measurement is perhaps the most controversial issue of quantum theory. Peculiarities and controversial aspects of quantum measurement are numerous and in spite of the fact that several books have been written on this subject no really essential progress seems to be made—till the last two decades. Let us now summarize some of the issues and peculiarities of the quantum measurement problem.

1. In the term *quantum measurement* the noun 'measurement' has a meaning very different from the one used in the classical world. As already mentioned, before quantum mechanics it was taken for granted that when we measure something, we are gaining knowledge of a pre-existing state—of an independent fact about the world. However, quantum mechanics, at least its standard interpretation, says otherwise: some things are not determined except when they are measured—it is only by being measured that they take on specific values. If we therefore attribute to the word "measurement" its ordinary meaning, i.e. the acquaintance of knowledge about some pre-existing objective reality, we reach various contradictions.

2. Some classical "measurement tasks" are not appropriate in the quantum setting because there they are not well defined. For example, in some cases it is not meaningful to pose the question "what is the value of the property P of the object O", but it is meaningful to ask whether a particular x is the value of P for the object O.

Example 9.1.4 *(Peres, 1993) There is a way to produce photons with various polarizations, but there is no way to measure polarization of a particular photon.*[21] *The question "What is the polarization of that photon?" cannot be answered and is considered in quantum physics as having no meaning. The legitimate question that can be answered experimentally is whether or not a particular photon has a specified polarization. As a consequence, if Alice prepares a sequence of photons and sends them to Bob without disclosing their polarization, then there is no instrument whatsoever to sort them into bins for polarization* $0-30°$, $30-60° \ldots$ *in the way that agrees with the polarization as produced by Alice. This fact is actually used in Section 6.2 to make a secure system for cryptographical key generation.*

3. It is not always possible to obtain by measurements full information about the unknown state of a quantum object. In addition, the unknown states of quantum objects are considered as having no definite value except when they are measured.[22]

4. There have been attempts to consider quantum measurement as the very basic concept of quantum physics (see Peres, 1993), and to use it to derive from it such concepts as quantum state. By Peres, a measurement consists of a *preparation*, a *test* and a *selection*, where preparations and tests are even more fundamental concepts. A quantum state is then defined as an equivalence class of preparations.[23]

5. Quantum measurement can also be seen as an irreversible addition to otherwise fundamentally reversible quantum evolution.

6. Measurements are considered as the key tools of science to get information that is then used to abstract theories and to get knowledge. One can say that natural sciences are based on observation of Nature. It is believed by most scientists that measurements in general

[21] However, we can estimate the polarization with certain fidelity.

[22] A different situation is, for example, when it is known how was a quantum object produced, or when there is a possibility of producing an unlimited number of copies of the same object.

[23] It is a good and fundamental question what kind of equivalence to consider for preparations.

help to get knowledge about objective reality that is not fully known to us. However, the mysteries and paradoxes quantum theory runs into make some believe that quantum theory is incompatible with the claim that quantum measurements can discover some unknown but pre-existing reality. By Peres (1993), "We have no satisfactory reason for ascribing objective existence to physical quantities as distinguished from the numbers obtained when we make measurements we correlate with them."

7. The problem of measurement concerns also the basic dichotomy between the classical and the quantum world. Indeed, from a broader point of view also the measuring device, or at least a part of it, can be considered as a quantum system. This fictitious process of shifting the microscopic level can be repeated. Some see as the last stage of this chain our consciousness.

8. Several types of measurements are of special importance. For example:

- **Repeatable measurements.** These are measurements M such that if M is applied to the result x obtained by M, we get x again. Projections P (with the property $P = P^2$) are examples of repeatable measurements. As discussed in Peres (1993), not all meaningful measurements are repeatable, and repeatable measurements are more an exception than the rule. In spite of that the term "measurement" is generally used for repeatable measurements. The non-repeatable measurements are also of importance because they may provide more information than ideal repeatable ones.

- **Maximal or complete measurements.** These are measurements that produce the same number of outcomes as the dimension of the corresponding Hilbert space. Each maximal test is uniquely determined by an orthonormal basis.

- **YES-NO measurements**. They are specified by a subspace and its orthonormal complement. The projection into one of the subspaces is interpreted as YES and into the second as a NO answer. A special case of a **primitive measurement** is when one of the subspaces is spanned by just one vector.

- **POV measurements**. **P**ositive **o**perator **v**alued **m**easurements (see Bush et al. 1997, Peres, 1993, and Section 9.2.8), are different forms of measurement. For example, the number of outcomes they produce can be larger than the dimension of the underlying Hilbert space. Formally, a POV measurement is given by a set $\{O_i\}_{i=1}^k$ of positive Hermitian operators such that $\sum_{i=1}^k O_i = I$.

9.1.5 Quantum paradoxes

> The folly of mistaking a paradox for a discovery, a metaphor for a proof, a torrent of verbiage for a spring of capital truths, and oneself for an oracle, is inborn in us.
>
> Paul Valéry (1985)

Design of paradoxes[24] has been an old methodology in physics to point out, in a irresistible way, some inconsistency of physical theories. Three such paradoxes, that have played an

[24] These paradoxes played and still play an important role even not all physicists see all "quantum para-

important role either in clarifying the role of information in physics or in the development of quantum mechanics itself, will now be briefly discussed.

Maxwell's demon

It seems that it was in connection with the attempts to explain Maxwell's[25] demon paradox, from 1867, that information processing considerations started to play a significant role in physics for the first time.

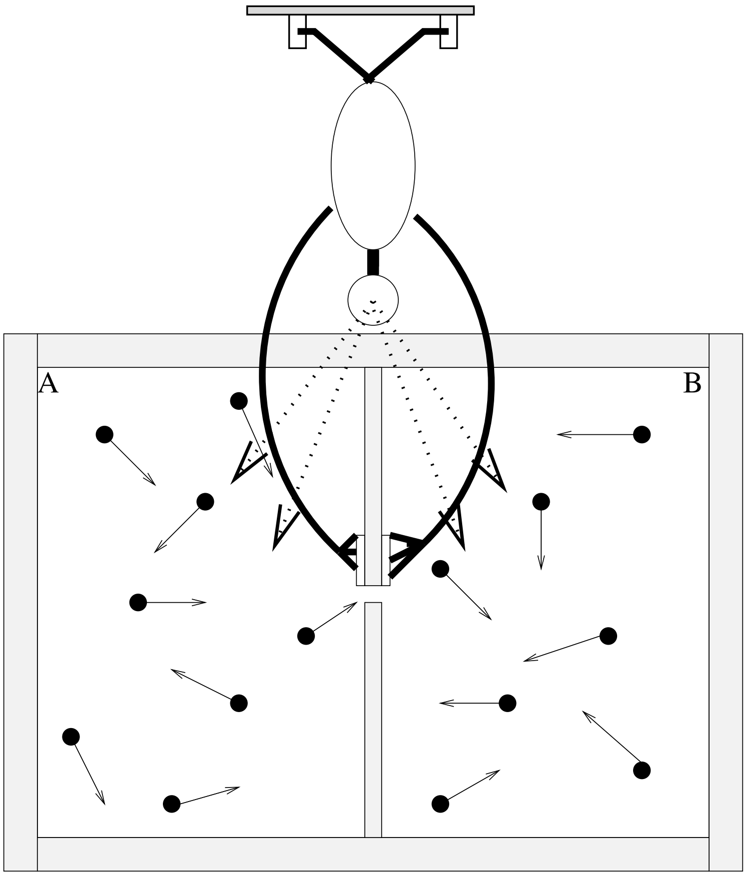


Figure 9.11: Maxwell's demon

doxes" as real paradoxes. For example, Peres' (1993) position is as follows: "There is a temptation to believe that each particle has a wave function which is its objective property.. . . Unfortunately, there is no experimental evidence whatsoever to support this naive belief. On the contrary, if this view is taken seriously, it leads to many bizarre consequences, called "quantum paradoxes". These so-called paradoxes originate solely from an incorrect interpretation of quantum theory. The latter is thoroughly pragmatic and, when carefully used, never yields two contradicting answers to well-posed questions. It is only the misuse of quantum concepts, guided by pseudorealistic philosophy, which leads to those "paradoxical results".

[25] James Clerk Maxwell (1831–1879), a Scottish physicist. In 1871 he became the first professor of experimental physics at Cambridge. On the basis of Faraday's laws for electricity and magnetism he developed a mathematical uniform theory of electricity and magnetism. Maxwell discovered equations, nowadays bearing his name, to describe phenomena of classical electromagnetism.

Maxwell's demon is a creation that operates a shutter to open and to close a trapdoor between two compartments A and B of a completely isolated chamber containing a gas of molecules with a random distribution of particles and velocities. Demon pursues the subversive policy of only opening the door when a faster molecule approaches it from the right, or a slow one from the left. Hence A cools down and B heats up. Working in this way for a while the demon separates hot molecules from cold and establishes a temperature difference between two compartments and decreases the entropy of the system without doing any work—apparently violating the second law of thermodynamics.

Maxwell's demon paradox created enormous controversy among physicists.[26] The first explanation was offered by Szilard in 1929 and it was based on the belief that measurements that the demon has to perform, on the location and speed of molecules, increase its entropy and this compensate for the decrease of the entropy in the system.

A real explanation came only after a deeper insight into thermodynamic cost of information processing was obtained by Landauer and Bennett (the later one first showed that demon measurements can be performed reversibly and therefore without an increase of entropy). It is based on modern knowledge that not information acquisition but information erasure requires energy. In a simplified form the explanation goes as follows.

The demon has to collect and store information in his memory about the locations and speeds of the molecules. Since his memory is finite he has to erase information from his memory from time to time, and it is during this process that the entropy increases as required by the second law of thermodynamics.

Schrödinger cat

One of the most puzzling phenomena in our physical world, that is basically quantum mechanical, is why there is no quantum superposition on "macroscopic scale objects" (with the exception of such phenomena as superconductivity). Or, does it actually exist and we are only not able to observe it? In addition, why is it that in the case of measurement quantum evolution does not continue as before? We can consider a given quantum system together with a measuring device and such a system should develop according to a unitary evolution. There is therefore an apparent contradiction here that were made very vivid through famous Gedanken experiment of Schrödinger's cat. (The experiment has numerous formulations and we consider one of them.)

Let us assume we have a completely isolated chamber with four key elements: an observer, a cat with a cat-food pot, a cup of poison, and an apparatus controlled by a beam of photons from a photon source, also inside the chamber. The beam of photons is directed, as in a Mach-Zehnder interferometer (see Figure 2.5), to a half-silvered mirror. When a photon gets through the mirror, nothing particular happens. The cat keeps having a good time. If the photon is reflected at the mirror it triggers a photo-cell (as a measuring device) and this causes poison to leak from the cup to the cat-pot and the cat dies immediately.

From the point of view of (an unfortunate) observer in the chamber there are two possibilities. Either the measuring device, the photo-cell, does not record the photon and the (lucky) cat is alive, or it does and the (poor) cat is dead. There are only these two possibilities and one of them has to happen. For an internal observer there are therefore two options concerning the cat: "alive or dead", and both have the same probability.

However, the situation looks different to an external observer. He "sees" the whole

[26] For a more detailed treatment of the Maxwell paradox and its implications concerning entropy and information see Leff and Fex (1990).

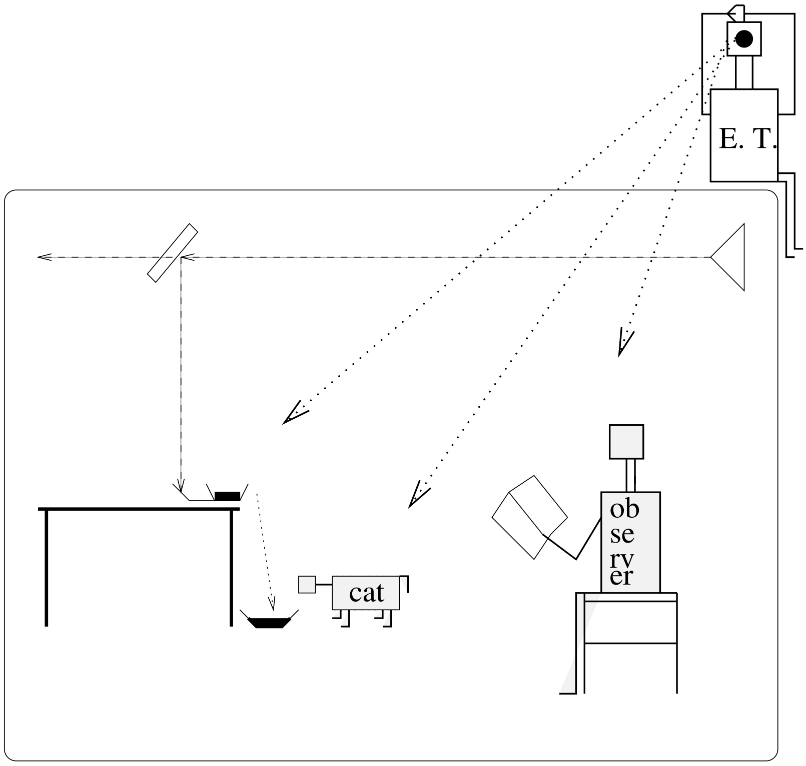


Figure 9.12: Schrödinger's cat

system in the chamber as a single quantum system in which only unitary evolutions occur, no measurement. From his point of view the photon is in a superposition of two states and the cat gets into the state $\frac{1}{\sqrt{2}}(|\text{alive}\rangle|\uparrow\rangle + |\text{dead}\rangle|\downarrow\rangle)$, that is, she is both alive and dead at the same time but neither of both, with the same probability. However, this contradicts our experience. Cats we see are either alive or dead.

Where does Schrödinger's paradox lead us to and how to deal with it? Some problems are easy to identify: there is no clear definition what a measurement is. There seems to be no way to draw line between those measurements that are "possible" and those that are "impossible". Some even considered that the problem with Schrödinger's cat lies in the fact that we have a conscious observer both inside and outside the chamber and that perhaps the laws of quantum physics do not apply to consciousness.

Schrödinger's Gedanken experiment led to the development of the so-called many-world interpretation of quantum mechanics (see Everett, 1957, 1977) and Section 9.1.4.[27]

Recently, an understanding of Schrödinger's cat mystery has developed that uses decoherence as the key element and goes briefly as follows. In order to specify fully a quantum state of a cat one needs to specify quantum states of all its components, atoms, electrons,... There is a huge number of quantum states that correspond to alive cats and a huge number of states that correspond to dead cats. All these states constantly evolve due to the inherent

[27] The many-world interpretation has recently been used by Deutsch (1997) when considering quantum physics as one of four main strands of explanation that he considers as ones that may constitute the first to-be-developed-yet "Theory of Everything".

interaction of their elements and interactions with the environment. Quantum superposition "both alive and dead but neither of both" can exist but only for an unnoticeable tiny fraction of time, because it is very unstable, and then evolves, due to the decoherence, into a mixed state: alive with a probability one-half and dead with the same probability. In short, Schrödinger's cat does not exist. Or rather it has an immesurable life-time before it evolves into a classical or Newtonian cat (Lindley, 1996).

EPR measurements

Einstein was a strong opponent of the key view of the Copenhagen interpretation of quantum physics, namely that quantum properties are not determined (or that they even do not have a meaning) until they are measured, and insisted that unmeasured quantities must exist in some state even though we might not know what the state is.

Einstein believed that quantum mechanics is incomplete, and that there must be a deeper and more detailed theory that would include all the necessary information to allow us to make full and certain predictions of the outcomes of measurements, not only the statements of possibilities and probabilities. (On the other hand, Bohr claimed that looking for such a theory is a misguided venture, motivated by our romantic thoughts of what physics should be rather than by pragmatic understanding of what physics really is.)

In order to discredit views of physics resulting from the Copenhagen interpretation, Albert Einstein,[28] Boris Podolsky and Nathan Rosen (EPR), in their famous paper "Can quantum mechanical description of reality be considered complete?", in 1935, developed an experimental set-up that helped to illustrate how strange consequences follow from quantum theory.

Their basic reasoning goes as follows. Let us imagine two particles, whose total momentum is constant, flying apart at the same speed. Once they are far apart you measure the position (or momentum) of the first particle and by that you immediately know the position (momentum) of the second particle.

However, Einstein and his colleagues made out of that two important conclusions.

1. By measuring precisely the position of the first particle we get precisely the position of the second particle. Since no measurement was involved on the second particle we can now measure its momentum precisely. However, this contradicts Heisenberg's principle.

2. If, without disturbing in any way a system (second particle), we can determine with certainty the value of a physical quantity (position or momentum), there has to be an element of physical reality that corresponds to this physical quantity. Therefore, both position and momentum of particles have to be elements of physical reality.

Bohr came up with surprising, but actually deep explanation why Einstein's reasoning, by which he derived a contradiction with Heisenberg's uncertainty principle, is wrong. According to Bohr one is not allowed to combine into one consideration outcomes of two incompatible measurements (of the position of one particle and of the momentum of the second particle).

[28] Albert Einstein (1879–1955), considered as one of the best physicists of all time, an American theoretical physicist of German origin. He wrote the basic papers on the special and general theory of relativity, showed the equality of gravitation and inertia, contributed to the development of quantum theory postulating light quanta. He received the 1921 Nobel prize for physics.

An important modification of the basic EPR Gedanken experiment, due to Bohm, shows the problem in a even clearer way. His basic Gedanken experiment dealt with two particles that fly apart in such a way that their spins add up to zero or, in a more modern setting, that they form a pair of entangled qubits, in the total state

$$\frac{1}{\sqrt{2}}(|0\rangle|1\rangle + |1\rangle|0\rangle)$$

and are spatially separated. In such a case a measurement of the spin of one of the qubits determines (observes) uniquely the spin of another qubit without the second one being disturbed by an indirect observation. Einstein called this phenomenon a "spooky-action-at-a-distance" because measurement in one place seems to have an instantaneous effect at the other place. The term "spooky" indicates that the influence was implied rather than directly seen.

Einstein and his colleagues concluded from their Gedanken experiment that explanations of the real phenomena which the current quantum physics offers (namely, its Copenhagen interpretation) are not complete and suggested the way, a program, to fix it—how a proper fundamental theory of Nature should look like. The EPR program asked for *completeness* ("In a complete theory there is an element corresponding to each element of reality."), *locality* ("The real factual situation of system A is independent of what is done with system B, which is spatially separated from the former.") and *reality* ("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 physical reality corresponding to this physical quantity.").

In particular, they suggested that the wave functions (quantum states) do not provide a complete description of the physical reality, and therefore there have to exist additional, so-called "hidden" variables, whose objective values would unambiguously determine the result of any quantum experiment.

9.1.6 The quantum paradox

Paradoxes discussed in Section 9.1.5 are, strictly taken, hardly real paradoxes, because they do not contradict themselves. They discuss weird and mysterious phenomena, and required fundamental changes in the way we think about reality, but they are not paradoxes.

There is, however, one genuine quantum paradox in the Copenhagen interpretation of quantum mechanics, that concerns the key concept of this interpretation—measurement.

On the one hand, quantum physics is considered as the fundamental theory of physics (and therefore, in principle, the workings of all things should be explicable in quantum mechanical terms) and, on the other hand, it assumes that quantum systems exist in undetermined states until they are measured by (classical) measuring devices that have only determined states, even though these devices should also be considered as quantum and therefore being, in principle, also always in undetermined states until their states are measured, ...

Some physicists, including several very prominent ones, even tended to accept the position that the process of dividing classical parts of measuring devices into classical and quantum parts ends at our consciousness. Namely, that it is not so much the physical act of measurement as the mental act of becoming aware of the result that, finally, creates the borderline between quantum uncertainty and specific knowledge.

A measurement is in the Copenhagen interpretation a step we have to take to get from the quantum world to the classical world, but the Copenhagen interpretation gives slightest explanation of how this step is performed. It only says what its outcomes are. There is nothing in quantum theory that explains or determines the exact mechanism of quantum measurement and the resulting state collapse. In particular, both measurements and state collapse are presented in such a way that they would not require involvement of forces of any kind.[29]

Fortunately, the problem of quantum measurement paradox seems to be on the way to being solved and by means of quantum mechanics itself. Two concepts play the key role. Decoherence and theoretical insights into the behaviour of complex systems. They allow us to understand how large assembles of weird quantum objects can behave in a reasonable way and how Nature gets around seemingly not understandable quantum measurement phenomena.

9.1.7 Interpretations of quantum theory

> Sir, I have found you an argument. I am not obliged to find you an understanding.
>
> Samuel Johnson (1709–1784)
> to a critic of his views on some matters.

The trouble with quantum mechanics, and the reason for its various "interpretations", is that despite its decades of phenomenal success, it is still not understood sufficiently well what "quantum things" mean. We are not able to see "inside them". Many of the best physicists have worked on this problem—but not without reach conflicting conclusions.

There are two reasons for this. At first we are not able to visualize quantum objects and phenomena and to translate mathematical formulas and explanations provided by quantum theory into recognizable pictures and understandable words. Secondly, quantum theory actually offers us certain ideas of how the world works. However, they do not conform to our prior expectations we should like, or we think, the world should be.

Interpretations are attempts to get around these basic difficulties. The goal is to interpret and explain quantum concepts and phenomena in such a way that we can say, or at least some of us, "I got it".

Finding and analysing interpretations of quantum theory is one of the main tasks of modern science and also of the philosophy of sciences. The fact that this task seems to be still far from being resolved is something we have to learn to live with, and not to blame quantum physics for it. As Lindley says (1996), "if science sometimes provides explanations without giving us what we would regard as an understanding, the deficiency belongs to us, not to science".

Quantum formalisms versus interpretations

von Neumann's (1932) idea to formulate quantum mechanics in terms of Hilbert space concepts established the most important **formalism** within which quantum mechanics has developed. The most compact formalism for quantum mechanics developed by Dirac, in

[29] Bohr solved "easily" the measurement problem by asserting that measurements can be made and one never had problems to determine whether a prospective measurement is really a measurement, and he just did not bother with the problem of how measurements can be made in general.

1930, lacked at that time mathematical rigor, especially because of the use of the delta functions, that were for that time mathematically unacceptable. This has changed in between and therefore Dirac's formalism represents an alternative for quantum mechanics. (For a more detailed, but concise treatment of the formalisms of quantum mechanics see Jammer (1974).)

The **Interpretation** of quantum mechanics formalisms is one of the deepest and most difficult problems of current science. Not only the physicists disagree on *which formalism is correct*; the philosphers of science disagree even on *what does it mean to have an interpretation.*

In order to introduce problems of interpretations of quantum mechanics let us take a widespread position that a physical theory is a partially interpreted formal system (Jammer, 1974). A physical theory is seen as having two components: an abstract formalism $\mathcal{F}$ and the rules of correspondence $\mathcal{R}$ (or an "interpretation" of $\mathcal{F}$). $\mathcal{F}$ is a deductive logical calculus without empirical meaning and contains, in addition to logical constructs, also nonlogical terms (as "state"). The rules of correspondence, $\mathcal{R}$, make $\mathcal{F}$ physically meaningful by assigning physical meaning to *some* of the nonlogical terms of $\mathcal{F}$. Those nonlogical terms that are not directly interpreted by $\mathcal{R}$ are called *theoretical terms.* They are only contextually defined through the role they play within the logical structure of $\mathcal{F}$. On one side, theoretical terms can be in principle removed from a theory, but on the other side, they are an instrument for new discoveries. Let $\mathcal{F}_{\mathcal{R}}$ denote $\mathcal{F}$ accompanying by $\mathcal{R}$.

One school of thoughts takes $\mathcal{F}_{\mathcal{R}}$ as a physical theory—a mathematical system suitable to describe as completely, concisely and precisely as possible our experimentally observable knowledge of the "physical reality". Other school of thoughts does not consider such approach as sufficient and requires to supplement $\mathcal{F}_{\mathcal{R}}$ with some unifying **principle** which "establishes some internal coherence among the descriptive features of $\mathcal{F}_{\mathcal{R}}$ and endorse it thereby with explanatory and predictive power". This is an interpretation of $\mathcal{F}_{\mathcal{R}}$ which is one of the most controversial problems of quantum mechanics and philosophy of science.

In addition to the rules of correspondence and the unifying principle, the third way to provide an interpretation is to construct a **model** of the theory.

Main interpretations

There are several ways of categorizing the existing interpretations. One criterion is whether a given interpretation is experimentally distinguishable from the orthodox one, or tries to go beyond it.

If an interpretation is not experimentally distinguishable from the orthodox one, then its benefit can be mainly on an aesthetic or logical or methodological level, as a change of a research paradigm. It is natural that for such interpretations it is not easy to supersede the orthodox one, unless it is much simpler, mainly for sociological reasons. The scientific community is in general very conservative and prefers to stick with a theory till the facts showing its obsolescence are much too strong.

The main interpretations of linear quantum mechanics, in which evolution is described by the linear Schrödinger equation, are the following ones: Copenhagen interpretation[30], many-worlds interpretation (Everett, 1957, 1977), and the hidden variable interpretations, Bohm (1952), or its more modern version—an ontological interpretation of Bohm and Hiley (1993).

[30]By Peres (1993) there are many noncompatible versions of the Copenhagen interpretation and by him. "There is no real conflict between Stapp (1972) and Ballentine (1970)—two important expositions of the Copenhagen interpretation—except that one calls Copenhagen interpretation what the other considers as the exact opposite of the Copenhagen interpretation."

For a recent analysis of these and other interpretations of linear quantum mechanics see Bub (1997). Interestingly enough, probably due to the fact that recent quantum experiments display ever more extreme forms of non-classical behaviours, the interest for other than Copenhagen interpretations seems to grow, but no real alternative has yet emerged.

There are also various models of nonlinear quantum mechanics in which evolution is described by a nonlinear Schrödinger equation—obtained from linear Schrödinger equation by introducing nonlinear modifications. For example, models due to de Broglie (1956) and Weinberg (1989). No experiments supported yet the existence of nonlinear evolutions, but some physicists do not consider their existence as completely ruled out.[31]

The Copenhagen interpretation is based on the following two principles.

1. A quantum system that has not yet been measured exists in a state of genuine indeterminacy. It makes no sense to say (and may even lead to contradictions) that it is in a specific but unknown state.

2. The act of measurement forces the system to adopt one of the possible classical values, with a probability that can be calculated from the appropriate quantum state of the system and its measurement.

The first principle of the Copenhagen interpretation actually denies the existence of an independent and unique objective reality on which all observers can agree—a fundamental concept on which natural sciences are based. The second principle relies on a magic of measurement no one was able to explain yet, and no one was able to understand why a measurement could make indeterminacy go away.

The Copenhagen interpretation of quantum mechanics is not so much philosophy as an "act of intellectual self-discipline". It does not make quantum phenomena, such as the two-slit experiment any easier to understand, it just tells us that we should not hope to understand it in the way we should like to. It resolves certain difficulties only by declaring them out of bounds.

The Copenhagen interpretation implies that quantum physical properties are not, as in the classical world, intrinsic and unchangeable characteristics of the things we are measuring but instead of that arise, in the quantum world, as a result of the act of a measurement and cannot be ascribed any useful or consistent meaning before a measurement is made.

The Copenhagen interpretation divides the worlds into physical objects of two types: things we measure (or might in principle measure) and things we measure with. Objects to be measured are quantum objects and they live in uncertain, indefinite states until they are measured. Objects we measure with are classical and always in a definite state.

There seem to be two inconsistencies in the Copenhagen interpretation, if looked at from outside (even if it is consistent under certain internal assumptions).

1. Quantum objects can be assigned a definite value of a property only when a measurement is made, but to make a measurement we need some sort of non-quantum—classical—devices. At the same time quantum mechanics aspires to be the fundamental theory of physics and therefore also its measuring devices should be, in principle, quantum and should adhere to all laws of quantum mechanics.

[31] Nonlinear models may have "weird" properties and exhibit "unphysical effects". As mentioned by Peres (1993), it may happen that a state $u(0)$ evolves after t steps into $u(t)$ and the state $v(0)$ into $v(t)$ but $u(0)+v(0)$ does not evolve into $u(t)+v(t)$. In addition, if this is the case, and other postulates of quantum mechanics are not changed, evolution in such a system can violate second law of thermodynamics.

2. The Copenhagen interpretation of quantum mechanics relies on a notion called "measurement" which seemed to be beyond the abilities of quantum mechanics itself to explain.[32] Quantum mechanics seemed to be unable to explain by which physical processes a measurement act makes things to change from undetermined to fully determined ones.

From the practical point of view, the Copenhagen interpretation works satisfactorily. From the philosophical or fundamental point of view many physicists consider the Copenhagen interpretation as unsatisfactory—in spite of its phenomenal practical success. It is therefore natural that various other interpretations of quantum mechanics have been worked out.

The **hidden variable interpretation** of quantum mechanics, due to Bohm (1952), tries to formalize ideas of Einstein and to deny the fact that indeterminacy in quantum mechanics is unavoidable reality. Bohm tried to restore for quantum mechanics the underlying reality, even the one that cannot be seen and detected. In Bohm's interpretation, measurements seem to follow probability laws only because we are ignorant of certain hidden properties of the things we are measuring. Were we to know the values of the hidden variables, we could say precisely what outcomes a measurement would produce.

Bohm has worked out a new deterministic formulation of quantum mechanics that is mathematically the same as the standard theory, but which is rearranged so that everything looks classical except one strange **quantum potential**, into which all nonclassical aspects of quantum theory end up.

Bohm gave specific and almost classical meaning to "wave" and "particle" halves of quantum mechanics. The role of waves is to guide particles to create superpositions. For example, in the two-slit experiment waves make particles go through one or another slit in order to create the superposition pattern, and it is the quantum potential that gave rise to guide waves. In Bohm's interpretation the classical world is a part of the quantum world and some paradoxes, such as Schrödinger's cat paradox, do not exist.

Bohm built quantum mechanics on classical foundations. Everything is deterministic. Every particle follows a predictable path and it is only because we do not have a precise knowledge of the initial conditions that we need to use probabilities to describe the values measurementswill end up with. In the case of photons it is the precise but unknown initial position and the initial momentum of each photon that form the "hidden variables".

Several problems arise with Bohm's interpretation. It can be shown that guide waves cannot use force to control particles. To get around this problem Bohm and Hiley recently suggested that guide waves carry (active) information about where particles should go and through this information they guide particles. However, this implies that both quantum potential and guide waves have to exhibit non-local phenomena because they need to gather information instantaneously from all parts of the current experiment.

Another problem arises in connection with the determinism in Bohm's theory. If the movement of particles is deterministic, how is it possible that a stream of them, in the two-slit experiment, creates a superposition pattern? Bohm's way out is the following one: particles in a stream are not completely identical. Each particle is fully determined and has its own momentum and direction. Guide waves send each particle in a completely deterministic way but since all of them have slightly different initial conditions they go along different routes.

[32]Bohr was aware of this problem but got away with it by claiming that physicists know how to measure things and in practice there has never been any doubt about whether an action is or is not a measurement.

How useful is Bohm's interpretation? We cannot obtain from it a result contradicting predictions of the Copenhagen interpretation. Moreover, since we are forbidden from ever seeing the hidden variables, and supposedly deterministic information they carry, it becomes a matter of metaphysical taste whether we think they are there or not (Lindley, 1996).

Some say that Bohm's interpretation adds mathematical complications without adding really new results. Others say that observations is all one gets at the Copenhagen interpretation, and the Bohm's interpretation at least offers you a chance to "dig deeper".

In the **many-worlds (universes) interpretation** of quantum physics, due to Everett (1957, 1977), the main problem of Copenhagen interpretation, how to explain measurement, and especially the fact that during any measurement only one potential value is preserved and all others are irreversibly lost, is solved in a very special way. Everett assumes that before a measurement all possible outcomes coexist and after the measurement all possible outcomes of each measurement are preserved in a special way, namely, that each carries on a separate existence in its own world (or universe). If a further measurement in any of these worlds leads to a creation of new worlds they are totally different from all so far existing worlds. (As a consequence, in a more complex quantum process the number of worlds existing in parallel can be huge.) Another view of the many-worlds interpretation of measurement, perhaps more modern, is as another unitary operator that produces an entangled state of the system and measuring device (Bennett, 1988b).

There is a variety of interpretations and variations of Everett's interpretation and they all fall under the name "many-worlds interpretation". A widely held belief is that there is no possibility of doing an experiment in one world that would reveal another world, and that Everett's interpretation can neither be proved, or disproved and that it is not experimentally distinguishable from other interpretations. (However, Deutsch, 1985, claims that it would be possible to make an experimental test of Everett's interpretation by using a quantum computer, provided "truly artificial intelligence programs are available".)

Everett's theory is based on the innocent looking postulate that "all isolated systems evolve according to the Schrödinger equation". Its two basic consequences of far-reaching importance are: (1) The entire universe evolves according to the Schrödinger equation, since it is by definition an isolated system. (2) There can be no definite outcome of any quantum measurement since this would violate the main postulate.

For a long time Everett's interpretation found only few followers. Recently an interest in this interpretation seems to be increasing. Deutsch (1996) is one of those that try to go on with this interpretation. For a recent analysis of objections to the many-worlds interpretation see Tegmark (1997).

It is of importance to realize that all Copenhagen, many-worlds and hidden variable interpretations are interpretations of the same theory, not of different theories. In addition, one has to realize that "weird things seem to happen in the quantum world and it is simply not possible to make weirdness go away by shuffling around with the terms of the theory and calling such an effort an interpretation" (Lindley, 1996).

Adherence to one or another interpretation of quantum physics is often related to what one believes to be more fundamental—mathematical language or human language. There are actually two diametrically opposite paradigms one can choose from.

1. The outside view (mathematical structure) is physically real and the inside view (human language we use to describe it) is only a useful approximation for describing our subjective perception.

2. The (subjectively perceived) inside view is physically real, and the outside view with

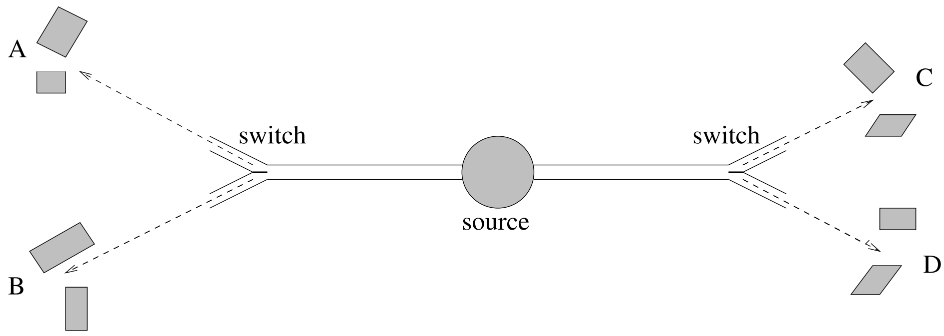


Figure 9.13: Aspect's experiment

its mathematical language is merely a useful approximation.

9.1.8 Incompleteness of quantum mechanics

Einstein, and many others, believed that such mysteries as entanglement are due to the fact that the current quantum theory is incomplete, and he believed that by using some "hidden variables" one could develop a complete theory without non-local influences where one could believe in the existence of an objective reality for quantum phenomena. Einstein believed that there can be some deeper theory that can explore quantum mechanical phenomena in terms of some fundamental and not yet known concepts in such a way that what seems currently to be indeterminacy would be seen only as a consequence of our current ignorance of some fundamental phenomena of the world. This has been demonstrated to be definitely wrong by ingenious theoretical considerations and the Gedanken experiment due to Bell (1964) and physical experiments by Alain Aspect (1982) and others. The basic ideas follow in a simplified form (Peres, 1993 and Lindley, 1996).

Assume that a pair of electrons whose spins add up to zero is created by some source and sent off in two separate directions. In each electron's path we install two Stern–Gerlach magnets and a switching device that can send an electron randomly to one of the two magnets—see Figure 9.13. All four magnets are set at different angles.

From each of the magnets A, B, C and D, the electron can come in one of two possible streams, and let us set for $Y \in \{A, B, C, D\}$, $Y = 1$ or $Y = -1$ depending on which stream a particular electron gets in. 1 and -1 are therefore all potential values of A, B, C, D, considered from now on as variables. In each particular experiment one of two variables A and B and one of C and D get one of the values 1 and -1.

Consider now the function of arguments A, B, C and D defined as

$$X = C(A + B) + D(A - B).$$

It is easy to verify that if A, B, C and D can take only values 1 and -1, then X can take only one of the values 2 or -2.

If the experiment is repeated many times then elements of each of the pairs (A, C), (A, D), (B, C) and (B, D) are assigned some of the values -1 or 1 in about the same number of cases. It makes therefore sense to talk about the average value EX of X and it is evident that this average value should be between -2 and 2, i.e. $-2 \leq EX \leq 2$. (This is actually

Bell's theorem and one of Bell inequalities[33].)

At the same time Bell was able to show, by considering joint probabilities, that quantum mechanics laws imply that the average can be as large as $2\sqrt{2} \approx 2.8$.

Two questions arise. Is this really true? Does nature behave in such a way that almost trivially true inequalities $-2 \leq EX \leq 2$ are not satisfied? And if yes, how is this possible? What went wrong with the derivation of the inequalities $-2 \leq EX \leq 2$?

In the famous and much disputed experiment of Alain Aspect (1982) and his colleagues, performed in 1981–1982 in Paris, they used photons and measured their polarization. The photon pairs were emitted by the source in the entangled state—their overall polarization state was neutral. The decision on how to operate optical switches was not made until the photons were in full flight—too late for a message from one photon (concerning the direction of the measurement) to reach the other photon. The experiment confirmed Bell's expectation. His inequalities were violated.

Where is the problem? Copenhagen interpretation has no difficulties to identify it. Bell's inequalities are not well put together. In the expression for X we mix values of variables obtained in two different and incompatible experiments. This is not allowed according to this interpretation. If we measure A and C at one experiment. then at that experiment B and D have no definite value and the fact that they may have only values 1 or -1 is of no importance. X is not put together according to the rules of the Copenhagen interpretation. Information obtained from different experiments cannot always be used to build a single picture of the physical world. We have to think of the entangled electrons as being in a single quantum system, until a measurement is made, and not as two separate particles that have some properties in common because of a common origin.

There are many other conclusions one can derive from Bell's theorem and Aspect's experiment. One of them is that we have to give up hope for the existence of objective reality in such a strong sense as we used to assume in classical physics with such a clear and simple correspondence between reality and theory concepts.

Aspect tested Bell's inequalities over a distance of 12 m. Since then various other experiments have been performed testing Bell's inequalities over longer distances. Recently, Tittel at al. (1998) reported violation of special "Bell's inequalities" by photons 10.9 km apart.

Observe that the existence of two entagled spatially separated particles led to a conflict with local realism for statistical predictions of quantum mechanics. On the other hand, recent experiments demonstrating the existence of three entangled spatially separated particles (Bouwmeester, 1998) leads to a conflict with the local realism for "nostatistical predictions of quantum mechanics".

The problem of the reliability of all these experiments has not yet been fully resolved. Most physicists seem to be inclined to take them as "definite evidence". Some claim that all experiments so far have relied on some unproven assumptions and therefore they still leave loopholes for hidden variable interpretations.

9.2 Hilbert Space Framework for Quantum Computing

Truth is great and its effectivness endures.

[33]More exactly this is so-called CHSH inequality (Clauser, Horne, Shimony and Holt, 1969). Bell actually used three variables, equality $A(B-C) = \pm(1-BC)$ and the corresponding inequality $|EA(B-C)| \leq 1 - EBC$.

Ptahhotpe, 24th century BC

The family of all possible (pure) states of a quantum system constitutes what is known as Hilbert space. The Hilbert space formalism is therefore the basic framework for formally precise definitions and a study of the quantum mechanical concepts, phenomena, systems, algorithms and processes. This is especially true for the physical foundations of quantum computing.

9.2.1 Hilbert spaces

The basic underlying concept is that of a vector space.

Definition 9.2.1 *A* **vector (linear) space** $\mathcal{S}$*, with a carrier* H*, over a field* $\mathcal{K}$ *with the carrier* K *is an algebra* $\mathcal{S} = \langle H, +, ^{-1}, \mathbf{0}, K, +_f, \times_f, 0, 1, \cdot\rangle$ *such that* $\langle H, +, ^{-1}, \mathbf{0}\rangle$ *is a commutative group,* $\mathcal{K} = \langle \mathcal{K}, +_f, \times_f, 0, 1\rangle$ *is a field, and* $\cdot : \mathcal{K} \times H \rightarrow H$ *is a scalar multiplication satisfying the following axioms for any* $a, b \in K$*,* ϕ *and* $\psi \in H$*:*

- $a \cdot (\phi + \psi) = a \cdot \phi + a \cdot \psi$, $(a +_f b) \cdot \phi = a \cdot \phi + b \cdot \phi$ *{distributive laws}*
- $(a \cdot (b \cdot \phi)) = (a \times_f b) \cdot \phi$
- $1 \cdot \phi = \phi$.[34]

Example 9.2.2 *(1) Vector spaces* $H = \mathbf{R}^n$ *or* $H = \mathbf{C}^n$ *with* $K = \mathbf{R}$ *or* $K = \mathbf{C}$*. (2)* H *is the set of all infinite sequences* $\{x_i\}_1^\infty$ *with* $x_i \in \mathbf{R}$ *such that* $\sum_{i=1}^{\infty} |x_i|^2 < \infty$ *and* $K = \mathbf{R}$.

Definition 9.2.3 *A* **complex inner-product space** $\mathcal{H}$ *is a vector space with a carrier* H *over the field of complex numbers, equipped with an inner product (called also scalar product or Hermitian scalar product)* $\langle .|.\rangle : H \times H \rightarrow \mathbf{C}$ *satisfying, for any* $\phi, \phi', \psi \in H$*, and any* $c_1, c_2 \in \mathbf{C}$*, the following axioms:*

- $\langle\psi|\phi\rangle = \langle\phi|\psi\rangle^*$
- $\langle\psi|\psi\rangle \geq 0$ *and* $\langle\psi|\psi\rangle = 0$ *if and only if* $\psi = \mathbf{0}$
- $\langle\psi|c_1\phi + c_2\phi'\rangle = c_1\langle\psi|\phi\rangle + c_2\langle\psi|\phi'\rangle$

The inner product introduces on H the norm

$$||\phi||_H = \sqrt{\langle\phi|\phi\rangle}$$

and the metric[35]

$$dist_H(\phi, \psi) = ||\phi - \psi||.$$

$||\phi||$ is called also the length of the vector ϕ and $\langle\phi|\phi\rangle$ is called the squared length of ϕ. (Instead of $||\phi||_H$ (and $dist_H(\phi, \psi)$) we usually write only $||\psi||$ ($dist(\phi, \psi)$, if H is clear from the context.)

Example 9.2.4 *If* $\mathcal{H} = \mathbf{C}^n$ *for a fixed* n *and the inner product is defined by*

$$\langle(x_1, \ldots, x_n)|(y_1, \ldots, y_n)\rangle = \sum_{i=1}^n x_i^* y_i,$$

then we speak about the n**-dimensional complex inner-product space**.

[34] If there is no danger of confusion the operators $\cdot$ and $\times_f$ are usually omitted. In addition instead of $\phi + \psi^{-1}$ we usually write $\phi - \psi$.

[35] Which in turn allows us to introduce on H a topology and such concepts as continuity.

Two basic notations are used for elements of inner product spaces. In so-called **von Neumann notation** a plain letter, say ψ, denotes an element (vector). In **Dirac notation** an element is denoted as $|\psi\rangle$ and called **ket vector**. We use mainly Dirac notation, which is usually more handy. However, in some cases, in order not to have an abuse of delimiters, a simpler von Neumann notation is used.

Exercise 9.2.5 *Show the following properties for any elements ϕ, ψ of an inner product space H, and any $c \in \mathbf{C}$: (a) $\langle c\phi|\psi\rangle = c^*\langle\phi|\psi\rangle$; (b) $||c\phi|| = |c|\ ||\phi||$; (c) $||\phi+\psi||^2 + ||\phi-\psi||^2 = 2||\phi||^2 + 2||\psi||^2$ (parallelogram law);*

An inner-product space with a carrier H is called **complete** if for any sequence $\{\phi_i\}_{i=1}^{\infty}$ with $\phi_i \in H$, and with the property $\lim_{i,j\to\infty} ||\phi_i - \phi_j|| = 0$ there is a unique element $\phi \in H$ such that $\lim_{i\to\infty} ||\phi - \phi_i|| = 0$. A complete inner-product space is called **Hilbert space**.[36]

The elements of $\mathcal{H}$ are usually called vectors, and if they have norm 1, then (pure) states. In the following we mostly identify $\mathcal{H}$ with its carrier H.

Orthonormal bases of Hilbert spaces

The following concept of orthogonality plays a key role in the theory of Hilbert spaces.

Definition 9.2.6 *Two vectors ϕ and ψ of a Hilbert space H are called orthogonal if $\langle\phi,\psi\rangle = 0$. A set $S \subseteq H$ is* **orthogonal** *if any two of its elements are orthogonal. S is* **orthonormal** *if it is orthogonal and all its elements have norm 1.*

In ordinary terms, orthogonal states (represented by orthogonal vectors) are things that are independent of each other; for example, both basic states of a spin-$\frac{1}{2}$ particle, all positions a particle can be located in or all configurations of a quantum automaton.

Exercise 9.2.7 *(a) Show that if two vectors ϕ, ψ of a Hilbert space H are orthogonal, then $||\phi+\psi||^2 = ||\phi||^2 + ||\psi||^2$; (b) If $\{\phi_1, \ldots, \phi_n\}$ is an orthonormal set in H, then for all $\phi \in H$: $\sum_{i=1}^{n} |\langle\phi_i|\phi\rangle|^2 \leq ||\phi||^2$ (Bessel's inequality).*

In a more general approach one can define the **angle** of two vectors ϕ and ψ as follows:

$$\theta_{\phi,\psi} = \begin{cases} 0, & \text{if } \phi = 0 \text{ or } \psi = 0; \\ \arccos(Re(\langle\phi|\psi\rangle/(||\phi|| \cdot ||\psi||))), & \text{otherwise.} \end{cases}$$

Exercise 9.2.8 *Show the following* **law of cosines** *for any $\phi, \psi \in H$:*

$$||\phi||^2 + ||\psi||^2 = ||\phi+\psi||^2 - 2||\phi||\,||\psi||\cos\theta_{\phi,\psi}.$$

[36] Many of the results stated in the following for Hilbert spaces hold also for inner-product spaces in general. However, to simplify presentation we shall talk about Hilbert spaces mostly. Note also that the completeness requirement has no direct physical meaning. Proofs of many general results about Hilbert spaces require the use of limits and to have limit elements within the space under consideration.

Definition 9.2.9 *A subspace G of a Hilbert space H is a subset of H closed under addition and scalar multiplication.*

An important property of Hilbert spaces is their decomposition into orthogonal subspaces. It holds:

Theorem 9.2.10 *For each closed subspace W of a Hilbert space H there exists a unique subspace $W^{\perp} = \{\phi \mid \langle\phi|\psi\rangle = 0$ for any $\psi \in W\}$ such that each $\phi \in H$ has a unique representation $\phi = \psi_1 + \psi_2$, $\psi_1 \in W$, $\psi_2 \in W^{\perp}$. In such a case we write $H = W \oplus W^{\perp}$.*[37]

Those orthonormal sets that are bases are of special importance.

Definition 9.2.11 *An orthonormal set $B \subseteq H$ is an orthonormal basis for H if none of its proper supersets is orthonormal.*

It can be shown that all bases of a Hilbert space H have the same cardinality, called dimension of H. If the cardinality is finite we speak about the finite-dimensional Hilbert space.[38] Otherwise, we talk about the infinite-dimensional Hilbert space. For example, in the case of spin states of a particle the corresponding Hilbert space has finite dimension. Namely, in the case of a spin-$\frac{1}{2}n$ particle the corresponding Hilbert space is $(n+1)$-dimensional. Infinite-dimensional Hilbert space is in the case of states being positions of particles or configurations of space unbounded Turing machine.

Exercise 9.2.12 *(a) Show that if n vectors are mutually orthogonal, then they are linearly independent; (b) Show that dimension of Hilbert space H can be defined as the maximal number of linearly independent vectors in H.*

If $\mathcal{B} = \{\beta_i\}_{i=1}^n$, $\mathcal{G} = \{\gamma_i\}_{i=1}^n$ are two orthogonal bases, and $\gamma_i = \sum_{j=1}^n c_{ij}\beta_j$, then $c_{ij} = \langle\gamma_i, \beta_j\rangle$. Bases $\mathcal{B}$ and $\mathcal{G}$ are called **mutually unbiased** if $|c_{ij}|^2 = \frac{1}{n}$.

Two mutually unbiased bases can be seen as being as different as possible. The standard and dual bases are mutually unbiased. Since each basis corresponds to a test, mutually unbiased bases correspond to tests with clearly distinguishable outcomes. Example: tests for the vertical and the horizontal polarization of photons or for the clockwise and counterclockwise polarization.

Definition 9.2.13 *If $\mathcal{B} = \{\phi_i\}_{i=1}^n$ is a base of an n-dimensional Hilbert space and for a vector ψ it holds $\psi = \sum_{i=1}^n \alpha_i\phi_i$, then the vector $(\alpha_1, \ldots, \alpha_m)$ is called a representation of ψ in the base $\mathcal{B}$.*

Exercise 9.2.14 *Let $\mathcal{B}_1 = \{\phi_i\}_{i=1}^n$ and $\mathcal{B}_2 = \{\psi_i\}_{i=1}^n$ be two orthonormal bases of an n-dimensional Hilbert space. Show that there is a matrix $M_{\mathcal{B}_1,\mathcal{B}_2}$, which maps $\mathcal{B}_1$-representations of vectors to their $\mathcal{B}_2$-representation.*

[37] It can be shown that $W^{\perp} = \{\phi \mid \langle\phi|\psi\rangle = 0$ for any $\psi \in W\}$.

[38] In general it is required for a complete vector space to be the Hilbert space that it is also separable, i.e. there exists a sequence $\{\phi_n\}_{n=1}^{\infty}$ of elements of H such that for any $\phi \in H$, and any $\varepsilon > 0$ there exists at least one $i \in \mathbf{N}$ such that $||\phi - \phi_i|| < \varepsilon$. This requirement of separability is always satisfied for finite dimensional Hilbert spaces. For infinite-dimensional Hilbert spaces H the requirement of separability is equivalent to the requirement that H has a countable orthonormal basis.

Self-duality of Hilbert spaces

Another important feature of Hilbert spaces is their self-duality. This concept refers to the space of continuous linear mappings from H to $\mathbf{C}$ (called **functionals**). If the addition and scalar multiplication are defined in the most obvious way, then the space of linear functionals for a given Hilbert space H is again a Hilbert space, so-called **conjugate space** denoted H^*, related to H as follows:

Theorem 9.2.15 *For each continuous functional $f \in H^*$, $f : H \to \mathbf{C}$ there exists a unique vector $\phi_f \in H$ such that $f(\psi) = \langle \phi_f | \psi \rangle$ for any $\psi \in H$. In addition, if $f_1, f_2 \in H^*$, then $\langle f_1 | f_2 \rangle_{H^*} = \langle \phi_{f_1} | \phi_{f_2} \rangle_H$ and $||f||_{H^*} = ||\phi_f||_H$.*

Since the mapping $f_\psi(\phi) = \langle \psi | \phi \rangle$ is a functional for any $\psi \in H$, the last theorem establishes a bijection between H and H^*. Instead of f_ψ, the notation $\langle \psi |$ is used and $\langle \psi |$ is called a **bra vector** corresponding to the ket vector $|\psi\rangle$. In this notation $\langle \psi | (|\phi\rangle) = \langle \psi | \phi \rangle$.

In the case of n-dimensional complex Hilbert spaces a "ket" $|\psi\rangle$ can be considered as an n-dimensional column vector and a "bra" $\langle \phi|$ as an n-dimensional row vector. The **scalar product** $\langle \phi | \psi \rangle$ is then a complex number—the result of a usual "row vector $\times$ column vector" product. The transformation $|\phi\rangle \leftrightarrow \langle \phi|$ corresponds to a transposition and conjugation. The **tensor product** $|\psi\rangle\langle\phi|$ is a matrix, the result of a usual column vector $\times$ row vector product.

Exercise 9.2.16 *Show that if trace(A), or short Tr(A), of a matrix A denote the sum of its diagonal elements, then (a) Tr$(|\psi\rangle\langle\phi|) = \langle\phi|\psi\rangle$; Tr$(a_1|\psi_1\rangle\langle\phi_1| + a_2|\psi_2\rangle\langle\phi_2|) = a_1\langle\phi_1|\psi_1\rangle + a_2\langle\psi_2|\phi_2\rangle$.*

Exercise 9.2.17 *Show: (a) $(|\psi_1\rangle + |\psi_2\rangle)\langle\phi| = |\psi_1\rangle\langle\phi| + |\psi_2\rangle\langle\phi|$; (b) $|\psi\rangle(\langle\phi_1| + \langle\phi_2|) = |\psi\rangle\langle\phi_1| + |\psi\rangle\langle\phi_2|$; (c) $(a|\psi\rangle)\langle\phi| = a(|\psi\rangle\langle\phi|) = |\psi\rangle(a\langle\phi|)$.*

9.2.2 Linear operators

Another key concept of Hilbert spaces and also of quantum computing is that of a linear operator.

Definition 9.2.18 *By a (**linear**) **operator** of Hilbert space H we understand a linear mapping $A : H \to H$ (the domain of which forms a subspace of H).*[39] *A linear operator A is called* **positive** *or* **semi-definite** *(notation $A \geq 0$) if $\langle\psi|A\psi\rangle \in \mathbf{R}^{\geq 0}$ for every vector $\psi \in H$. If A_1, A_2 are linear operators, then we write $A_1 \geq A_2$ if and only if $A_1 - A_2 \geq 0$. Let A be an operator of a Hilbert space H such that the closure of the domain of A is H. The* **adjoint operator** *of A is an operator A^* the domain of which consists exactly of those $\psi \in H$ for which the linear mapping $\phi \to \langle\psi|A\phi\rangle$ is continuous on the domain of A, and for all $\psi \in \mathrm{domain}(A^*)$, $\langle\psi|A\phi\rangle = \langle A^*\psi|\phi\rangle$. If $A = A^*$, then the operator A is called* **self-adjoint** *or* **Hermitian**. *By $\mathcal{L}(H)$ we denote the class of all self-adjoint operators of H.*

[39] T is a linear mapping if $T(\alpha\phi + \beta\psi) = \alpha T(\phi) + \beta T(\psi)$.

Each linear operator A is represented by a matrix M_A. (Often we take A to be the matrix M_A.) In the case of an n-dimensional Hilbert space, the matrix has degree n. If an operator A is self-adjoint, then the corresponding matrix M_A is **Hermitian**, i.e., $M_A = M_A^*$.

If A is a self-adjoint operator in H_n, then there exist n **eigenvalues** λ such that the (eigenvalue) equation $A\phi = \lambda\phi$ has a nonzero-vector solution ϕ_λ such that $A\phi_\lambda = \lambda\phi_\lambda$. ϕ_λ is called an **eigenvector** corresponding to the eigenvalue λ. The set of eigenvalues of an operator A is called its **spectrum**.

Using the basic properties of scalar products one can easily deal with the following problems:

Exercise 9.2.19 *Show that all eigenvalues of self-adjoint operators are real.*

Exercise 9.2.20 *Show that two eigenvectors corresponding to different eigenvalues of a self-adjoint operator are orthogonal.*

The n eigenvalues of A need not all be distinct. Let there be exactly k different eigenvalues $\lambda_1, \ldots, \lambda_k$ of A and let $m(i)$ be the multiplicity of λ_i. In such a case the set of eigenvectors corresponding to λ_i forms a subspace, the so-called **eigenspace**, denoted by H_{λ_i} of H_n of the dimension $m(i)$. An eigenvalue of multiplicity 1 (> 1) is called nondegenerate (degenerate). A matrix M is called nondegenerate (degenerate) if (not) all of its eigenvalues are nondegenerate.

For any Hilbert space H and any of its self-adjoint operators A there is an orthonormal basis of H consisting of eigenvectors of A.

A special role among operators of a Hilbert space have projections. If $H = W \oplus W^\perp$ is a decomposition of a Hilbert space H into two orthogonal subspaces W and $W^\perp$, then each $\phi \in H$ has a unique representation $\phi = \phi_W + \phi_{W^\perp}$, where $\phi_W \in W$ and $\phi_{W^\perp} \in W^\perp$. In such a case the mapping

$$P_W\phi = \phi_W$$

is an operator called the **projection** onto the subspace W. If W is a subspace spanned by a single vector ϕ, then instead of $P_{\{\phi\}}$ we write simply P_ϕ.

Exercise 9.2.21 *Show that an operator A is a projection onto some subspace if and only if A is* **idempotent** *(i.e., $A = A^2$), and self-adjoint.*

Let us denote by $\mathcal{P}(H)$ the set of all projections of a Hilbert space H.

Exercise 9.2.22 *Show: (a) $\langle P_W\phi|\psi\rangle = \langle\phi|P_W\psi\rangle$ holds for any subspace W and any vectors ϕ, ψ; (b) the square of the length of the projection of ϕ into W, i.e. $||P_W\phi||^2$ is $\langle\phi|P_W\phi\rangle$.*

Therefore, if $|\psi\rangle = \sum_{i=1}^n \langle\phi_i|\psi\rangle|\phi_i\rangle$ for an orthogonal basis $\{|\phi_1\rangle, \ldots, |\phi_n\rangle\}$, then $P_{\phi_i}|\psi\rangle = \langle\phi_i|\psi\rangle|\phi_i\rangle$.

If $\phi \in H$, $||\phi|| = 1$, then the operator denoted by $|\phi\rangle\langle\phi|$ and defined by

$$|\phi\rangle\langle\phi|(|\psi\rangle) = \langle\phi|\psi\rangle|\phi\rangle$$

is actually the projection P_ϕ onto the one-dimensional space spanned by the vector $|\phi\rangle$.

Exercise 9.2.23 *Let W be a subspace of a Hilbert space H. Show that a projection operator onto W can be expressed using projection operators $|\phi\rangle\langle\phi|$.*

In the case of finite-dimensional Hilbert spaces, if $\lambda_1, \ldots, \lambda_n$ are (not necessarily different) eigenvalues of an $n \times n$ matrix M_A of a linear operator A, and $\phi_1, \ldots, \phi_n$ are corresponding eigenvectors that form an orthonormal base, then A has the following, the so-called **spectral representation**

$$A = \sum_{i=1}^{n} \lambda_i |\phi_i\rangle\langle\phi_i| = \sum_{i=1}^{n} \lambda_i P_{\phi_i}$$

An important functional is that of the trace *Tr*. We shall consider this operator in the following, unless explicitly mentioned otherwise, only for finite dimensional Hilbert spaces. The **trace** of an operator T is usually defined by

$$Tr(T) = \sum_{\phi \in \mathcal{B}} \langle\phi|T\phi\rangle$$

where $\mathcal{B}$ is an orthonormal basis of H. This definition is O.K. because it can be shown, see also execises below, that the trace of an operator is invariant under a change of basis. Let $\mathcal{T}(H)$ be the set of all operators of H with finite traces.

Exercise 9.2.24 *Show: (a) $Tr(cA) = cTr(A)$ for any $c \in \mathbf{C}$ and any linear operator A; (b) $Tr(A_1 + A_2) = Tr(A_1) + Tr(A_2)$ for any linear operators A_1, A_2; (c) the trace functional is cyclically invariant, namely: $Tr(A_1A_2) = Tr(A_2A_1)$ and $Tr(A_1A_2A_3) = Tr(A_2A_3A_1)$ for any linear operators A_1, A_2, A_3.*

Exercise 9.2.25 *Show that if P is a projection operator into a k-dimensional space, then $Tr(P) = k$.*

Another important case is when a linear operator is bounded.

Definition 9.2.26 *A linear operator A is called* **bounded** *if there is an $s \in \mathbf{R}^{\geq 0}$ such that $||A\psi|| \leq s||\psi||$ for all $\psi \in H$.*

If A is a linear bounded operator then its **norm** *$||A||$ is defined by*

$$||A|| = \sup\{\frac{||A\psi||}{||\psi||}\,|\,)0 \neq \psi \in H\}.$$

Exercise 9.2.27 *Show that $|\langle A\phi|\psi\rangle| \leq ||A||^2$ and $|\langle\phi|A\phi\rangle| \leq ||A||^2$ for any bounded operator A and any states ϕ, ψ.*

Exercise 9.2.28 *Show that $||A + B|| \leq ||A|| + ||B||$ and $||AB|| \leq ||A||\ ||B||$ for any two bounded operators A, B.*

All linear operators of finitely dimensional Hilbert spaces are bounded.

For two linear operators (matrices) A, B, the expression

$$[A, B] = AB - BA$$

is called a **commutator**. If $[A,B]=0$, we say that operators A,B **commute**. In the case of commuting observables the order of the corresponding measurements is not of importance. Two operators A,B are said to **anticommute** if $AB+BA=0$, notation $\{A,B\}=0$.

Exercise 9.2.29 *Determine when two operators $|\phi_1\rangle\langle\phi_1|$ and $|\phi_2\rangle\langle\phi_2|$ commute, if $|\phi_1\rangle$ and $|\phi_2\rangle$ are pure states.*

One can show that two Hermitian operators commute if and only if there is a basis in which they both have diagonal form.

9.2.3 Mixed states and density matrices

There are two types of **states** in a Hilbert space: **(pure) states** and **mixed states** (**mixtures**).

There are several ways of defining pure states. In physics literature a position is often taken that each vector represents a state and two vectors represent the same state if one is a scalar multiple of the other. A similar way is to consider **rays**, sets of vectors that can be obtained from each other by a scalar multiplication (by a real number), and to say that states are rays. The most common in quantum computing literature is to consider as states vectors of norms 1.

A pure state is a state of an isolated quantum system. Evolution of pure states is described by unitary matrices.

A probability distribution $\{(p_i,\phi_i)\,|\,1\leq i\leq k\}$ on pure states $\{\phi_i\}_{i=1}^k$, with probabilities $0<p_i\leq 1$, $\sum_{i=1}^k p_i=1$, is called a **mixed state** or **(probabilistic) mixture**, and denoted $[\psi\rangle=\{(p_i,\phi_i)\,|\,1\leq i\leq k\}$ or

$$[\psi\rangle = p_1\phi_1\oplus\ldots\oplus p_k\phi_k=\bigoplus_{i=1}^k p_i\phi_i.$$

If all p_i are the same the term **random mixture** is sometimes used. In general, a quantum system is not in a pure state. There may be several reasons for that. A source to produce a quantum state is not perfect and it produces the state ϕ_i with probability p_i. It may also be the case that we have only partial knowledge about a system. For example, in an interpretation of quantum mechanics the result of the measurement of a pure state

$$|\psi\rangle=\sum_{i=1}^n\alpha_i|\phi_i\rangle$$

with respect to the observable given by an orthonormal basis $\{\phi_i\}_{i=1}^n$ can be considered as the mixed state

$$[\psi\rangle=\bigoplus_{i=1}^n|\alpha_i|^2|\phi_i\rangle.$$

In general, quantum processes are not always fully specified. Mixtures and density matrices are a suitable framework to deal with such cases.

To each mixed state $[\psi\rangle$ corresponds a **density matrix** $\rho_{[\psi\rangle}$. If $[\psi\rangle=|\phi\rangle=\sum_{i=1}^n c_i|\phi_i\rangle$ is a pure state, then $\rho_{[\psi\rangle}=|\phi\rangle\langle\phi|$, i.e. $\rho_{[\psi\rangle}(i,j)=c_i^*c_j$. For example, if $|\phi\rangle=\alpha|0\rangle+\beta|1\rangle$, then

$$|\phi\rangle\langle\phi|=\begin{pmatrix}|\alpha|^2 & \alpha\beta^*\\ \alpha^*\beta & |\beta|^2\end{pmatrix}.$$

Remark 9.2.30 The traditional meaning of density matrices is related to their representation of mixed states—eigenvalues of density matrices correspond to probabilities for finding the system in their corresponding eigenvectors. More modern approaches see density matrices as an alternative to vectors of Hilbert space to describe states of quantum systems.

Correspondence between mixed states and density matrices is not one-to-one. For example, if

$$[\psi_1\rangle = \frac{1}{2}|0\rangle \oplus \frac{1}{2}|1\rangle \text{ and } [\psi_2\rangle = \frac{1}{2}(|0\rangle + |1\rangle) \oplus \frac{1}{2}(|0\rangle - |1\rangle)$$

are two mixed states over the standard basis in H_2, then

$$\rho_{[\psi_1\rangle} = \rho_{[\psi_2\rangle} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} = \frac{1}{2}\mathbf{I}.$$

Exercise 9.2.31 *Show that if $|\phi_1\rangle, |\phi_2\rangle$ are orthonormal states of H_2, then the density matrix for the state $[\psi\rangle = \frac{1}{2}|\phi_1\rangle \oplus \frac{1}{2}|\phi_2\rangle$ has the form $\frac{1}{2}\mathbf{I}$.*

Density matrices are linear operators, so-called **density operators**, of Hilbert space and have various interesting and important properties. Observe that if all pure states of a mixed state are states of a Hilbert space H then the corresponding density matrix is an element of the space $H^* \otimes H$.

Exercise 9.2.32 *Show: (a) each density operator ρ is self-adjoint, positive and $Tr(\rho) = 1$. (b) a density operator ρ is the density operator of a pure state if and only if $\rho^2 = \rho$.*

Exercise 9.2.33 *Show an example of a mixed state with two pure states and another example of a mixed state with three pure states such that the density matrices of both of these mixed states are the same.*

In a more general setting, density matrices are arbitrary positive and self-adjoint operators with trace 1.

Density matrices completely specify all practically distinguishable properties of mixed states.

Example 9.2.34 *Consider two spin-$\frac{1}{2}$ particles which are far apart and in the entangled state $\frac{1}{\sqrt{2}}(|0\rangle|1\rangle + |1\rangle|0\rangle)$. If an observer of the first particle measures the particle with respect to the basis $\{|0\rangle, |1\rangle\}$, then the outcomes $|0\rangle$ or $|1\rangle$ are obtained with the same probability. The observer of the second particle then sees his particle as being in the state $|1\rangle$ in the first case and in the state $|0\rangle$ in the second case. In total he can say that the second particle is in the mixed state $(\frac{1}{2}, |0\rangle) \oplus (\frac{1}{2}, |1\rangle)$.*

If the observer of the first particle measures his particle with respect to the dual basis $\{|0'\rangle, |1'\rangle\}$, then again he gets $|0'\rangle$ with probability $\frac{1}{2}$ and $|1'\rangle$ with the same probability. In such a case the second particle is in the state $|1'\rangle$ or $|0'\rangle$ depending on the outcome of the first measurement. One can say again that the second particle is in the mixed state $(\frac{1}{2}, |0'\rangle) \oplus (\frac{1}{2}, |1'\rangle)$. However, to both of these mixtures the same density matrix $\frac{1}{2}\mathbf{I}$ corresponds, which fully characterizes the state of the second particle.

Example 9.2.35 *Let us assume that to polarize a stream of photons we use for each photon one of the following two methods.*

(a) A photon is randomly polarized either in the vertical or horizontal direction.

(b) A photon is randomly polarized either in the right-to-left or left-to-right direction.

In both cases we have the same density matrix $\frac{1}{2}\mathbf{I}$ *and the receiver has no way to find out which of these two methods was used.*

To each mixed state ρ correspond pure states in the following sense.

Definition 9.2.36 *Let* ρ *be a mixed state of a Hilbert space* H. *A* **purification** *of* ρ *is any pure state* $|\psi\rangle$ *in some compound Hilbert space* $H \otimes H'$ *such that* $\rho = Tr_{H'}|\psi\rangle\langle\psi|$ *(i.e., purification of* ρ *is any pure state having* ρ *as the reduced state for a subsystem).*

If ρ is nondegenerate and has n different eigenvalues λ_i, $1 \leq i \leq n$ and the corresponding eigenvectors are ϕ_i, then the spectral representation of ρ has the form

$$\rho = \sum_{i=1}^{n} \lambda_i P_{\phi_i},$$

where all $\lambda_i > 0$, because ρ is positive. In such a case ρ can be seen as representing a mixture of mutually orthogonal pure states ϕ_i with probabilities λ_i. This decomposition into orthogonal states is unique, but, as already mentioned, ρ is also decomposable into a mixture of (in general) non-orthogonal pure states ϕ_i.

If some eigenvalues of ρ are degenerate, then

$$\rho = \sum_{i} \lambda_i P_{\lambda_i}$$

and the eigenvalues correspond to mutually orthogonal subspaces. The subspace H_λ corresponding to the degenerate eigenvalue λ is generated by the set of solutions of the eigenvalue equation $\rho\omega = \lambda\omega$. In this case, it is not possible to represent ρ as a mixture of a unique set of orthogonal pure states.

Pure states, represented by idempotent density operators, are homogeneous in the sense that no idempotent density operator is expressible as a nontrivial convex sum of two or more different density operators.

Exercise 9.2.37 *Show that if* ρ *is idempotent and* $\rho = p_1\rho_1 + p_2\rho_2$, $p_1 + p_2 = 1$, $p_1, p_2 > 0$, *then* $\rho = \rho_1 = \rho_2$.

Remark 9.2.38 A pure state, represented by an idempotent density operator is pure in the sense that it is homogeneous or extremal with respect to the set of states represented by density operators. A mixed state represents a mixture only in the sense that it is non-homogeneous, not in the sense that it represents a unique mixture of pure states.

Exercise 9.2.39 *(Llyod, 1997) Show that if for two mixtures* $[\phi\rangle = \{(|\phi_i\rangle, p_i) \mid 1 \leq i \leq k\}$ *and* $[\psi\rangle = \{(|\psi_i\rangle, q_i) \mid 1 \leq i \leq k\}$ *we define* $\langle\phi|\psi\rangle = \sum_{j=1}^{k} \sqrt{p_j q_j}\langle\psi_j|\phi_j\rangle$, *then this "scalar product" indeed has all the properties a scalar product should have.*

A density matrix of pure states of n qubits has dimension 2^n. It can be reduced to a subset of m qubits by a tracing operation in which the rest of the system, represented by a 2^{n-m} dimensional Hilbert space $\mathcal{F}$, is **traced out**. The resulting matrix is denoted $\rho_{|A} = Tr_{\mathcal{F}}\rho$ and it is defined by $\rho_{|A}(i,j) = \sum_{k=1}^{2^{n-m}} \rho(ik, jk)$.

Schmidt decomposition theorem

Pure states of a compound Hilbert space $H_1 \otimes H_2$ can be decomposed in the following special and useful way into states of H_1 and H_2.

Theorem 9.2.40 *If H_1 and H_2 are Hilbert spaces of dimensions d_1 and d_2, $|\phi\rangle$ is a pure state of $H_1 \otimes H_2$ and $\rho = |\phi\rangle\langle\phi|$, $\rho_1 = Tr_{H_2}\rho$, $\rho_2 = Tr_{H_1}\rho$, then*

$$|\phi\rangle = \sum_{i=1}^{r} \sqrt{\lambda_i}|\phi_i, \psi_i\rangle, \tag{9.10}$$

where $r \leq \min\{d_1, d_2\}$ is the dimension of the space of non-zero eigenvectors of ρ_1, λ_i are common eigenvalues of ρ_1 and ρ_2 and $\{|\phi_i\rangle\}$, $\{|\psi_i\rangle\}$ are orthonormal eigenvectors of ρ_1 and ρ_2.

Proof. (Lo and Chau, 1997) Clearly $|\phi\rangle = \sum_{i=1}^{d_1} |\phi_i, \psi'_i\rangle$, where $\{\phi_i\}_{i=1}^{d_1}$ is an orthonormal basis of H_1. Therefore

$$\begin{aligned} Tr_{H_2}|\phi\rangle\langle\phi| &= Tr_{H_2}\sum_{i=1}^{d_1}\sum_{j=1}^{d_1}|\phi_j,\psi'_j\rangle\langle\psi'_i,\phi_i| = \sum_{i=1}^{d_1}\sum_{j=1}^{d_1}\sum_{k=1}^{d_2}\langle\theta_k,\psi'_j\rangle|\phi_j\rangle\langle\phi_i|\langle\psi'_i|\theta_k\rangle \\ &= \sum_{i=1}^{d_1}\sum_{j=1}^{d_1}\sum_{k=1}^{d_2}\langle\psi'_i|\theta_k\rangle\langle\theta_k|\psi'_j\rangle|\phi_j\rangle\langle\phi_i| = \sum_{i=1}^{d_1}\sum_{j=1}^{d_1}\langle\psi'_i|\psi'_j\rangle|\phi_j\rangle\langle\phi_i| \end{aligned}$$

where $\{\theta_k\}_{k=1}^{d_2}$ is an orthonormal basis of H_2. Moreover,

$$\rho_1 = \sum_{i=1}^{r}\lambda_i|\phi_i\rangle\langle\phi_i|$$

and therefore $\langle\psi'_i|\psi'_j\rangle = \delta_{ij}\lambda_i$ for all i,j. Hence $|\psi_i\rangle = \sqrt{\lambda_i}|\psi'_i\rangle$ is an orthonormal set in H_2 and therefore the Schmidt decomposition holds. In addition, by computing $Tr_{H_1}|\phi\rangle\langle\phi|$ one gets $\rho_2 = \sum_{i=1}^{r}\lambda_i|\phi_i\rangle\langle\phi_i|$ and therefore $|\phi_i\rangle$ is the eigenstate of ρ_2 corresponding to λ_i.

Superoperators

Transformations on density matrices are performed by **superoperators**—linear operators that map linear operators of a Hilbert space H_1 on linear operators of another Hilbert space H_2. The informal idea of physically implementable superoperators is captured by the formal concept of **completely positive maps**. They are superoperators T that are positive (they map positive semi–definite Hermitian matrices into positive semi-definite Hermitian matrices), trace preserving and such that for any finite dimensional Hilbert space H also the operator $T \otimes I_H$ has the same property, where I_H is the identity superoperator on H. (This means that positivity must remain if Hilbert spaces H_1 and H_2 are extended by adding more qubits.)

Examples of superoperators are encoders, decoders and quantum channels. Superoperators can also be seen as unitary operators in a larger Hilbert space.

9.2.4 Probabilities and observables

Probabilistic interpretation of quantum physics is one of its main features.

A self-adjoint operator A of a Hilbert space H, with spectral representation

$$A = \sum_{i=1}^{k} \lambda_i P_{\lambda_i},$$

where $\lambda_1, \ldots, \lambda_k$ are all its distinct eigenvalues, can be seen as an observable that represents decomposition of H into subspaces H_{λ_i} corresponding to eigenvalues λ_i, and with λ_is the numerical value associated to the subspace H_{λ_i}.

If the spectrum of A is simple, then the probability assigned to the eigenvalue λ_i by a unit vector ψ is defined as follows:

$$prob_\psi(\lambda_i) = |\langle\omega_i|\psi\rangle|^2 = ||P_{\omega_i}\psi||^2,$$

where ω_i is the eigenvector corresponding to λ_i. Therefore, $prob_\psi(\lambda_i)$ is the square of the length of the projection of ψ into ω_i. If A is degenerate, then

$$prob_\psi(\lambda_i) = ||P_{\lambda_i}\psi||^2,$$

where P_{λ_i} is the projection operator into the subspace H_{λ_i}.

In general we define for any state $|\psi\rangle$ and any subspace W:

$$prob_\psi(W) = Tr(\rho P_W),$$

where $\rho = |\psi\rangle\langle\psi|$ is the density operator. In general we define for a density operator ρ and $|\psi\rangle \in H$:

$$prob_\rho(W) = Tr(\rho \mathcal{P}_W).$$

Exercise 9.2.41 *Show for density operators* ρ, ρ_1, ρ_2 *and subspaces* W_1, W_2 *(a)* $prob_\rho(W_1 \oplus W_2) = prob_\rho(W_1) + prob_\rho(W_2)$*; (b)* $prob_{\rho_1\otimes\rho_2}(W_1 \otimes W_2) = prob_{\rho_1}(W_1) prob_{\rho_2}(W_2)$.

The **expectation value** of A in the state ψ is defined by

$$Exp_\psi(A) = \sum_{i=1}^{k} \lambda_i prob_\psi(\lambda_i).$$

Often we use notation $\langle\psi|A|\psi\rangle$ or $\langle A\rangle_\psi$ instead of $Exp_\psi(A)$ or even $\langle A\rangle$ if ψ is clear from the context.

Exercise 9.2.42 *Show that: (a)* $Exp_\psi(A) = \langle\psi|A\psi\rangle$*; (b) Show that* $Exp_\psi(A) = Tr(P_\psi A)$.

In the case of a mixed state ψ given by the pure states $\psi_1, \ldots, \psi_k$ and probabilities $p_1, \ldots, p_k$, with $\sum_{i=1}^{k} p_i = 1$ we have

$$Exp_\psi(A) = \sum_{i=1}^{k} p_i Exp_{\psi_i}(A) = \sum_{i=1}^{k} p_i \langle\psi_i|A\psi_i\rangle.$$

Exercise 9.2.43 *Show for the mixed state ψ given by pure states $\psi_1, \dots, \psi_k$ and probabilities $p_1, \dots, p_k$ that $\sum_{i=1}^k p_i Exp_\psi(A) = Tr((\sum_{i=1}^k p_i P_\psi)A)$.*

The concept of observables is one of the most specific and important in quantum computing. The key point is that any pair consisting of an observable A and a state T determines a probability distribution of values of the observable A on state T.

In the classical setting, examples of "observables" are position, velocity, energy, momentum, and so on. Two of them, position and momentum, can be seen as "canonical" because other classical observables can be expressed in terms of these two. In the Hilbert space framework of quantum mechanics, observables are operators of a special type.

9.2.5 Evolution of quantum states

The dynamic evolution of states in a Hilbert space is described by a linear Schrödinger time-dependent equation in the form

$$i\hbar \frac{\partial \psi(t)}{\partial t} = H\psi(t), \text{ or } i\hbar \frac{\partial |\psi(t)\rangle}{\partial t} = H|\psi(t)\rangle,$$

where each $\psi(t)$ is a pure state. This evolution is characterized by a unitary operation $U(t) = e^{-iHt/\hbar}$ such that $|\psi(t)\rangle = U(t)|\psi(0)\rangle$.

Exercise 9.2.44 *Determine the matrix $U(t)$ for the case $H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and determine its eigenvalues.*

It follows from the above principle that unitary time evolution never turns a pure state into a mixed state. Moreover, it holds for any $t_1, t_2 \geq 0$ that $U(t_1 + t_2) = U(t_1)U(t_2)$ and for any $|\phi\rangle \in H$, $t_0 \in \mathbf{R}$

$$\lim_{t \to t_0} U(t)|\phi\rangle = U(t_0)|\phi\rangle.$$

The unitary operator $U(t)$ defines therefore a one-parameter group of transformations of the Hilbert space.

Since $P_\psi|\phi\rangle = \langle\psi|\phi\rangle|\psi\rangle$ for any vectors $|\phi\rangle$ and $|\psi\rangle$, we have

$$P_{U(t)|\psi\rangle}|\phi\rangle = \langle U(t)\psi|\phi\rangle U(t)|\psi\rangle = U(t)P_\psi U^{-1}(t)|\phi\rangle.$$

On this base one can show that evolution of a mixed state, represented by the density operator $U(t)$, is given by

$$U(t) \to \rho(t) = U(t)\rho(0)U^{-1}(t).$$

9.2.6 Measurements

Experiments show, and this has been formulated as one of the main principles of the Copenhagen interpretation, that on the macroscopic level we can receive, as outcomes of measurements, only classical values (states) and not quantum superpositions. However, on the quantum level the outcome is the quantum state obtained by the "collapse" of the measured state.

The basis of the main approach to the measurement in quantum computing is von Neumann's projection postulate: observation of a state ψ by an observable A provides, as the result of the measurement, one of the eigenvalues of A and makes the state ψ collapse and renormalize. For finite-dimensional Hilbert spaces this has been formally stated as follows:

Let H be a Hilbert space of dimension n corresponding to a quantum system S. Let

$$A = \sum_{i=1}^{n} \lambda_i |\phi_i\rangle\langle\phi_i|$$

be a spectral representation of an observable A, where λ_i are its eigenvalues and ϕ_i is its eigenvector corresponding to λ_i and $\{\phi_i\}_{i=1}^{n}$ is an orthonormal basis of H.

Observation of a state $\psi = \sum_{i=1}^{n} |\phi_i\rangle$ by A produces the value λ_i with probability $|\sum_{j=1}^{k} c_{i_j}|^2$, where $c_{i_1}, \ldots, c_{i_k}$ are all such values that $\lambda_{i_1} = \lambda_{i_2} = \ldots = \lambda_{i_{k-1}} = \lambda_{i_k} = \lambda_i$, and in addition the post-measurement state of the system is $\sum_{j=1}^{k} \lambda'_{i_j} |\phi_{i_j}\rangle\langle\phi_{i_j}|$, where $\lambda'_{i_j} = \frac{\lambda_{i_j}}{\sqrt{\sum |c_{i_j}|^2}}$.

Example 9.2.45 (Polarization of photons) *Consider a polarized light. To test vertical/horizontal polarization, which corresponds to the computational basis* $\{|0\rangle, |1\rangle\}$, $|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $|1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, *we need to use the observable*

$$\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

which has ± 1 *as eigenvalues and* $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ *are the corresponding eigenvectors.*

To test for diagonal polarization $\pm 45°$, *which corresponds to the basis* $\{|0'\rangle, |1'\rangle\}$, *we need to use the observable*

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

with eigenvalues ± 1 *and eigenvectors* $\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$ *and* $\begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$.

Finally, to test for circular polarization, which corresponds to the basis $|0''\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle)$, $|1''\rangle = \frac{1}{\sqrt{2}}(i|0\rangle + |1\rangle)$ *we need to use the observable*

$$\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},$$

with eigenvalues ± 1 *and eigenvectors* $\begin{pmatrix} \frac{1}{\sqrt{2}} \\ i\frac{1}{\sqrt{2}} \end{pmatrix}$, $\begin{pmatrix} \frac{i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$.

9.2.7 Tensor products and Hilbert spaces

In order to consider compound quantum systems formally, the concept of the direct sum and of the tensor product of Hilbert spaces, vectors and matrices are of importance.

Let H be a Hilbert space with a basis $\mathcal{A} = \{|\phi_1\rangle, \ldots |\phi_n\rangle\}$ and let H' be a Hilbert space with the basis $\mathcal{B} = \{|\psi_1\rangle, \ldots, |\psi_m\rangle\}$, and with $\mathcal{A} \cap \mathcal{B} = \emptyset$.

The **direct sum** $H \oplus H'$ is the Hilbert space of dimensionality $\dim(H \oplus H') = \dim(H) + \dim(H')$, and with the base $\{|\phi_1\rangle, \ldots, |\phi_n\rangle, |\psi_1\rangle, \ldots, |\psi_m\rangle\}$.

If $|\phi\rangle = \sum_{i=1}^{n} c_i|\phi_i\rangle \in H$ and $|\psi\rangle = \sum_{j=1}^{m} d_j|\psi_j\rangle \in H'$, then $|\phi\rangle \oplus |\psi\rangle$ is defined to be $\sum_{i=1}^{n} c_i|\phi_i\rangle + \sum_{j=1}^{m} d_j|\psi_j\rangle$.

If now A is a linear operator (matrix) on $\mathcal{H}$ and B a linear operator on $\mathcal{H}'$ then we define:

$$A \oplus B = \begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & B \end{pmatrix}$$

This means that $A \oplus B$ is applied component-wise in the following sense:

$$(A \oplus B)(|\phi\rangle \oplus |\psi\rangle) = A(|\phi\rangle) \oplus B(|\psi\rangle).$$

The **tensor product** $H_1 \otimes H_2$ of Hilbert spaces H_1 and H_2 is of the dimension $\dim(H_1 \otimes H_2) = \dim(H_1)\dim(H_2)$ and with the basis $\mathcal{A} \times \mathcal{B} = \{(|\phi_i\rangle|\psi_j\rangle)|\phi_i\rangle \in \mathcal{A}, |\psi_j\rangle \in \mathcal{B}\}$.

Instead of $(|\phi_i\rangle, |\psi_j\rangle)$ we usually write $|\phi_i\rangle|\psi_j\rangle$ or simply $|\phi_i\psi_j\rangle$. In such a case we define

$$|\phi\rangle \otimes |\psi\rangle = \sum_{i=1}^{n}\sum_{j=1}^{m} c_i d_j |\phi_i\psi_j\rangle.$$

If $|\phi_1\rangle, |\phi_2\rangle$ are states of H_1 and $|\psi_1\rangle, |\psi_2\rangle$ are states of H_2, then $|\phi_1\rangle \otimes |\psi_1\rangle$ and $|\phi_2\rangle \otimes |\psi_2\rangle$ are states of H and their scalar product is defined by

$$\langle \phi_1 \otimes \psi_1 | \phi_2 \otimes \psi_2 \rangle = \langle \phi_1|\phi_2\rangle\langle\psi_1|\psi_2\rangle.$$

The tensor product $A \otimes B$ of $n \times n$ matrices $A = \{a_{ij}\}_{i,j=1}^{n}$ and $B = \{b_{ij}\}_{i,j}^{n}$ is an $nm \times nm$ matrix s defined by

$$A \otimes B = \begin{pmatrix} a_{11}B & \ldots & a_{1n}B \\ \vdots & & \vdots \\ a_{n1}B & \ldots & a_{nn}B \end{pmatrix}$$

Exercise 9.2.46 *Show that: (a) the operation $\otimes$ on matrices is associative; (b) $Tr(A \otimes B) = Tr(A)\,Tr(B)$ for matrices A and B.*

Exercise 9.2.47 *Show, for matrices A, A_1, A_2, B, B_1, B_2: (a) $(A \otimes B)(|\phi\rangle \otimes |\psi\rangle) = (A(|\phi\rangle) \otimes B(|\psi\rangle)$; (b) $(A_1 + A_2) \otimes B = (A_1 \otimes B) + (A_2 \otimes B))$.*

9.2.8 Generalized measurements-POV measurements

von Neumann's projection measurements are not general enough and a maximal projection measurement is not always the most efficient way to get information about the state of a quantum system. More general, and at the same time more common, are indirect measurements at which an auxiliary quantum system, an ancilla, is first prepared in a standard state and connected with the original system and then a measurement on both systems is performed. Peres (1993) is the basis for the following motivation and demonstration of POVM. For more about POVM see Bush et al. (1997).

Quantum information gain

Let us have a quantum source that produces a pure state $|\phi_i\rangle$ with probabilities p_i, $1 \leq i \leq n$. and let us assume that we know probability $P_{\mu i}$ such that the measurement of $|\phi_i\rangle$ produces the value μ.

Given μ, Bayes Theorem can be used to determine a posteriori probability $Q_{i\mu}$ that given μ it was the state ϕ_i the measurement of which produced μ. Indeed,

$$Q_{i\mu} = \frac{p_i P_{\mu i}}{q_\mu},$$

where $q_\mu = \sum_{j=1}^{n} P_{\mu j} p_j$ is the apriori probability of the outcome μ at the measurement.

Before the measurement only the probabilities p_i were known and Shannon's entropy of the system was $-\sum_{i=1}^{n} p_i \lg p_i$. After μ was obtained, the new entropy is $H_\mu = -\sum_{i=1}^{n} Q_{i\mu} \lg Q_{i\mu}$.

In some cases H_μ may be larger than the initial entropy. However, on the average a quantum measurement reduces the Shannon entropy and the average information gain is

$$I_{av} = -\sum_{i=1}^{n} p_i \lg p_i - \sum_{\mu} q_\mu H_\mu.$$

The generalized measurements represent a way to increase information gain.

POV measurements

Let a quantum system $\mathcal{S}$ is in the mixed state represented by a density matrix ρ and let H be the corresponding Hilbert space. Let us consider an auxiliary system $\mathcal{S}_a$, an ancilla, in the state ρ_a. The resulting density matrix of the combined system $S \otimes S_a$ is $\rho \otimes \rho_a$.

Let the maximal measurement is performed on $\rho \otimes \rho_a$ in $\mathcal{S} \otimes \mathcal{S}_a$. In such a case different outcomes of the measurement correspond to projection P_μ to different orthogonal subspaces and it holds

$$P_\mu P_\nu = \begin{cases} P_\mu, & \text{if } \mu = \nu; \\ 0, & \text{if } \mu \neq \nu, \end{cases} \qquad \sum_\mu P_\mu = I.$$

In such a case the probability that the outcome μ is obtained by measuring ρ is given by

$$P_{\mu i} = Tr[P_\mu(\rho_i \otimes \rho_a)] = \sum_{mr,ns} P_\mu[mr,ns]\rho[n,m]\rho_a[s,r],$$

and therefore

$$P_{\mu i} = Tr(A_\mu \rho_i),$$

where A_μ is defined by

$$A_\mu[m,n] = \sum_{rs} P_\mu[mr,ns]\rho_a[s,r],$$

and A_μ is an operator on H.

The set of all operators A_μ is called a **positive operatot valued measure** (POVM), because each A_μ is a positive operator. The matrices A_μ do not commute in general and satisfy the relation

$$\sum_\mu A_\mu = I.$$

One difference between von Neumann's projection measurement and POVM is that in the later case the number of possible outcomes may be larger than the dimensionality of the underlying Hilbert space H. The probability of the outcome μ is given by $Tr(A_\mu \rho)$ instead of von Neumann's $Tr(P_\mu \rho)$.[40]

Information gain via POVM

The question is that if a mixed state is given by p_i and ρ_i (for a pure state ϕ_i), what is its Shannon entropy before and after measurement with POVM as well as how to choose a POVM so that the gain is maximized.

Example 9.2.48 *Four matrices $\frac{1}{4}(I \pm \sigma_x)$, $\frac{1}{4}(I \pm \sigma_z)$ form a POVM. Consider the case of the mixed state consisting of four states $|0\rangle$, $|1\rangle$, $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle$, all with probability $\frac{1}{4}$. In this case Shannon entropy before measurement is $\lg 4$ and after testing $\lg 4 - \frac{1}{2}\lg 2$.*

It can be shown that that the average information gain is

$$I_{av} \leq S(\sum_{i=1}^{n} p_i \rho_i) - \sum_{i=1}^{n} p_i S(\rho_i)$$

and the optimal POVM consists of matrices of rank 1, see Peres (1993) for more details.

[40] POVM measurement is also related with the general view of a physical process which starting in a state ψ produces a random classical outcome μ and causes a collapse of ψ into the state ψ_μ of another Hilbert space H'. This is then formalized through the mapping $\mu \to M_\mu$, where $M_\mu : H \to H'$ is a positive operator called the measurement operator associated with the initial Hilbert space H into the final Hilbert space H'. The only requirement on M_μ is that $\sum_\mu M_\mu^* M_\mu = I$. The mapping $\mu \to E_\mu = M_\mu^* M_\mu$ is then called POVM. The probability of μ for a given mixed state ρ is then $Tr(M_\mu^* \rho)$.

9.3 Deterministic and Randomized Computing

Basic theoretical understanding of the computational possibilities of classical computing has been developed using various models of Turing machines with time and space as main computational resources and the corresponding complexity classes. The main potentials and limitations of computing have emerged into two main computational thesis. These basic concepts and results are summarized in this section. For more see, for example, Gruska (1997).

9.3.1 Computing models

Several basic models and modes of computing will be introduced in this section.

Deterministic models

The very basic model of the universal computer, invented in 1937 by A. M. Turing, is that of **one–tape Turing machine** (TM). A Turing machine $\mathcal{M}$, see Figure 9.14, consists of a **bi–infinite tape**, divided into an infinite number of cells in both directions, with one distinctive starting cell, or 0-cell. Cells of the tape can contain any symbol from a finite **tape alphabet** Γ, or a symbol $\sqcup$ representing the empty cell; a **read–write head** is positioned at any moment of the discrete time on a cell; a **finite control** unit is always in one of the states of a finite set Q of states and implements a transition function

$$\delta : Q \times \Gamma \rightarrow Q \times \Gamma \times \{\leftarrow, \downarrow, \rightarrow\}.$$

The interpretation of $\delta(q, \sigma) = (q', \sigma', d)$ goes as follows: if $\mathcal{M}$ is in the state q and the head reads σ, then $\mathcal{M}$ enters the state q', replaces σ by σ' in the cell the head is currently on, and the head moves in the direction d (to the right (left) if $d = \rightarrow$ ($\leftarrow$) or does not move at all if $d = \downarrow$. Formally, $\mathcal{M} = \langle \Gamma, Q, q_0, \delta \rangle$ or $\mathcal{M} = \langle \Sigma, \Gamma, q_0, Q_t, \delta \rangle$ if a set $\Sigma \subset \Gamma$ is considered as the **input alphabet** and $Q_t \subseteq Q$ contains the so-called **terminating states**. We can assume that $Q_t = \{ACCEPT, REJECT\}$. It is assumed that once $\mathcal{M}$ gets into a terminating state than it remains in such a state.

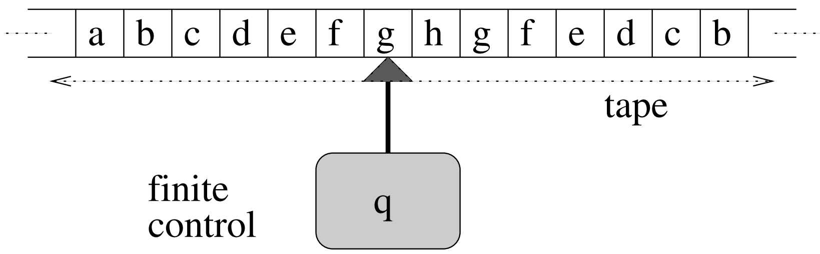


Figure 9.14: One-tape Turing machine

A **computation** of $\mathcal{M}$ is defined formally using the concept of **configuration** (q, w, w'), where $q \in Q$ and $w, w' \in \Gamma - \{\sqcup\}$, which provides a complete description of the current **global state** of $\mathcal{M}$: the state q the machine $\mathcal{M}$ is in; the contents ww' of the tape (with leftmost and rightmost infinite number of $\sqcup'$s discarded), and with the position of the head on the first symbol of w'. If $\mathcal{M}$ moves in one (several) steps from a configuration C to a configuration C', we write $C \vdash_{\mathcal{M}} C', (C \vdash^*_{\mathcal{M}} C')$. Each configuration (q_0, ε, w) is called

initial, and each configuration (q, w, w'), $q \in Q_t$, is called **terminating**. A **(terminating) computation** is a finite sequence of configurations $C_0, C_1, C_2, \ldots, C_k$ such that C_0 is the initial configuration, $C_i \vdash_{\mathcal{M}} C_{i+1}$ if $0 \leq i < k$, and C_k is a terminating configuration.

If $\mathcal{M}$ starts with an initial configuration (q_0, ε, x) and its computation ends in a halting configuration (q, w, w'), then we write $\mathcal{M}_t(x) = ww'$ and $\mathcal{M}_s(x) = q$. If, for some input string x, $\mathcal{M}$ does not halt at all, i.e., its computation is infinite, we write $\mathcal{M}(x) = \nearrow$.

Exercise 9.3.1 *Show that for each TM $\mathcal{M}$ we can design another Turing machine $\mathcal{M}'$ that simulates $\mathcal{M}$ and in no case it can move into a state both from right and left.*

Turing machines are a natural computer model to study, on the one hand, language acceptance and decision problems and, on the other, computation of string-to-string and integer-to-integer functions.

Definition 9.3.2 *(1) Let $\mathcal{M} = \langle \Sigma, \Gamma, Q, q_0, \delta \rangle$ be a TM with the input alphabet Σ. Then*

$$L(\mathcal{M}) = \{w \mid w \in \Sigma^*, \mathcal{M}_s(w) = ACCEPT\}$$

is the language, over Σ, **accepted** *by $\mathcal{M}$. In addition, if $\mathcal{M}_s(x) \in \{ACCEPT, REJECT\}$ for any $x \in \Sigma^*$, then $L(\mathcal{M})$ is said to be the language* **decided (recognized)** *by $\mathcal{M}$.*

(2) A language $L \subseteq \Sigma^*$ is said to be recursively enumerable if there is a TM $\mathcal{M}$ that accepts $L = L(\mathcal{M})$, and it is called **recursive**, *if there is a TM that decides (recognizes) L.*

Definition 9.3.3 *(1) A (partial) string-to-string function $f : \Sigma^* \to \Sigma^*$ is said to be (partially) computable by a TM $\mathcal{M} = \langle \Sigma, \Gamma, Q, q_0, \delta \rangle$, $\Sigma \subseteq \Gamma$, if $\mathcal{M}_t(x) = f(x)$ for any $x \in \Sigma^*$ from the domain of f and $\mathcal{M}(x) = \nearrow$, otherwise.*

(2) If there is a TM $\mathcal{M}$ that (partially) computes a function $f : \Sigma^* \to \Sigma^*$, then f is called **(partially) recursive**.

(3) A function $f : \mathbf{N}^t \to \mathbf{N}^s$ is called **(partially) recursive**, *if there is a TM $\mathcal{M}$ such that $f(x_1, \ldots, x_t) = (y_1, \ldots, y_s)$, if and only if*

$$\mathcal{M}_t(1^{x_1+1}01^{x_2+1}0\ldots01^{x_t+1}) = 1^{y_1+1}0\ldots01^{y_s+1}.$$

Exercise 9.3.4 *A TM, as defined above, can perform in one step three actions: a state change, a writing and a head move. Show that to each TM $\mathcal{M}$ we can design a TM $\mathcal{M}'$ which performs in each step at most two of these three elementary actions and: (a) accepts the same language as $\mathcal{M}$; (b) computes the same function as $\mathcal{M}$.*

Exercise 9.3.5 *Design a Turing machine that: (a) multiplies two integers; (b) recognizes whether a given word is a palindrome; (c) recognizes whether a given integer is prime.*

A natural and important generalization of one-tape Turing machines are **multitape Turing machines** (MTM)—see Figure 9.15. A k-tape Turing machine has k tapes, one of them is the input tape, and there is a head on each of the tapes. Formally, a k tape MTM is defined as $\mathcal{M} = \langle \Sigma, \Gamma, Q, q_0, \delta \rangle$, where

$$\delta : Q \times \Gamma^k \to Q \times \Gamma^k \times \{\leftarrow, \downarrow, \rightarrow\}^k.$$

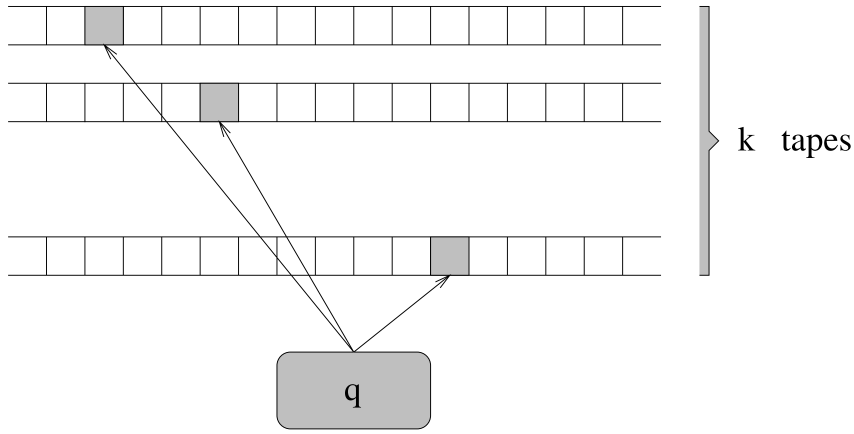


Figure 9.15: Multitape Turing machines

If, at the beginning of a computation step, $\mathcal{M}$ is in the state q, its head on the $i$the tape reads the symbol σ_i and $\delta(q, \sigma_1, \ldots, \sigma_k) = (q', \sigma_1', \ldots, \sigma_k', d_1, \ldots, d_k)$, then $\mathcal{M}$ moves, in one step, into the state q', the head on the ith tape replaces σ_i by σ_i' and moves in the direction d_i. One-tape TM can be seen as a special case of MTM.

It is straightforward to introduce basic concepts concerning **time resources** for computation on MTM. If an MTM $\mathcal{M}$ starts with a string w on its input tape and with all other tapes empty and yields in m steps a terminating configuration, then m is the time of the computation of $\mathcal{M}$ on the input w. Denote by $Time_{\mathcal{M}}(n)$ the maximal number of steps of $\mathcal{M}$ for inputs of length n. $\mathcal{M}$ is said to operate within the time bound $f(n)$ for a function $f : \mathbf{N} \to \mathbf{N}$, or to be $f(n)$-**time bounded**, if $\mathcal{M}$ halts within $f(|w|)$ steps, for any input $w \in \Sigma^*$. If a language L is accepted by an $f(n)$-time bounded MTM, then we write $L \in Time(f(n))$. Thus, $Time(f(n))$ is the family of languages that can be decided by an $f(n)$-time bounded MTM—a **time complexity class**.

Similarly, one can introduce analogous concepts for space as a computational resource of MTM. $Space_{\mathcal{M}}(n)$ is the maximal numbers of cells, on any of the tapes, $\mathcal{M}$ uses when computing with inputs of length n. $\mathcal{M}$ is said to be $s(n)$-space bounded, where $s : \mathbf{N} \to \mathbf{N}$, if $\mathcal{M}$ uses at most $s(|w|)$ cells on any of the tapes for any input w. $Space(s(n))$ is the family of languages that can be accepted with space bound $s(n)$—a **space complexity class**.

Of key importance for classical computing is the fact that there is a **universal** MTM U that can efficiently, i.e., in polynomial time, simulate any other MTM.

The basic idea of such a universal MTM is very simple. If U gets on its input tape a word w and a description (encoding) $\langle\mathcal{M}\rangle$ of an MTM $\mathcal{M}$ (a description of its transition function), in the form of a word in the tape alphabet of U, then in order to simulate one step of $\mathcal{M}$ on an input w, U scans $\langle\mathcal{M}\rangle$ to determine the step $\mathcal{M}$ would perform on w and then U performs on w this particular step. Time for a simulation of one step of U can be made proportional to the length of $\langle\mathcal{M}\rangle$.

There are many ways an MTM can be encoded by a word. Let us illustrate one of them for the case of a one-tape TM $\mathcal{M} = \langle\Gamma, Q, q_1, Q_t\delta\rangle$, where $\Gamma = \{a_1, a_2, \ldots, a_n\}$, $Q = \{q_1, \ldots, q_m\}$, $Q_t = \{q_{m+1}, q_{m+2}\}$. $\mathcal{M}$ can be encoded by a binary string.

For example,

$$\text{a transition } \delta(q_i, a_j) = (q_k, a_l, d) \text{ can be encoded as } 0^i 10^j 10^k 10^l 10^{d'},$$

where $d' = 1$ if $d = \leftarrow$, $d' = 2$ if $d = \downarrow$, and $d' = 3$ if $d = \rightarrow$. The whole TM $\mathcal{M}$ can then be encoded by the string

$$\langle\mathcal{M}\rangle = 10^n 110^m 11 \underbrace{0^{i_1}10^{j_1}10^{k_1}10^{l_1}10^{d'_1}}_{\text{first transition}} 11 \underbrace{\ldots\ldots\ldots\ldots}_{\text{other transitions}} 111.$$

and a $w \in \Gamma^*$, $w = a_{i_1} \ldots a_{i_k}$ can be encoded by $\langle w\rangle = 0^{i_1}10^{i_2}1\ldots 10^{i_k}$.

Another very basic model of Turing machines is the oracle Turing machine—see Figure 9.16. This model is to capture the idea of computing with a help (of an oracle or a subroutine). Let us illustrate this concept again on one-tape oracle Turing machines.

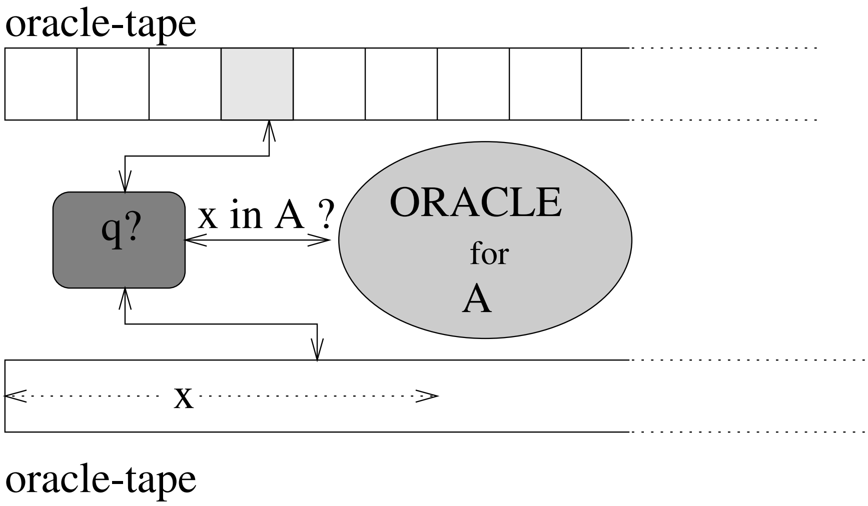


Figure 9.16: Oracle Turing machine

A (one-tape) **oracle Turing machine** $\mathcal{M}$, with the oracle–tape alphabet Δ and a language $A \subseteq \Delta^*$ as an **oracle**, is a Turing machine with two tapes, an ordinary read-write tape and a special write–only **oracle-tape**. In addition, $\mathcal{M}$ has one special **pre-query state**, say $q?$, and two **post–query states** q^+ and q^-, such that whenever $\mathcal{M}$ comes to the state $q?$, then the next state is either q^+ or q^-, depending on whether the contents x of the oracle–tape at that moment is, or is not, in A. In other words, when $\mathcal{M}$ gets into the "query" state $q?$, this can be seen as $\mathcal{M}$ asking the oracle the question about the membership of the word x, written on the oracle–tape, in A. In addition, and this is crucial, it is assumed that the oracle's answer is "free" and immediate (because the oracle is supposed to be all-powerful—as oracles should be). In other words, a transition from the state $q?$ to one of the states q^+ or q^- takes one step, as for all other transitions.

Denote by $\mathcal{M}^A$ an oracle Turing machine $\mathcal{M}$ with the oracle A—the same TM can be connected with different oracles—and let $L(\mathcal{M}^A)$ denote the language accepted by such an oracle Turing machine with the oracle A.

Nondeterministic model

All models of TM introduced so far can be considered as realistic models of classical computing. This is not the case for so-called nondeterministic Turing machines—a mathematically very natural generalization of TM that plays a very important role in the study of the potentials and limitations of classical computing.

A one-tape **nondeterministic Turing machine** (NTM) $\mathcal{M} = \langle \Gamma, Q, q_0, \delta \rangle$ is defined formally in a similar way to a one-tape deterministic TM, except that instead of a transition function we have a **transition relation**

$$\delta \subset Q \times \Gamma \times Q \times \Gamma \times \{\leftarrow, \downarrow, \rightarrow\}.$$

This means that a step of an NTM is not in general uniquely determined and several alternative steps may be offered to choose (nondeterministically). As a consequence, a configuration C of an NTM may have several potential next configurations, and $\mathcal{M}$ can go, nondeterministically, from C to one of them. We can therefore see the overall computational process of an NTM not as a sequence (of subsequent) configurations, but as a tree of (branching) configurations—see Figure 9.17.

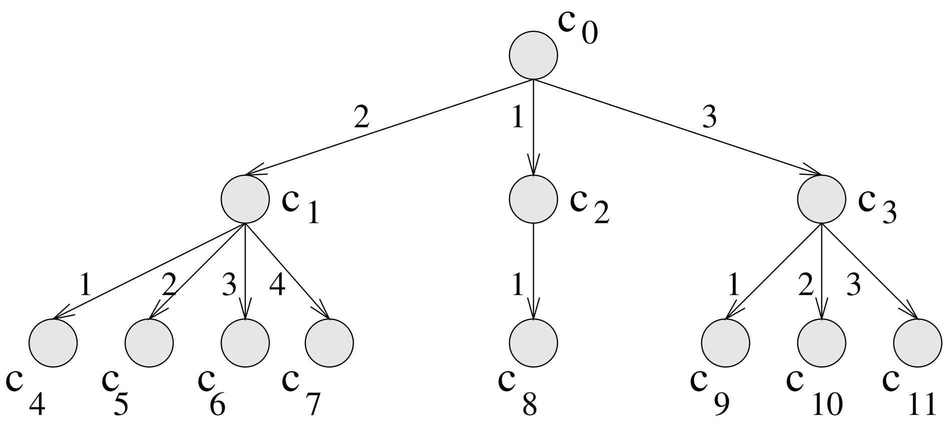


Figure 9.17: Tree of configurations of NTM

We say that an NTM $\mathcal{M}$ accepts an input w (in time $t(|w|)$ and space $s(|w|)$), if there is at least one path in the configuration tree, with q_0w being the configuration at the root, that ends in the accepting state (and it has the length at most $t(|w|)$, and none of the configurations on the path is longer than $s(|w|)$). This can be used to define, in a natural way, when an NTM computes a relation or a function with certain time and space bounds. For an NTM $\mathcal{M}$ let $L(\mathcal{M})$ be the language accepted by $\mathcal{M}$.

Exercise 9.3.6 *Show that for each NTM $\mathcal{M}$ we can design an NTM $\mathcal{M}'$ that can make exactly two moves in each nonterminating configuration, accepts the same language as $\mathcal{M}$, and there is an integer k such that $\mathcal{M}$ accepts an input w in t steps if and only if $\mathcal{M}'$ accepts w in kt steps.*

Another way to see an NTM is as having a transition function δ in the form

$$\delta : Q \times \Gamma \times Q \times \Gamma \times \{\leftarrow, \downarrow, \rightarrow\} \rightarrow \{0, 1\}.$$

The fact that $\delta(q, \sigma, q', \sigma', d) = 1\ (= 0)$ is interpreted that if $\mathcal{M}$ is in the state q and its head reads σ, then $\mathcal{M}$ may (may not) perform the following move: changes its state q into q', replaces σ by σ' and moves the head in the direction d.

Exercise 9.3.7 *Design an NTM that decides in polynomial time whether a given Boolean formula is satisfiable.*

Complexity classes for NTM: Denote by NTime($t(n)$) (NSpace($s(n)$)) the family of languages accepted by $t(n)$-time bounded ($s(n)$-space bounded) NTM and denote

$$\mathbf{NP} = \bigcup_{k=0}^{\infty} \mathrm{NTime}(n^k), \qquad \mathbf{NPSPACE} = \bigcup_{k=0}^{\infty} \mathrm{NSpace}(n^k).$$

9.3.2 Randomized computations

Another important machine model for the study of the potentials and limitations of computing nare **probabilistic Turing machines** (PTM). They are defined in a similar way to NTM. The transition function of a PTM $\mathcal{M} = \langle \Gamma, Q, q_0, \delta \rangle$ is a mapping

$$\delta : Q \times \Gamma \times Q \times \Gamma \times \{\leftarrow, \downarrow, \rightarrow\} \rightarrow [0,1]$$

that has to satisfy the following **local probability condition**: For each $q \in Q$, $\sigma \in \Gamma$

$$\sum_{q',\sigma',d'} \delta(q, \sigma, q', \sigma', d') = 1.$$

$\delta(q, \sigma, q', \sigma', d')$ is considered as the probability that if $\mathcal{M}$ is in the state q and reads σ, then it makes the following move: the state q is changed into q', σ is replaced by σ' on the square scanned by the head, and the head moves in the direction d'.

The transition function δ induces the probability for any two configurations C and C' that C moves into C' in one step.

A computation of a PTM $\mathcal{M}$ can be again seen as a configuration tree in which to each configuration transfer a probability is assigned—see Figure 9.18. From that one can derive the probability that a computation reaches a particular node—this is the product of probabilities assigned to all edges (transfers) on the path from the root to the particular node.

In contrast to NTM, probabilistic TM are considered to be a realistic model of computation.

Exercise 9.3.8 *A quantum computation step can be seen as a matrix–vector multiplication. Show that in a similar way we can see: (a) deterministic; (b) nondeterministic; (c) probabilistic computations and describe which conditions have to satisfy, in all these cases, the corresponding matrices.*

Algorithms that make use of random numbers (generators) are often very simple, and their efficiency is either comparable to or better than that of deterministic algorithms for the same problem.

Example 9.3.9 (Randomized QUICKSORT) *To sort a sequence $S = (a_1, \ldots, a_n)$ of n different elements one can use the following algorithm:*

Algorithm 9.3.10 (RQUICKSORT)

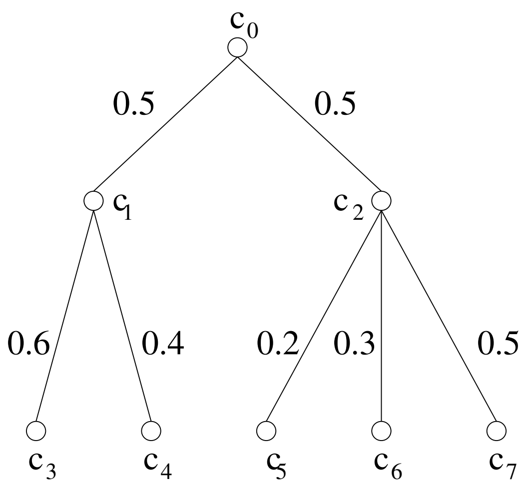


Figure 9.18: Configuration tree of a PTM-80%

1. *Choose randomly* $1 \leq j \leq n$.
2. *Partition* S *into sequences* S_1 *and* S_2 *of elements* $a_i \leq a_j$ *and* $a_i > a_j$, *respectively.*
3. *Sort recursively* S_1 *and* S_2.

If the random choice of j*'s corresponds to the uniform distribution, then RQUICKSORT requires on average* $\Theta(n \lg n)$ *time.*

Exercise 9.3.11 *Design a randomized algorithm to determine, given a sequence of* n *elements and* $1 \leq k \leq n$, *the* k*th smallest element.*

9.3.3 Complexity classes

The amount of computational resources needed to solve an algorithmic problem on a computer is clearly of practical importance. It is also intuitively clear that by limiting computational resources the class of problems that can be solved with the given resources gets smaller (or at least not larger).

One of the most important insights that theory of computation brought up is that the study of the power of computational resources is extremely stimulating, deep and leads to very fundamental results concerning the power and limitations of information processing.

Complexity classes of deterministic computing

The very basic classes of classical deterministic computing are:

$$\begin{array}{rclcrcl} \mathbf{P} & = & \bigcup_{k=0}^{\infty} \mathit{Time}(n^k), & \qquad & \mathbf{NP} & = & \bigcup_{k=0}^{\infty} \mathit{NTime}(n^k), \\ \mathbf{PSPACE} & = & \bigcup_{k=0}^{\infty} \mathit{Space}(n^k), & & \mathbf{NPSPACE} & = & \bigcup_{k=0}^{\infty} \mathit{NSpace}(n^k) \end{array}$$

If $\mathcal{L}$ is a family of languages, then $\mathbf{co} - \mathcal{L}$ is the family of languages complements of which are in $\mathcal{L}$. Of special importance is the class $\mathbf{co} - \mathbf{NP}$.

The basic relations between these complexity classes have the form

$$\mathbf{P} \subseteq \mathbf{NP} \subseteq \mathbf{PSPACE} = \mathbf{NPSPACE} \tag{9.11}$$

Important open problems of complexity theory are to determine whether particular inclusions in (9.11) are proper.

Classes **P** and **NP** are of special interest for theory of computation. The class **P** of languages (problems) decidable (solvable) in polynomial time was the first one to be consider as the class of those problems for which a feasible solution exists.

The class **NP** may seem to be quite artificial, because the underlying model of nondeterministic Turing machines is not a realistic one. Actually, the opposite is true. Its great importance stems from the fact that **NP** represents the class of problems for which one can easily (in polynomial time) check correctness of a prospective solution. This may be made more formal as follows.

Definition 9.3.12 *A binary relation $R \subseteq \Sigma^* \times \Sigma^*$ is called* **polynomially decidable** *if there is a deterministic Turing machine that decides the language $\{x\#y \mid (x,y) \in R\}$, with $\#$ being a marker not in Σ, in polynomial time. Moreover, a binary relation R is called* **polynomially balanced**, *if there is an integer k such that $(x,y) \in R$ implies $|y| \leq |x|^k$. (In other words, the size of y is polynomially bounded by the size of $|x|$, and therefore, if $(x,y) \in R$, y can be only polynomially larger than x.)*

It holds:

Theorem 9.3.13 *A language L is in* **NP** *if and only if there is a polynomially decidable and polynomially balanced relation R such that $L = \{x \mid (x,y) \in R \text{ for some } y\}$.*

Theorem 9.3.13 throws light on the substance of the class **NP**. If a problem (language) L is in **NP**, then each "yes" instance of the problem $x \in L$) has a short, of polynomial size, **certificate (witness, proof of membership)** y that x is in$L$. One can easily, in polynomial time, verify the correctness of a chosen certificate for x.

Another importance of the class **NP** comes from the fact that it contains so-called **NP-complete problems**, and many of them are among the most important algorithmic problems in many areas of science and engineering. In addition, **NP**-complete problems are the hardest to solve in **NP**. The basic definitions go as follows:

Definition 9.3.14 *A language $L_1 \subseteq \Sigma_1^*$ is polynomial time reducible to the language $L_2 \subseteq \Sigma_2^*$, if there is a polynomial time computable function $f : \Sigma_1^* \to \Sigma_2^*$ such that if $x \in \Sigma_1^*$, then $x \in L_1 \iff f(x) \in L_2$.*

A language L_0 is called **hard** *for* **NP** *if each language from* **NP** *is polynomial time reducible to L_0. L_0 is called* **NP-complete** *if L_0 is in* **NP** *and it is hard for* **NP**.

This means that if one would be able to show about one **NP**-complete language L_0 that $L_0 \in \mathbf{P}$, then $\mathbf{P} = \mathbf{NP}$.

Example 9.3.15 *Let us mention two* **NP***-complete problems: (1) to decide whether a given Boolean formula is satisfiable; (2) to decide whether a given graph has a Hamiltonian cycle.*

Prime recognition is a problem that is known to be in **NP**, it is not known to be in **P** and it does not seem to be **NP**-complete, see Gruska (1997).

Relativized complexity classes

If $\mathcal{L}$ is a complexity class and A is a language, then $\mathcal{L}^A$ denotes the **relativized class** $\mathcal{L}$, with respect to the oracle A; namely the class of languages accepted by an oracle Turing machine $\mathcal{M}^A$, where $\mathcal{M}$ is a Turing machine accepting a language from $\mathcal{L}$. If $\mathcal{A}$ is a family of languages, then $\mathcal{L}^{\mathcal{A}} = \bigcup_{A\in\mathcal{A}} \mathcal{L}^A$.

In the case where there is the relation $\mathcal{L}_1 \subseteq \mathcal{L}_2$ between two complexity classes $\mathcal{L}_1$ and $\mathcal{L}_2$ and all attempts to resolve the question whether $\mathcal{L}_1 = \mathcal{L}_2$ have failed, then the study of the relativized complexity classes $\mathcal{L}_1^A$ and $\mathcal{L}_2^A$, for different oracles A, is an attempt to get more insights into the relation between $\mathcal{L}_1$ and $\mathcal{L}_2$.

Two main questions to consider are:

1. Are there oracles A and B such that $\mathcal{L}_1^A \neq \mathcal{L}_2^A$ and $\mathcal{L}_1^B = \mathcal{L}_2^B$?
2. Is it true that $\mathcal{L}_1^R \neq \mathcal{L}_2^R$ for a randomly chosen oracle R?

For example, there are oracles A, B, C, D such that: (1) $\mathbf{P}^A = \mathbf{NP}^A$; (2) $\mathbf{P}^B \neq \mathbf{NP}^B$; (3) $\mathbf{NP}^C = \mathbf{PSPACE}^C$; (4) $\mathbf{P}^D \neq \mathbf{PSPACE}^D$.

The main outcome of these results is the understanding that methods that relativize (i.e. if they can be shown to get a relation between complexity classes then they can be used to show the same relation between the relativized complexity classes) cannot be used to separate such complexity classes as **P**, **NP** and **PSPACE**.

Randomized complexity classes

Randomized computing was introduced in order to solve some algorithmic problems more efficiently. The most fundamental question concerning the power and limitations of randomized computing is the one whether polynomial time randomized computing is indeed more powerful than deterministic polynomial time computing.

There are several ways to define polynomial time complexity classes for randomized computing. The main ones are classes **RP**, **PP**, $\mathbf{ZPP} = \mathbf{RP} \cap \mathbf{co-RP}$, and **BPP**. These classes can be defined as classes of languages L for which there is a randomized algorithm $\mathcal{A}$ such that it holds:

RP	$x \in L \Rightarrow Pr(\mathcal{A}(x) \text{ accepts}) \geq \frac{1}{2}$	$x \notin L \Rightarrow Pr(\mathcal{A}(x) \text{ accepts}) = 0$
PP	$x \in L \Rightarrow Pr(\mathcal{A}(x) \text{ accepts}) > \frac{1}{2}$	$x \notin L \rightarrow Pr(\mathcal{A}(x) \text{ accepts}) \leq \frac{1}{2}$
BPP	$x \in L \Rightarrow Pr(\mathcal{A}(x) \text{ accepts}) \geq \frac{3}{4}$	$x \notin L \Rightarrow Pr(\mathcal{A}(x) \text{ accepts}) \leq \frac{1}{4}$

Class **ZPP** can also be defined as the class of problems that can be solved by PTM with certainty in expected polynomial time.

Classes **ZPP**, **RP** and **PP** fall nicely into the hierarchy of deterministic complexity classes presented in the previous subsection. Indeed, it holds:

Theorem 9.3.16 $\mathbf{P} \subseteq \mathbf{ZPP} \subseteq \mathbf{RP} \subseteq \mathbf{PP} \subseteq \mathbf{PSPACE}$.

There are two simple models of probabilistic Turing machines that are sufficient and often more convenient for definition and study of polynomial time randomized complexity classes.

The first model (see Figure 9.19a), is that of a one-tape Turing machine enhanced by a special "randomized tape" and with the head which moves one square right after each reading from the random tape. Before any computation the randomized tape is assumed to be filled up with an infinite random binary string. At some steps of computation the symbol under the current position of the head on random tape is read—this corresponds to a "random coin-tossing" at such a step—and computation continues depending on the outcome of the reading.

Another elegant approach to the definition of the randomized complexity classes is through special non-deterministic Turing machines $\mathcal{M}$ in which there are exactly two next configurations to each non-terminating configuration and terminating states of $\mathcal{M}$ are divided into accepting and rejecting states. In addition, these Turing machines are assumed to have all computational paths of the same length. In this setting the classes **RP**, **PP** and **BPP** can be defined as follows:

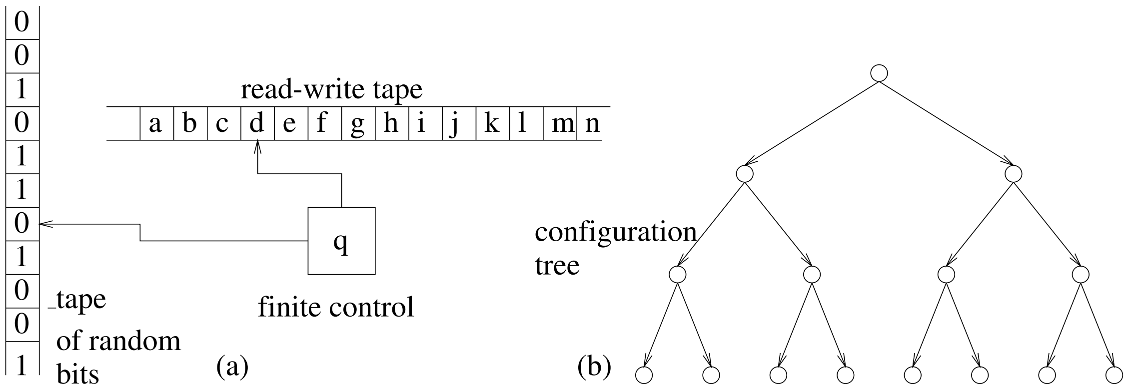


Figure 9.19: A Turing machine with a tape of random bits and a configuration tree of a special NTM

RP A language L is in **RP** (**r**andom **p**olynomial time) if there is a polynomial time-bounded NTM $\mathcal{M}$ such that if $x \in L$, then **at least half** of all computations of M on x terminate in the accepting state, and if $x \notin L$, then **all** computations of M on x terminate in the rejecting state. (*Monte Carlo acceptance* or *one-sided Monte Carlo acceptance.*)

PP A language L is in **PP** (**p**robabilistc **p**olynomial time) if there is a polynomial time bounded NTM $\mathcal{M}$ such that $x \in L$, if and only if **more than half** the computations of $\mathcal{M}$ on x terminate in the accepting state. (*Acceptance by majority.*)

BPP A language L is in **BPP** (**b**ounded error (away from $\frac{1}{2}$) **p**robabilistic **p**olynomial time) if there is a polynomial time-bounded NTM $\mathcal{M}$ such that

1. If $x \in L$, then at least $\frac{3}{4}$ of the computations of $\mathcal{M}$ on x terminate in the accepting state.

2. If $x \notin L$, then at least $\frac{3}{4}$ of the computations of $\mathcal{M}$ on x terminate in the rejecting state.

(*Acceptance by clear majority* or *two-sided Monte Carlo acceptance.*)

The class **BPP** is currently considered as the most important complexity class which contains languages (problems) solvable in polynomial randomized time and a problem (language) is considered as feasible if and only if it is in **BPP**.

It is clear that $\mathbf{P} \subseteq \mathbf{BPP}$ but neither of the relations $\mathbf{NP} \subseteq \mathbf{BPP}$ nor $\mathbf{BPP} \subseteq \mathbf{NP}$ has been shown yet. There is a belief, and a certain evidence, but not a proof, that $\mathbf{P} \subsetneq \mathbf{BPP} \subseteq \mathbf{NP}$. One technical result supporting this hypothesis is that each language in **BPP** has a polynomial time Boolean circuit but a similar result for **NP** does not seem to hold because for **NP**-complete problems no polynomial-size circuits are known.

Remark 9.3.17 *Prime recognition is known to be in* **ZPP**.

9.3.4 Computational theses

Relevance, robustness and the importance of the current theory of computability is based on the belief that the following thesis, relating an intuitive notion of effective computation with a formal concept of Turing machines, holds.

> **Church–Turing's computability thesis.** An abstract process can be naturally called an **effective** procedure if and only if it can be realized by a Turing machine.

There is plenty of convincing evidence in favour of the Church–Turing's thesis and no really hard argument against it.

Another key intuitive notion is that of **feasible computation**. The first quite successful attempt to find a formal equivalent of this concept was the one to identify decision problems for which feasible algorithms exist with the class **P** of decision problems and with the class **FP** of functions computable in polynomial time on multiple Turing machines.

The next step of the development has been an attempt to extend this view of feasible computation to the one realized in polynomial time by randomized algorithms. As the result of algorithmic observations and theoretical results in computational complexity, the following view has emerged.

> **Modern Church-Turing thesis.** Any (physically) reasonable model of computation can be **efficiently** simulated on a probabilistic Turing machine.

In the thesis we take the word "reasonable" to mean "in principle physically realisable" and by "efficiently realizable on a probabilistic Turing machine" to be in the class **BPP**.