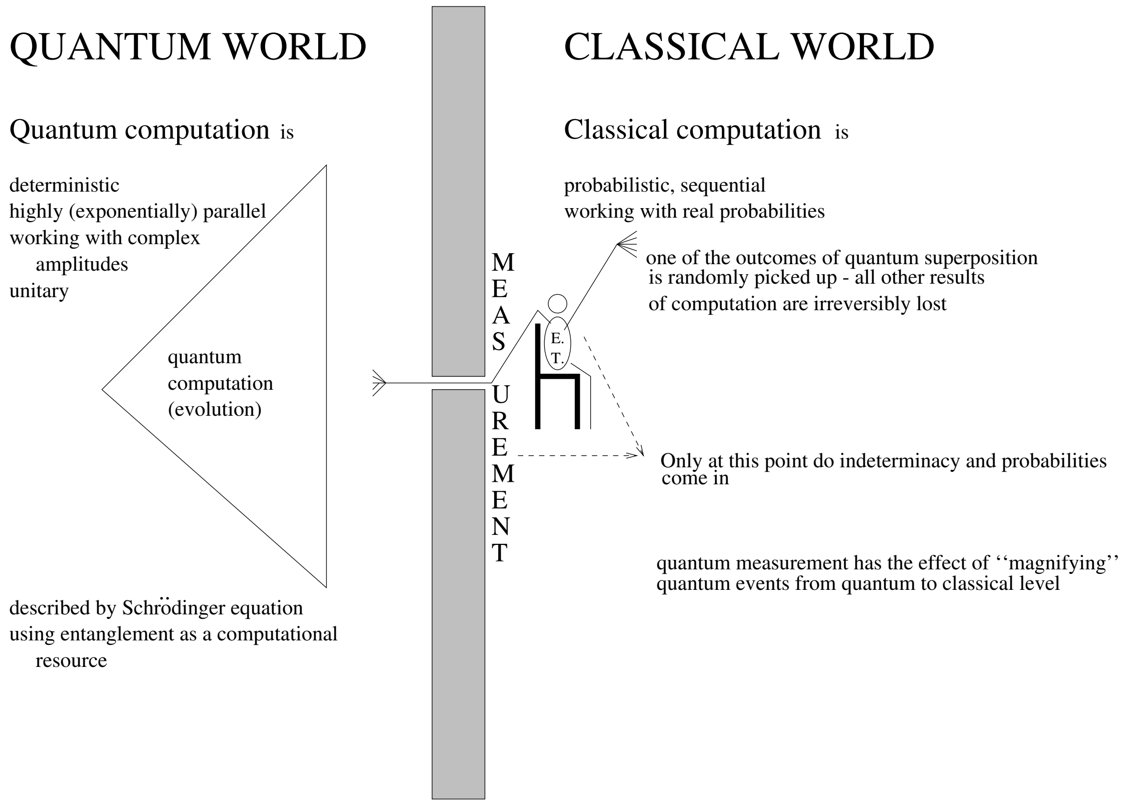
QUANTUM COMPUTING

Jozef Gruska



Chapter 1

FUNDAMENTALS

INTRODUCTION

The power of quantum computing is based on several phenomena and laws of the quantum world that are fundamentally different from those one encounters in classical computing: complex probability amplitudes, quantum interference, quantum parallelism, quantum entanglement and the unitarity of quantum evolution. In order to understand these features, and to make a use of them for the design of quantum algorithms, networks and processors, one has to understand several basic principles which quantum mechanics is based on, as well as the basics of Hilbert space formalism that represents the mathematical framework used in quantum mechanics.

The chapter starts with an analysis of the current interest in quantum computing. It then discusses the main intellectual barriers that had to be overcome to make a vision of the quantum computer an important challenge to current science and technology. The basic and specific features of quantum computing are first introduced by a comparison of randomized computing and quantum computing. An introduction to quantum phenomena is done in three stages. First, several classical and similar quantum experiments are analysed. This is followed by Hilbert space basics and by a presentation of the elementary principles of quantum mechanics and the elements of classical reversible computing.

LEARNING OBJECTIVES

The aim of the chapter is to learn

1. the main reasons why to be interested in quantum computing;
2. the prehistory of quantum computing;
3. the specific properties of quantum computing in comparison with randomized computing;
4. the basic experiments and principles of quantum physics;
5. the basics of Hilbert space theory;
6. the elements of classical reversible computing.

> You have nothing to do but mention the quantum theory, and people will take your voice for the voice of science, and believe anything.
>
> Bernard Shaw (1938)

Quantum computing is a big and growing challenge, for both science and technology. Computations based on quantum world phenomena, processes and laws offer radically new and very powerful possibilities and lead to different constraints than computations based on the laws of classical physics. Moreover, quantum computing seems to have the potential to deepen our understanding of Nature as well as to provide more powerful information processing and communication tools. At the same time the main theoretical concepts and principles of quantum mechanics that are needed to grasp the basic ideas, models and theoretical methods of quantum computing, are simple, elegant and powerful. This chapter is devoted to them.

Introduction of the basic concepts in this chapter will be detailed and oriented mainly to those having no, or close to no, knowledge of quantum physics and quantum information processing.

1.1 Why Quantum Computing

> Do not become attached to things you like, do not maintain aversion to things you dislike. Sorrow, fear and bondage come from one's likes and dislikes.
>
> Buddha

Quantum computing is without doubt one of the hottest topics at the current frontiers of computing, or even of the whole science. It sounds very attractive and looks very promising.

There are several natural basic questions to ask before we start to explore the concepts and principles as well as the mystery and potentials of quantum computing.

1. **Why to consider quantum computing at all?** The development of classical computers is still making enormous progress and no end of that seems to be in sight. Moreover, the design of quantum computers seems to be very questionable and almost surely enormously expensive. All this is true. However, there are at least four very good reasons for exploring quantum computing as much as possible.

- Quantum computing is a **challenge**. A very fundamental and very natural challenge. Indeed, according to our current knowledge, our physical world is fundamentally quantum mechanical. All computers are physical devices and all real computations are physical processes. It is therefore a fundamental challenge, and actually our duty, to explore the potentials, laws and limitations of quantum mechanics to perform information processing and communication.

All classical computers and models of computers, see Gruska (1997), are based on classical physics (even if this is rarely mentioned explicitly), and therefore they are not fully adequate. There is nothing wrong with them, but they do not seem to explore fully the potential of the physical world for information processing. They are good and powerful, but they should not be seen as reflecting our full view of information processing systems.[1]

Moreover, theoretical results obtained so far provide evidence that quantum computation represents the first real challenge to the modern, efficiency oriented, version of the Church-Turing thesis:

> *Any reasonable model of computation can be efficiently simulated by probabilistic Turing machines.*

- Quantum computing seems to be a **must** and actually our **destiny**. As miniaturization of computing devices continues, we are rapidly approaching the microscopic level, where the laws of the quantum world dominate. By Keyes (1988), an extrapolation of the progress in miniaturization shows that around 2020 computing should be performed at the atomic level. At that time, if the development keeps continuing as hitherto, one electron should be enough to store one bit, and the energy dissipation of $1kT \ln 2$ should be sufficient to process one bit.[2,3] Thus, not only scientific curiosity and challenges, but also technological progress requires that the resources and potentials of quantum computing be fully explored.[4]

- Quantum computing is a **potential**. There are already results convincingly demonstrating that for some important practical problems quantum computers are theoretically exponentially more powerful than classical computers. Such results, as Shor's factorization algorithm, can be seen as *apt killers* for quantum computing and have enormously increased activity in this area. In addition, the laws of quantum world, harvested through quantum cryptography, can offer, in view of our current knowledge, unconditional security of communication, unachievable by classical means.

- Finally, the development of quantum computing is a **drive** and gives new impetus to explore in more detail and from new points of view concepts, potentials, laws and limitations of the quantum world and to improve our knowledge of the natural world. The study of information processing laws, limitations and potentials is nowadays in general a powerful methodology to extend our knowledge, and this seems to be particularly true for quantum mechanics. Information is being identified as one of the basic

[1] At this point it should be made clear that quantum computers do not represent a challenge to the basic Church–Turing thesis concerning computability. They cannot compute what could not be computed by classical computers. Their main advantage is that they can solve some important computational tasks much more efficiently than classical computers.

[2] In such a case it will be necessary to include in the design and description of computers quantum theory and such quantum phenomena as superposition and entanglement, to obtain correct predictions about computer behaviour. However, the clear necessity to go deeper into the quantum level for improving performance of computers does not immediately imply that the way pursued under the current interpretation of the term "quantum computing" is the only one, or even the best one.

[3] The single electron transistor is already under development, see page 311.

[4] At the same time one should note that while quantum physics has been already for a long time essential to the understanding of the operations of transistors and other key elements of modern computers, computation remained to be a classical process. In addition, at the first sight there are good reasons for computing and quantum physics to be very far apart because determinism and certainty required from computations seem to be in strong contrast with uncertainty principle and probabilistic nature of quantum mechanics.

and powerful concepts of physics and quantum entanglement is an important communication resource. Several profound insights into the natural world have already been obtained on this basis.[5]

Remark 1.1.1 *The above ideas are so new and important, that they deserve an additional analysis.*

Historically, the fundamental principles of physics first concerned the problems of matter—what things are made of and how they move. Later, the problems of energy started to be reflected in the leading principles of physics—how energy is created, expressed and transformed. As the next stage an alternative seems to be to look to information processing for a new source of fundamental principles and basic laws. For example, concerning the particles, the questions of the movement of particles may be superseded by how particles can be utilized for information processing. Finally, let us observe some similarities between energy and information. Both of them have many representations, but basic principles, and also equations, hold independently of the form in which energy or information is presented.

The increasing importance of information processing principles for current science has been first, correctly, reflected in the views and understanding (due to Landauer, 1991), that "information is physical" and in the corresponding changes of emphases on the essence and ways to deal with information processing problems. However, it could be the case that this is only the first step and perhaps even more fundamental changes in the principles of physics could be obtained from the view that "physics is informational".[6]

These new views of the role of information in quantum physics also bring new potentials, challenges and questions for quantum physics. Is the well known "weirdness" of the quantum world due to the fact that physical reality is governed by even more basic laws of the information processing world? Is quantum theory a theory of the physical or of the information world? Can the study of quantum information help to deal with the most basic problems quantum theory has?

As an example of a change of research aims in physics under the influence of computer science research paradigms, consider quantum evolution. Traditionally, quantum physics has been concerned with the study or design of particular quantum systems and the study of various related fundamental problems. In addition to these problems quantum computing brought up new general and fundamental questions. Namely, what are the best, from well defined quantitative point of views, quantum evolutions to solve particular algorithmic or communication tasks. Or a problem of the maximum quantum computation power achievable in a quantum system of a certain dimension and disturbance level (Steane, 1998b), and of the way to reach such a maximum.

New fundamental questions in quantum mechanics are raised also in connection with the following problem: how secure are, or can be, quantum cryptographic protocols? For

[5]For example, manifestations of quantum nonlocality that go beyond entanglement (see Bennett et al. 1998), the use of quantum principles for secure transmission of classical information (quantum cryptography), the use of quantum entanglement for reliable transmission of quantum states over a distance (quantum teleportation), the possibility of preserving quantum coherence in the presence of irreversible noise processes (quantum error correction and fault tolerant computation). In addition, by Steane (1997), one has to realize that historically much of fundamental physics has been concerned with discovering fundamental particles of Nature and the equations which describe their motions and interactions. It now appears that a different program may be equally important. Namely, to discover the ways Nature allows, and prevents, information to be expressed and manipulated, rather than particles to move.

[6]A lot of research is still needed to determine the position and real role information plays in physics. The extreme views go even so far that information is a physical quantity, similar as energy in thermodynamics (Horodecki, 1991, and Landauer, 1991, 1995), or even that information is deeper than reality—a substance that is more fundamental than matter and energy.

example, the question how much information can be extracted from a quantum system for a given amount of expected disturbances? These questions are of fundamental importance far beyond quantum cryptography. To answer these questions, new theoretical insights and also new experiments seem to be needed.

In addition, an awareness has been emerging also in the foundations of computing that fundamental questions regarding computability and computational complexity are in a deep sense questions about physical processes.[7] If they are studied on a mathematical level then the underlying models have to reflect fully the properties of our physical world. This in particular implies that computational complexity theory has to be, in its most fundamental form, based on models of quantum computers.[8]

2. **Can quantum computers do what classical ones cannot?** The answer depends on the point of view. It can be YES. Indeed, the simplest example is generation of random numbers. Quantum algorithms can generate truly random numbers. Deterministic algorithms can generate only pseudo-random numbers. Other examples come from the simulation of quantum phenomena. On the other hand, the answer can be also NO. A classical computer can produce truly random numbers when attached to a proper physical source.

3. **Where lie the differences between the classical and quantum information processing?** Some of the differences have already been mentioned. Let us now discuss some others.

Classical information can be read, transcribed (into any medium), duplicated at will, transmitted and broadcasted. Quantum information, on the other hand, cannot be in general read or duplicated without being disturbed, but it can be "teleported" (as discussed in Section 6.4).

In classical randomized computing, a computer always selects one of the possible computation paths, according to a source of randomness, and "what-could-happen-but-did-not" has no influence whatsoever on the outcome of the computation. On the other hand, in quantum computing, exponentially many computational paths can be taken simultaneously in a single piece of hardware and in a special quantum way and "what-could-happen-but-did-not" can really matter.

Acquiring information about a quantum system can inevitably disturbs the state of the system. The tradeoff between acquiring quantum information and creating a disturbance of the system is due to quantum randomness. The outcome of a quantum measurement has a random element and because of that we are unable always faithfully infer the (initial) state of the system from the measurement outcome.

---

[7]An understanding has emerged that each specific computation is performed by a physical system evolving in time and, consequently, that one of the basic problems of computing, namely "what is efficiently computable?" is deeply related to one of the basic problems of physics, namely "which dynamical systems are physically realizable?"

[8]The following citations reflect a dissatisfaction with the fact that the development of complexity theory ignored one of its most fundamental tasks. The fact that this had been so is in one way explainable but, in another way, hardly forgivable.

A. Ekert (1995): *Computers are physical objects and computations are physical processes. The theory of computation is not a branch of pure mathematics. Fundamental questions regarding computability and computational complexity are questions about physical processes that reveal to us properties of abstract entities such as numbers or ideas. Those questions belong to physics rather than mathematics.*

J. Beckman et al (1996): *The theory of computation would be bootless if the computations that it describes could not be carried out using physically realizable devices. Hence it is really a task of physics to characterize what is computable, and to classify the efficiency of computations. The physical world is quantum mechanical. Therefore, the foundations of the theory of computation must be quantum mechanical as well. The classical theory of computation should be viewed as an important special case of a more general theory.*

Perhaps the main difference between classical and quantum information processing lies in the fact that quantum information can be encoded in mutual correlations between remote parts of physical systems and quantum information processing can make essential use of this phenomena—called **entanglement**—not available for classical information processing.

Another big difference between the classical and quantum worlds that strongly influences quantum information processing stems from the fact that the relationship between a system and its subsystems is different in the quantum world than in the classical world. For example, the states of a quantum system composed of quantum subsystems cannot be in general decomposed into states of these subsystems.

4. **Can quantum computers solve some practically important problems much more efficiently?** Yes. For example, integer factorization can be done in polynomial time on quantum computers what seems to be impossible on classical computers. Searching in unordered database can be done provably with less queries on quantum computer.

5. **Where does the power of quantum computing come from**? On one side, quantum computation offers enormous parallelism. The size of the computational state space is exponential in the physical size of the system and the energy available. A quantum bit can be in any of a potentially infinite number of states and quantum systems can be simultaneously in superposition of exponentially many of the basis states. A linear number of operations can create an exponentially large superposition of states and, in parallel, an exponentially large number of operations can be performed in one step.

Secondly, it is the branching and quantum interference that create parallel computation and constructive/destructive superpositions of states and can amplify or destroy the impacts of some computations. Due to this fact, we can, in spite of the peculiarities of quantum measurements, utilize quantum parallelism.

Thirdly, it is mainly the existence of so-called "entangled states" that makes quantum computing more powerful than classical and allows even very distant parts of systems to be strongly tied. This creates a base for developing and exploring quantum teleportation and other phenomena that are outside of the realm of the classical world.

*****

After all this excitement let us start to deal with more prosaic and "harder "questions.

6. **Where are the drawbacks and bottlenecks of quantum computing?** There are, unfortunately, quite a few. Let us mention here only two of them.

- Quantum computing can provide enormous parallelism. However, there are also enormous problems with harnessing the power of its parallelism. According to the basic principles of quantum mechanics, a (projection) measurement process can get out of (large) quantum superposition only one classical result, randomly chosen, and the remaining quantum information can be irreversibly destroyed.
- An interaction of a quantum system with its environment can lead to the the so-called decoherence effects and can greatly influence, or even completely destroy, subtle quantum interference mechanisms. This appears to make long reliable quantum computations practically impossible.

7. **How feasible are (powerful) quantum computers and really important quantum information processing applications?** It is too early to give a definite answer. On one side, there is a strong scientific belief, based on long term experiences of science, that something very important will come out of the research in quantum computing.

On the other hand, one has to admit that many of the current exciting results concerning quantum computing should be seen as **Gedanken experiments**. Namely, one works with systems (experiments) that perhaps do not exist, or cannot be performed in the real world, or only with enormous difficulty, but do not contradict any known law within a (certain) consistent theory of quantum mechanics.[9] Such considerations, systems and results are usually taken as being in principle acceptable.

In addition, in the recent years quite impressive progress has been made on the experimental level and ways have been found to deal with many problems that seemed to prevent the utilization of the power of quantum computing. Especially experimental quantum cryptography has made formidable progress to show that long distance optical fiber, open-air and even earth-satellites quantum key generation seems to be feasible.

Finally, it seems quite safe to assume that either quantum computing will meet its expectations or something new and important will be learned and our knowledge of Nature will be enhanced.

8. **Are not current computers quantum?** No, in spite of the fact that current computers use elements, for example semiconductors, whose functioning cannot be explained without quantum mechanics. Current computers are in some very restricted sense quantum mechanical because everything can be seen as being quantum mechanical. In spite of that, current computers are not considered as fully quantum mechanical. The main difference between a classical and a quantum computer is on the information storage and processing level. In classical computers information is recorded in macroscopic two-level systems, called **bits**, representing two bit values. In quantum computers information is recorded and processed at microscopic level using two-level quantum systems, called **quantum bits**, that can be in any quantum superposition of quantum states corresponding to two classical bits.

9. **Can quantum computers eventually replace classical ones?** Nobody knows, but this does not seem to be so, at least not in the near future. Both classical and quantum computers have their strong and weak points, and it seems currently that they can support, but not replace, each other.

1.2 Prehistory of Quantum Computing

> The past is but the beginning of a beginning, and all that is and has been is but the twilight of the dawn.
>
> Herbert Georg Wells (1866-1940)

Since 1945 we have been witnessing a rapid growth of the raw performance of computers with respect to their speed and memory size. An important step in this development was the invention of transistors, which already use some quantum effects in their operation. However, it is clear that if such an increase in performance of computers continues, then after 50 years, our chips will have to contain 10^{16} gates and operate at a 10^{14} Hz clock rate

[9] The term "Gedanken experiment" is used in several meanings. Sometimes it is required that the corresponding systems or experiments are in principle possible. Sometimes it is sufficient that no physical law is known that would not allow such an experiment.

(thus delivering 10^{30} logic operations per second)[10]. It seems that the only way to achieve that is to learn to build computers directly out of the laws of quantum physics.

In order to come up seriously with the idea of quantum information processing, and to develop it so far and so fast, it has been necessary to overcome several intellectual barriers.

The most basic one concerned an important feature of quantum physics—reversibility (see Section 1.7).[11] None of the known models of universal computers was reversible. This barrier was overcome first by Bennett[12] (1973), who showed the existence of universal reversible Turing machines, and then by Toffoli (1980, 1981) and Fredkin and Toffoli (1982), who showed the existence of universal classical reversible gates.[13]

The second intellectual barrier was overcome by Benioff (1980, 1982, 1982a) who showed that quantum mechanical computational processes can be at least as powerful as classical computational processes. He did that by showing how a quantum system can simulate actions of the classical reversible Turing machines. However, his "quantum computer" was not fully quantum yet and could not outperform classical ones.

The overcoming of these basic intellectual barriers had significant and broad consequences. Relations between physics and computation started to be investigated on a more general and deeper level. This has also been due to the fact that reversibility results implied the theoretical possibility of zero-energy computations.[14] A Workshop on Physics and Computation started to be organized and in his keynote speech at the first of these workshops, in 1981, R. Feynman (1982)[15] asked an important question: *Can (quantum) physics be (efficiently) simulated by (classical) computers?* At the same time he showed good reasons to believe that the answer is negative. Namely, that it appears to be impossible to simulate a general quantum physical system on a probabilistic Turing machine without an exponential slowdown[16]. Moreover, he speculated that one could deal with the problem by allowing computers to run according to the laws of quantum mechanics. In other words, that quantum computers could be exponentially more powerful than classical ones and could

[10]Due to these facts, the concern was voiced quite a while ago on the possible negative effects that quantum phenomena could induce in the "classical" operations of computers. For example, what fundamental limits could Heisenberg's uncertainty principle impose on memory chips whose bits are stored in single electron states? This approach was later superseded, as we shall see, by more optimistic, more constructive and more ambitious aims to harness the power of quantum mechanics to perform computations.

[11]Reversibility is actually not an exclusive phenomenon of the quantum world. Reversibility also occurs in the classical physics. It is only the physics of large systems (classical but also quantum) that is not reversible. The fact is that classical computationally reversible systems suggested by Bennett and others, as discussed later, were not practically realizable. This brought up the idea of considering quantum reversible information processing systems.

[12]For earlier references see Section 9.5 in Appendix.

[13]Bennett (1988) traces the need to think seriously about the thermodynamics of mental processes (and computation was thought of this way in the nineteenth century), back to the famous paradox of "Maxwell's demon" from 1871, which seemed to violate the second law of thermodynamics, see Appendix, Section 9.1.5.

[14]Actually, the original motivation for studying the reversibility of computation came from the interest in determining the ultimate thermodynamic costs of elementary information processing operations, especially because heat removal has always been a major engineering concern in the design of classical computers, limiting the density with which active components could be packed. In the beginnings of the modern computer era there was a folklore belief, going back to a von Neumann's lecture in 1949 (see Burks, 1966), that at least $kT\ln 2$ of energy is needed per bit operation. Attempts to prove this misleading folklore belief led Landauer to the discovery of reversible computing.

[15]Richard P. Feynman (1918-1988), an American physicist. His main scientific contributions were in quantum electrodynamics and in the study of interactions of elementary particles. He gave a mathematical description of helium. Feynman received the 1965 Nobel prize for physics for his contributions to quantum electrodynamics. He has also been known for his extraordinary capabilities to explain physical phenomena and his lectures and textbooks represent an additional important contributions to modern physics.

[16]Actually, this is nowadays intuitively pretty obvious because n interacting 2-state quantum systems may have up to 2^n basis states.

be a first reasonable model of computation that does not obey the modern Church–Turing thesis.[17]

The third intellectual barrier that had to be overcome was a lack of a proper model for a universal quantum computing device capable of simulating effectively any other quantum computer. The first step to overcome this barrier was done by Deutsch (1985) who elaborated Feynman's ideas and developed a (theoretically) physically realisable model of quantum computers, a quantum physical analogue of a probabilistic Turing machine, which makes full use of the quantum superposition principle, and on any given input produces a random sample from a probability distribution. Deutsch conjectured that it might be more efficient than a classical Turing machine for certain computations. He also showed the existence of a universal quantum Turing machine (that could consequently simulate any physical process and experiment) and also a model of quantum networks—a quantum analog of classical sequential logical circuits. However, his model of the universal Turing machine had the drawback that the simulation of other quantum Turing machines (QTM), could be exponential.[18] This problem was then overcome by Bernstein and Vazirani (1993) and Yao (1993). They showed the existence of universal quantum Turing machines capable of simulating other quantum Turing machines in polynomial time. (For a full proof see Bernstein and Vazirani (1997).) The paper of Bernstein and Vazirani (1993) laid the foundations of quantum complexity theory. In addition, Yao (1993) showed that QTM and quantum circuits compute in polynomial time the same class of functions. This result implies that the concept of quantum computation in polynomial time is robust enough and independent of the machine models.

In parallel with the development of the basic models of quantum computing an effort was put into overcoming the fourth intellectual barrier. Can quantum computing be really more powerful than classical computing? Are there some good reasons to assume that quantum computing could bring an essential (exponential) speed-up of computations for at least some important information processing problems? This was a very important issue because it was clear that any design of a quantum computer would require overcoming a number of large scientific and engineering barriers and therefore it was needed to know whether the proposed model of quantum computer offers, at least theoretically, any substantial benefit over the classical computers.

In spite of the fact that this problem has not yet been completely resolved there is already strong evidence that this is so.

It was first shown by Deutsch and Jozsa (1992), that there are problems unknown to be in **P** that could be solved in polynomial time on quantum computers, and therefore belong to the class **QEP** of problems solvable with certainty in polynomial time on quantum computers. By recasting the original Deutsch–Jozsa problem, in the framework of so-called "promise problems", Berthiaume and Brassard (1992, 1992a, 1992b, 1994) proved the first separation results in the relativized quantum complexity theory. For example, they showed that there is an oracle A such that $\mathbf{QEP}^A \not\subseteq \mathbf{ZPP}^A$—they proved the existence of an oracle for which there are computational problems that QTM can solve in polynomial time with certainty, but each probabilistic Turing machine to solve these problems with certainty needs exponential time for some inputs. These results were first improved by Bernstein and

[17] R. Freivalds called my attention to the fact that Yu. Manin already in 1980 in his book "Computable and uncomputable" pointed out explicitly the potential advantages of quantum computing (exponential number of basis states to work with simultaneously) and emphasized a need to design a theory of quantum automata that would be abstract enough and would have a proper balance between mathematical principles and fundamental principles of quantum mechanics without specification of some physical realizations.

[18] Deutsch centered his attention on the computability and not on complexity issues.

Vazirani (1993) and later by Simon (1994). He proved the following result that was at that time the strongest argument in favor of the superiority of quantum computers over classical ones.

Theorem 1.2.1 *There exists an oracle relative to which there is a problem solvable in polynomial time (with bounded error probability) on a quantum computer, but any probabilistic Turing machine with bounded error probability solving this problem (using the oracle) will require exponential time (at least $2^{n/2}$ steps) on infinitely many inputs (of length n).*

Results of Bernstein and Vazirani (1993) and Simon (1994) provide formal evidence that, in the relativized setting, QTM are more powerful than PTM.[19]

However, all these problems were quite artificial. Very important and much needed steps along these lines have been the results of Shor (1994, 1997) who, building on the works of the above mentioned authors, especially on Simon's method, showed how to factor integers, and how to compute discrete logarithms in polynomial time on potential quantum computers—two problems of crucial importance for public-key cryptography.

Due to these results quantum computing, that till then used to be considered as a curiosity for few visionaries, started to be of broader scientific, and not only scientific, interest. An intensive search started to discover physical principles and processes that could eventually make quantum computation practical. Moreover, several groups of experimental physicists around the world have begun projects to explore experimentally the basic principles of quantum computing.

The next question to address was whether one can build a practically successful quantum computer. Could quantum computing be brought from a visionary stage to an experimental stage (and later to an engineering stage)?

This question is still to be answered. An intensive effort to deal with quantum computer design problems has brought some remarkable success, but also revealed new problems.

On one hand success came in an unexpected area. Quantum cryptography—in which one tries to exploit quantum phenomena[20] to transmit quantum information in such a way that undetectable eavesdropping is impossible, has already reached an experimental stage. There has also been success in the effort to find sufficiently simple reversible quantum gates that could be used to build potential quantum computers. The classical universal reversible gates have three inputs and outputs. Sleator and Weinfurter (1995), Barenco (1995) and DiVincenzo (1995) have shown universal two bit quantum gates. This has been an important result because the problem to control interaction of three particles seems to be much more complex than for the case of two particles. In addition, Barenco (1995) and Lloyd (1995) have shown that almost any quantum two-bit gate is universal. These results greatly simplified the search for physical implementations of quantum computational networks.

On the other hand, it has also turned out that the first models of quantum computers were oversimplified and that for quantum computing to come to an experimental or even engineering stage many fundamental problems still need to be solved. The necessity of examining impacts of inaccuracies, emissions and coupling with the environment of any realistic device on the capability of quantum computing to meet their promises has long been emphasized by Landauer (1994). Especially problems decoherence causes made many to believe that it is in principle impossible to design reliably enough functioning quantum

[19] However, it is necessary to make clear that the question whether quantum computers allow one to obtain essentially more computational power has not yet been completely satisfactorily answered.

[20] Heisenberg's uncertainty principle—see Section 9.1.2.

computer.[21]

The situation started to look almost hopeless. A breakthrough came after overcoming another intellectual barrier: it was realised that the situation is not as bad as it looks and that physics does not need to rely on itself only in the search for how to overcome problems of the imperfections of operations, emission and of the decoherence. Mathematics and informatics seem to be able to help significantly. The first important and encouraging result was due to Bernstein and Vazirani (1993). They showed that quite weak precision requirements are sufficient for quantum computing—only logarithmic precision for inputs and gates is needed. Discovery of error-correcting codes by Shor (1995), and soon by many others, allowed one to cope with decoherence and operational imperfections during transmission and storage of quantum information. (In behind there was a key discovery that quantum noise/errors, in principle continuous, can be viewed and dealt with as being discrete.) The discovery of quantum fault-tolerant computations by Shor (1996) allowed one to cope with decoherence and imprecisions during processing of quantum information.[22] The discovery of "concatenated codes" (Knill and Laflamme, 1996) and "quantum repeaters" (Briegel, 1998), allows one to cope with the problem of storage and transmission of quantum information for a long time and long distance with desirable reliability.

Quantum cryptography has also contributed to an awareness that quantum computing is full of pitfalls, not fully understood yet. In 1993, Brassard, Crépeau, Jozsa and Langlois surprised the community by the claim (proof) that a quantum bit commitment protocol provably unbreakable by both parties is possible. It took three years to find out, by Lo and Chau (1997, 1997a) and Mayers (1998), that proposed protocols are, in principle, insecure.

Another intellectual barrier was overcome by contributions of Cirac and Zoller (1995). They showed, at least on the laboratory level, that in the search for technology to build quantum processors and computers one does not need to wait till some "unobtainium" is available, but that one can start with the existing technologies with which there are already rich experimental experiences. (Of course, this is not the whole story. One also has to realize that even if it might be possible to build small quantum computers, scaling up to machines large enough to make really important computations could present fundamental difficulties.)

1.3 From Randomized to Quantum Computation

A comparison of probabilistic Turing[23] machines (PTM), with quantum Turing machines (QTM) will allow us to see, in an easy and transparent way similarities and differences

[21] Pessimism that technology cannot be made reliable enough to realize useful computations is not a new phenomenon in the short history of modern computers. For example, in the autobiography of K. Zuse (1984), there is a story about sceptical reactions to his talk in 1938 in which he anticipated -based on discussions with Schreyer-that about 2000 tubes would be needed to build an electronic computer. (At that time the biggest electronic devices were broadcasting stations with few hundreds of valves.) Similarly, the idea that ENIAC with its 16000 tubes could work for a sufficiently long time was for that time an enginnering phantasy that would hardly get through a granting agency of "peace time".

[22] Actually Landauer's constant challenge of "visionaries" to show a really workable path to the future has been of immense significance for making correct research agenda in quantum computing. Quantum computing is an excellent example of the rapid progress in science and technology that can be achieved by optimists and visionaries if they closely cooperate with, and listen to, sceptics and pessimists directing constructively the effort of visionaries and optimists on the key problems to attack.

[23] Alan M. Turing (1912-1954) an English mathematician. He wrote fundamental papers on computability and artificial intelligence and invented a computation model bearing his name. During the Second World War Turing participated in the cryptanalysis project ULTRA in Bletchley Park and in the design of the first powerful electronic computer Colossus. After the war he supervised the design and building of ACE, a large electronic digital computer at the National Physical Laboratory.

between these two basic models of classical and quantum computing. In this way we can also demonstrate the advantages and problems quantum computing has.

There are good reasons to start our introduction to quantum computing by comparing probabilistic and quantum Turing machines. Probabilistic Turing machines represent nowadays the most important model of classical computing. Polynomial time computation on probabilistic Turing machines stands for a formal equivalent of "feasibility" in classical computing. In addition, similarly to classical Turing machines, quantum Turing machines were historically the first really fully quantum and powerful model of quantum computing.

1.3.1 Probabilistic Turing machines

Formally, a (one-tape) **probabilistic Turing machine**, on a finite set Q of states and the finite alphabet Σ, is given by a transition function

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

assigning to each possible transition a probability in such a way that for each configuration[24] c_0 and all its successor-configurations $c_1, \ldots, c_k$, the following **local probability condition** is satisfied: If p_i, $1 \leq i \leq k$, is the probability, assigned by δ, of the transition from c_0 to c_i, then (see Figure 1.1a):

$$\sum_{i=1}^{k} p_i = 1.$$

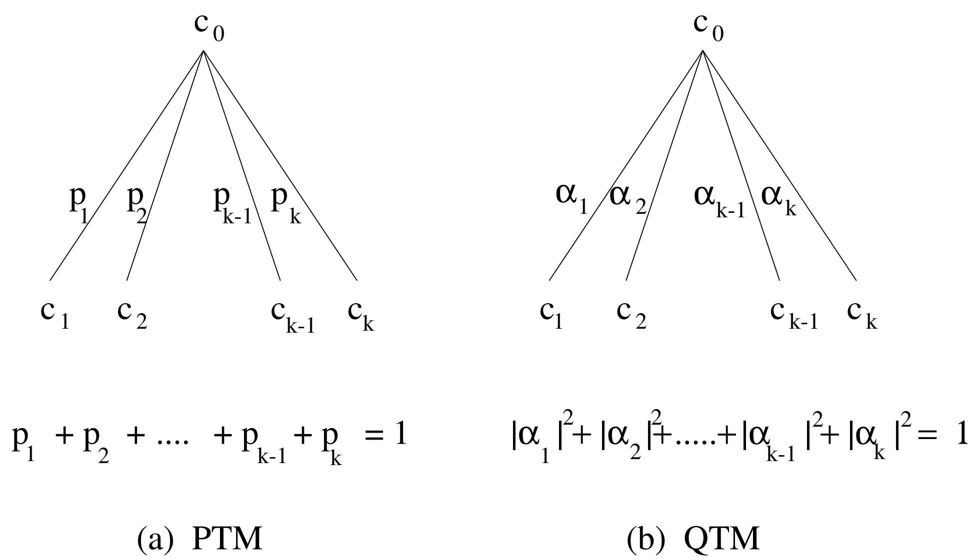


Figure 1.1: Local probability conditions

This condition is often written in the following form: if $(\sigma_1, q_1) \in \Sigma \times Q$, then

$$\sum_{(\sigma,q,d)\in\Sigma\times Q\times\{\leftarrow,\downarrow,\rightarrow\}} \delta(\sigma_1, q_1, \sigma, q, d) = 1.$$

[24] A configuration is a full description of the global state of a PTM. It can be seen as having the form $w_1 q w_2$, where $w_1 w_2$ is the current content of the tape, q is the current state and the current position of the head of the PTM is on the cell with the first symbol of w_2.

On the base of the transition function δ of a PTM $\mathcal{M}$ we can assign probabilities to all edges, to all nodes and also to all configurations of each level of any configuration tree of T. The probability assigned to an edge $c \rightarrow c'$ of such a tree is given directly by δ and represents the probability that computation goes, in one step, from c to c'. From that we can assign a probability to each node N of any configuration tree, see Figure 1.2a, as the product of all probabilities assigned to the edges on the path from the root to N. (The probability assigned to the root is defined to be 1.) The probability assigned to an arbitrary node N is therefore the probability that a computation starting at the root reaches the node N.

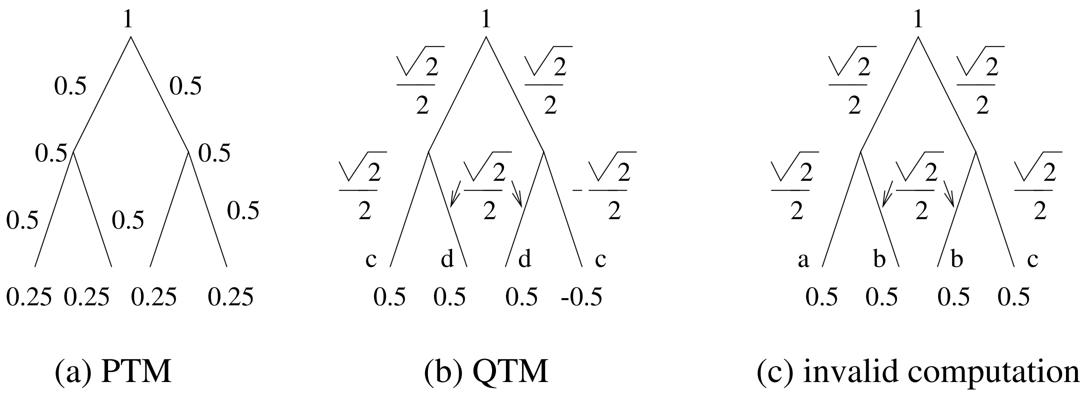


Figure 1.2: Configuration trees with probabilities and the probability amplitudes

It may happen that at a certain level of a configuration tree there are several occurrences $c^{(1)}, \ldots, c^{(m)}$ of the same configuration c, see Figure 1.3a. In such a case, if p_i is the

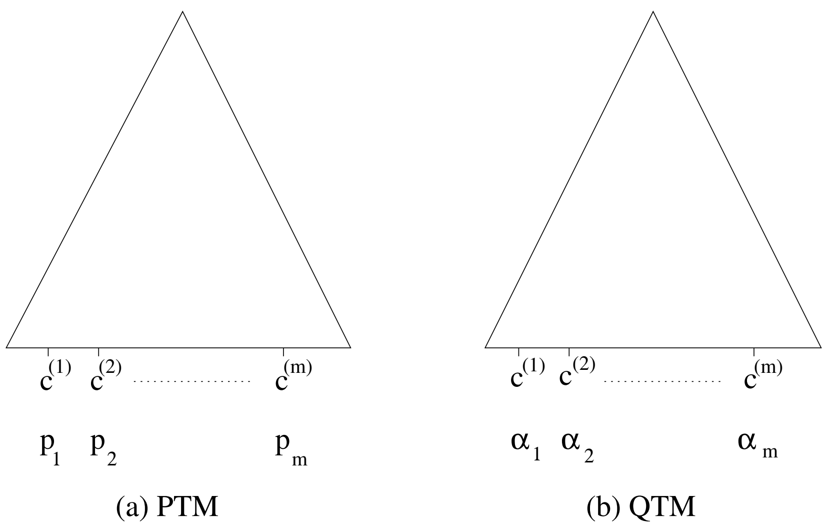


Figure 1.3: Multiple occurrences of the same configuration

probability assigned to the occurrence $c^{(i)}$ of the configuration c, then the total probability that the configuration c occurs at that level of the configuration tree is given by the sum

$$\sum_{i=1}^{m} p_i.$$

Now, if $c_1, \ldots, c_k$ are all distinct configurations occurring at a certain level of the configuration tree, and $p_1, \ldots, p_k$ are their global probabilities of occurrence at that level, then the following **global probability condition** has to be satisfied:

$$\sum_{i=1}^{k} p_i = 1.$$

Exercise 1.3.1 *Show that if a PTM satisfies the local probability condition, then it also satisfies the global probability condition.*

The local probability condition can also be seen as assigning to each configuration a "linear superposition" of successor configurations

$$F(c) = p_1c_1 + p_2c_2 + \ldots + p_kc_k,$$

with $\sum_{i=1}^{k} p_i = 1$, where F is some kind of "global transition function" of T and p_i is the probability of having c_i as the next configuration of c.[25] Let us now consider any superposition $\sum_{i=1}^{k} p_ic_i$ of configurations with $\sum_{i=1}^{k} p_i = 1$. If we replace, in such a superposition, any particular configuration c by the superposition of its successor configurations, as above, and make corresponding multiplications by constants and corresponding additions, we get again a superposition of configurations with coefficients summing up to 1.

All this implies that the transition function δ of a PTM $\mathcal{M}$ actually determines a so-called **transition matrix** $M_{\mathcal{M}}$, rows and columns of which are labeled by configurations of $\mathcal{M}$ (and therefore the matrix can be infinite) and $M_{\mathcal{M}}(i,j)$ is the probability that the configuration c_i is the successor configuration of c_j. Such a transition matrix clearly has all entries nonnegative and the sum of its entries in each column is 1.

In this case, if we multiply $M_{\mathcal{M}}$ with a column vector of the same dimension and only nonnegative elements, the sum of which is 1, we get again a column vector with only nonnegative elements the sum of which is 1. We can therefore see $M_{\mathcal{M}}$ as a mapping that maps any superposition of configurations satisfying the global probability condition to another superposition of configurations satisfying again the global probability condition.

The time evolution of a probabilistic Turing machine $\mathcal{M}$ can therefore also be described by a sequence of probability distributions, represented by superpositions, which begins with the superposition containing only the initial configuration, and such that the ith distribution provides the likelihood of each possible configuration after the $(i-1)$st step of the evolution of $\mathcal{M}$. Each next superposition (probability distribution) can be obtained from the previous one by multiplying with the matrix $M_{\mathcal{M}}$. The matrix $M_{\mathcal{M}}$ can therefore be seen as representing the evolution of the PTM $\mathcal{M}$.

In each particular computation of a PTM only one path is taken from the whole set of paths of the configuration tree, in accordance with the assigned probabilities. To simulate a

[25]This assignment of linear superpositions to PTM has no real meaning for PTM but helps us to make a better analogy with QTM.

PTM we need therefore to keep track of only a constant amount of information. We could also imagine a PTM $\mathcal{M}$ as being put into a box with a glass top through which we could watch (not influence) the particular steps taken by $\mathcal{M}$, one after another. At the end of the computation we could see the result obtained. A PTM computation can therefore be observed and the act of their observation has no effect on its further computation. (Why should it have?)

Concerning the outcomes, a PTM can be seen as defining a random sample, a probability distribution on the final configurations for each initial configuration.

In order to study the computational power of PTM we need to impose some restriction on the probabilities allowed. Otherwise one could hide hard-to-compute numbers or properties into them. It is well known that in order to study computational complexity problems of randomized computing, it is sufficient to allow only probabilities from the set $\{0, \frac{1}{2}, 1\}$.

After this lengthy review of probabilistic Turing machines and their behaviour we are in a better position to discuss quantum Turing machines and their behaviour.

1.3.2 Quantum Turing machines

Formally, a (one-tape) **quantum Turing machine**, with a finite set Q of states and the finite alphabet Σ, is given by a transition function

$$\delta : \Sigma \times Q \times \Sigma \times Q \times \{\leftarrow, \downarrow, \rightarrow\} \longrightarrow \mathbf{C}_{[0,1]}$$

assigning a so-called **amplitude** (or **probability amplitude**)—a complex number, the absolute value of which is in the interval $[0, 1]$—to each transition in such a way that for each configuration[26] c_0 and all its successor configurations $c_1, \ldots, c_k$ the following **local probability condition** is satisfied: if α_i is the amplitude assigned to the transition from c_0 to the configuration c_i, then (see Figure 1.1b)

$$\sum_{i=1}^{k} |\alpha_i|^2 = 1$$

and therefore $|\alpha_i|^2$ can be seen (and will be seen) as a probability of transition from c_0 to c_i. However, as discussed later, this is not the only condition a transition function of a QTM has to satisfy.

The transition function of a QTM can be used to assign amplitudes (not probabilities) also to all edges, nodes and all configurations of the same level of a configuration tree. The amplitude assigned to an edge is given directly by δ. The amplitude assigned to a node is the product of the amplitudes assigned to all edges on the path from the root to that node, assuming again that the amplitude 1 is assigned to the root (see Figure 1.2b).

As for the case of PTM, let us assume that at a particular level of the configuration tree there are several occurrences, say $c^{(1)}, c^{(2)}, \ldots, c^{(m)}$, of the same configuration c. (See Figure 1.3b). Let now α_i be the amplitude of the configuration $c^{(i)}$ at that level. In such a case the total amplitude of c at that level is defined to be

$$\beta = \sum_{i=1}^{m} \alpha_i.$$

[26]The concept of configuration is defined in a similar way as for PTM.

So far all that looks quite similar to the case of PTM. The only difference being that in the case of PTM we have worked with probabilities and now we are working with (probability) amplitudes. An essential difference between PTM and QTM concerning their computations comes now. If $c_1, c_2, \ldots, c_k$ are all mutually different configurations at a certain level of the configuration tree, then their total amplitudes $\beta_1, \ldots, \beta_k$ have to satisfy the following **global probability condition**

$$\sum_{i=1}^{k} |\beta_i|^2 = 1,$$

and $|\beta_i|^2$ is said to be the probability of the occurrence of the configuration c_i at that level of computation.

It is not true that if a QTM satisfies all local probability conditions, then it also satisfies all global probability conditions. A counter example is shown in Figure 1.2c.

What does all this actually imply? As we shall see soon, the way probabilities are assigned to configurations at particular levels of computations represents an enormous difference with respect to the case of PTM. Indeed, one of the important consequences is the existence of constructive and destructive interferences.

Exercise 1.3.2 *If α, β are complex numbers, then*

$$|\alpha + \beta|^2 = |\alpha|^2 + |\beta|^2 + 2|\alpha||\beta| \cos\theta,$$

where θ is the angle which the vectors α and β subtend at the origin of the complex plane. Analyse the value of $|\alpha + \beta|^2$ for cases: (a) $\alpha = \beta$; (b) $\alpha = -\beta$; (c) $\alpha = i\beta$.

It may happen, that for the amplitudes $\alpha_1, \ldots, \alpha_k$ of all occurrences of a configuration at a level of the configuration tree the following inequality holds: $|\sum_{i=1}^{k} \alpha_i|^2 > \sum_{i=1}^{k} |\alpha_i|^2$ and in such cases we speak about **positive interference** or about **constructive interference**.

For example, the configuration d in the configuration tree in Figure 1.2b has at the last level two occurrences, each with the amplitude $\frac{1}{2}$, and therefore with the probability $\frac{1}{4}$. However, the total probability that this configuration occurs at that level is actually $(\frac{1}{2} + \frac{1}{2})^2 = 1$, and therefore not $2 \times \frac{1}{4}$ but $4 \times \frac{1}{4}$, twice as much than it would seem.

It may also happen that $|\sum_{i=1}^{k} \alpha_i|^2 < \sum |\alpha_i|^2$. In such a case we speak about **negative interference** or **destructive interference**.

For example, the configuration c in the configuration tree in Figure 1.2b has two occurrences, one with the amplitude $\frac{1}{2}$, second with the amplitude $-\frac{1}{2}$. Their sum is therefore 0. This implies, in spite of the fact that there are two paths leading to the configuration c, that the probability that c occurs at that level is 0 !!!

The total probability of the occurrence of a configuration at a step of computation is the probability that an observer will get that configuration as the result of the computation at that level (if an observation is performed). If such a probability is 0, there is no way to get the results, in spite of the fact there are computational paths leading to them! Why is that so? This is surely puzzling. But this is the way quantum world is. Such are the rules that say how much and which information one can get from the quantum world to the classical world by an "observation" or a "measurement"—one of the most puzzling phenomenon of quantum physics, to be discussed in more detail in the Sections 1.4.3, 1.5.3 and 1.6.2 as well as in Appendix, Section 9.1.4.

From the fact that we can have positive and negative interferences, one of the basic tricks of quantum computing follows. One has to program quantum computers in such a way that correct and desirable answers, due to positive interference, have large probability, and incorrect, or not desirable answers, due to destructive inference, have very small, or zero, probability.

What does this all imply? Is there some other condition, more transparent than the global probability condition a quantum Turing machine has to satisfy? Yes, there is.

In the same way as for PTM, to each QTM $\mathcal{M}$ we can associate a matrix $M_{\mathcal{M}}$ of configuration transitions such that $M_{\mathcal{M}}(i,j)$ is the amplitude of having the configuration c_i as the successor configuration of the configuration c_j. Entries of $M_{\mathcal{M}}$ are therefore complex numbers and the local probability condition implies that Euclidean norm of its column vector is 1.

For a QTM $\mathcal{M}$ it has to hold that the matrix $\mathcal{M}$ is **unitary**, i.e.,

$$M_{\mathcal{M}} M^*_{\mathcal{M}} = M^*_{\mathcal{M}} M_{\mathcal{M}} = I,$$

where $M^*_{\mathcal{M}}$ is the conjugate transpose of $M_{\mathcal{M}}$, i.e. the transposition of $M_{\mathcal{M}}$ and conjugation of its elements, and I is the unit matrix.

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

(a) (b) (c)

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \qquad \frac{1}{2}\begin{pmatrix} 1-i & 1+i \\ 1+i & 1-i \end{pmatrix}$$

(d) (e)

$$\begin{pmatrix} i\cos\theta & \sin\theta \\ \sin\theta & i\cos\theta \end{pmatrix} \quad \begin{pmatrix} e^{i\alpha}\cos\theta & -ie^{i(\alpha-\theta)}\sin\theta \\ -ie^{i(\alpha+\theta)}\sin\theta & e^{i\alpha}\cos\theta \end{pmatrix}$$

(f) (g)

Figure 1.4: Examples of unitary matrices of degree 2

Exercise 1.3.3 *(a) Show that if A and B are unitary matrices of the same dimension, then the matrix* $\begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & B \end{pmatrix}$ *is also unitary: (b) Show that if A and B are unitary matrices of the same dimension, then so is the matrix $A \cdot B$.*

Exercise 1.3.4 *Show the following properties of Pauli matrices: (a) $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = I$; (b) $\sigma_k\sigma_l = i\sigma_m$, where (k,l,m) is a cyclic permutation of (x,y,z); (c) all Pauli matrices have eigenvalues 1 and -1.*

The requirement of unitarity is far from obvious. Examples of unitary matrices of degree two, including Pauli matrices[27], are shown in Figure 1.4[28]. A general form of the unitary matrix of degree two can be found on page 63.

It can be shown that if the transition matrix of a QTM is unitary, then all global probability conditions are satisfied. Another important consequence of unitarity is that each QTM is reversible (and the same is true for each quantum evolution). This means that from a given superposition of configurations in a step of a computation we can uniquely deduce the superposition of configurations in the previous step.

Exercise 1.3.5 *(a) Verify the unitarity of the matrices shown in Figure 1.4; (b) show the unitarity of any $n \times n$ matrix A with $A[i,j] = \frac{2}{n}$ if $i \neq j$ and $A[i,i] = -1 + \frac{2}{n}$, for any $1 \leq i \leq n$.*

Exercise 1.3.6 *Show that: (a) the determinant of any unitary matrix is ± 1; (b) all unitary matrices of a degree n form a group, with respect to multiplication, usually denoted $U(n)$; (c) all unitary matrices of degree n and with determinant equal to 1 also form a group, usually denoted $SU(n)$; (d)* all eigenvalues of unitary matrices have absolute value 1.*

Another essential difference between a PTM and a QTM could be seen when we would try "to observe" the evolution of a QTM and to find out the results of its computations.

In the case of a PTM, at each particular computation a single path through the configuration tree has to be chosen, and we could watch (though not influence) the path being taken. The result would be obtained with the probability attached to the final configuration.

On the other hand, a QTM always follows all paths of the configuration tree simultaneously! Since the number of nodes at the levels of a configuration tree can grow exponentially, this means that a QTM can, simultaneously, take an exponentially large number of paths and can be, at particular steps of computation, in a superposition of exponentially many configurations (with respect to the number of computational steps), at the same time! In addition, the computational evolution of a QTM, and of any quantum computation, is fully determined by its unitary matrix and it is deterministic.

Moreover, there is no way "to watch" the computations of a QTM. We could "put it into a box and let it run" but we cannot watch it, or open the box before computation is done—at least not without serious consequences. This would be an observation (a measurement) and, according to the laws of quantum mechanics, it could immediately lead to a disruption of the computation and could result in a loss of (quantum) information! At the end of the computation we can try to observe (measure) the result. However we can not get in general the whole resulting superposition of configurations. Only one of them, randomly chosen, with the probability determined by the corresponding global amplitude. In addition, once such information is obtained, all other results of the computation (all other configurations) are irreversibly lost. Finally, it is only at this point, in general, in a measurement, or an observation, where probabilities and indeterminancy enter quantum computation.

[27] Wolfgang Pauli (1900–1958), an American physicist of Austrian origin, with positions in Hamburg, Zürich and Princeton. He received the 1945 Nobel prize for his exclusion principle (formulated in 1924), according to which no two electrons in an atom may be in the same quantum state. In 1930 Pauli derived the existence of the neutrino, before it was experimentally observed. He also made fundamental contributions in quantum electrodynamic, quantum field theory and paramagnetism.

[28] Matrices σ_x, σ_y and σ_z are called *Pauli matrices* and they play an important role in the theory of spin-$\frac{1}{2}$ electrons. They were introduced by W. Pauli in 1927 to describe angular momentum and magnetic momentum of electrons.

A QTM can therefore get in linear time an exponential number of results, but unfortunately we cannot read them all out. (Disappointing, but this is the way it is. In spite of that, QTM can be more efficient than classical ones.)

This seems to imply that we actually cannot get more with a QTM than with a probabilistic Turing machine that would provide us, randomly, with one result! Fortunately, this is not true. There are sometimes clever ways to make a QTM use its enormous parallelism to get to the single needed result, with a high probability, which seems not to be obtainable efficiently without this quantum parallelism. This is, however, not a simple task, what will be demonstrated in Section 3.2.

Both PTM and QTM produce their results with certain probabilities. Therefore they actually define probability distributions on possible outputs.

In order to study the computational power of a QTM we also need to make some restriction on the probability amplitudes allowed. Otherwise one could hide hard-to-compute numbers or properties into them. It has been shown, as discussed in Chapter 5, that in order to study the computational complexity problems of quantum computing it is sufficient to allow only amplitudes from the set $\{-1, -\frac{4}{5}, -\frac{3}{5}, 0, \frac{3}{5}, \frac{4}{5}, 1\}$.

Remark 1.3.7 Figure 1.5 summarizes, very informally, the main features of classical versus quantum computations. As we shall see in Section 1.6, quantum computation, as determined by quantum evolution only, is a deterministic process, contradicting widespread naive beliefs, and probabilities appears only when "creatures" of the classical world try "to observe" the outcomes of quantum processes. On the other hand, modern complexity theory considers probabilistic computations as the main mode of efficient computing.

It is one of the profound problems in science to determine what classical and quantum worlds actually are and where a borderline between them is—if there is any. Interesting enough, even some of the founders of quantum mechanics have been very careful about it. For example, Bohr avoided to referring explicitly to two types of worlds. He only emphasized a need to use two different languages to talk about quantum and classical phenomena.

1.4 Hilbert Space Basics

> The whole science is nothing more than a reformulation of everyday thinking.
>
> Albert Einstein: *Physics and reality (1936)*

Hilbert[29] space is a mathematical framework suitable for describing the concepts, principles, processes and laws of quantum mechanics. Pure states of quantum systems are considered to be vectors of a Hilbert space. One can say that to each isolated quantum system corresponds a Hilbert space. Some even go farther by claiming that there is no reality on the quantum level; such a reality emerges only in the case of a measurement, and what we know about the quantum level are only computational procedures, expressed in terms of

[29] David Hilbert (1862–1943), a German mathematician and logician. Hilbert was perhaps the most influential mathematician of his period. After work on the theory of invariants, he developed a new approach to abstract algebra and functional analysis. Of key importance for quantum mechanics are his abstract spaces bearing his name. He made important contributions to algebraic number theory, functional analysis, integral equations and to variation calculus. He also worked on several fundamental problems of physics.

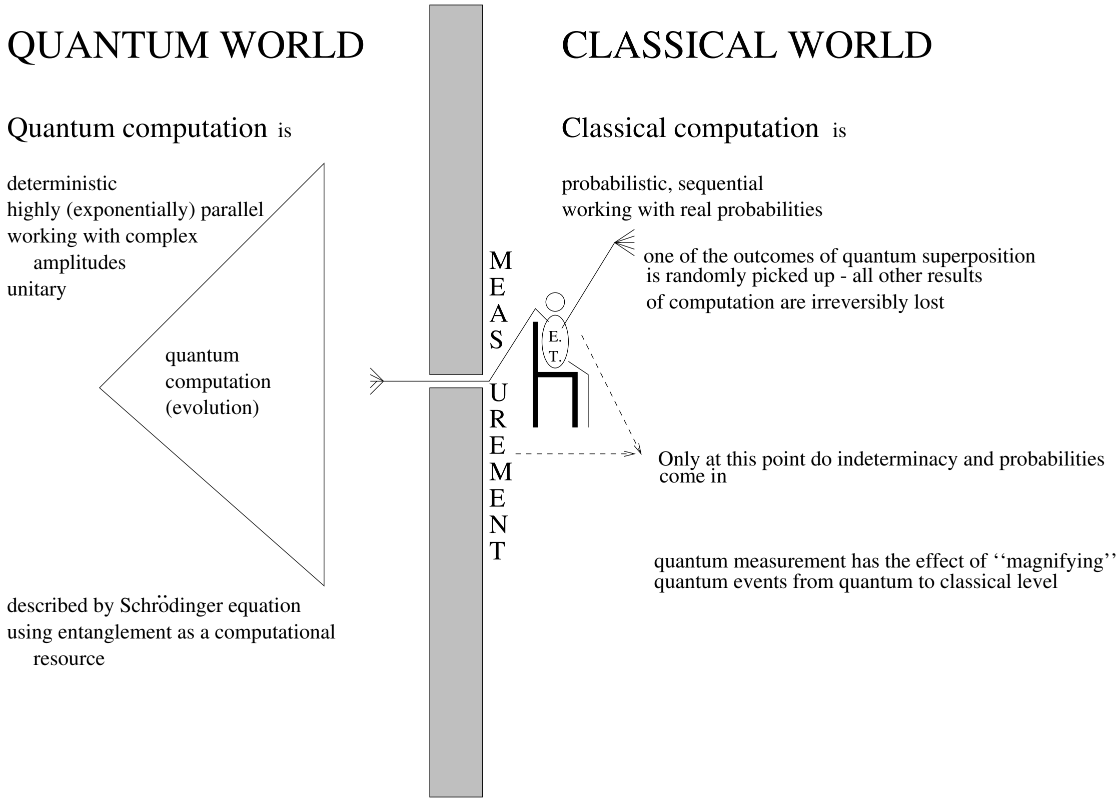


Figure 1.5: Quantum and classical worlds

Hilbert space concepts, to compute evolutions of quantum systems and probabilities of the measurement outcomes.

Let us start with the two most important examples of Hilbert spaces for quantum mechanics.[30]

Example 1.4.1 (Hilbert spaces $l_2(D)$) *For any countable set D, let $l_2(D)$ be the space of all complex valued functions on D bounded by the so-called l_2-norm, i.e.*

$$l_2(D) = \{x \mid x : D \to \mathbf{C}, \left(\sum_{i \in D} x(i)x^*(i)\right)^{1/2} < \infty\}^{31}.$$

We say that $l_2(D)$ is a Hilbert space with respect to the **inner product** $\langle \cdot | \cdot \rangle : l_2(D) \times l_2(D) \to \mathbf{C}$,[32] *defined by*

$$\langle x_1 | x_2 \rangle = \sum_{i \in D} x_1^*(i) x_2(i).$$

[30] von Neumann's idea to formulate quantum mechanics in terms of Hilbert spaces was one of the most important steps to the development of modern quantum physics.

[31] x^* denotes the conjugate of the complex number x; i.e., $x^* = a - bi$ if $x = a + bi$, where a, b are real.

[32] In this book we use the following notation for sets of numbers: $\mathbf{C}$—the set of complex numbers; $\mathbf{R}$—the set of reals; $\mathbf{Q}$—the set of rationals; $\mathbf{Z}$—the set of whole numbers; $\mathbf{N}$ ($\mathbf{N}^{\geq 0}$)—the set of (nonnegative) integers.

Elements of $l_2(D)$ are usually called vectors (to be indexed by elements of D). The notation $l_2 = l_2(\mathbf{N})$ is usually used in the case $D = \mathbf{N}$.

Our second example of a Hilbert space is actually the main one considered in quantum mechanics. However, for dealing with the very basic concepts of quantum computing it is not as much needed.

Example 1.4.2 (Hilbert space L_2) [33] *Let (a,b) be an interval, with finite or infinite bounds, on the real axis. By $L_2((a,b))$, or simply L_2, we denote the set of all complex valued functions such that $\int_a^b |f(x)|^2\,dx$ exists, equipped with the inner product*

$$\langle f|g\rangle = \int_a^b f^*(t)g(t)\,dt < \infty.$$

If f and g are such that $|f|^2$ and $|g|^2$ are integrable functions[34] *on (a,b), then so are functions cf and $f+g$, for any complex number c, and therefore L_2 is a linear space.*[35]

Hilbert spaces are discussed in more detail in Appendix, Section 9.2. In this section only the very basic concepts and results are summarized.

Definition 1.4.3 *An* **inner-product space** *H is a complex vector space, equipped with an inner product $\langle\cdot|\cdot\rangle : H \times H \rightarrow \mathbf{C}$ satisfying the following axioms for any vectors ϕ, $\psi, \phi_1, \phi_2 \in H$, and any $c_1, c_2 \in \mathbf{C}$.*

$$\begin{aligned}
\langle\phi|\psi\rangle &= \langle\psi|\phi\rangle^*,\\
\langle\psi|\psi\rangle &\geq 0 \text{ and } \langle\psi|\psi\rangle = 0 \ \text{ if and only if } \ \psi = \mathbf{0},\\
\langle\psi|c_1\phi_1 + c_2\phi_2\rangle &= c_1\langle\psi|\phi_1\rangle + c_2\langle\psi|\phi_2\rangle.^{36}
\end{aligned}$$

The inner product introduces on H the norm (length)

$$||\psi|| = \sqrt{\langle\psi|\psi\rangle}$$

and the metric (Euclidean distance)

$$\mathrm{dist}(\phi,\psi) = ||\phi - \psi||.$$

This allows us to introduce on H a metric topology and such concepts as continuity.

Exercise 1.4.4 *Show the following properties of the norm for $\phi,\psi \in H$, $a \in \mathbf{C}$:(a) $||\phi|| \geq 0$ for all $\phi \in H$; (b) $||\phi|| = 0$ if and only if $\phi = \mathbf{0}$ (c) $||\phi + \psi|| \leq ||\phi|| + ||\psi||$ (triangle inequality); ; (d) $||a\phi|| = |a|\ ||\phi||$; (e) $|\langle\phi,\psi\rangle| \leq ||\phi||\,||\psi||$ (Schwarz inequality).*

---

[33] Hilbert studied spaces l_2 and L_2, in his work on linear integral systems, and that is why von Neumann all spaces of such types named as Hilbert spaces.

[34] With respect to Lebesgue measure.

[35] To be more precise L_2 is to be the set of Lebesgue integrable functions on (a,b) and we do not consider as different a pair of functions that differ only on a set of measure zero. In such a linear space the zero element is a function that is equal to zero almost everywhere on (a,b).

[36] Caution! In more mathematically oriented literature the third axiom for the inner product has often the form: $\langle c_1\phi_1 + c_2\phi_2, \psi\rangle = c_1\langle\phi_1|\psi\rangle + c_2\langle\phi_2|\psi\rangle$. In order to transfer results of one of these two axiomatic approaches into another conjugate values have to be taken.

Unit norm vectors of an inner-product space are also called **(pure) states** of H.[37]

Exercise 1.4.5 *Show the following properties for the distance mapping for any* $\phi, \psi, \chi \in H$*: (a)* $dist(\phi,\psi) \in \mathbf{R}^{\geq 0}$*; (b)* $dist(\phi,\psi) = dist(\psi,\phi)$*; (c)* $dist(\phi,\psi) = 0$ *if and only if* $\phi = \psi$*; (d)* $dist(\phi,\psi) \leq dist(\phi,\chi) + dist(\chi,\psi)$*.*

Definition 1.4.6 *An inner-product space* H *is called* **complete***, if for any sequence* $\{\phi_i\}_{i=1}^{\infty}$*, with* $\phi_i \in H$*, and with the property that* $\lim_{i,j\to\infty} ||\phi_i - \phi_j|| = 0$*, 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 a* **Hilbert space***. Two Hilbert spaces* H_1 *and* H_2 *are said to be* **isomorphic***, notation* $H_1 \simeq H_2$*, if the underlying vector spaces are isomorphic and their isomorphism preserves the inner product. (For example, surprisingly,* $l_2 \simeq L_2$*—Riesz-Fischer Theorem.)*

For each ϕ of a Hilbert space H the mapping $f_\phi : H \to \mathbf{C}$ defined by

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

is a linear mapping on H in the sense that $f_\phi(c\psi) = cf_\phi(\psi)$ and $f_\phi(\psi_1 + \psi_2) = f_\phi(\psi_1) + f_\phi(\psi_2)$. One can even show that we get all linear mappings from H to $\mathbf{C}$ by this construction. Namely, it holds:

Theorem 1.4.7 *To each continuous linear mapping* $f : H \to \mathbf{C}$ *there exists a unique* $\phi_f \in H$ *such that* $f(\psi) = \langle\phi_f|\psi\rangle$ *for any* $\psi \in H$*.*

The space of all linear mappings (called also **functionals**) of a Hilbert space H forms again a Hilbert space, the so-called **dual Hilbert space** (or **conjugate Hilbert space**) H^* with the inner product $\langle f|g\rangle = \langle\phi_f|\phi_g\rangle$, for $f, g \in H^*$.

The mapping $f_\phi(\psi) = \langle\phi|\psi\rangle$ is a functional for any $\phi \in H$. Therefore, the last theorem establishes a bijection between H and H^*, and $H \cong H^*$. On the base of this relation the handy "ket-bra" notation, due to Dirac[38], can be introduced.

A vector ψ of a Hilbert space H is denoted $|\psi\rangle$ and referred to as a **ket-vector**. The corresponding functional, given by the previous theorem, is denoted $\langle\psi|$ and referred to as a **bra-vector**. $\langle .|$ can therefore be seen as the operator that maps each state ϕ into a functional $\langle\phi|$ such that $\langle\phi|(|\psi\rangle) = \langle\phi|\psi\rangle$ for every state.

In the case of an n-dimensional Hilbert space, a ket vector $|\psi\rangle$ can be seen as an n-dimensional column vector and a bra-vector as an n-dimensional row vector. The **inner product** $\langle\phi|\psi\rangle$ is then a usual "row vector $\times$ column vector" product. The transformation $|\phi\rangle \leftrightarrow \langle\phi|$ corresponds to transposition and conjunction. The **outer product** $|\phi\rangle\langle\psi|$ is then an $n \times n$ matrix—a usual "column vector $\times$ row vector" product.

Exercise 1.4.8 *Let* $\mathcal{A}$ *be a set of states of a Hilbert space* H *each which has norm* 1 *and there is an* ε *such that* $||\phi - \psi|| \geq \varepsilon$ *if* $\phi, \psi \in \mathcal{A}$*,* $\phi \neq \psi$*. Is it true that the set* $\mathcal{A}$ *has to be finite?*

[37] The idea of representing pure quantum states of a quantum system by unit vectors of a Hilbert space, one of the key ideas of modern quantum theory, is due to von Neumann (1932).

[38] Paul Adrian Maurice Dirac (1902–1984), an English physicist. He formulated a version of quantum mechanics that took into account the theory of relativity. He shared the 1933 Nobel prize with E. Schrödinger. Together with E. Fermi he determined the laws of statistical mechanics of a system of atoms and he envisioned the existence of positrons.

1.4.1 Orthogonality, bases and subspaces

The concepts of an orthonormal basis and of an orthogonal decomposition of a Hilbert space are fundamental for the Hilbert space theory.

Definition 1.4.9 *Two vectors ϕ and ψ of a Hilbert space are called* **orthogonal**, *notation $\phi \perp \psi$, 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.

Exercise 1.4.10 *Show that if ϕ, ψ are distinct elements of an orthonormal set, then $||\phi - \psi|| \leq \sqrt{2}$.*

Exercise 1.4.11 *Show that if n nonzero vectors of a Hilbert space are mutually orthogonal, then they are linearly independent.*

Remark 1.4.12 The key role of orthogonality for quantum computing is that whenever a measurement is performed on a quantum system, then those quantum states that lead to distinguishable outcomes have to be mutually orthogonal. No measurement or observation of a quantum system is able to distinguish faithfully between non-orthogonal states.

Definition 1.4.13 *A set $\mathcal{B}$ of vectors of a Hilbert space H forms an (orthonormal) basis of H if it is orthonormal and each vector v in H can be uniquely expressed in the form*

$$\psi = \sum_{\phi \in B} \alpha_\phi \phi, \quad \textit{with} \quad \alpha_\phi \in \mathbf{C} \ .$$

It can be shown that all bases of a Hilbert space H have the same cardinality, called the dimension of H. In addition, two Hilbert spaces, H_1 and H_2, of the same dimension are isomorphic. A d-dimensional Hilbert space will be denoted by H_d.

Example 1.4.14 *Two examples of orthonormal bases in H_3:*
(a) $\{(0,0,1),(0,1,0),(1,0,0)\}$; *(b)* $\{(\frac{2}{\sqrt{6}},\frac{1}{\sqrt{6}},\frac{1}{\sqrt{6}}),\ (0,-\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}),\ (\frac{1}{\sqrt{3}},-\frac{1}{\sqrt{3}},-\frac{1}{\sqrt{3}})\}$

Exercise 1.4.15 *Show that the dimension of a finite dimensional Hilbert space H is the maximum of linearly independent vectors of H.*

Finite dimensional Hilbert spaces correspond, for example, to such properties of particles as spin value or polarization. Infinite dimensional Hilbert spaces correspond to such properties of particles as position or momentum. In the case of quantum automata the corresponding Hilbert space has as an orthonormal basis the set of all configurations of the automaton. In the case of space bounded computations, the corresponding Hilbert space is finite dimensional; otherwise it is infinite dimensional.

There are several other equivalent and important definitions of the orthonormal basis of Hilbert spaces:

Theorem 1.4.16 *Let B be an orthonormal set of a Hilbert space H. The following statements are equivalent:*

1. *B is an orthonormal basis of H.*
2. *If $\langle x|\phi\rangle = 0$ for all $\phi \in B$, then $x = \mathbf{0}$.*
3. *For all $\phi \in H$, $|\phi\rangle = \sum_{\psi\in B}\langle\phi|\psi\rangle|\psi\rangle$.*
4. *For all $\phi, \psi \in H$, $\langle\phi|\psi\rangle = \sum_{\gamma\in B}\langle\phi|\gamma\rangle\langle\gamma|\psi\rangle$.*
5. *For all $\psi \in H$, $||\psi||^2 = \sum_{\phi\in B}|\langle\phi|\psi\rangle|^2$ (the so-called Parseval's identity).*

Exercise 1.4.17 *If $\mathcal{B} = \{|\theta_i\rangle \mid i \in I\}$ is a basis of a Hilbert space, then $\mathcal{B}' = \{\langle\theta_i| \mid i \in I\}$ is the basis of the dual Hilbert space and if $|\phi\rangle = \sum_{i\in I}\alpha_i|\phi_i\rangle$, $\alpha_i = \langle\phi_i|\phi\rangle$, then $\langle\phi| = \sum_{i\in I}\beta_i\langle\phi_i|$, $\beta_i = \langle\phi|\phi_i\rangle$.*

Definition 1.4.18 *A* **subspace** *G of an inner-product space H is a subset of H closed under addition and scalar multiplication.*

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

Theorem 1.4.19 *For each closed subspace W of a Hilbert space H there exists a unique subspace $W^\perp$ such that $\langle\phi|\psi\rangle = 0$, whenever $\phi \in W$ and $\psi \in W^\perp$ and each $\psi \in H$ can be uniquely expressed in the form $\psi = \phi_1 + \phi_2$, with $\phi_1 \in W$ and $\phi_2 \in W^\perp$. In such a case we write $H = W \oplus W^\perp$ and we say that W and $W^\perp$ form an orthogonal decomposition of H.*

In a natural way we can make a generalization of an orthogonal decomposition

$$H = W_1 \oplus W_2 \oplus \ldots \oplus W_n,$$

of H into mutually orthogonal subspaces $W_1, \ldots, W_n$ such that each $\psi \in H$ has a unique representation as $\psi = \phi_1 + \phi_2 + \ldots + \phi_n$, with $\phi_i \in W_i, 1 \leq i \leq n$.

1.4.2 Operators

Definition 1.4.20 *A linear operator on a Hilbert space H is a linear mapping $A : H \to H$. The set of all linear operators of a Hilbert space H will be denoted $\mathcal{L}(H)$. $\mathcal{L}(H_1, H_2)$ will stand for the set of all linear operators from a Hilbert space H_1 into the Hilbert space H_2.*

An application of a linear operator A to a vector $|\psi\rangle$ is denoted $A|\psi\rangle$ or $A(|\psi\rangle)$. A is also a linear operator of the dual Hilbert space H^*, mapping each linear functional $\langle\phi|$ of the dual space to the linear functional, denoted by $\langle\phi|A$. (If $\langle\phi|A$ is applied to a vector, then A is applied first and then $\langle\phi|$). A linear operator A is called **positive** or **semi-definite**, notation $A \geq 0$, if $\langle\psi|A\psi\rangle \geq 0$ for every $|\psi\rangle \in H$.

Each linear operator A of a countable Hilbert space H with a basis $\mathcal{B} = \{|\theta_i\rangle| \, i \in I\}$ can be represented by a matrix, in general infinitely dimensional, whose rows and columns are labeled by elements of I and with $\langle\theta_i|A|\theta_j\rangle$ in the i-th row and j-th column. In such a case a row indexed by $i \in I$ is the vector $\langle\theta_i A|$ and a column indexed by j is the vector $A|\theta_j\rangle$.[39]

[39]Linear operators are often identified with the corresponding matrices. This correspondence is always related to some basis and if no basis is explicitly mentioned, then it should be clear implicitly from the context which basis is considered.

The **norm** $||A||$ of a linear operator A is defined as $||A|| = \sup_{||\phi||=1} ||A|\phi\rangle||$. A linear operator A is called **bounded** if its norm $||A||$ is finite. A linear operator is bounded if and only if it is continuous.

Exercise 1.4.21 *Show that if M is a complex matrix of degree d such that $|M(i,j)| \leq \varepsilon$ for all i, j, then $||M|| \leq d\varepsilon$.*

Projections have a special role among linear operators. If $H = W_1 \oplus W_2$ is an orthogonal decomposition of a Hilbert space H into subspaces W_1 and W_2, then, as mentioned above, each $\psi \in H$ has a unique representation $\psi = \psi_1 + \psi_2$, where $\psi_1 \in W_1$ and $\psi_2 \in W_2$. In such a case the mappings

$$P_{W_1}(\psi) = \psi_1 \quad \text{and} \quad P_{W_2}(\psi) = \psi_2$$

are called **projections** onto the subspaces W_1 and W_2, respectively.

Example 1.4.22 *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 a projection onto the one-dimensional subspace spanned by the vector $|\phi\rangle$.

Of special importance are **adjoint** and **self-adjoint** operators. The adjoint operator T^* to a bounded linear operator T is an operator such that for any ϕ and $\psi \in H$, $\langle\psi|T\phi\rangle = \langle T^*\psi|\phi\rangle$. An operator T is self-adjoint if $T = T^*$. Instead of $\langle\psi|T\phi\rangle$, notation $\langle\psi|T|\phi\rangle$ is often used. Hence

$$\langle T^*\psi|\phi\rangle = \langle\psi|T|\phi\rangle = \langle\psi|T\phi\rangle.$$

In addition, the identity $\langle\phi, T^*, \psi\rangle = \langle\phi, T, \psi\rangle^*$ holds.

Exercise 1.4.23 *Let A, A_1, A_2 be linear operators. Show: (a) $||A_1A_2|| \leq ||A_1||\,||A_2||$; (b) $||A_1|| - ||A_2|| \leq ||A_1 + A_2|| \leq ||A_1|| + ||A_2||$.*

Exercise 1.4.24 *Let A, B be self-adjoint operators. Which of the following operators are surely self-adjoint: (a) $A + B$; (b) AB; (c) $AB + BA$; (d) $i(AB - BA)$?*

To self-adjoint operators correspond **Hermitian matrices**, i.e., matrices A such that $A = A^*$.

Exercise 1.4.25 *Show that if A is a linear operator and A^* exists, then $||A^*|| = ||A||$.*

Theorem 1.4.26 *Hermitian matrices have the following properties.*

1. *All eigenvalues of a Hermitian matrix are real.*
2. *The determinant of a Hermitian matrix equals the product of its eigenvalues.*
3. *The* **trace** *(the sum of diagonal elements) of a Hermitian matrix A, notation* $\mathrm{Tr}(A)$, *equals the sum of its eigenvalues.*

Actually, the last two properties hold for all matrices. In addition, if $A \geq 0$ is a Hermitian matrix, then there exists a unique matrix B such that $B \cdot B = A$. This B is then denoted $\sqrt{A}$.

A self-adjoint operator A of a finite dimensional Hilbert space H has the so-called **spectral representation**. If $\lambda_1, \ldots, \lambda_k$ are its distinct eigenvalues, then A can be expressed in the form

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

where P_i is the projection operator into the subspace of H spanned by the eigenvectors corresponding to λ_i.

Example 1.4.27 (Finite-state inner-product spaces) *In various cases we will work within complex inner-product spaces $l_2(D) = \mathbf{C}^D$, with a finite set D (of classical states), equipped with the inner-product as in Example 1.4.2. To each $a \in D$ will correspond a unit norm basis vector $|a\rangle$ and $\{|a\rangle \mid a \in D\}$ will form an orthonormal basis of $l_2(D)$. Any bijection $B : D \to D$ may be regarded as a (classical) unitary transformation on $l_2(D)$, acting by the rule $B|a\rangle \to |B(a)\rangle$. An important case for us will be when D is a set of configurations of an automaton.*

1.4.3 Observables and measurements

In quantum physics in order to extract quantum information from a quantum system we have to **observe** the system—to perform a **measurement** of the system. Two basic types of observables are the so-called **sharp observables** and **unsharp observables**. To the sharp observables correspond the so called **projection measurements**, or PV measurements, and to the unsharp observables correspond the so-called **positve operator valued** (POV) measurements (POVM).

Quantum projection measurements are quantum tests whose outcomes are labeled by real numbers. A quantum test, as performed in laboratory, consists of two phases. A preparation phase, described algorithmically, during which a physical system is set up. During this phase both the testing instrument (observable) and the to-be-measured state are fixed. This phase is completely deterministic. The second phase is the testing itself. During the testing (measuring) one of the potential outcomes of the test is produced with a probability that can be computed according to the rules under which the preparation phase was done. Testing is a randomized process.

Quantum tests can produce non-numerical values. For example, the polarizations of photons. In order to make a quantum measurement out of a quantum test, different numerical labels have to be associated with different outcomes of the test.

An observable is a property of the physical system that can be measured. In the classical physics, position, speed and momentum are examples of observables. In quantum theory a (sharp) observable is a self-adjoint operator.

In quantum mechanics we used to consider the measurement of states mainly with respect to sharp observables. The numerical outcome of the measurement of a pure state $|\psi\rangle$ with respect to an observable A is one of the eigenvalues of A and the side impact of such a measurement is a "collapse" of $|\psi\rangle$ into a state $|\psi'\rangle$. In the measurement the eigenvalue λ_i is obtained with probability

$$Pr(\lambda_i) = ||P_i|\psi\rangle||^2 = \langle\psi|P_i|\psi\rangle,$$

and the new state $|\psi'\rangle$, into which $|\psi\rangle$ collapses, has the form

$$|\psi'\rangle = \frac{P_i|\psi\rangle}{\sqrt{\langle\psi|P_i|\psi\rangle}}.$$

This means that a measurement of $|\psi\rangle$ with respect to A irreversibly destroys $|\psi\rangle$, unless $|\psi\rangle$ is an eigenvector of A.

Sometimes it is important to know the average value of an observable A when a state $|\psi\rangle$ is measured. This average value, denoted $\langle A\rangle_\psi$, is defined to be $\langle\psi|A\psi\rangle = \langle\psi|A|\psi\rangle$.

Exercise 1.4.28 *Show that if $|\psi_1\rangle, \ldots, |\psi_n\rangle$ are eigenvectors of a self-adjoint operator A that form an orthonormal basis and $\lambda_1, \ldots, \lambda_n$ are the corresponding eigenvalues, then for every vector $|\psi\rangle = \sum_{i=1}^n \alpha_i|\psi_i\rangle$, it holds $\langle\psi|A|\psi\rangle = \sum_{i=1}^n |\alpha_i|^2\lambda_i$.*

In the last 10–15 years it has been shown that unsharp observables and the corresponding POV measurements, see Section 9.2.8, are of key importance in many cases. In quantum information processing they play an important role in quantum cryptography (Chapter 6) and in quantum information theory (Chapter 8).

1.4.4 Tensor products in Hilbert spaces

If a quantum system $\mathcal{S}$ is composed of two quantum subsystems $\mathcal{S}_1$ and $\mathcal{S}_2$ and to them correspond Hilbert spaces H, H_1 and H_2, then H is the so-called **tensor product** of H_1 and H_2, written as

$$H = H_1 \otimes H_2$$

and this means that vectors of H are tensor products, defined below, of vectors from H_1 and H_2.

The tensor product of vectors $x = (x_1, x_2, \ldots, x_m)$ and $y = (y_1, y_2, \ldots, y_n)$, notation $x \otimes y$, is an mn-dimensional vector with elements

$$(x_1y_1, \ldots, x_1y_n, x_2y_1, \ldots, x_2y_n, \ldots, x_my_1, \ldots, x_my_n)$$

Exercise 1.4.29 *Show that the tensor product of vectors is an associative operation.*

Exercise 1.4.30 *Show that if x, y are vectors, $c \in \mathbf{R}$, then: (a) $c(x \otimes y) = (cx) \otimes y = x \otimes (cy)$; (b) $x \otimes z + y \otimes z = (x + y) \otimes z$; (c) $x \otimes y + x \otimes z = x \otimes (y + z)$.*

The tensor product is also defined for matrices

$$A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nn} \end{pmatrix} \quad B = \begin{pmatrix} b_{11} & \ldots & b_{1m} \\ \vdots & & \vdots \\ b_{m1} & \ldots & b_{mm} \end{pmatrix}$$

as follows

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

and $A \otimes B$ is therefore a matrix of degree mn.

If $\mathcal{B}_1, \ldots, \mathcal{B}_k$ are orthonormal bases of Hilbert spaces $H_1, \ldots, H_k$, then

$$\mathcal{B}_1 \otimes \ldots \otimes \mathcal{B}_k = \bigotimes_{i=1}^{k} \mathcal{B}_i = \{x_1 \otimes \ldots \otimes x_k \mid x_i \in B_i\}$$

is the orthonormal basis of the Hilbert space

$$H = \bigotimes_{i=1}^{k} H_i.$$

Example 1.4.31 *If H_2 is the two-dimensional vector space with the basis $\mathcal{B}_2 = \{|0\rangle, |1\rangle\}$, where $|0\rangle = (1,0)^T$, $|1\rangle = (0,1)^T$, then*

$$\bigotimes_{i=1}^{n} \mathcal{B}_2 = \{|x_1\rangle \otimes |x_2\rangle \otimes \ldots \otimes |x_n\rangle \mid x_1 \ldots x_n \in \{0,1\}^n\}$$

is the orthogonal basis of the 2^n-dimensional vector space

$$H_{2^n} = \bigotimes_{i=1}^{n} H_2$$

and instead of $|x_1\rangle \otimes |x_2\rangle \otimes \ldots \otimes |x_n\rangle$ the following notations are often used:

$$|x_1\rangle|x_2\rangle \ldots |x_n\rangle \quad or \quad |x_1, x_2, \ldots, x_n\rangle \ or \ |x_1 \ldots x_n\rangle.$$

If $\{\alpha_i\}_{i \in I}$ is a basis of a Hilbert space H, $\{\beta_j\}_{j \in J}$ is a basis of a Hilbert space H', then in $H \otimes H'$ the inner product is induced by the mapping $\langle \alpha_i \beta_j | \alpha_k \beta_l \rangle = \langle \alpha_i | \alpha_k \rangle \langle \beta_j | \beta_l \rangle$.

1.4.5 Mixed states and density operators

Pure states are fundamental objects for quantum mechanics in the sense that the evolution of any closed quantum system can be seen as a unitary evolution of pure states. However, to deal with unisolated[40] and composed quantum systems the concept of the mixed state is of importance.

A probability distribution $\{(p_i, \phi_i) \mid 1 \leq i \leq k\}$ 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 **mixture**, and denoted by $[\psi\rangle = \{(p_i, \phi_i) \mid 1 \leq i \leq k\}$ or

$$[\psi\rangle = (p_1, \phi_1) \oplus \ldots \oplus (p_k, \phi_k). \tag{1.1}$$

where $\oplus$ symbol is used just to separate particular pure states and their probabilities.

Example 1.4.32 *A mixed state (1.1) can be obtained as the output of a source which produces the pure quantum state $|\phi_i\rangle$ with probability p_i.*

We often need to consider a subsystem $\mathcal{A}$ of a quantum system $\mathcal{B}$. As we shall see later, some global states $|\phi\rangle$ of $\mathcal{B}$ cannot be expressed in the form $|\phi\rangle = |\phi_{\mathcal{A}}\rangle \otimes |\phi_{\mathcal{B}-\mathcal{A}}\rangle$, where $|\phi_{\mathcal{A}}\rangle$ is a state of $\mathcal{A}$ and $|\phi_{\mathcal{B}-\mathcal{A}}\rangle$ is a state of $\mathcal{B} - \mathcal{A}$. However, $|\phi\rangle$ can be expressed in the form $|\phi\rangle = \sum_{i=1}^{l} a_i |\psi_i\rangle|\phi_i\rangle$, where $\{\psi_i\}_{i=1}^{l}$ is an orthonormal basis of $\mathcal{B} - \mathcal{A}$ and $\{\phi_i\}_{i=1}^{l}$

[40] A more modern term for an unisolated quantum system is “open” system.

are states of $\mathcal{A}$. For an observer $\mathcal{O}$ who interacts with the subsystem $\mathcal{A}$ only, the subsystem seems to be in just one of the states $|\phi_i\rangle$, in each with some probability. In other words, for the observer $\mathcal{O}$ the subsystem $\mathcal{A}$ is in a mixed state.

Remark 1.4.33 Observe that if $k > 1$, then $p_1|\phi_1\rangle + \ldots + p_k|\phi_k\rangle$ is not a state because $||p_1\phi_1 + \ldots + p_k\phi_k|| < 1$. For example, if $|\phi_1\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, $|\phi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$, then $\frac{1}{2}|\phi_1\rangle + \frac{1}{2}|\phi_2\rangle = \frac{1}{\sqrt{2}}|0\rangle$.

To each mixed state $[\psi\rangle$ corresponds a **density operator** $\rho_{[\psi\rangle}$. If $[\psi\rangle = |\phi\rangle$ for a pure state $|\phi\rangle$, then $\rho_{[\phi\rangle} = |\phi\rangle\langle\phi|$. If $[\psi\rangle = \bigoplus_{i=1}^{k}(p_i, |\phi_i\rangle)$, where $|\phi_i\rangle$ are pure states, then

$$\rho_{[\psi\rangle} = \sum_{i=1}^{k} p_i|\phi_i\rangle\langle\phi_i|.$$

Exercise 1.4.34 *Let $|\phi\rangle$ and $|\psi\rangle$ be two states and $|\psi\rangle = U|\phi\rangle$ for a unitary operator U. Express the $|\psi\rangle\langle\psi|$ in terms of the matrices $|\phi\rangle\langle\phi|$ and U.*

The representation of pure states depends on the choice of the basis, the same is true for density operators, which are uniquely represented in the matrix form through **density matrices**. Denote by $M_\rho^{\mathcal{B}}$ the matrix representation of the density operator ρ, with respect to the basis $\mathcal{B}$. The concepts of density operator and density matrix are often considered as synonymous and the basis is considered to be clear from the context, unless described explicitly.

As discuss in more detail on page 370, the same density matrix can correspond to different mixed states.

Instead of "a density matrix ρ" one can sometimes read also "a state ρ" because ρ is undistinguishable from mixed states to whom ρ corresponds.

In general, ρ is a density operator[41] if ρ is Hermitian, $\rho \geq 0$ and $Tr(\rho) = 1$. This implies $\rho = \rho^*$.

A general form of a density matrix is:

$$\rho = \sum_{i,j=1}^{n} p_{ij}|\phi_i\rangle\langle\phi_j|,$$

where $p_{ij} = p_{ij}^*$, $\sum_{i=1}^{n} p_{ii} = 1$ and $|\phi_i\rangle$ are pure states.

To each density operator ρ there is a basis $\mathcal{B}_\rho$ in which the matrix $M_\rho^{\mathcal{B}_\rho}$ is diagonal. This is used to define $f(\rho)$ for functions $f : \mathbf{C} \to \mathbf{C}$ as follows:

$$M_{f(\rho)}^{\mathcal{B}_\rho}(i,j) = f(M_\rho^{\mathcal{B}_\rho}(i,j)).$$

For example, in this sense one understands such operations as $\sqrt{\rho\rho^*}$ or $\lg\rho$.

The degree of ignorance embodied in a mixed state $[\phi\rangle = \bigoplus_{i=1}^{k}(p_i, \phi_i)$ is represented by its (quantum) **von Neumann entropy**

$$S(\rho_{[\phi\rangle}) = -Tr\rho_{[\phi\rangle}\lg\rho_{[\phi\rangle}.$$

[41] Density operators were introduced by von Neumann (1932).

One can show that $S(\rho_{[\psi\rangle}) \leq H(p_1, \ldots, p_k)$, where H is the classical Shannon entropy.

One of the profound differences between the quantum and classical systems lies in the relation between a system and its subsystems. As discussed below a state of a Hilbert space $H = H_A \otimes H_B$ cannot be always decomposed into states of its subspaces H_A and H_B. We also cannot define any natural mapping from the space of linear operators on H into the space of linear operators on H_A (or H_B). However, density operators are much more robust and that is also one reason for their importance. A density operator ρ on H can be "projected" into H_A by the operation of **tracing out** H_B, to give the density operator (for finite dimensional Hilbert spaces):

$$\rho_{H_A} = Tr_{H_B}(\rho) = \sum_{|\phi\rangle, |\phi'\rangle \in \mathcal{B}_{H_A}} |\phi\rangle \left(\sum_{|\psi\rangle \in \mathcal{B}_{H_B}} \langle \phi, \psi | \rho | \phi', \psi \rangle \right) \langle \phi' |,$$

where $\mathcal{B}_{H_A}$ ($\mathcal{B}_{H_B}$) is an orthonormal basis of the Hilbert space H_A (H_B).

The following remark may clarify the meaning of tracing out. If $H = H_A \otimes H_B$ and ρ is a density matrix of H, then $\rho_{H_A} = Tr_{H_B}\rho$ is the unique density matrix of H_A such that

$$Tr(\rho_{H_A} O) = Tr(\rho(O \otimes I))$$

for each observable (Hermitian matrix) O of H_A.

Informally, one often says that in order to get the density matrix of a subsystem given the density matrix of the whole system one should trace over the degrees of freedom of the rest of the system.

Let us denote by $\rho(H, \psi, H')$ the density matrix associated to Hilbert spaces H, H' and a state $|\psi\rangle$ of $H \otimes H'$ as follows

$$\rho(H, \psi, H') = Tr_{H'} |\psi\rangle\langle\psi|.$$

In physics literature $\rho(H, \psi, H')$ is usually denoted as $\rho(H, \psi)$ or $\rho(H)$ or ρ_H and it is said to be the density matrix of H.

In isolated quantum systems the basic elements are pure states, and evolutions of these states are performed by unitary transformations (matrices). On the other hand, in unisolated (open) quantum systems, pure states can naturally evolve into mixed states (which can also be described as pure states of a larger system, composed of a given system and the environment). For unisolated quantum systems the basic elements are therefore mixed states, or density matrices, and evolutions are performed by the so-called **superoperators**[42] (certain positive and trace preserving linear transformations over density matrices), acting on density matrices (see page 97, and also Davies (1976), for more detail).

Exercise 1.4.35 *(a) Show that any* 2×2 *matrix* A *can be written as* $A = x_0 I + x_1 \sigma_x + x_2 \sigma_y + x_3 \sigma_z$ *for proper* x_0, x_1, x_2, x_3 *and Pauli matrices* $\sigma_x, \sigma_y, \sigma_z$*; (b) find necessary and sufficient conditions for parameters* x_0, x_1, x_2, x_3 *such that* A *is a density matrix.*

The degree of resemblance, or "distance" of two states (density matrices) ρ_1 and ρ_2 is called **fidelity** $F(\rho_1, \rho_2)$. In the case both ρ_1 and ρ_2 are pure states, $|\phi_1\rangle\langle\phi_1|$ and $|\phi_2\rangle\langle\phi_2|$, respectively,

$$F(\rho_1, \rho_2) = |\langle\phi_1|\phi_2\rangle|^2, \tag{1.2}$$

[42]Sometimes the term "quantum operator" or even simply "operator" is used instead of "superoperator".

which represents the closeness of $|\phi_1\rangle$ and $|\phi_2\rangle$ in the natural geometry of Hilbert space. In the case only ϕ_1 is a pure state, then.

$$F(\phi_1, \rho_2) = F(|\phi_1\rangle\langle\phi_1|, \rho_2) = |\langle\phi_1|\rho_2|\phi_1\rangle|^2,$$

which is just the average of (1.2) over mixed states with density matrix ρ_2. Several other concepts of fidelity are considered in Chapter 8.

Physical meaning of density matrices

The concept of the density matrix is one of the fundamental ones of quantum theory and views of its physical meaning keep evolving.

The very basic view (see Bennett, 1998), is that a density matrix corresponding to a mixed state represents all and only that information which can be learned about the mixed state by sampling or about a pure state of a compound system by observing its part only. In other words mixed states with the same density matrix are indistinguishable (by measurement) and pure states ψ of a compound system $S_1 \otimes S_2$ with the same $\rho(S_1, \psi, S_2)$ are indistinguishable (by observing S_1 only). Moreover, if ψ is a pure state of a compound system $S_1 \bigotimes S_2$, then for any mixed state $[\psi_0\rangle = \oplus_{i=1}^{k}(p_i, \psi_i)$ of S_1 such that $\rho_{[\psi_0\rangle} = \rho(S_1, \psi, S_2)$, there is a measurement on S_2 which generates $[\psi_0\rangle$ in the following sense: if the measurement yields the outcome i with probability p_i, then the system S_1 is left in the pure state ψ_i.[43]

Remark 1.4.36 The importance of mixed states and especially of density matrices for quantum mechanics theory in general, and for quantum information processing in particular, is growing. For quantum computation this lies in the following. Real quantum computers can rarely be in pure states, rather they are in mixed states, and interact with their environment leading to non-unitary evolutions. Research in quantum error-correcting codes and in fault-tolerant computation has shown that non-unitary evolutions, in the case of weak interactions with the environment, do not need to imply a loss of quantum computational power. Computation with mixed states has been shown (see Section 2.3.5), to be as powerful as with pure states. Finally, some of the technologies being used for experimenting in quantum computing work with highly mixed states.

Remark 1.4.37 The Hilbert space is a very nice and useful abstraction of physical reality. However, one should not forget that real quantum computing is not performed in a Hilbert space but in laboratories.

1.5 Experiments

> Every experiment is like a weapon which must be used in its particular way—a spear to thrust, a club to strike. Experimenting requires a man who knows when to thrust and when to strike, each according to need and fashion.
>
> Philippus Aureolus Paracelsus (1493-1541)

[43] A new meaning of density matrices which gives it the same onthological status as the wavefunction describing a pure state is discussed by Aharonov and Anandan (1998).

The assignment of complex amplitudes to quantum events, superposition rules and the special way of considering probabilities in the quantum world are quite puzzling. Naturally a question arises: why is it so?

One of the possible answers is: this is the way it works in quantum world; such are its rules (see Feynman et al. 1964)/indexFeynman, Richard P..

In this section we take a little more didactic approach and describe, in a very simplified “textbook form”[44], several basic quantum experiments that suggest and justify the above mentioned principles and rules (see Feynman et al. 1964).

1.5.1 Classical experiments

In order to understand better the special properties of particles, let us compare their behaviour with the behaviour of some elements of the classical world, with bullets and waves, in special experimental setups.

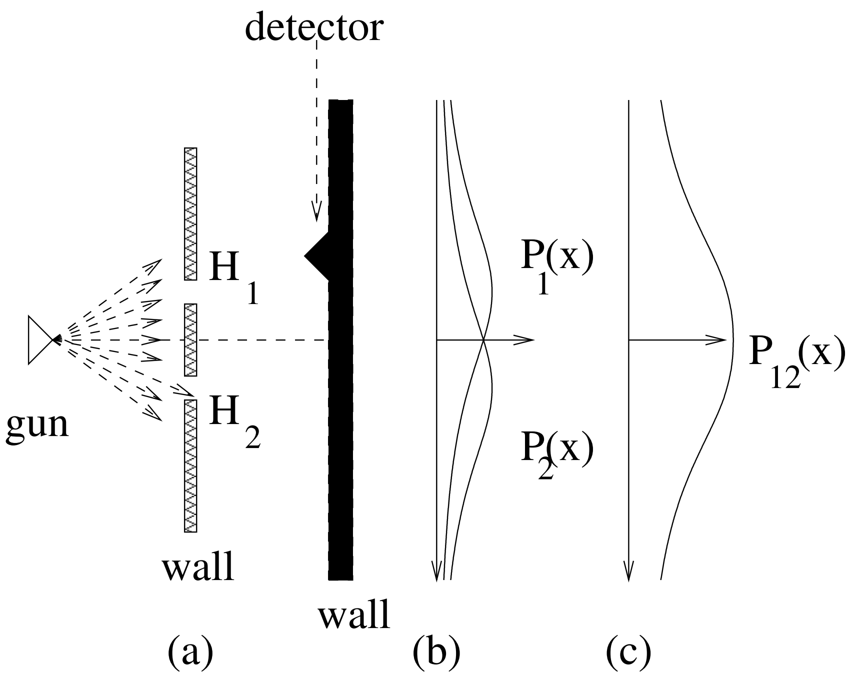


Figure 1.6: Experiment with bullets

Consider at first the experiment depicted in Figure 1.6. An automatic gun is shooting bullets with a random distribution to all directions. In front of it there is a wall with two holes; each of them is just large enough for exactly one bullet to go through at a time. However, any bullet can get reflected in a hole and can change its direction afterwards in an arbitrary way. At the second wall behind, there is a detector that can be moved up and down and used to count the number of bullets hitting different places on the second wall. In this way, we can experimentally find the answer to the following question: What is the probability that a bullet reaches a given position at the second wall? The results of the

[44]In Bush et al. (1997), such descriptions of quantum experiments are labeled as “caricatures”.

measurements are shown, for the case that only one hole, say H_1 (H_2), is open, and for a suitable distances of the holes and walls, in Figure 1.6b, by the curve $P_1(x)$ ($P_2(x)$). The probability $P_{12}(x)$, for the case where both holes are open, is depicted in the Figure 1.6. The results are as expected by common sense: $P_{12}(x) = \frac{1}{2}(P_1(x) + P_2(x))$ for all positions x.

The second experiment, this time with waves on the water, is illustrated in Figure 1.7a. A generator vibrates and makes waves. They move toward the first wall, with two holes again, and then to the second wall with a detector, which can detect the intensity $I(x) = |h(x)|^2$ of the wave (or its energy), i.e. the square of the height-amplitude $h(x) = e^{i\phi(x)}\sqrt{I(x)}$ of the wave. In the next Figure 1.7b, the curve shows the level of the intensity $I_1(x)$ ($I_2(x)$) for the case that only the hole H_1 (H_2) is open. The results are again as expected. The intensity curve, for the case that both holes are opened, is shown in Figure 1.7c, and from the wave theory it is well known, that this is a consequence of the wave interference phenomena. In this case actually $I_{12}(x) = |h_1(x) + h_2(x)|^2$. This means, that in some cases we have positive and in some cases negative interference. It is also well known in which positions there are local minima and where local maxima (as well as how large they are).

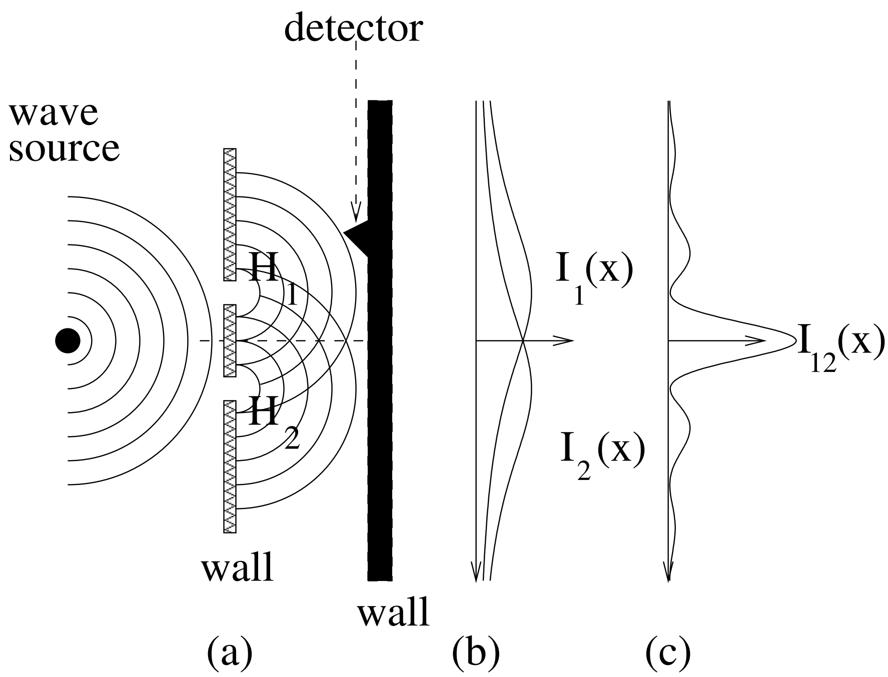


Figure 1.7: Experiments with waves

1.5.2 Quantum experiments—single particle interference

Our third experiment, depicted in Figure 1.8a, has a setting similar to the first one. Again, we have a source.[45] However, this time it produces electrons. The first wall has two tiny holes, which are usually called slits, each just large enough for one electron to get through

[45]This is a modification of Young's famous two-slit experiment with photons from 1801.

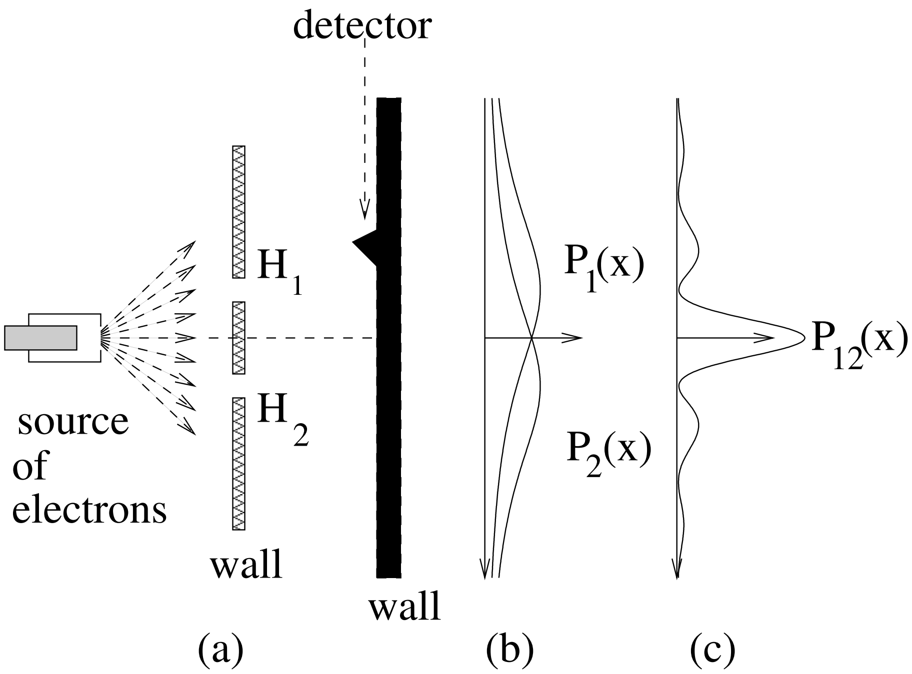


Figure 1.8: Two slit experiment

at a time, perhaps changing its direction behind the slit.[46] The second wall has, again, a detector that can be moved up and down and can count the number of electrons reaching a given position of the second wall. In this way, we can experimentally determine the probabilities that electrons reach given positions on the second wall. The results are shown in Figure 1.8b, by the curve $P_1(x)$ ($P_2(x)$) for the case that electrons reach the position x at the second wall and that only one slit, namely H_1 (H_2), is open. Again, the results are as expected, the maxima are exactly at points where the straight lines from the source through the slits reach the second wall. However, contrary to our intuition, in the case that both slits are open we get the curve $P_{12}(x)$, shown in Figure 1.8c, similar to that in Figure 1.7c. Very surprisingly, at some places one observes fewer electrons when both slits are open than in the case only one slit is open![47] A similarity of Figures 1.7 and 1.8 indicates that electrons, particles, sometimes behave as waves!

There seem to be two surprising conclusions one can draw from these experimental results. By opening the second slit, it suddenly seems that electrons are somehow prevented from doing what they could do before! It seems that by allowing an electron to take an alternative route we have actually stopped it from traversing either route.

Electrons are particles, but they seem to have a wave-like behaviour as they pass through the holes! Each particle seems to behave as if it is going through both holes at once and

[46]One could ask how large an electron is in order to know how large slits should be. There are several possible answers to such a question: (a) electron is seen, to our knowledge, as a point object with no dimension; (b) we do not know exactly. (When it is under the collision of other high energy particles, an electron behaves like a point particle with radius less than, say, 10^{-16} cm.)

[47]This is, actually, only a rough idea of a potential experiment. In order really to function, its components would need to have extraordinarily small sizes. However, from similar and really feasible experiments, it is pretty well known how the results would look in the experiment considered here.

afterwards creating waves that interfere, as in the second, wave, experiment.[48] However, we cannot predict the precise path for any electron.

Observe that by opening the second, slit the number of electrons arriving at some places increases four times and at some places decreases to zero. The places with local minima (maxima) are known to be in the distance $\frac{l\lambda}{L}$, where l is the distance of two walls, L is distance of two slits and λ is the wavelength.

The similarity of the curves $P_{12}(x)$ in Figures 1.7 and 1.8, can be seen as a motivation for trying to assign a complex amplitude to the fact that an electron reaches a position, and then to find the corresponding probabilities in a similar way to how we found the intensities in the previous example and so to have interference responsible for contra-intuitive outcomes.

Indeed, it turns out that one can assign to the event that an electron reaches a position x going through the first (second) slit a complex amplitude $\psi_1(x)$ ($\psi_2(x)$) in such a way that $P_1(x) = |\psi_1(x)|^2$, $P_2(x) = |\psi_2(x)|^2$ and $P_{12}(x) = |\psi_1(x) + \psi_2(x)|^2$. Now it is easy to see the reason for oscillations in the case of $P_{12}(x)$. Indeed, we get

$$\begin{aligned} P_{12}(x) &= |\psi_1(x) + \psi_2(x)|^2 = |\psi_1(x)|^2 + |\psi_2(x)|^2 \\ &\quad +2Re(\psi_1(x))Re(\psi_2(x)) + 2Im(\psi_1(x))Im(\psi_2(x)) \\ &= P_1(x) + P_2(x) + 2Re(\psi_1(x))Re(\psi_2(x)) + 2Im(\psi_1(x))Im(\psi_2(x)), \end{aligned}$$

where $Re(y)$ and $Im(y)$ denote the real and imaginary parts of y. Oscillations are due to the term $2Re(\psi_1(x))Re(\psi_2(x))) + 2Im(\psi_1(x))Im(\psi_2(x))$.[49]

It is also important to mention that the results of this experiment do not depend on the frequency with which electrons are shot. The same interference pattern would be obtained, in the end, if each electron is shot only after the previous one hits the wall or if there is an interval of several years between two consecutive electrons.[50] It is also important to realize that quantum physics has no explanation where a particular electron reaches the detector wall. All quantum physics can offer are statements on the probability that an electron reaches a certain position on the detector wall.

In order to illustrate another quantum phenomenon, to be discussed later, we consider a modification of the previous experiment—with a measurement during the experiment.

In the experiment depicted in Figure 1.9a, the basic setting is similar to the previous experiment. However, this time we have in addition a source of light on the right hand side of the first wall, just in the middle between the two slits. If we now watch the experiment we can detect (as indicated by the small black circle in the figure), through which slit a particular electron passes through the first wall. If it goes through the slit H_1, some light appears for a moment in its neighborhood, as a reflection; if its passes trough the slit H_2, we see some light near that slit.

Again, we can determine the probabilities that electrons reach positions on the second wall for the case where only one slit is open, Figure 1.9b, and for the case where both slits are open, Figure 1.9c. The curves in Figure 1.9b are similar to those in Figure 1.7b, as

[48] In our experiment, electrons behave as being little packets of waves and they well demonstrate the "wave-particle" duality in quantum mechanics. According to this duality principle, fundamental quantum objects are neither waves nor particles, but sometimes one thing or the other, or perhaps always a little of both—see Lindley (1996).

[49] However, since there is such a similarity between the interference produced by waves, as the elements of the classical world, and the interference exhibited by quantum phenomena, one may wonder how it is possible that quantum computers can be essentially more powerful than classical ones. As discussed in more detail in Section 2.1, it is mainly quantum entanglement that has no counterpart in the classical world and makes quantum computation so powerful.

[50] A similar, one-particle interference, has been observed with photons, neutrons and atoms.

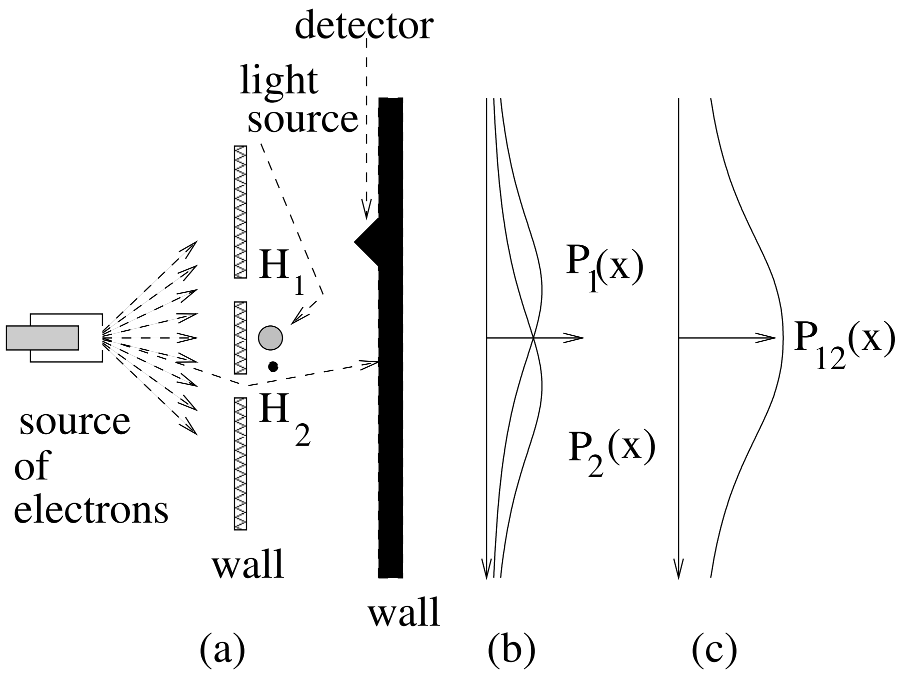


Figure 1.9: Two-slit experiment with an observation

expected. However, the curve for the case where both slits are open is different from that in Figure 1.8c, and actually similar to that in Figure 1.6c. This is again a counter-intuitive phenomenon. The explanation is, that the resulting behaviour of electrons is due to the fact that we have been observing (or at least could observe) their behaviour by putting a light source next to the slits. In this case an **observation** or a **measurement** results in the lost of interference.

We again have a particular case of a very well known phenomenon in the quantum world. A quantum system behaves differently if it is observed from when it is not! Moreover, the interference pattern disappears would we change our original electron experiment in some other way in order to find out which way electrons go. This can also be seen as another example of the uncertainty principle of quantum mechanics, see Section 9.1.2. Either you observe an interference pattern or you can find out which way the electron went, but you cannot do both. Seeing the interference pattern and detecting an electron are both measurements that cannot be performed in the same experiment—one has to choose one or the other. Observe also that detecting which slit an electron went through is sort of a “particle measurement”; recording the interference pattern is a “wave measurement”. One can do one or the other, but not both.

A slight modification of the previous experiment is that instead of a light source we have a single photon. In such a case there is an interaction between the electron and photon. If the photon has a short wavelength, then the interference pattern disappears. However, if a photon of a longer wavelength is used (thereby reducing the momentum kick conveyed to the electron), then the interference pattern is restored when the wavelength is greater then the slit distance.

In the case of an additional photon between the slits, it is useful to turn the experiment on its head and to ask, on the basis of outcomes, whether there was an attempt to measure the way the electron went. The resulting pattern of measurement answers the question. If one see the interference, then there was no measurement. This observation, namely, that we can detect the existence of a measurement from the outcome is the basis of the quantum cryptography discussed in Chapter 6.

The above example suggest how laws of quantum physics can be used to detect eavesdropping during quantum communications. Let Alice be shooting electrons and Bob is observing the probability pattern electrons produce. If he see no superposition then he knows that somebody tried to observe flying electrons; otherwise he can be sure that nobody tried to interfere with electrons.

Another modification of the basic two-slit experiments is obtained by placing another particle with spin before the first wall and close to one of the slits. In such a case electron passing through the slit interacts with the particle and flips the spin (see DiVincenzo and Terhal, 1998). This interaction with the environment is called **decoherence** and in this case decoherence causes the interference pattern to disappear.[51]

A different kind of experiment, the so called "delay choice experiment" employs as the basic component an often used device, a **beam splitter**, which sends the incoming photon one way or another, with equal probability, see Figure 1.10. If the light from the laser is divided in this way, the two beams are physically distinct, but nevertheless coherent. The two routes along which photons travel can be as long as needed. Assume that on one route we install a photon detector, distant enough that we can wait until the photon has passed through the beam splitter to switch the detector on or off. With the detector off we have a standard means of creating an interference pattern. With the detector on, we are actually asking which way the photon went, and therefore we lose the interference pattern.

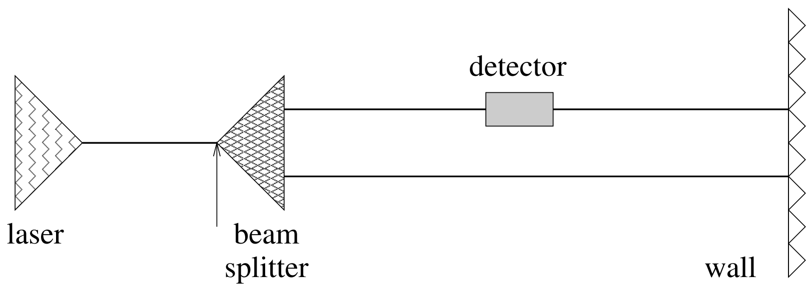


Figure 1.10: Delay choice experiment

1.5.3 Quantum experiments—measurements

As the last example of quantum experiments, let us consider, in an idealized form, one of the other famous experiments of quantum physics which demonstrates that some quantum phenomena are not determined except when they are measured.

The experiment was performed in 1921 by Otto Stern and Walter Gerlach in Berlin (see Figure 1.11). They shot a beam of atoms with random magnetic orientation (thought of, for this experiment, as little bar magnets, with North and South Poles) through a magnetic field, graded by intensity from top to bottom. The magnetic field is created by two vertically

[51] In the essence we keep discussing the same problem over and over because each measurement is an interaction with environment and each interaction with environment can be sees as a measurement.

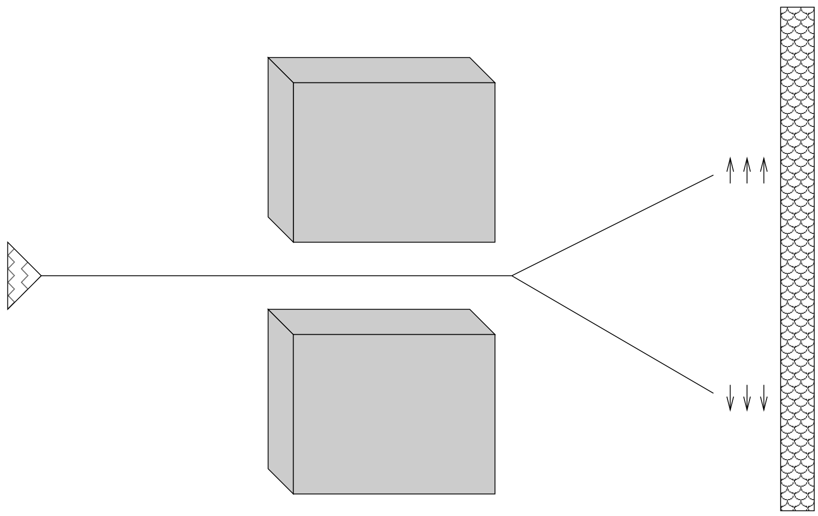


Figure 1.11: Stern-Gerlach experiment with spin-$\frac{1}{2}$ particles

positioned magnets, to sort atoms according to their magnetic orientation, and the result are recorded on a screen. It was expected that the atoms emerging from the magnetic field would be smeared out, when they hit the screen, into a range of all possible deflections, with all possible orientations of their magnetic axes (as it would be the case with real magnets). Instead of that, it was discovered that atoms emerging from the magnetic field struck the screen in exactly two places, each with only one orientation, say "up" or "down", each with equal probability, so they came up in a "half-up and half-down manner". Moreover, the same phenomenon appeared when the magnets themselves were turned ninety degrees so that their magnetic field was horizontal and not vertical. (See Figure 1.12a). Again, the atoms hit the screen in exactly two spots, to the left and right of the original line of the beam and again with the same probability. We can say they came out in a "half-left and half-right" manner.

Actually, no matter how the magnetic field was lined up, it always split the beam of atoms into two. As if each atom was forced somehow to take up either one or the other of just two possible orientations, dictated by the alignment of the magnets.

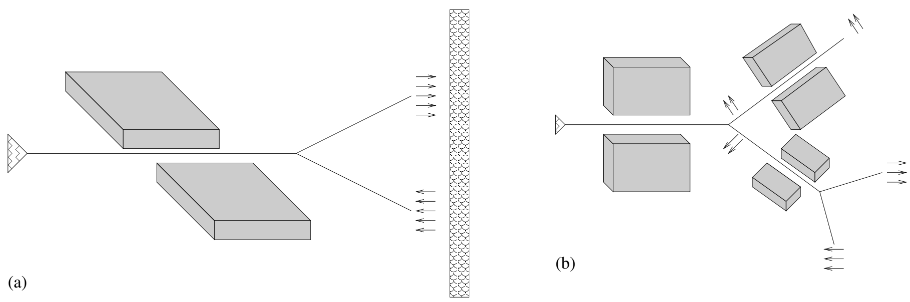


Figure 1.12: Several Stern-Gerlach magnets

It can be demonstrated that magnets do not physically 'sort' atoms passing through by directly manipulating their magnetic axes. The quantum theory explanation is the following one: Passing an atom through a magnetic field amounts to a measurement of its magnetic alignment, and until you make such a measurement there is no sense in saying what the atom's magnetic alignment might be; only when you make a measurement do you obtain one of only two possible outcomes, with equal probability, and those two possibilities are defined by the direction of the magnetic field that you use to make the measurement.

Finally, let us see what happens if a one of the streams of atoms, which came out of the vertically aligned magnet passes again through another Stern-Gerlach magnet—see Figure 1.12b. If the second magnet is again vertically aligned, the stream just gets through without being divided into two. However, if the stream goes through horizontally aligned magnets, then it again comes out in two streams, in a "half-right and half-left manner".

Quantum mechanics provides the following explanation: In the first magnet a measurement is performed with respect to one orientation, up or down. If the second magnet performs the same measurement then the same outcomes happen. However, if atoms with orientation "up or down" are measured in the second magnet with respect to the orientation "left or right", then half of them get oriented left and half get oriented right.

The same experiment can be performed with electrons and other particles. The quality of a particle measured by Stern-Gerlach magnets is called **spin**. [52] A particle is called a spin-$\frac{n}{2}$ particle if the Stern-Gerlach magnets sort the particles into exactly $n+1$ possible outcomes and with equal probability for each outcome.

The Stern-Gerlach experiment has become an important part of the empirical foundations of quantum theory. It helped to discover the basic principles of quantum measurements.

Remark 1.5.1 For a more detailed treatment of basic experiments of quantum mechanics, with respect to unsharp observables (see Bush et al. 1997).

1.6 Quantum Principles

> Quantum mechanics, that mysterious, confusing discipline, which none of us really understands, but which we know how to use.
>
> Murray Gell-Mann (1929-)

In order to explain the basic principles of quantum computing, and to develop quantum algorithms and networks, it is not necessary to explore why the strange behaviour of quantum systems appears as discussed in Section 1.5. It is sufficient to use some basic empirically-based principles of quantum behaviour.[53] In order to formulate these principles we need to introduce some basic concepts: events, states, amplitudes of events, evolution, compound quantum systems and measurement.

Remark 1.6.1 *We are adhering here to the standard (or canonical or orthodox) interpretation of quantum mechanics. To the one actually used by physicists to predict and analyze experimental results. One can also say that we are using the* **Copenhagen interpretation**

[52] Of course, there is much more to say about spins, see, for example, Peres (1993).

[53] See Feynman et al. (1964) and Berthiaume (1997). However, it is not possible to derive the whole of quantum theory from empirically based postulates, and additional postulates have to be introduced with mathematical intuition as the main guide. (For example, the requirement of completeness for Hilbert spaces.)

of quantum mechanics in the sense of Dirac and von Neumann[54]*, in which quantum states are considered as a complete description of reality, measurements are projections, and each physical quantity is represented by a Hermitian operator, called observable.*

1.6.1 States and amplitudes

A **quantum state**[55] is a complete description of a quantum system. By an **event** of a quantum experiment (system, process) we understand a pair

(final state, initial state).

For example, an event in the first experiment with electrons in Figure 1.8a, is that an electron leaves a source s and arrives at the position x.

One of the main goals of quantum mechanics is to predict whether events happen or not, and if they do, then with which probability.

Principle 1.6.2 *The probability p that an event happens is given as $p = |\alpha|^2$, where α is a complex number called the probability amplitude, or simply* **amplitude**[56]*, of the event.*

The amplitude of an event with the initial state IS and a final state FS is usually denoted by

$$\langle \mathrm{FS} \,|\, \mathrm{IS} \rangle,$$

using Dirac's bra-ket notation.

For example, $\langle x|s\rangle$ is to be read as "$\langle x|s\rangle$ is the amplitude of detecting an electron at the position x provided the electron left the source s".

Let us now go back to the first experiment with electrons: for an electron to travel from s to x, it has to pass through one of the slits and to do so it has two options: to go through the slit H_1 or through the slit H_2.

As it is in the case above, an event can often be decomposed into several sequential subevents, and can be combined from several optional and independent ways of happening. The following principle, similar to those with probabilities, expresses how to determine the amplitude of an event given the amplitudes of its subevents.

Principle 1.6.3 *(a) If an event can be decomposed into two sequential subevents, then the amplitude of the event is the product of amplitudes of the subevents.*

(b) If an event consists of several alternative and independent subevents, then the amplitude of the event is the sum of the amplitudes of all subevents considered separately.

For example, from this principle we can derive in our example

$$\begin{aligned} \langle x|s\rangle &= \langle x|\mathrm{wall}\rangle\langle \mathrm{wall}|s\rangle \\ &= \langle x|1\rangle\langle 1|1\rangle\langle 1|s\rangle + \langle x|2\rangle\langle 2|2\rangle\langle 2|s\rangle, \end{aligned}$$

[54]John von Neumann (1903-1957), an American mathematician and physicist of Hungarian origin, one of the leading scientists of his period. He made lasting contributions to almost all areas of modern mathematics and its applications as well as to theoretical physics, especially quantum physics. Von Neumann was one of the leading scientists in the development of the first very powerful electronic computer as well as of the first atomic bomb.

[55]In quantum mechanics, a state is a **ray** in the corresponding Hilbert space. By a ray, an equivalence class of vectors is understood, which differs by multiplication with a non-zero real number. Often the vectors with the norm 1 are considered as representing such equivalence classes. Instead of the term "quantum state", the term "wave-function" is often used in (older) physics-oriented literature.

[56]In some sense, probability amplitudes can be seen as "complex square roots of probabilities".

where $\langle x|i\rangle$ is the amplitude of an electron arriving at x, given that it came out of the slit i. Similarly, $\langle i|s\rangle$ is the amplitude of the event of having an electron arriving at the slit i after leaving the source s. In addition, $\langle i|i\rangle$ expresses the amplitude that an electron arriving to the slit i leaves through the same slit.

Remark 1.6.4 $\langle x|1\rangle\langle 1|s\rangle$ is the amplitude that an electron arrives to x through the slit H_1 after leaving s. However, from that we cannot conclude that $|\langle x|1\rangle\langle 1|s\rangle|^2$ is the probability that an electron actually passes through the slit H_1 to reach x. On the other hand, if we choose to detect the electron at the slit H_1, by magnifying the effect of its presence in the slit to the classical level, then we can see $|\langle 1|s\rangle|^2$ as the probability that the electron is actually presented at the slit H_1. For an interference to occur, we need to ensure that the passage of the electron through the slit remains on the quantum level.

In our experiment, slits are natural elements for expressing events. We can even consider general events $\langle i, j\rangle$, $i, j \in \{1, 2\}$. In such a case:

$$\langle 1|1\rangle = \langle 2|2\rangle = 1,$$
$$\langle 2|1\rangle = \langle 1|2\rangle = 0.$$

A similar situation, as in the above experiment, occurs in quantum systems in general. Whenever an appropriate set of basis states (conditions) is chosen, then any event can be decomposed into events entering and leaving these basis states. This motivates the following definition.

Definition 1.6.5 *The set $B = \{i \,|\, i$ denotes a state$\}$ is a set of* **basis states**, *if for all $i, j \in B$, $\langle i, j\rangle = \delta_{ij}$, i.e.,*

$$\langle i|j\rangle = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{otherwise;} \end{cases}$$

and for any initial state X and any final state Y it holds

$$\langle Y|X\rangle = \sum_{i\in B}\langle Y|i\rangle\langle i|X\rangle. \tag{1.3}$$

The next principle says that such basis states always exist.

Principle 1.6.6 *Any event can be described in terms of a set of states of a basis, often called basis states if the basis is fixed, by giving the transition amplitudes to and from those basis states.*

Our experiment, shown in Figure 1.8, seems to have only one natural set of basis states—electrons are at the slits—but actually, for any experiment and any quantum system, there are infinitely many sets of basis states, and all these sets of basis states have the same cardinality.
[57]

By the first principle, the square of the norm of an amplitude gives the probability of the corresponding event. By the last principle, the amplitude of an event is the sum of

[57] Basis states might be various possible locations of a particle, or some other properties of a particle, such as its spin value. For any nonnegative integer n there is a spin-$\frac{n}{2}$ particle. Its spin has $n+1$ possible values and the corresponding Hilbert space has dimension $n+1$. For example, pions are spin-0 particles, and electrons are spin-$\frac{1}{2}$ particles.

amplitudes of subevents corresponding to the given set of basis states. The set of basis states is to be "complete" (none can be added), two basis states have to be orthogonal (this refers to the fact that $\langle i|j\rangle = 0$ if $i \neq j$, $i, j \in B$), and, finally, the sum of probabilities for reaching a basis state from any initial state has to be one. We have therefore a motivation for the next principle.

Principle 1.6.7 *The amplitude of an event is a sum of amplitudes of events corresponding to the given set of basis states. In addition, for any set B of basis states and for any initial state X*

$$\sum_{i \in B} |\langle i|X\rangle|^2 = 1.$$

The proper mathematical framework to deal with quantum systems is that of Hilbert spaces, as discussed in Section 1.4 and in more detail in Appendix 9.2. The basic principle is:

Principle 1.6.8 *To each quantum system there corresponds a Hilbert space whose dimension equals the maximum number of reliably distinguishable states of the system.*

In the Hilbert space framework, to each state X of the system a (column) **state vector**, or **ket-vector**, $|X\rangle$ and also a (row) **bra-vector**$\langle X|$ correspond. In addition, a scalar product "$\cdot$" of a bra-vektor and a ket-vector is defined in such a way that $\langle X| \cdot |Y\rangle = \langle X|Y\rangle$ for any states X, Y of the system.

For ket- and bra-vectors similar relations hold as in 1.3. They are summarized in the next principle:

Principle 1.6.9 *For any states X, Y of a quantum system, and any set B of basis states it holds:*

$$|X\rangle = \sum_{i \in B} |i\rangle\langle i|X\rangle,^{58} \tag{1.4}$$

and

$$\langle Y| = \sum_{i \in B} \langle Y|i\rangle\langle i|. \tag{1.5}$$

In addition, if $|X\rangle = \sum_{i \in B} \alpha_i |i\rangle$, then $\langle X| = \sum_{i \in B} \alpha_i^* \langle i|$.

The base $\mathcal{B}$ and the sequence of complex numbers $\{\langle i|x\rangle\}_{i \in \mathcal{B}}$ form the so-called **quantum information** of the state $|x\rangle = \sum_{i \in \mathcal{B}} \langle i|x\rangle|i\rangle$.

Exercise 1.6.10 *Show that if $\langle Y| = \sum_{i \in B} \beta_i \langle i|$ and $|X\rangle = \sum_{i \in B} \alpha_i |i\rangle$, then $\langle Y|X\rangle = \sum_{i \in B} \alpha_i \beta_i$.*

Observe that equations 1.4 and 1.5 can be seen as being obtained from 1.3 by abstracting either from the final condition or from the initial condition.

[58] In more mathematically-oriented literature the equality 1.4 is usually written in the form $|X\rangle = \sum_{i \in B} \langle i|X\rangle|i\rangle$.

1.6.2 Measurements—the projection approach

The next principle deals, in a simplified form, with perhaps the most complex and questionable of the basic problems of quantum mechanics—with observations, or measurements, or the extraction of information from quantum states.

Before quantum mechanics, it was taken for granted that when one measures something one is gaining knowledge of a pre-existing state—an independent fact about the world. The Copenhagen interpretation says otherwise: some quantum things or properties are not determined except when they are measured and it is only by being measured that they take a specific (classical) value.[59]

The basic problem of quantum measurement is the following one: given a measuring instrument or an observation device, or simply a test T, and a quantum state $|\phi\rangle$, what kind of information can T get out of $|\phi\rangle$?

In Section 1.4, and in a more detailed survey of Hilbert spaces in Appendix 9.2, T is considered to be a self-adjoint operator. Now we will take an equivalent but more transparent approach and model an observation through a special decomposition, called observable, of the Hilbert space corresponding to the given quantum system.

Definition 1.6.11 *Let H be a Hilbert space of states of a quantum system. An* **observable** $\mathcal{O} = \{E_1, \ldots, E_k\}$ *is a set of disjoint and mutually orthogonal subspaces such that*

$$H = E_1 \oplus E_2 \oplus \ldots \oplus E_k \quad \{\textit{an orthogonal sum}\}$$

and an injective mapping $\mu : \{E_1, \ldots, E_k\} \rightarrow \mathbf{R}$.

The following principle states the effect of an observable $\mathcal{O}$ on a state $|\phi\rangle$.

Principle 1.6.12 *Let $|\phi\rangle$ be a state and $\mathcal{O} = \{E_1, \ldots, E_k\}$ be an observable. $|\phi\rangle$ can be expressed uniquely, as a linear superposition of its components (projections) along each of the E_i's:*

$$|\phi\rangle = \sum_{i=1}^{k} \alpha_i |\phi_{E_i}\rangle,$$

where $|\phi_{E_i}\rangle$ is a state in E_i (the projection of $|\phi\rangle$ into E_i), and $\langle\phi_{E_i}|\phi_{E_i}\rangle = 1$ for all i. (Uniqueness is up to a phase factor.)

An observation (measurement) of $|\phi\rangle$ by $\mathcal{O}$ has the following consequences.

1. *One of the subspaces $E_1, \ldots, E_k$, say E_i, is selected and the value $\mu(E_i)$ is produced. The probability that a subspace E_i is selected is $|\alpha_i|^2$.*

2. *After the observation, the state $|\phi\rangle$ "collapses" into the (renormalized) state $|\phi_{E_i}\rangle$.*[60]

3. *The only classical information given by $\mathcal{O}$ is the value of the function μ. In the case $\mu(E_i) = i$, this is just information which of the subspaces $E_1, \ldots, E_k$ was selected (or into which of subspaces the state $|\phi\rangle$ was projected). All information not in $|\phi_{E_i}\rangle$ is irreversibly lost.*

---

[59] For a more detailed discussion of the measurement problem see Appendix 9.1.4.

[60] A state $|\phi\rangle$ is seen as having "norm" 1 if $\langle\phi|\phi\rangle = 1$. If $\langle\phi|\phi\rangle = k > 0$, then for the state $|\phi'\rangle = \frac{1}{\sqrt{k}}|\phi\rangle$ it holds that $\langle\phi'|\phi'\rangle = 1$. $|\phi'\rangle$ is said to be obtained from $|\phi\rangle$ by "renormalization".

A measurement with the observable $\mathcal{O} = \{E_1, \ldots, E_k\}$ therefore causes the quantum system to behave randomly and to "destroy" its pre-measurement state $|\psi\rangle$ unless it falls entirely into one of the subspaces E_i, $1 \leq i \leq k$.

Example 1.6.13 *Let us assume that we have a quantum system with exactly two basis states $|0\rangle$ and $|1\rangle$. (And therefore a two dimensional Hilbert space corresponds to it.)*[61]

The so-called **standard observable** *for a state $|\phi\rangle = \alpha|0\rangle + \beta|1\rangle$, with $|\alpha|^2 + |\beta|^2 = 1$, is $\mathcal{B} = \{E_0, E_1\}$, where E_i, $i = 0, 1$ is the linear subspace generated (spanned) by the vector $|i\rangle$. An example of another, so-called "dual" observable for ϕ, is $\mathcal{O} = \{E_0', E_1'\}$, where E_0' is the linear subspace generated by the vector $|0'\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and E_1' is the subspace generated by the vector $|1'\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$.*[62]

A measurement in an n-dimensional Hilbert space can produce up to n different values. If it can provide exactly n such values, we speak about **maximal measurement**. Each maximal measurement corresponds to a choice of an orthonormal basis in H_n. If $\mathcal{B} = \{\phi_i\}_{i=1}^n$ is such an orthonormal basis and a state $|\psi\rangle$ is written in the form

$$|\psi\rangle = \sum_{i=1}^{n} \alpha_i |\phi_i\rangle \tag{1.6}$$

then we say that 1.6 expresses the state being measured, the measurement being made (given by the observable defined by the bases) and the probabilities of various measurement outcomes: $\{|\alpha_i|^2\}_{i=1}^n$. (In such a case it is implicitly assumed that the numerical value is provided by the mapping $\mu(E_i) = i$.)

Remark 1.6.14 Each outcome of a quantum measurement is given with a certain probability that is uniquely determined by the state being measured and the observable being used. However, these probabilities are not a "cause of our ignorance". One cannot improve them using more sophisticated measurements. Results of quantum mechanics are probabilistic not because of our insufficient understanding of the reality, but because quantum theory itself has "nothing more to say", or even because Nature has "nothing more to say".

The concept of an observation or a measurement is a formalization of one of the basic principles of quantum mechanics: the particular result you see in any measurement depends not just on the things (state) you are measuring, but on the way you set up your measuring device (observable) to measure it. The phenomenon you measure, such as the spin of particles, is a strictly undetermined quality until a measurement yields a measurable value of it. At the same time, it is going beyond the bounds of quantum mechanics to think that one can deduce from any measurement in which of the prior states the measured system "really was". Measurements assign values to quantities that were till that moment undetermined.

Remark 1.6.15 There is a very natural and simple way to define a proper observable in the case of quantum finite automata (see Section 4.1.3), that fully corresponds to what is done in the case of computations of classical finite automata. In this case the corresponding Hilbert space is the one spanned by all configurations of the automaton, and they form its basis.

[61] We can well imagine $|0\rangle$ as the (column) vector $(1,0)^T$ and $|1\rangle$ as the (column) vector $(0,1)^T$.

[62] We will see in Section 2.2 how one can make use of such an observable. The basic idea is that the same state observed through different observables can give a definite answer in one case and a completely random answer in another case.

This Hilbert space is then partitioned into three subspaces: the one spanned by accepting configurations, the second spanned by rejecting configurations and the third spanned by other configurations. A measurement, that is to be performed after each "evolution step", produces one of the answers "accept", "reject" and "non-halting". In the case that in a particular state all halting (accepting or rejecting) configurations have amplitudes 0, the measurement has no effect on the state because the state collapses into itself.

Remark 1.6.16 One can see the following representation of a quantum state

$$|\psi\rangle = \sum_{i \in \mathcal{B}} \alpha_i |i\rangle$$

as encapsulating the information one needs to determine probabilities of different outcomes of the measurement with respect to the observable given by subspaces spanned by vectors $|i\rangle$. A quantum state is therefore not a property of a quantum system (some particle) in isolation but a combined description of the "thing being measured" and the "measurement being made". (This is, of course, one of the "working interpretations" of the Copenhagen interpretation.[63] Some disagree and consider a quantum state as existing per se.)

The concepts of a state and a measurement are so crucial for quantum mechanics that one can say that the essence of quantum theory is to provide a mathematical representation of states (and thereby of observables) and rules for computing the probabilities of various outcomes of possible measurements.

Remark 1.6.17 In the last 15 years new measurement theoretical and operational foundations have been laid out for a realistic interpretation of quantum mechanics as a theory for individual quantum systems and, the so-called *operational quantum physics* (see Bush et al. 1997), has been developed. Fundamental to this approach is a distinction between sharp and unsharp observables. According to this interpretation quantum mechanics describes physical systems existing independently, once they have been prepared or identified by an observable.

1.6.3 Evolution of quantum systems

> A quantum system is a useful abstraction which frequently appears in the literature, but does not really exists in nature.
>
> Asher Peres (1993)

In a very general sense, a quantum system is whatever admits a closed dynamical description within quantum theory.

So far, we have considered only the static case, namely that the initial condition (state vector) $|X\rangle$ does not change after being set. We could then consider the amplitudes $\langle Y|X\rangle$ for different Y. In the evolution of a quantum system, in particular in a computational process, some transformation of the initial state has to be performed. On the physical side,

[63] For this book it does not seem to matter which of the interpretations of the Copenhagen interpretations is considered.

some apparatus A is used. On the mathematical side, some operator A is used that maps one state into another.

The main question of interest now is the following one: what is the amplitude of the event $\langle Y|A$ applied to$X\rangle$? Or, in a more common notation, what is the value of

$$\langle Y|A|X\rangle.$$

Fortunately, all such operators A of quantum systems are well understood. They are linear operators of a special form:

Principle 1.6.18 *To an evolution in an isolated quantum system there corresponds a transformation by a unitary operator in its Hilbert space. Relative to a given basis $\mathcal{B}$, a unitary operator P is represented by a unitary matrix $M_P^{\mathcal{B}}$ such that, for $i,j \in B$, $M_P^{\mathcal{B}}[i,j]$ is the amplitude of the transition from the state j to the state i.*

Since unitary matrices preserve the norm of state vectors they can be seen as performing rotations on quantum states. This actually means that all we can do with quantum states is to "rotate" them.

In principle, to any unitary matrix there exists a quantum system evolving according to that matrix. However, the real design of such systems can give rise to formidable technological problems.

On the other hand, a variety of experimental techniques have already been developed to realize unitary evolutions of two-state quantum systems. This is of practical importance because there is a technique, see Section 5.1, to decompose large unitary matrices into a product of simpler ones, which represent the evolutions of two-state systems.

In a more detailed way, the evolution of a quantum system S is described by the linear Schrödinger equation.[64]

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

where $\hbar$ is the Planck constant, $H(t)$ is an observable of S called Hamiltonian of the system (which is a quantum analogue of Hamiltonian of the classical mechanics—the total energy of the system—and can be represented by a Hermitian matrix) in time t, and $|\psi(t)\rangle$ is the state of $\mathcal{S}$ at the time t. $\psi(t)$ is also called the **wave function** of $\mathcal{S}$.

In the case where the Hamiltonian is time independent, the formal solution of the Schrödinger equation has the form

$$|\psi(t)\rangle = \mathrm{U}(t)|\psi(0)\rangle$$

and

$$\mathrm{U}(t) = e^{-Ht/\hbar}$$

is the evolution operator which can be represented by a unitary matrix.

Exercise 1.6.19 *State which of the following matrices are unitary or Hermitian:*

$$(a) \begin{pmatrix} \frac{1+i}{2} & \frac{1-i}{2} \\ \frac{1-i}{2} & \frac{1+i}{2} \end{pmatrix} \; (b) \begin{pmatrix} 2 & 1-3i \\ 1+3i & 5 \end{pmatrix} \; (c) \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}.$$

[64] Erwin Schrödinger (1887-1961) an Austrian theoretical physicist. He formulated the basic equation for quantum evolution that now bears his name. For this equation and the total contribution to quantum mechanics he shared, with P. Dirac, the 1933 Nobel prize for physics. Schrödinger was the first to notice, in 1935, entanglement as a phenomenon of quantum physics.

Exercise 1.6.20 *Determine all eigenvalues and eigenvectors of the matrices:*

$$(a) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (b) \begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix} \quad (c) \begin{pmatrix} a & b \\ b & a \end{pmatrix} \quad (d) \begin{pmatrix} \cos\theta & e^{i\phi}\sin\theta \\ e^{i\phi}\sin\theta & -\cos\theta \end{pmatrix}.$$

There is also another way to see that linear operators representing quantum evolution in finite dimensional quantum systems have to be unitary.

A quantum evolution operator A has to map quantum states into quantum states. This implies that for any state x it has to hold

$$\langle Ax|Ax\rangle = \langle x|x\rangle = 1$$

and therefore

$$\langle x|x\rangle = \langle A^*Ax|x\rangle$$

which yields

$$A^*A = I.$$

The last equation also implies that if $Ax = y$, then $A^*y = x$ and therefore each quantum process is reversible and A^* is the operator for the reverse process corresponding to the process given by A.

If A is a finite matrix, then $AA^* = I \iff A^*A = I$. However, this is not the case for infinite matrices. For example, if

$$A = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

then $AA^* = I$, but $A^*A \neq I$.

Finally, observe that the equality $AA^* = I$ is equivalent to the assertion that row vectors of A are orthonormal and the property $A^*A = I$ is equivalent to the claim that column vectors of A are orthonormal.

Exercise 1.6.21 *Show that if a matrix A is such that $A^*A = I$, then A is unitary if and only if the mapping defined by A is surjective.*

Exercise 1.6.22 *Show that if A_1, B_1, A_2, B_2 are unitary matrices, then* $||B_1B_2 - A_1A_2|| \leq ||B_1 - A_1|| + ||B_2 - A_2||$.

The quantum evolution of a pure state $|\phi_0\rangle$ can also be seen as a sequence[65] of pure states $|\phi_0\rangle, |\phi_1\rangle, |\phi_2\rangle, \dots$ and unitary operators $U_1, U_2, U_3, \dots$ such that $|\phi_i\rangle = U_i|\phi_{i-1}\rangle$, for $i = 1, 2, 3, \dots$.

Similarly, the evolution of a density matrix ρ_0 can be seen as a sequence of density matrices $\rho_0, \rho_1, \rho_2, \dots$ and unitary operators $U_1, U_2, U_3, \dots$ such that $\rho_i = U_i\rho_{i-1}U_i^*$ for $i = 1, 2, 3, \dots$.

[65] Each quantum evolution is in principle a continuous process and therefore such a discretization is only a useful computational simplification and abstraction.

The following property of the tracing out operation is also of importance:

$$Tr_{H_B}((U_A \otimes U_B)\rho(U_A \otimes U_B)^*) = U_A(Tr_{H_B}\rho)U_A^*,$$

where U_A (U_B) is a unitary operation on the Hilbert space H_A (H_B) and ρ is a density matrix of $H = H_A \otimes H_B$. This implies that we can commute two important operations on density matrices: an evolution step and a tracing out operation, in the following sense. We can either first perform an evolution step of ρ in H and then trace out H_B, or first trace out H_B and then make an evolution step on the resulting density matrix in H_A.

1.6.4 Compound quantum systems

There are two basic situations when composition and decomposition of quantum systems and of the corresponding Hilbert spaces has to be handled.

The first one is the problem of measurement. Let $\mathcal{S}$ be a quantum system and $H_{\mathcal{S}}$ the Hilbert space associated to $\mathcal{S}$. Intuitively, a measurement is an event performed in order to receive information about the value of an observable A in a state $|\phi\rangle$. In any case, a measurement has to be performed using some apparatus $\mathcal{A}$ to which another quantum system, and Hilbert space, say $H_{\mathcal{A}}$, corresponds. Initially these two systems can be assumed to be disjoint and noninteracting. However, during a measurement process they have to interact and to form some compound quantum system to which a composition of the corresponding Hilbert spaces $H_{\mathcal{S}}$ and $H_{\mathcal{A}}$ corresponds. In the Hilbert space $H_{\mathcal{A}}$ it has to exist a counterpart to A, say B, and it is assumed that to each of the possible values of A a value of B corresponds and that by observing that value of the measurement of B we get some information about the value of A in the state $|\phi\rangle$.

The second situation is at the design of quantum algorithms and quantum networks. For various reasons, it is desirable that systems with which we work or which we design can be seen as being composed in some systematic way from small, especially two-dimensional, Hilbert spaces.

The final principle deals with the relation between a compound quantum system and the tensor product of the corresponding Hilbert spaces.

Principle 1.6.23 *Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two quantum systems and let H_1 and H_2 be corresponding Hilbert spaces. Let the compound system of $\mathcal{S}_1$ and $\mathcal{S}_2$ be $\mathcal{S}$. It holds:*

1. *The tensor product $H = H_1 \otimes H_2$ is the Hilbert space associated to $\mathcal{S}$.*
2. *Observables of $\mathcal{S}$ are self-adjoint operators in H.*
3. *Evolutions in $\mathcal{S}$ are determined by unitary operators of H.*

Remark 1.6.24 The state spaces of n particles classically combines through the cartesian product and quantumly through the tensor product. In order to understand quantum computation, it is crucial to see the difference between cartesian and tensor product. The cartesian product of two subspaces $X \times Y$ has dimension $dim(X \times Y) = dim(X) + dim(Y)$. For the tensor product $X \otimes Y$ we have $dim(X \otimes Y) = dim(X) \times dim(Y)$.

1.6.5 Quantum theory interpretations

The physical interpretation of the abstract formalism of quantum theory has been a very controversial issue from the early days of quantum mechanics. Everyone agrees with Hilbert space **formalism** of quantum mechanics—the controversial issue is that of **interpretation**: how to relate constructs of the formalism to the observable physical reality. The best protagonists of quantum theory addressed this issue, but solutions they reached are conflicting and ambiguous.

The crucial issue is that of the measurement. The main way out has been to introduce "observers" who interact with the quantum world, sense outcomes of its evolution, and report the result into the classical world. This implies that an observer (or a measuring device) has to obey the laws of the quantum world when collecting information about the measured state and, at the same time, it has to be an object of the classical world, or at least to use the language of the classical world when reporting the outcome of a "measurement" or an "observation". This dichotomy is the root of the quantum measurement dilemma. It seems also that there can be no unambiguous ways to deal with this dichotomy, only some better or worse approximations.

Behind the quantum-to-classical dichotomy there is a mystery: are such concepts as quantum state a genuine physical reality, or only abstract concepts and tools invented to compute probabilities of those classical events we are able to perceive? The basic underlying problem is whether quantum theory is a theory of physical reality or only of our (limited) perception of the physical world.[66]

Remark 1.6.25 The most general type of measurements, the so-called generalized measurement or POVM measurement, are discussed in Section 9.2.8. They allow in some cases to extract more information from a quantum state then projection measurements and therefore they are needed to consider as a possible tools of an eavesdropper when security of quantum cryptographic protocols is investigated.

1.7 Classical Reversible Gates and Computing

Everything happens to everybody sooner or later if there is time enough.

Georg Bernard Shaw (1856–1950)

The original motivation for the study of reversibility in classical computing came from the observation that heat dissipation is one of the major obstacles for miniaturization of classical computers and the fact that the second law of thermodynamics implies that irreversible state changes during computation must dissipate heat.[67]This is nowadays known (see Bennett, 1998b), as:

> **Landauer's principle.** *To erase a bit of classical information within a computer, 1 bit of entropy must be expelled into the computer's environment (typically*

[66]Surprisingly, we have no really satisfactory reason for assigning objective existence to physical quantities as distinguished from numbers which we correlate with them. There are no reasons to suppose that a particle has at any moment a definite, but unknown, position which may be revealed by a measurement. On the contrary, we run into contradictions by assuming that (see Peres, 1993).

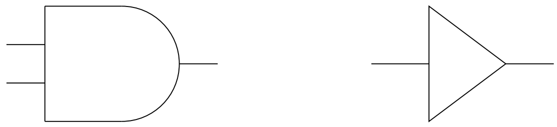
[67]In modern computers heat dissipation is about $kT10^8$ per logical operation. The heat must be removed by external means. For example, by constant cooling of all components of classical computers by the thermal coupling of the circuits to a heat reservoir, i.e. air. However, for quantum computing such a cooling by thermal coupling is not an option because it would lead to decoherence effects—see Section 7.2.2—and would destroy the superpositions of states, an important source of the power of quantum computing.

in the form of waste heat).

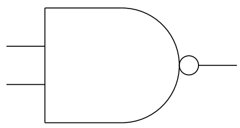
The importance of the investigations of the classical reversible computing follows from the fact that such computations are special cases of quantum computations.

1.7.1 Reversible gates

It is well known that two Boolean operations, or two gates, for instance NOT and AND, Figure 1.13a, b, are sufficient to design Boolean circuits for any Boolean function $f \in \mathcal{B}_n^m = \{g \,|\, g : \{0,1\}^n \to \{0,1\}^m\}$. Even, a single NAND gate, Figure 1.13c, is universal in the sense that it alone is sufficient to design a circuit for any Boolean function.



(a) AND (b) NOT



A	B	A NAND B
0	0	1
0	1	1
1	0	1
1	1	0

(c) NAND

Figure 1.13: Classical Boolean gates

Unfortunately, both AND and NAND are **irreversible** Boolean operations. By that we mean that from the output value(s) of the gate one cannot determine unambiguously the input values; information gets irreversibly lost "during the gate operation".

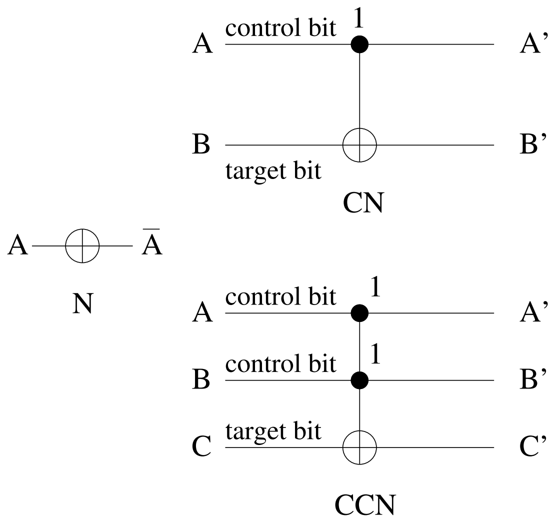
We talk about a **reversible** gate, Boolean function, operation or computation, as one with always enough information in the outputs to deduce the inputs unambiguously. Such operations, gates and computations are crucial for quantum computing because of the reversibility of the evolution in quantum physics.

Three reversible gates have turned out to be of special importance: the usual NOT gate (N), CONTROL NOT gate (CN or CNOT or XOR) (see Figure 1.14a), CONTROL CONTROL NOT gate (CCN or CCNOT), see Figure 1.14b, also called **Toffoli gate**.

In the CN gate $A' = A$, i.e. the input A gets through unchanged. The filled circle on the first wire represents a control in the following sense: if $A = 0$ then $\oplus$ on the second wire just lets the signal B get through and therefore $B' = B$. If $A = 1$, then $\oplus$ on the second wire acts as a NOT gate and $B' = \bar{B}$. In the CCN gate $A' = A$ and $B' = B$. The $\oplus$ on the last wire acts as a NOT gate but only if $A = B = 1$.

It has been shown that any reversible Boolean function in $\mathcal{B}_n^m, n \geq 3$, can be computed by a reversible network composed from the gates N, CN and CCN. It has actually been shown that the single 3-input and 3-output Toffoli gate, Figure 1.14, and the so-called **Fredkin gate**, Figure 1.15b are universal.[68]

[68]Observe that in Fredkin's gate the output has always the same number of 1's as the input.



A	B	A'	B'
0	0	0	0
0	1	0	1
1	0	1	1
1	1	1	0

A	B	C	A'	B'	C'
0	0	0	0	0	0
0	0	1	0	0	1
0	1	0	0	1	0
0	1	1	0	1	1
1	0	0	1	0	0
1	0	1	1	0	1
1	1	0	1	1	1
1	1	1	1	1	0

Figure 1.14: Reversible gates N, CN and CCN (or Toffoli gate)

Exercise 1.7.1 *Find all functions* $f : \{0,1\}^3 \rightarrow \{0,1\}$ *such that the mapping* $(a,b,c) \rightarrow (a,b,f(a,b,c))$ *is injective.*

Example 1.7.2 *It is actually easy to show universality of Fredkin and Toffoli gates. In the Toffoli gate, if* $A = 1$ *then* $B' = B \oplus C$*; if* $B = 0$*, then* $B' = A \wedge C$*; if* $A = 1, C = 1$*, then* $B' = \bar{B}$*. For the Fredkin gate, if* $C = 0$ *then* $B' = A \wedge B$*; if* $B = 0$ *and* $C = 1$*, then* $B' = \bar{A}$*. Thus, gates AND and NOT are realizable using both Fredkin and Toffoli gates.*

The fact that for each reversible Boolean function there is a Boolean circuit composed of copies of a single reversible gate is certainly of interest. However, how important is this really for computing in general? In most of the computational tasks

$$\text{given } a \text{ compute } f(a),$$

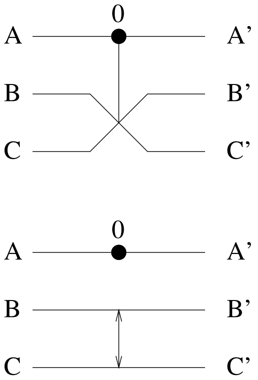
f is not one-to-one and therefore no reversible computation of f is possible.

However, after an inessential modification of the original computational task, namely: compute $f_0(a)$, where

$$f_0 : a \rightarrow (a, f(a)),$$

In short, the Fredkin gate realizes the function $f(a,b,c) = (a, ab \vee \bar{a}c, \bar{a}b \vee ac)$ and the Toffoli gate the function $f'(a,b,c) = (a, b \oplus (a \wedge c), c)$ or $f'(a,b,c) = (a,b,c \oplus (a \wedge b))$. Observe that there are two ways the Toffoli gate is considered—see Figure 1.15c.

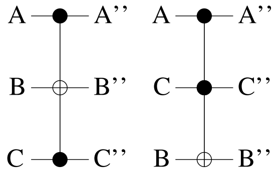
The CN gate, The Toffoli gate and the Fredkin gate were first presented by C. A. Petri in 1965, but their publication in 1967, in German and in a not too widespread Proceedings, went apparently unnoticed by most of those working on reversible computing. However, in view of the above fact, it would perhaps be historically more proper to talk about Petri-Toffoli and Petri-Fredkin gates. Petri has also shown the universality of these two gates for classical reversible computing.



(a) Fredkin gate notations

A	B	C	Fredkin gate A'	B'	C'	Toffoli gate A''	B''	C''
0	0	0	0	0	0	0	0	0
0	0	1	0	1	0	0	0	1
0	1	0	0	0	1	0	1	0
0	1	1	0	1	1	0	1	1
1	0	0	1	0	0	1	0	0
1	0	1	1	0	1	1	1	1
1	1	0	1	1	0	1	1	0
1	1	1	1	1	1	1	0	1

(b) Fredkin and Toffoli gates



(c) Toffoli gate notations

Figure 1.15: Fredkin and Toffoli universal reversible gates

we have a one-to-one function f_0 to compute. In the case of the Boolean functions $f : \{0,1\}^m \to \{0,1\}^m$ we write the corresponding one-to-one function $f_0 : \{0,1\}^{2m} \to \{0,1\}^{2m}$ as $f_0 : (x, b) \to (x, b \oplus f(x))$, where $\oplus$ is applied bitwise.

Example 1.7.3 *The two bit adder computation mapping:*

$$(a, b) \to (a \oplus b, a \wedge b) \qquad \{(sum,\ carry)\}$$

is not one-to-one. However, the modified computational task: $(a, b) \to (a, a \oplus b, a \wedge b)$ *is to compute a one-to-one mapping and can be realized using one CN gate and one CCN gate, with the constant input* $c = 0$*, as shown in Figure 1.16.*

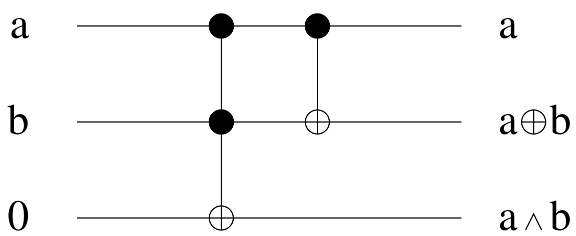


Figure 1.16: A reversible implementation of a two-bit adder

Exercise 1.7.4 *(a) Design a reversible circuit for a three-bit adder; (b) design a reversible circuit for multiplication of two-bit integers; (c) design a reversible circuit with six inputs and outputs such that one of the outputs determines whether the input* 5*-bit word is a palindrome.*

1.7.2 Reversible Turing machines

The concept of reversibility can be defined for any type of computer. For example, a one-tape deterministic Turing machine is called reversible if each configuration uniquely determines the previous one.

It can be shown that if a TM $\mathcal{M} = \langle \Sigma, Q, q_0, \delta \rangle$ is reversible, then there is another TM $\mathcal{M}' = \langle \Sigma, Q, q_0', \delta' \rangle$ (that can much differ from $\mathcal{M}$), such that if $\mathcal{M}$ transforms a configuration c to a configuration c', that is $\mathcal{M}(c) = c'$, then $\mathcal{M}'(c') = c$.

Exercise 1.7.5 *Find a necessary and sufficient condition for the transition function of a one-tape Turing machine $\mathcal{M} = \langle \Sigma, Q, q_0, \delta \rangle$ to be reversible.*

Bennett (1973) has shown that if a function f is computable by a one-tape Turing machine in time $t(n)$, then there is a three-tape reversible Turing machine computing, with the constant time overhead, the mapping:

$$a \to (a, j(a), f(a)),$$

where $j(a)$ is some "garbage". ($j(a)$ accumulates some "history" of computation—information that would normally be "thrown away".)

Unfortunately, the amount of garbage in the original Bennett's construction was very large and therefore the price for reversibility in terms of space was huge.

Fortunately, as Bennett (1973) has already shown, there is a reversible way to get rid of the garbage and therefore the garbage does not have to accumulate as computation proceeds. The basic idea of removing garbage can be sketched as follows:

Basic computation: $f : a \to (a, j(a), f(a))$.

Fanout: $(a, j(a), f(a)) \to (a, j(a), f(a), f(a))$.

"Uncomputing" f: $(a, j(a), f(a), f(a)) \to (a, f(a))$.

All these three computational tasks are clearly reversible. Moreover, fanout does not require any additional garbage space.

Example 1.7.6 *For any function $f : \{0,1\} \to \{0,1\}$, the mapping $(x, 0) \to (x, f(x))$ is one-to-one and therefore we can compute $f^4(x)$ reversibly as follows:*

$$\begin{aligned} (x,0,0,0,0) &\to (x, f(x), 0, 0, 0) \to (x, f(x), f^2(x), 0, 0) && (1.7)\\ &\to (x, f(x), f^2(x), f^3(x), 0) \to (x, f(x), f^2(x), f^3(x), f^4(x)) && (1.8)\end{aligned}$$

$f^4(x)$ can now be copied and the "garbage" $f(x), f^2(x), f^3(x)$ can be removed by "uncomputing" as illustrated in Figure 1.17.

In Bennett (1989), it is shown that any TM can be simulated by a reversible TM with constant overhead in time and only quadratic increase in space.

Let us now discuss in more detail (see Vazirani, 1997), reversible computation on Turing machines. For simplicity, let us consider only an one-tape Turing machine M computing a function f. In order to design a reversible Turing machine M' to perform the computation $x \to (x, f(x))$, an additional tape, a "history tape", is added. Each step of M will be simulated by two steps of M'. In the first step, the current state and tape symbol read by M are written, as one (new) symbol, on the history tape. This symbol is then read and the corresponding step of M is performed. Such a computation can be performed in time $2t(n)$,

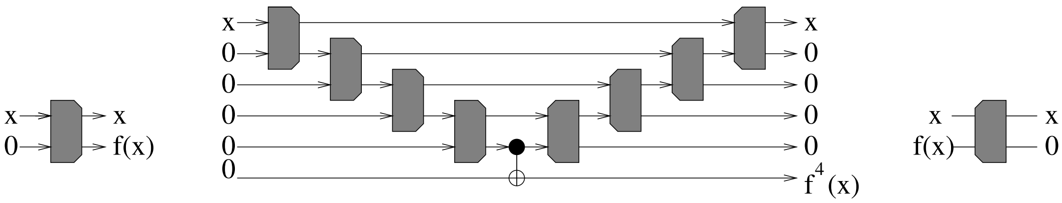


Figure 1.17: Reversible computation with garbage removal

where $t(n)$ is the computational time of M. However, the space requirement increases to $t(n) + s(n)$, where $s(n)$ is the space requirement of M.

Space requirements can be reduced, using an elegant simulation technique, to $\mathcal{O}(s(n)\lg t(n))$ by only slightly increasing the time to $t^{1+\varepsilon}(n)$ for an arbitrarily small ε (see Bennett, 1989). The basic idea goes as follows.

The first half of the computation of M is performed and the resulting configuration c is stored on the history tape. After that, the second half of the computation of M is performed. To perform the reversible computation step by step, computation starts with c and in each step the current state and symbol read are saved on the history tape. This allows to go from the end configuration backwards, step by step, till c is reached. At that moment the contents of the history tape are erased, except for c. Computation then starts from the very beginning and again in each step both the state and the symbol read are saved on the history tape. When c is reached, the steps are performed backwards, one by one, until the initial configuration is reached. In this way, the space is halved at the cost of doing four times more computation.

This technique can be improved by simulating computations in both halves recursively in the same way. The resulting resource requirements are: $Space(t(n)) = Space(\frac{t(n)}{2}) + s(n)$, yielding $Space(t(n)) = \mathcal{O}(s(n)\lg t(n))$. Time requirements are $Time(t(n)) = 6\,Time(\frac{t(n)}{2}) + \mathcal{O}(n)$, which yields $Time(t(n)) = t(n)^{\lg 6}$. An improvement can be obtained by breaking the computation not into two but into k parts. In such a case we get the following recurrence for time $Time(t(n)) \leq 4k\,Time(\frac{t(n)}{k}) + \mathcal{O}(n)$, which yields $t^{\lg 4k}(n)$ for time and in a similar way we get the bound $\mathcal{O}(ks(n)\lg_k t(n))$ for space requirements.

1.7.3 Billiard ball model of (reversible) computing

The billiard ball model, developed by Fredkin and Toffoli (1982), is an excellent example of a toy scientific model of no immediate practical application but of large scientific impact.

Balls of radius 1 travel on a unit grid in two directions, see Figure 1.18a. The direction of their movements can be changed either by an elastic collision, Figure 1.18b, or by a reflection at a "mirror", see Figure 1.18c. For example, Figures 1.18d and 1.18e show the billiard ball model implementation of a shift and a delay of the signals. Figure 1.20 shows a billiard ball implementation of the switch gate from Figure 1.19.

Switch gates are of importance because with four of them the Toffoli gate can be implemented as shown in Figure 1.21.

Remark 1.7.7 In practice (irreversible) computers in use today dissipate orders of magnitude more heat per bit processing than the theoretical lower bound $1kT\ln 2$, given by Landauer's principle. However, if computer hardware continues to shrink in size as so far,

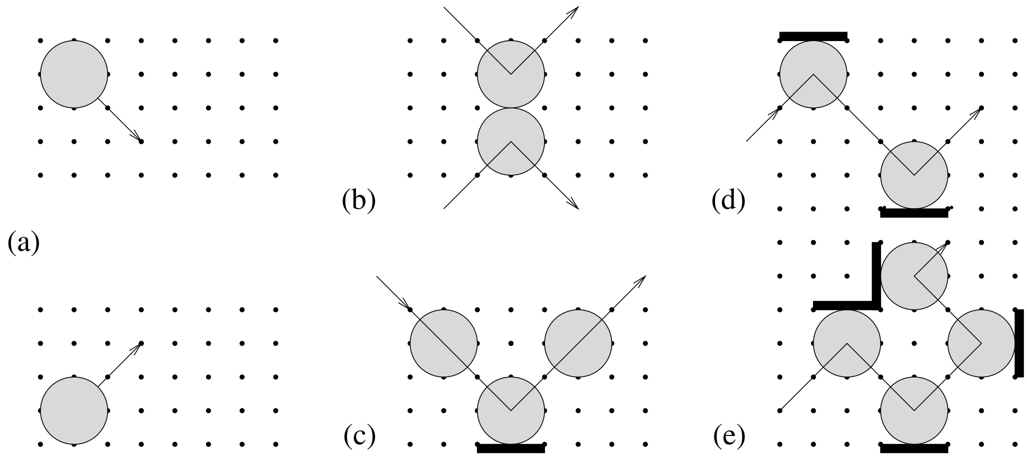


Figure 1.18: Billiard ball model of reversible computation

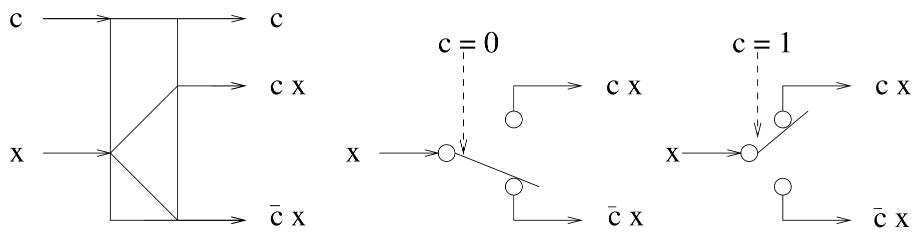


Figure 1.19: Switch gate

then the only feasible option to beat Landauer's lower bound seems to be reversible computation.

After Bennett's discovery of the universality of reversibility, the question arose whether such no-energy dissipating computers can ever be built. The billiard ball model is clearly unfeasible. Very small imprecisions would soon cause the balls to leave the track. As an alternative, Bennett (1973) considered a computational model in molecular dynamics and he was able to show that energy dissipation per bit operation would be much smaller (about 20 to 200 kT Joules). However, that model was seen also as not feasible. In spite of their having no applications, all these results showed the limits and paved the way to the current reversible CMOS devices that dissipate very little energy.

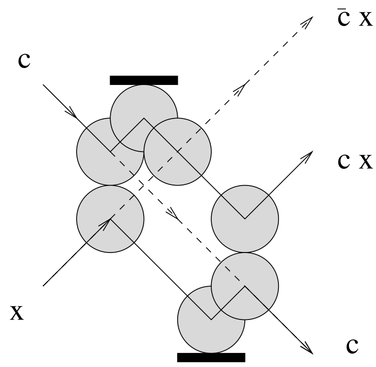


Figure 1.20: A billiard ball implementation of the switch gate

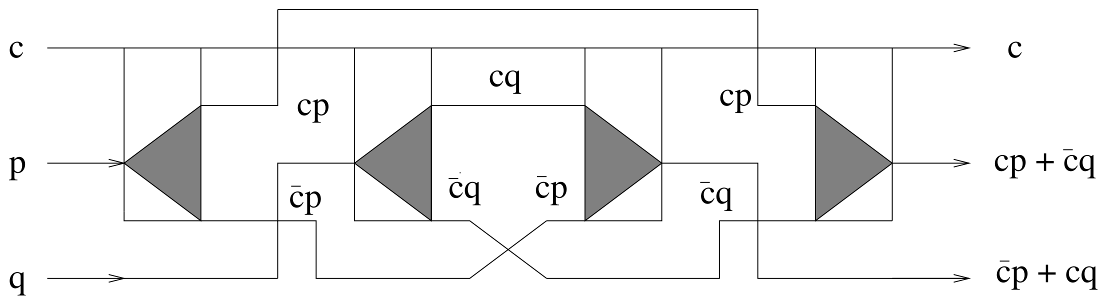


Figure 1.21: A realization of the Toffoli gate using 4 switches