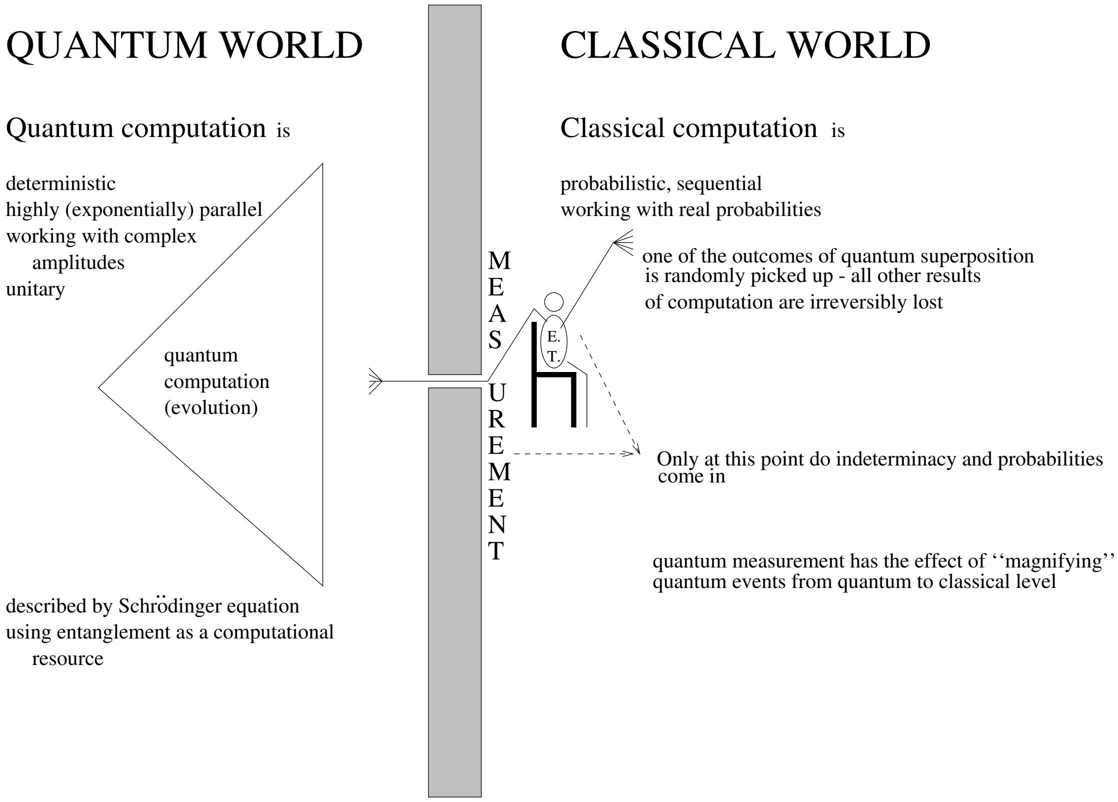
QUANTUM COMPUTING

Jozef Gruska



On the book web pages

http://mcgraw-hill.co.uk/gruska

one findes.

1. Basic information about the contents of the book.
2. Ordering and price information.
3. Second part of the Appendix. (A survey of basic concepts from complexity theory and models of computing. Additional exercises. Historical and bibliographical refrences.)
4. eps-versions of figures from the book.
5. Corrections.
6. Additions.

To my parents
for their love and care.

To my wife
for her ever increasing care, support and patience.

To my children
with best wishes for their future.

To my grandson
with best wishes for quantum computing age.

Contents

PREFACE

Come forth into the light of things.
Let Nature be your teacher.

W. Wordsworth (1770–1850)

A simplified view of the history of computing shows that computing was thought of mainly as mental processes in the 19th century; it is thought of mainly as machine processes in the 20th century, and it will be thought of mainly as Nature processes in the 21st century.

We cannot tell, of course, how much of this vision will be true. Currently, we see vigorous, interesting and, we expect, very important attempts to go much deeper than before into Nature in order to discover its potential for information processing. In this way we also hope to deepen our understanding of Nature. Perhaps the two areas of greatest potential are quantum computing and molecular computing.

The amount of theoretical research and experimental developments in quantum computing grows rapidly. At the same time, interest grows within the science and technology community, especially in physics and theoretical computing, and this interest in turn gives rise to a need for a systematic presentation and summary of the main concepts, methods and achievements through textbooks and courses. This book is addressed to this need.

There has long been a tradition in computing to look into the natural world for inspiration. There have been many attempts to understand, mimic and harnest information processing tools and power of the brain. Already finite automata have been developed as an abstraction of neurons activities. Neural networks represent another model inspired by the brain. Information processing of genetic mechanisms is a further source of inspiration. All these attempts are of interest and importance. However, the current attempts in molecular and quantum computing seem to go even deeper into Nature in the quest of exploring its information processing potential.

Our world is quantum mechanical. It is therefore natural, necessary, interesting and important to explore the foundations and potentials of quantum information processing. At the same time, we must explore the technologies and methods that allow the experimental realization of quantum information processing systems.

Quantum computing is a very new, fascinating, promising and puzzling scientific adventure in which we witness a merging and mutual influence of two of the most significant developments in science and technology of 20th century—quantum mechanics and computing. An adventure that may lead not only to **the** computer revolution, but also to a new scientific and technological basis for information processing in the 21st century.

> *It has been known, but not realized enough, from the birth of modern quantum mechanics theory that the most basic processes of Nature are actually quantum information processing processes and that amount of information processing going on everywhere around us in a tiny portion of matter and time is incomparable larger than all information processing classical technology has ever provided. In addition, it has not been realised, till the birth of modern quantum information processing research, that information processing capabilities of Nature cannot be matched by classical information processing tools, and due to severe limitations on retrieval of information from quantum to classical world, it has not been clear at all whether and how we can harness enormous information processing power of Nature for classical information processing.*

At the same time, as it is often the case with the very fundamental and powerful theories and ideas, the very basic concepts of quantum computing are surprisingly simple and elegant even though they seem to deal with mysterious and puzzling phenomena. Moreover, the technical—mathematical—tools needed to present an introduction to quantum computing are mostly those that are included in basic science education. The book demonstrates and utilises this fact in a way that is readable and understandable by the broad science and technology community.

It is hard to foresee exactly where the research and development in quantum computing will take us. However, we can safely say that something important will come out and that quantum computing is a challenge not only for informatics and physics—theoretical and also experimental — but also for science, technology and society in general.

For informatics as a science, quantum computing may bring the most radical change in its main research aims, scope and paradigms. Indeed, so far informatics has been developed, largely, with the global aims of serving current and foreseeable information processing technology. Quantum computing (with molecular computing) is perhaps the first significant challenge, chance and necessity for informatics to free itself from this short-term role of the servant of technology and to start to concentrate more on its most basic long term aims: to study the laws and limitations of the information processing world, to contribute to the development of new global theories and to deepen our understanding of various worlds: for example physical, biological, and chemical.

For informatics as a technology, the development of quantum information processing technologies can make a revolutionary contribution to the potential and security of information processing and communication systems.

For theoretical physics, quantum computing can be seen as a new challenge and also as an important new source of aims, stimuli, scientific methods and paradigms for dealing with one of the most basic problems of current science (physics). Namely, how to deepen and extend one of the most basic, powerful and fascinating theory in physics—quantum theory. It also brings an opportunity to understand more the role of information as an important resource and fundamental concept in physics and for the understanding of the physical world.

For experimental physics, especially for atomic physics and quantum optics, large needs of quantum computing to store, communicate and process quantum information faithfully bring radically new challenges of astounding complexity and importance.

The merging of insights, methodologies and research paradigms from quantum physics and theoretical computing has been vital for the development of quantum computing and it is expected to be even more so in the future. Historically the first ideas of quantum computing came from people in physics with knowledge of research paradigms, concepts and methods of theoretical computing. However, some key results, and actually the main "apt killers", came from the people in computing making use of only very basic concepts of quantum physics. The results obtained so far in quantum computing demonstrate that currently even the people in theoretical computing or physics with rudimentary knowledge of other area still have a chance to make significant contributions to the field. However, in the future, one could expect a growing need for a thorough knowledge of both of the areas behind quantum computing.

In this book, very basic concepts, models, methods and results of quantum computing are presented in a systematic way. Emphasis is much more on computational aspects, models, methods and problems than on the details of the underlying physics or on technological features which are of (enormous) importance for implementation of quantum information processing systems. The book is therefore primarily oriented towards readers with a computing/mathematics background, but should also be of interest, use, and importance to those with a background in physics and other areas of science and technology. The book assumes very little knowledge of physics and presents quantum computing concepts, models, and methods in a systematic and quite abstract way. The very basic concepts from quantum physics and from its mathematical model—Hilbert spaces—are dealt with in the introductory chapter and also in the Appendix.

The book consists of eight chapters, an extensive Appendix, a list of literature and a detailed index.

The introductory chapter, *Fundamentals*, presents, quite informally, the basic ideas and concepts of quantum computing, including the description of some basic experiments and principles of quantum mechanics and also of elements of Hilbert spaces. The chapter starts with a thorough discussion of why to consider quantum computing and provides an introduction of basic ideas of quantum computing via a comparison of quantum and randomized computing on the level of Turing machines. It also deals in some detail with the concept of reversibility in classical computing.

The second chapter, *Elements*, presents a very detailed treatment of such basic concepts of quantum computing as quantum states, bits, registers, gates, networks and evolutions. A special attention is given to the quantum entanglement, the key inherently quantum resource of quantum information processing. Quantum registers, evolution of their states and measurements, are also discussed and demonstrated. A variety of examples of quantum gates and networks is provided and the universality of quantum gates is dealt with in detail.

The third chapter, *Algorithms*, presents the basic results concerning the design of efficient quantum algorithms. A variety of important quantum algorithms is presented, starting with pioneering algorithms for simple promise problems and including Shor's integer factorization and discrete logarithm computation algorithms, Grover's search algorithm and algorithms for related search and counting problems. A general framework to design efficient quantum algorithms is also presented and illustrated. Finally, a framework is introduced to show lower bounds for quantum algorithms and limitations of quantum computers to speed-up some computations.

The fourth chapter, *Automata*, presents and analyses quantum versions of several basic models of computing: finite automata, Turing machines and cellular automata. These models bring new insights into quantum computing and at the same time a variety of new specific automata-theoretic problems.

In the fifth chapter, *Complexity*, several key problems concerning complexity of quantum computation and communication are dealt with: design of efficient universal quantum Turing machines, basic quantum time and space computational complexity classes and their mutual relations as well as their relation to classical computational complexity classes, and quantum communication complexity. Finally, computational power of nonlinear quantum mechanics is shortly discussed.

In the sixth chapter, *Cryptography*, at first several quantum key generation protocols are presented and their security is analysed. In addition, several quantum protocols for such fundamental problems as bit commitment and oblivious transfer, or for two party communication are presented. Very important, but tricky, complex and deep questions concerning security of cryptographic protocols are dealt with in detail. This includes a proof of unconditional security for quantum key generation as well as the proof of impossibility of such security for quantum bit commitment protocols. Finally, quantum teleportation, superdense coding and some of their applications are described. These areas of quantum information processing are already of more than theoretical interest. Especially in quantum cryptography, experimental progress has been formidable and one has good reasons to expect significant applications in the near future.

The seventh chapter, *Processors*, starts with the presentation of very early ideas how to design quantum processors. It then analyses several key problems one encounters in attempts to build quantum processors. These include: decoherence and inaccuracies and several ways how to cope with these problems: quantum stabilization and error-correction methods for making storage and communication of quantum information feasible; quantum fault tolerant techniques for making (long) processing of quantum information feasible.

Finally, several basic problems are pointed out which a technology has to be able to deal with well in order to create a potential base for the design of (experimental) quantum processors.

In the last chapter, *Information*, several quantum information theory and communication problems are dealt with. These include the development of the quantum counterparts of the basic concepts of the classical information theory, quantum data compression, communication through noiseless and noisy channels and capacities of quantum channels. Entanglement creation, manipulation, concentration, distillation and quantification concepts, protocols and results are also presented and analyzed.

The *Appendix* has five parts: In the first part, several basic problems concerning quantum physics are briefly and informally introduced and discussed. In the second part, there is a more detailed presentation of the basic concepts and results of Hilbert space theory related to quantum computing. In the third part, on the book web pages only, a short survey of the basic concepts and main results of computational complexity theory is provided for those less familiar with the topic. In the fourth part, also on the book web pages only, additional exercises are listed. Finally, in the last part, again on the book web pages only, additional historical and bibliographical references are provided.

The idea of writing this book came up quite naturally with my attempt to write an additional, on web only, chapter on quantum computing, as a supplementary chapter 12, to my book *Foundations of Computing*, 1997. The writing went well and manuscript got soon much too big just for a chapter. In addition, I have realised the attractiveness of the subject for computing people and its maturity which already allows, in spite of the very short history of the field, a systematic presentation of its main concepts and results of clearly lasting importance.

Naturally, the book is only an introduction to quantum computing, written from one perspective, that of computing. Several subjects that are dealt with only briefly in this book, such as design of quantum algorithms, quantum cryptography, quantum information, quantum error-correcting codes, quantum fault tolerant computing and quantum entanglement would already merit monographs. In addition, many subjects requiring more expertise in quantum physics, and at the same time not central for an introduction to the main problems of current quantum computing, have been covered only briefly, or not at all.

Referencing. A large effort has been made that results and ideas presented are properly credited and referenced. This has been a hard task because the field develops very fast. This is therefore to apologize for all omisions, imperfections or even misclaims and to ask those feeling that an addition or correction should be done along these lines to let me know and I will try to do that on the book web pages.

Acknowledgments. I have started to work on this book while visiting University Paris VI, in 1997, and, especially, University of Nice, Laboratoire d'Informatique Signaux et Systèmes at Sophia-Antipolis, during long stays in 1997 and 1998, within the PAST program. Excellent conditions provided by these universities and help by many people there, especially by Irène Guessarian and Bruno Martin are much appreciated.

I have also to acknowledge excellent conditions, full understanding and support provided by Faculty of Informatics, Masaryk University, Brno and also support of GAČR, Grant 201/98/0369 and of the Slovak Literary Agency.

This is also to thank Vladimír Bužek, Patrick Cegielski, Vladimír Černý, Christop Dürr, Rūsiņš Freivalds, Mika Hirvensalo, Juraj Hromkovič, Bernd Kirsig, Manfred Kudlek, Bruno Martin, Michele Mosca, Jozef Nagy, Masanao Ozawa, Pavol Petrovič, Jiří Rosický, Martin Stanek, Mark-Oliver Stehr, John Watrous and Thomas Worsch for reading, correcting and commenting earlier drafts of this book or its parts. Special thanks go to V. Bužek for discussions and also advices and help. Support of Roland Volmar team is also to acknowledge.

Expertize and helpfulness of our TeX and LaTeX expert Petr Sojka was much useful and it is much to appreciate. To appreciate is also help with figures, index and manuscript checking by Robert Batůšek, Petr Tobola andf especially by Petr Macháček.

Finally, very smooth cooperation with David Hatter from McGraw-Hill and his continuous, but very enjoyable and constructive, wisdom, pressure, understanding and help, has been very much appreciated during the whole process of the manuscript preparation. Cooperatiom with Steve Gardiner and his production stuff is also to appreciate.

Using the book as the textbook. The following is a possible structure of a one semester course: (1) Introduction (1.1–1.3, 1.7); (2) Hilbert spaces basics (1.4, 9.2); (3) Quantum principles (1.5–1.6 + Appendix 9.1) or Computational complexity (Appendixx —on book web pages 9.3); (4) Quantum bits, registers, gates and networks (2.1–2.3); (5) Basic quantum algorithms (3.1); (6) Shor's algorithms I (3.2); (7) Search algorithms (3.3); (8) Quantum algorithms design methodologies and limitations (3.4 and 3.5); (9) Quantum finite automata and Turing machines (4.1 and 4.2); (10) Quantum key generation

(6.1 and 6.2); (11) Quantum cryptographic protocols and teleportation (6.3 and 6.4); (12) Quantum error correction codes (7.2– 7.4); (13) Quantum fault-tolerant methods (7.5); (14) Quantum processors (7.1 and 7.6); (15) Quantum information theory (8.1– 8.2); (16) Quantum entanglement theory (8.3– 8.4).

Additional subjects: quantum computational complexity (5.1–5.4), quantum cellular automata (4.3).

November 8, 2011 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,

[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 313.

[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.

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 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

[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.

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 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"

[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.*

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.

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)

[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.

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 (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

[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.

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 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

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

[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.

"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 $\mathbf{QEP}^A \not\subset \mathbf{ZPP}^A$—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 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

[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.

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 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.)

[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.

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 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):

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

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

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

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.

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

[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.

[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.

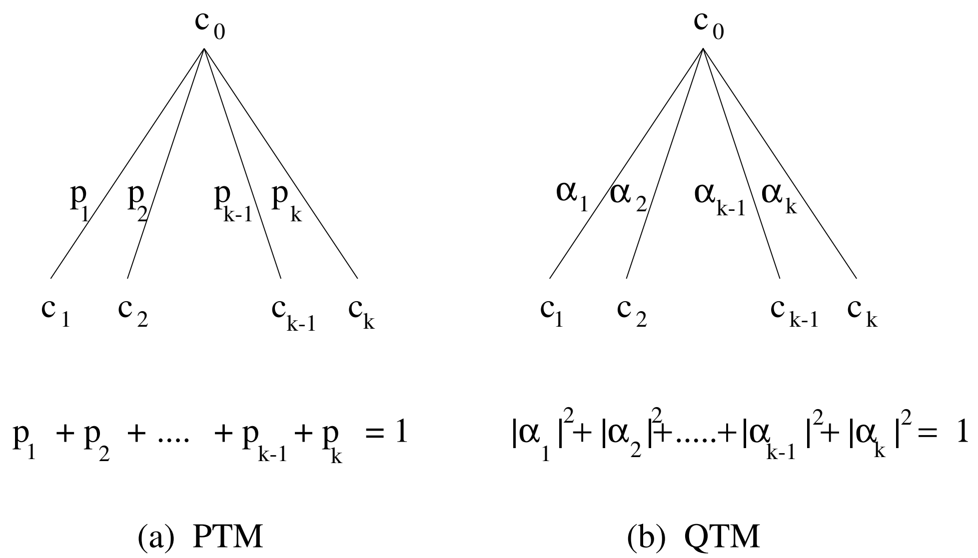


Figure 1.1: Local probability conditions

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:

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

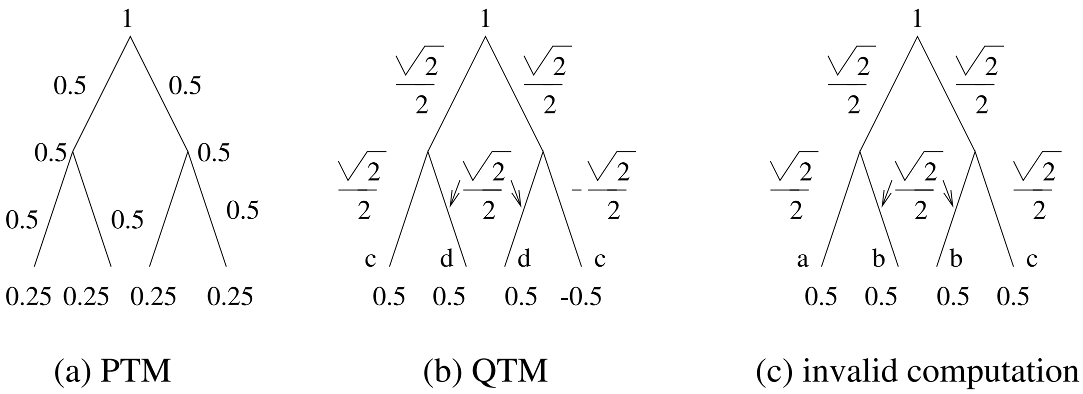


Figure 1.2: Configuration trees with probabilities and the probability amplitudes

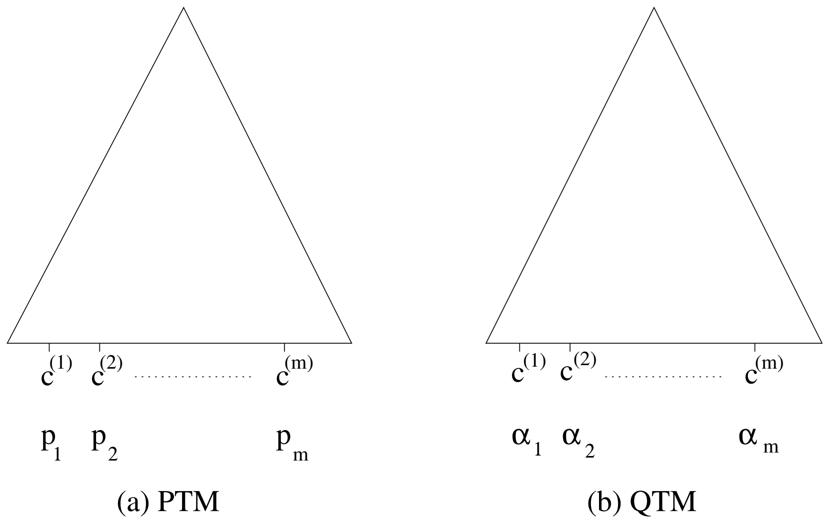


Figure 1.3: Multiple occurrences of the same configuration

> **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

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

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 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

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

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

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

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 $c_1, c_2, \dots, c_k$ are all mutually different configurations at a certain level of the configuration tree, then their total amplitudes $\beta_1, \dots, \beta_k$ 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 $\alpha_1, \dots, \alpha_k$ 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 64.

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.

[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.

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.

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)*

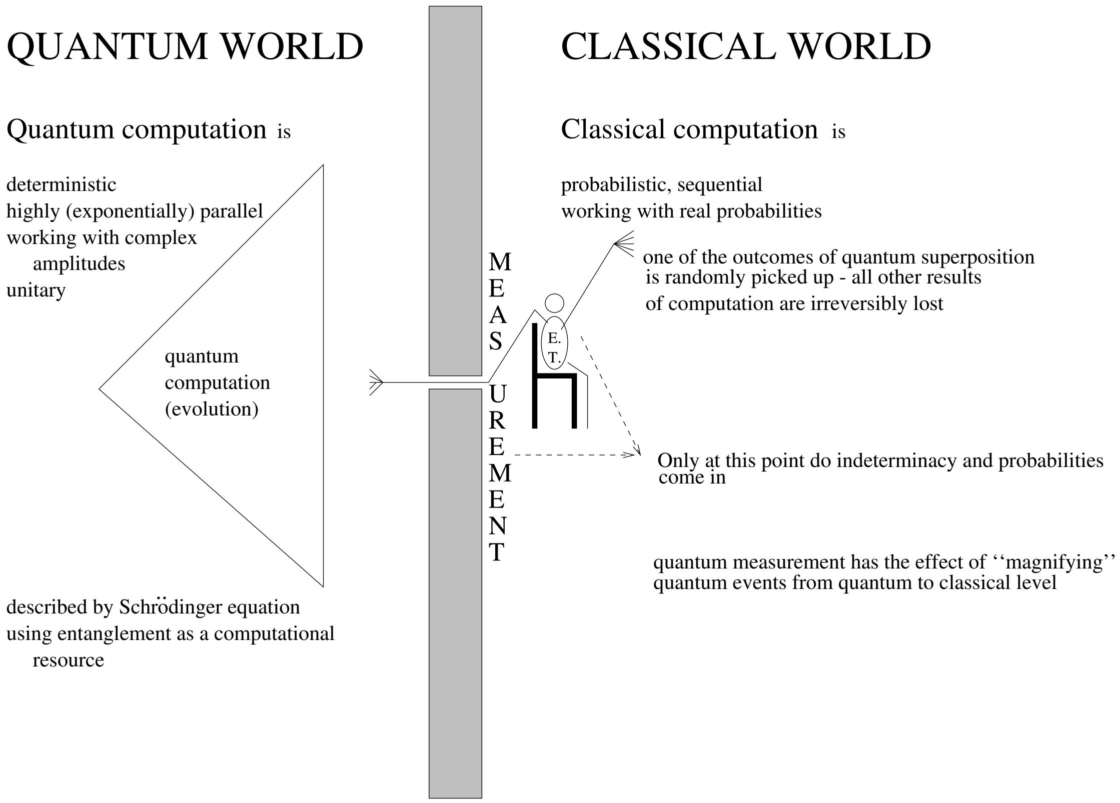


Figure 1.5: Quantum and classical worlds

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 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*

[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.

[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.

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** $\langle \cdot \mid \cdot \rangle : l_2(D) \times l_2(D) \to \mathbf{C}$,[32] *defined by*

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

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 $\langle\cdot|\cdot\rangle : H \times H \to \mathbf{C}$ satisfying the following axioms for any vectors ϕ, $\psi, \phi_1, \phi_2 \in H$, and any $c_1, c_2 \in \mathbf{C}$.*

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

[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.

[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.

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).*

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(ψ, ϕ); (c) dist$(\phi, \psi) = 0$ if and only if $\phi = \psi$; (d) dist$(\phi, \psi) \leq$ dist(ϕ, χ) + dist(χ, ψ).*

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.

---

[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.

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?*

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 $\phi_i \in W_i, 1 \le i \le n$.

1.4.2 Operators

Definition 1.4.20 *A linear operator on a Hilbert space H is a linear mapping $A : H \rightarrow H$. 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]

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) $i(AB - BA)$?*

[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.

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

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

as follows

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

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

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

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

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*

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

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

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

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

$$|x_1\rangle|x_2\rangle \dots |x_n\rangle \quad \textit{or} \quad |x_1, x_2, \dots, x_n\rangle \ \textit{or} \ |x_1 \dots x_n\rangle.$$

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 $\{\phi_i\}_{i=1}^l$ 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 372, 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

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

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

$\rho = \rho^*$.

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 $f : \mathbf{C} \rightarrow \mathbf{C}$ 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**

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

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,

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

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

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

[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).

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)

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.

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 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.

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

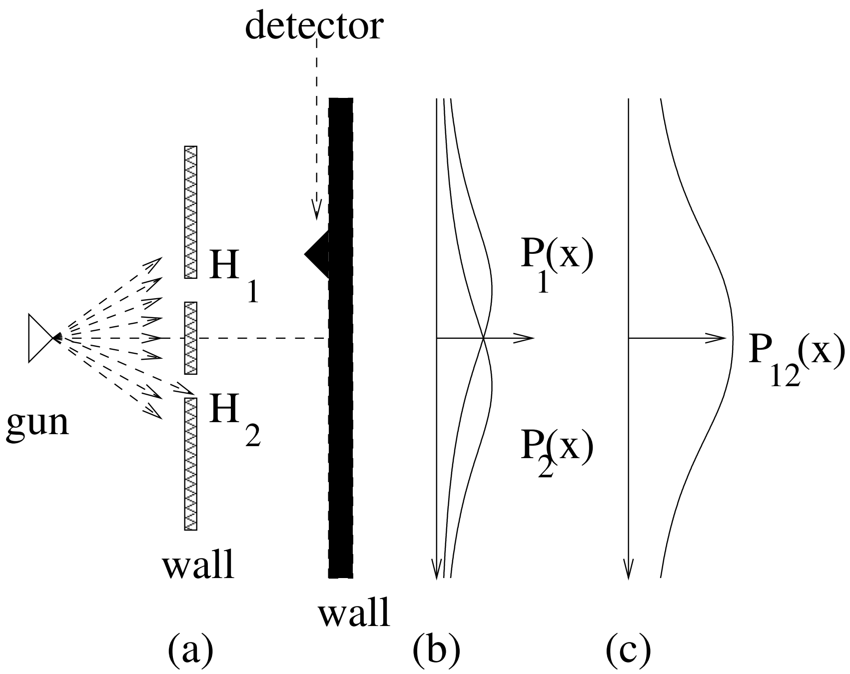


Figure 1.6: Experiment with bullets

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).

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 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

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

[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.)

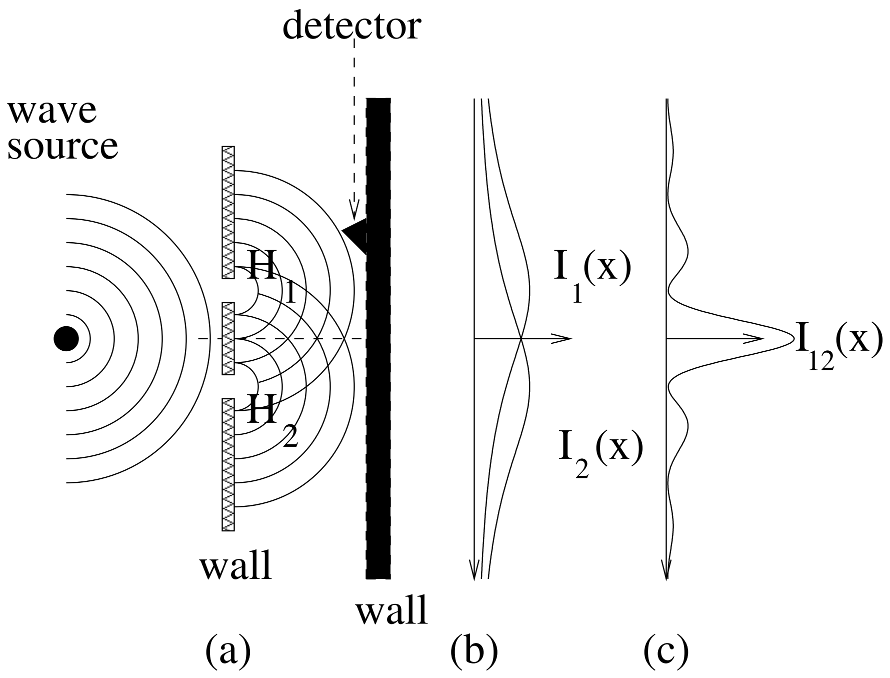


Figure 1.7: Experiments with waves

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 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

[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.

[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).

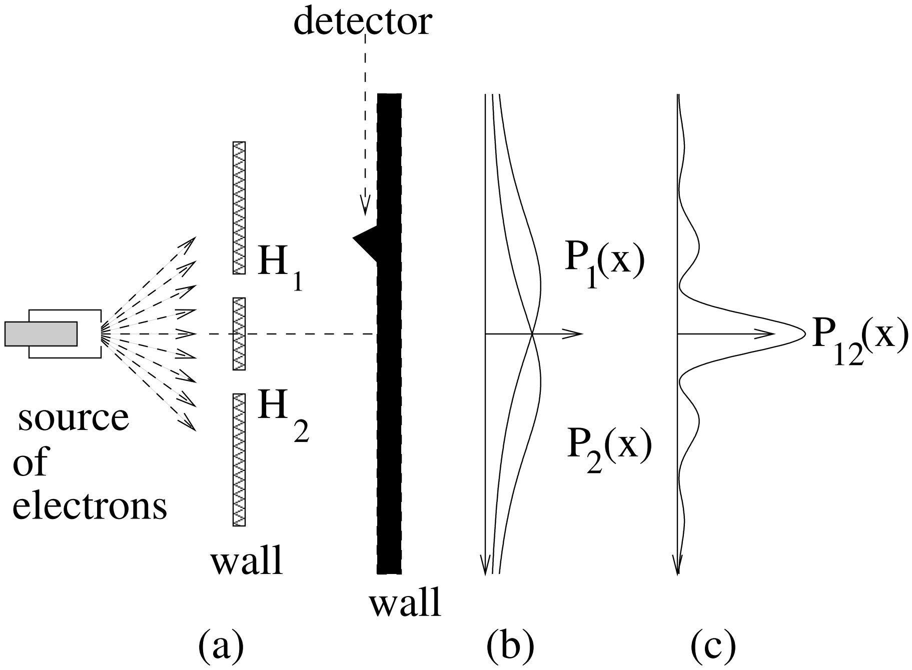


Figure 1.8: Two slit experiment

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,

[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.

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.

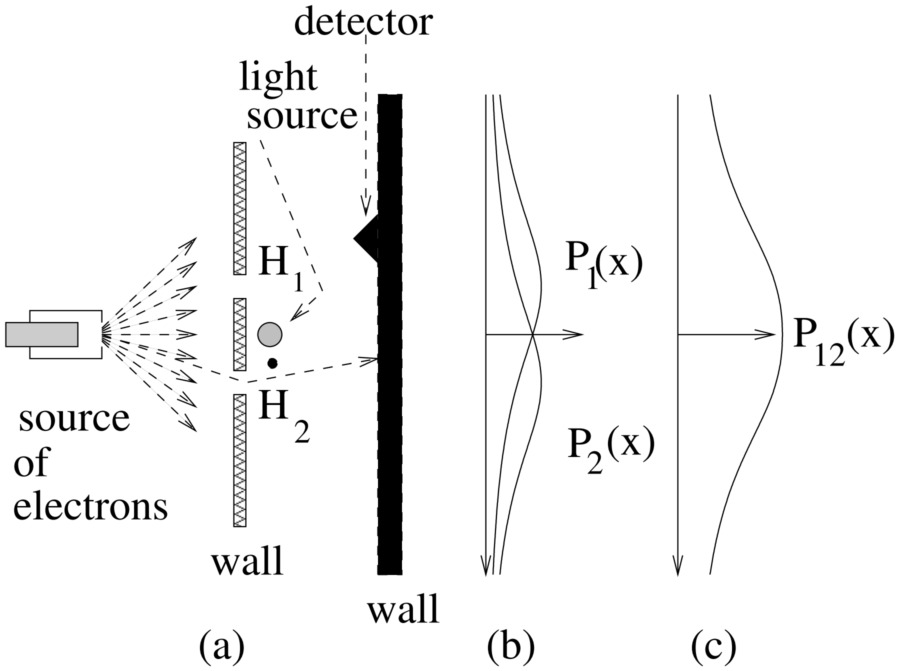


Figure 1.9: Two-slit experiment with an observation

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 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.

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

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.

[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.

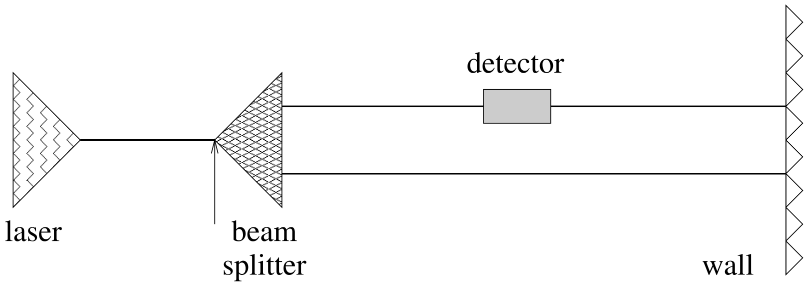


Figure 1.10: Delay choice experiment

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.

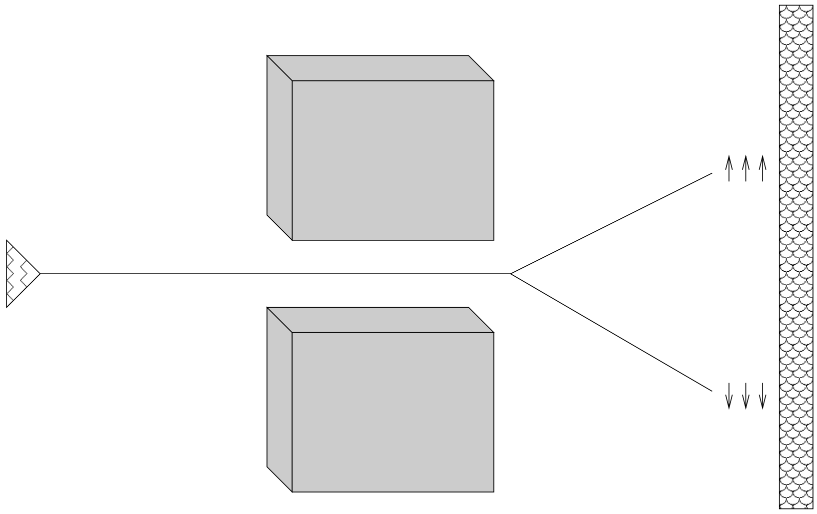


Figure 1.11: Stern-Gerlach experiment with spin-$\frac{1}{2}$ particles

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 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.

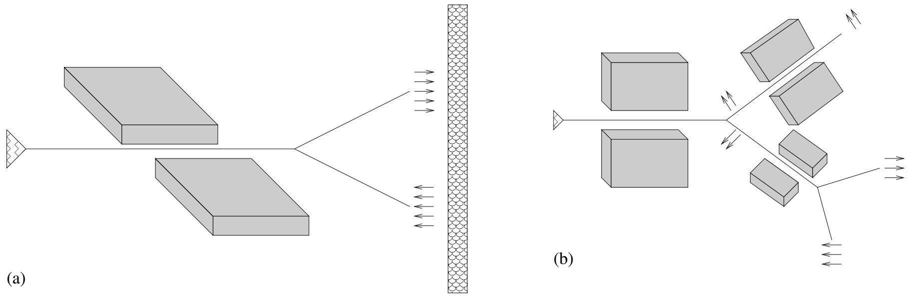


Figure 1.12: Several Stern-Gerlach magnets

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).

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

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** *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

$$\text{(final state, initial state)}.$$

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

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

[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.)

[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".

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

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

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.,*

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

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 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:*

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

[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.

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

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.

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 $\mu : \{E_1,\ldots,E_k\} \to \mathbf{R}$.

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:*

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

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

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.*

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, $1 \le i \le k$.

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".

[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".

[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.

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. 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.

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

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, 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

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

[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.

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:*

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

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

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

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 $|\phi_0\rangle, |\phi_1\rangle, |\phi_2\rangle, \ldots$ and unitary operators $U_1, U_2, U_3, \ldots$ such that $|\phi_i\rangle = U_i|\phi_{i-1}\rangle$, for $i = 1, 2, 3, \ldots$.

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

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.*

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

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

[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).

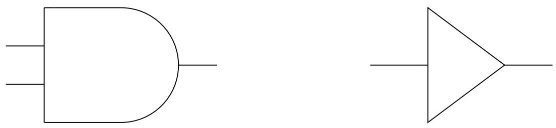
computers and the fact that the second law of thermodynamics implies that irreversible state changes during computation must dissipate heat.[67] This is nowadays known (see Bennett, 1998b), as:

> **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 in the form of waste heat).*

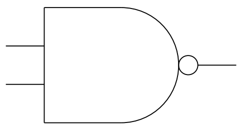
The importance of the investigations of the classical reversible computing follows from the fact that such computations are special cases of quantum computations.

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 $f \in \mathcal{B}_n^m = \{g \mid g : \{0,1\}^n \to \{0,1\}^m\}$. 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

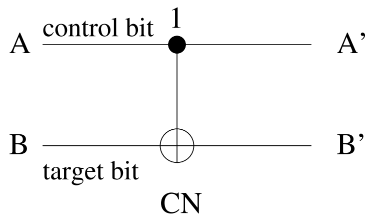
Figure 1.13: Classical Boolean gates

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".

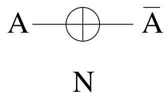
We talk about a **reversible** gate, Boolean function, operation or computation, as one with always enough information in the outputs to deduce the inputs unambiguously. Such operations, gates and computations are crucial for quantum computing because of the reversibility of the evolution in quantum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

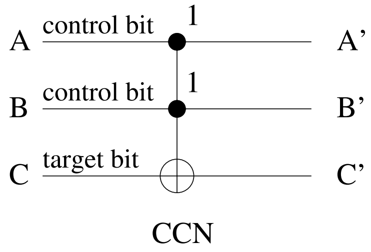
Three reversible gates have turned out to be of special importance: the usual NOT gate (N), CONTROL NOT gate (CN or CNOT or XOR) (see Figure 1.14a), CONTROL CONTROL NOT gate (CCN or CCNOT), see Figure 1.14b, also called **Toffoli gate**.

[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.



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)

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]

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.*

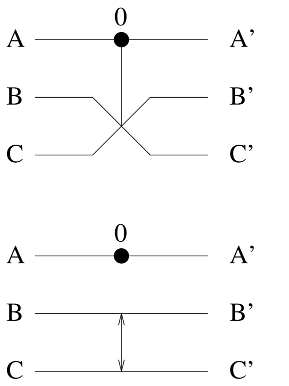
Example 1.7.2 *It is actually easy to show universality of Fredkin and Toffoli gates. In the Toffoli gate, if $A = 1$ then $B' = B \oplus C$; if $B = 0$, then $B' = A \wedge C$; if $A = 1, C = 1$, then $B' = \bar{B}$. For the Fredkin gate, if $C = 0$ then $B' = A \wedge B$; if $B = 0$ and $C = 1$, then $B' = \bar{A}$. 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

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

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.

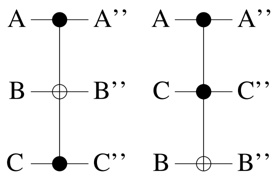
The CN gate, The Toffoli gate and the Fredkin gate were first presented by C. A. Petri in 1965, but their publication in 1967, in German and in a not too widespread Proceedings, went apparently unnoticed by most of those working on reversible computing. However, in view of the above fact, it would perhaps 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

			Fredkin gate			Toffoli gate		
A	B	C	A’	B’	C’	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

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

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

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

Example 1.7.3 *The two bit adder computation mapping:*

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

is not one-to-one. However, the modified computational task: $(a, b) \rightarrow (a, a \oplus b, a \wedge b)$ *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.*

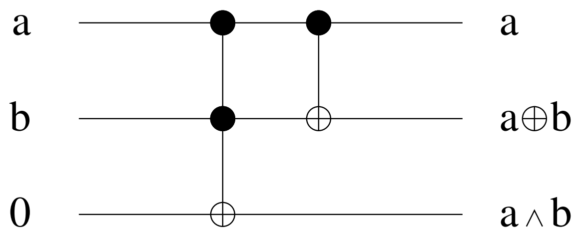


Figure 1.16: A reversible implementation of a two-bit adder

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.

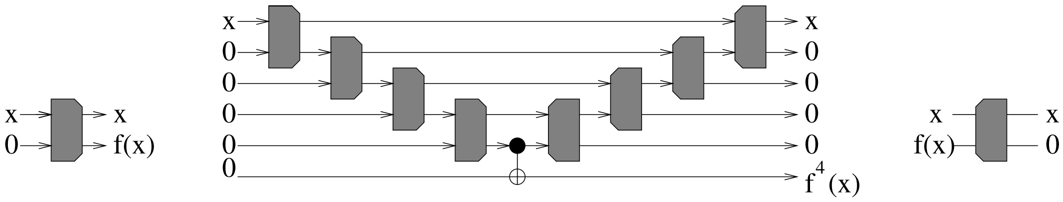


Figure 1.17: Reversible computation with garbage removal

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 $x \rightarrow (x, f(x))$, 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 $2t(n)$, 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: $\mathit{Space}(t(n)) = \mathit{Space}(\frac{t(n)}{2}) + s(n)$, yielding $\mathit{Space}(t(n)) = \mathcal{O}(s(n) \lg t(n))$. Time requirements are $\mathit{Time}(t(n)) = 6\,\mathit{Time}(\frac{t(n)}{2}) + \mathcal{O}(n)$, which yields $\mathit{Time}(t(n)) = t(n)^{\lg 6}$. 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 $\mathit{Time}(t(n)) \leq 4k\,\mathit{Time}(\frac{t(n)}{k}) + \mathcal{O}(n)$, 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

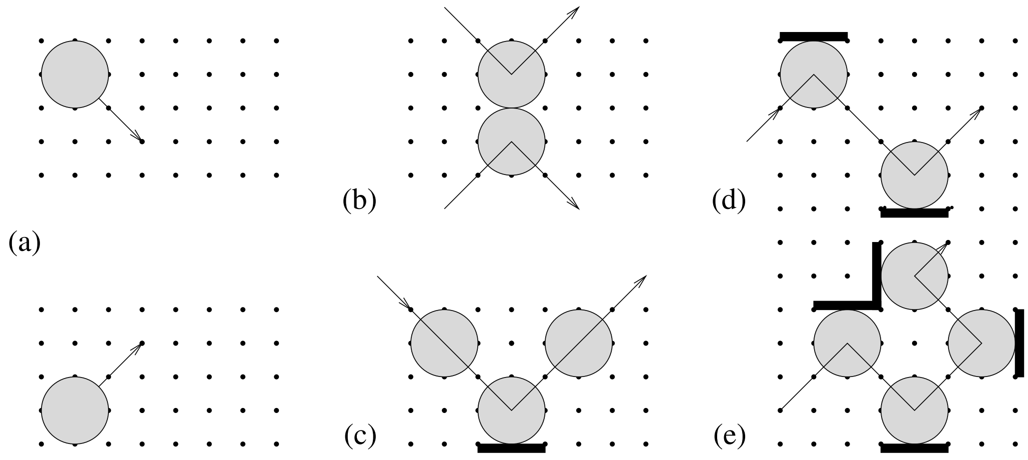


Figure 1.18: Billiard ball model of reversible computation

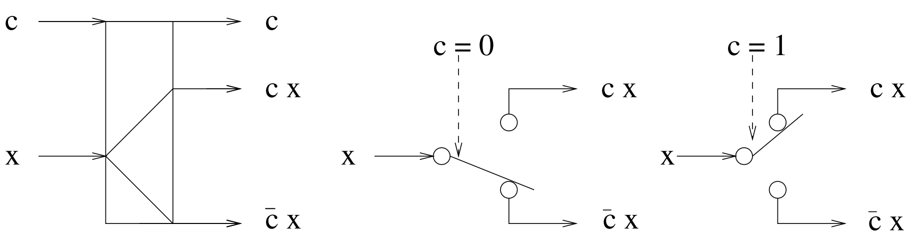


Figure 1.19: Switch gate

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, 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.

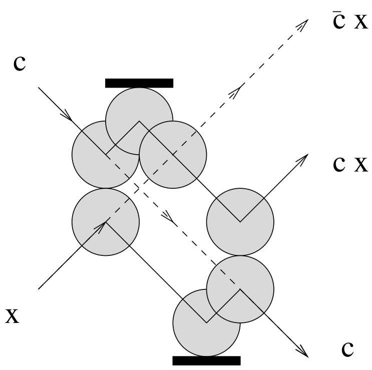


Figure 1.20: A billiard ball implementation of the switch gate

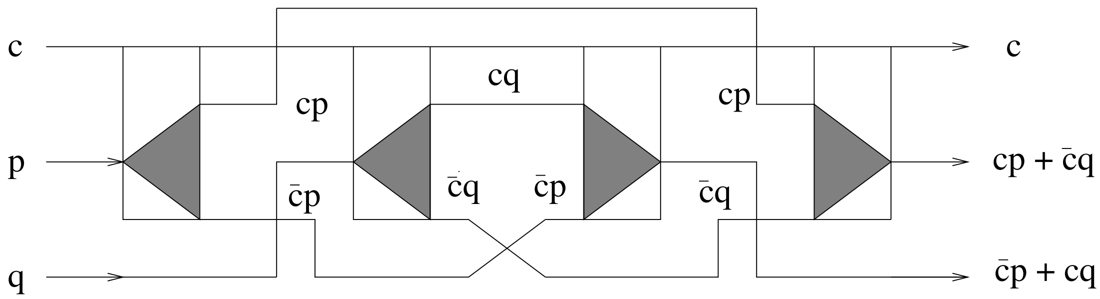


Figure 1.21: A realization of the Toffoli gate using 4 switches

Chapter 2

ELEMENTS

INTRODUCTION

The basic elements of quantum computing are easy to identify: quantum bits, quantum registers, quantum gates and quantum networks. However, at this point an analogy with classical computing ends. Quantum bits, registers, gates and networks are very different, have other properties and larger power than their classical counterparts.

A quantum bit can be in any state within an infinite set of states. A quantum register of n quantum bits can be, at the same time, in any of the infinitely many superpositions of 2^n basis states. The parallelism a quantum register can exhibit is striking. The key new feature is that a quantum register can be in an entangled state. On one side, entangled states with their non-locality features are a hallmark of quantum mechanics. On the other side, quantum entanglement is an important resource of quantum information processing.

There is a larger variety of quantum gates than of classical gates. There are already infinitely many one-input/output quantum gates. In addition, almost any two-input/output quantum gate is universal. A simple two input/output gate together with one input rotation gates form a set of universal gates.

LEARNING OBJECTIVES

The aim of the chapter is to learn:

1. the basic concepts concerning quantum bits and registers;
2. the concept of quantum entanglement and the examples of its power;
3. the basic examples of quantum gates and of quantum circuits;
4. some examples of universal quantum gates and a method to show universality of quantum gates;
5. the basic quantum arithmetical circuits;
6. the concept of quantum superoperator circuits.

> What canst thou see elsewhere which thou canst not see here? Behold the heaven and the earth and all the elements: for of these are all things created.
>
> Thomas à Kempis (1380–1471)

Quantum algorithms and circuits work with quantum versions of the same basic elements as classical ones: quantum bits, quantum registers and quantum gates. However, all these new elements have specific properties and power. A good knowledge of them is a base for understanding more elaborate algorithms and networks presented in the next chapter, as well as for the understanding of the most of all other topics we deal with in this book.

2.1 Quantum Bits and Registers

Four key problems of quantum information processing are: how to represent, how to store, how to transmit and how to manipulate quantum information. Two key elements to deal with these problems are quantum bits and quantum registers.

2.1.1 Qubits

Let $\mathcal{S}$ be a two-dimensional quantum system with two orthonormal states, denoted $|0\rangle$ and $|1\rangle$, that can be considered as forming a natural, or standard, or preferred, basis of $\mathcal{S}$.

Definition 2.1.1 *A* **qubit** *(quantum bit) is a quantum state*

$$|\phi\rangle = \alpha|0\rangle + \beta|1\rangle,^{1} \tag{2.1}$$

where $\alpha, \beta \in \mathbf{C}$ *and* $|\alpha|^2 + |\beta|^2 = 1$.[2]

[1] One can also say that a qubit is a unit vector in a two-dimensional inner-product space. For the representation of qubits we often assume that a particular basis, say $\{|0\rangle, |1\rangle\}$, has been fixed.

[2] The term qubit was coined by B. Schumacher (1995). A more classical term is a "two-level system", or a "two-state system".

For the purpose of quantum information processing two basis states $|0\rangle$ and $|1\rangle$ are usually taken as encoding classical bit values 0 and 1. A classical bit can be seen as a qubit promised to be in one of the basis states. In general, we can call any two-state system a physical bit and when the system is quantum and the two states are orthogonal quantum states, we can refer to it as a qubit. Therefore, any two-state quantum system is a potential candidate for a qubit.

In physics literature the following notation is often used for the states of the standard basis of various two-level quantum systems: in the case of spin$-\frac{1}{2}$ particles, and in the case of vertical or horizontal polarization of photons, $|\uparrow\rangle$ or $|\updownarrow\rangle$, is taken instead of $|0\rangle$ and $|\downarrow\rangle$, or $|\leftrightarrow\rangle$ is used instead of $|1\rangle$; in the case of diagonal

Exercise 2.1.2 *Show that any qubit state* $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ *can be expressed in the form* $|\psi\rangle = \cos\theta|0\rangle + e^{i\phi}\sin\theta|1\rangle$ *in the sense that* $|\alpha|^2 = |\cos\theta|^2$ *and* $|\beta|^2 = |e^{i\phi}\sin\theta|^2$.

Remark 2.1.3 The term "qubit" is often used in a more abstract way—as a variable that can take on any qubit state (2.1). In such a case we can talk about a (particular) state of a qubit. This is in accordance with the use of the term "bit" in classical computing where we talk about a value, or a state, of a bit.

The above definition leaves the actual physical medium of a qubit completely undefined—as long as objects are treated according to the quantum principles discussed in the previous chapter.

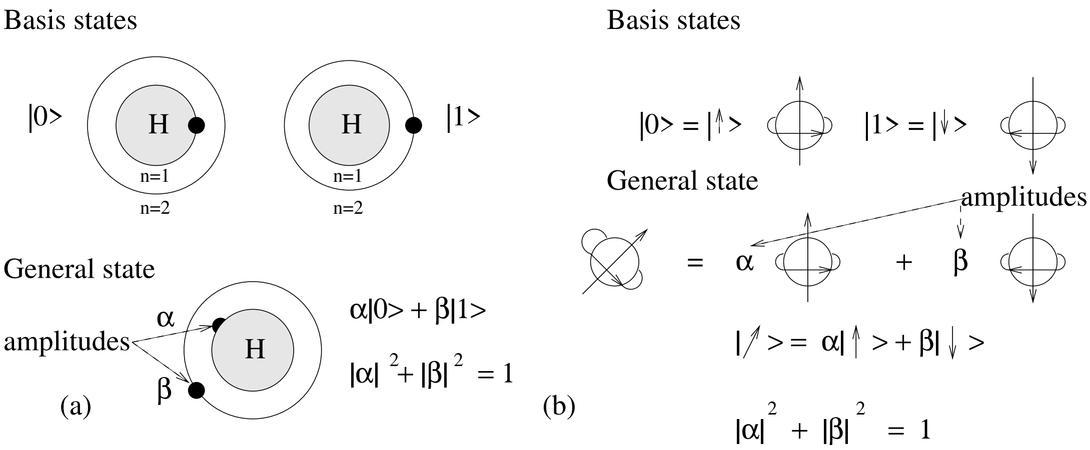


Figure 2.1: Qubit representations by energy levels of an electron in a hydrogen atom and by a spin-$\frac{1}{2}$ particle. The condition $|\alpha|^2 + |\beta|^2 = 1$ is a legal one if $|\alpha|^2$ and $|\beta|^2$ are to be the probabilities of being in one of two basis states (of electrons or photons).

There are many ways to realize qubits—there are many interesting/important two-dimensional quantum systems known in physics. For example, by the polarization of a photon or by the ground ($n = 1$) and excited ($n = 2$) states of an electron in the hydrogen atom (Figure 2.1a). One of the most often and best-explored two-level quantum systems is that of spin-$\frac{1}{2}$ particles with two basis states: spin-up (notation $|\uparrow\rangle$ or $|0\rangle$) and spin-down (notation $|\downarrow\rangle$ or $|1\rangle$) (Figure 2.1b).

From the implementation point of view the most promising candidates for qubits are so far photons, trapped ions and spins of atomic nuclei.

States $|0\rangle$ and $|1\rangle$ of a qubit can be seen, and are often referred to, as representing classical states (bits).[3] The main difference between classical bits and qubits is that while a classical bit can be set up only to one of the two states, namely 0 or 1, a qubit can take any quantum linear superposition of $|0\rangle$ and $|1\rangle$, i.e., in principle can be in any of uncountably

polarizations of photons (45° and 135°), $|\nearrow\rangle$ ($|\searrow\rangle$) is used for $|0\rangle$ ($|1\rangle$) in the case of circular polarization $|\circlearrowleft\rangle$ ($|\circlearrowright\rangle$) is used for $|0\rangle$ ($|1\rangle$).

[3]This is principally incorrect but often useful simplification, especially for getting an analogy with classical computing.

many states. This means that a large, even infinite, amount of information could potentially be encoded in amplitudes of a single qubit by appropriately choosing α and β.[4] Of course, this does not mean that in any implementation a qubit can take any of its potentially infinite number of theoretically possible states.[5]

One way to represent states of qubits geometrically is as points on the surface of a unit Riemann[6] sphere, where North and South poles correspond to the basis states (that correspond to bits) (see Figure 2.2a).[7] Qubits can be represented also by points on a Bloch[8] sphere (see Allen and Eberly, 1975), called also Poincaré sphere, and Figure 2.2b), using the spherical coordinate system. This representation is based on the fact that any qubit can be represented (see Exercise 2.1.2), as $\cos\frac{\theta}{2}|0\rangle + e^{i\phi}\sin\frac{\theta}{2}|1\rangle$. [9]

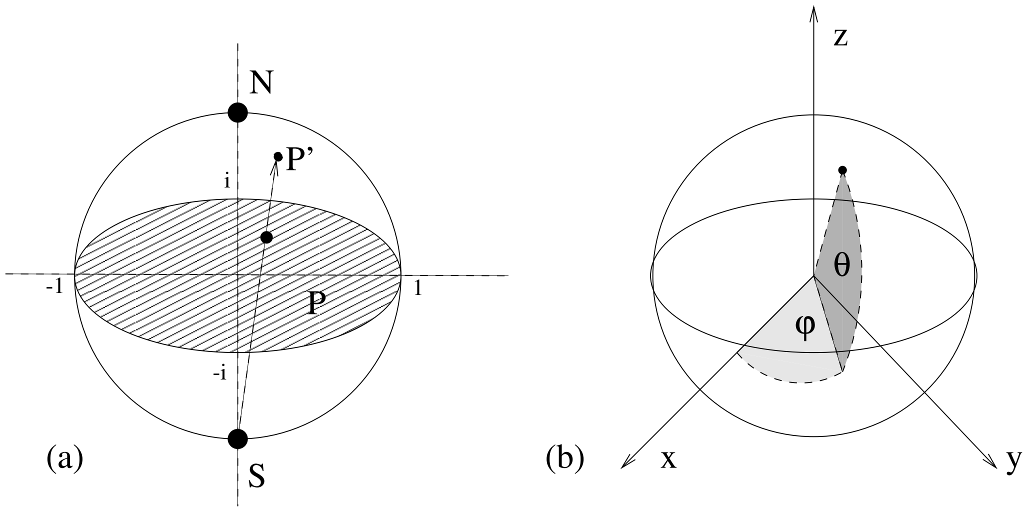


Figure 2.2: Representation of qubits on Riemann and Bloch spheres

[4]For this reason, in order to study computational complexity problems of quantum computing a restriction will have to be made on α and β. In principle, it will be required that they can be computed in polynomial time and have logarithmic size.

[5]For example, a **trine** is a qubit that can be in only one of the following three states (see Peres and Wootters, 1991): $|0\rangle$, $\frac{1}{2}|0\rangle + \frac{\sqrt{3}}{2}|1\rangle$ or $\frac{1}{2}|0\rangle - \frac{\sqrt{3}}{2}|1\rangle$.

[6]Georg Friedrich Bernhard Riemann (1826–1866), a German mathematician. His main contributions was in the theory of complex variable functions and their representations (on Riemann surfaces), non-Euclidean geometry (representation of elliptic spaces and an extension of Gauss's work on differential geometry to n-dimensional objects), and electromagnetic theory.

[7]The Riemann sphere is a sphere of unit radius whose equatorial plane is the complex plane whose centre is the origin of the plane. One qubit state $|\phi\rangle = \alpha|0\rangle + \beta|1\rangle$ can be represented by a point on a Riemann sphere as follows. If $\beta \neq 0$ we mark in the complex plane the point P that represents the number $\frac{\alpha}{\beta}$ and then we project P from the South Pole onto the sphere to get the point P' that then represents $|\phi\rangle$. If $\alpha = 0$ one gets the North Pole this way; if $\beta = 0$ the South Pole is the limit (Penrose, 1994).

[8]Felix Bloch (1905–1983), an American physicist of Swiss origin. His main contributions were in the quantum theory of solid bodies, ferromagnetism and quantum electrodynamics. Bloch developed methods to measure magnetic moments of atom nucleus.

[9]θ is the angle from z axis and ϕ the angle in the $x - y$ plane from the x axis—a phase. Representation of qubits by points on a Bloch sphere is of interest also because it provides an isomorphism between qubit operations and solid-body rotations.

Just as in the previous chapter, we can represent basis states as vectors as follows: $|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $|1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. In such a case we have representations:

$$|\nearrow\rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, |\searrow\rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}, |\circlearrowleft\rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} \end{pmatrix} \text{ and } |\circlearrowright\rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{i}{\sqrt{2}} \end{pmatrix}.$$

Qubit measurements

Unfortunately, what goes into a qubit does not necessarily come out. A single qubit is in principle not fully identifiable, i.e., given an unknown state $|\psi\rangle$ of a qubit, it is in general not possible to identify it fully by a projection measurement. Quantum physics has strict rules on how to extract information out of an unknown quantum state. The outcome of any projection measurements of a qubit must be formulated in classical terms. More exactly, we can get out of any projection measurement of one qubit only one classical bit of information. Therefore, even though there is a continuum of possible quantum states of a single qubit, these states cannot all be distinguished reliably from each other. No (von Neumann) measurement can extract more than one expected bit of information from any given qubit.[10] From an information theory point of view, from a qubit one can obtain by a (projection) measurement exactly the same amount of (classical) information as a classical bit has, even if it has infinitely many potential states.

Example 2.1.4 *A measurement of a qubit state $|\phi\rangle = \alpha|0\rangle + \beta|1\rangle$, corresponding to the observable $\{E_0, E_1\}$, where E_0 (E_1) is the subspace spanned by the state $|0\rangle$ ($|1\rangle$), or, in other words, with respect to the standard basis $\{|0\rangle, |1\rangle\}$, provides as the output bit* 0 (1) *with the probability $|\alpha|^2$ ($|\beta|^2$) and the state $|\phi\rangle = \alpha|0\rangle + \beta|1\rangle$ collapses into the state $|0\rangle$ ($|1\rangle$). All other information about the superposition is irreversibly lost. For an observer a qubit represents therefore a probability distribution.*

However, the qubit $|\phi\rangle = \alpha|0\rangle + \beta|1\rangle$ can be measured also with respect to infinitely many other bases. For example, with respect to the often used dual basis $\mathcal{D} = \{|0'\rangle, |1'\rangle\}$, where

$$|0'\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \qquad |1'\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle).$$

Since

$$|0\rangle = \frac{1}{\sqrt{2}}(|0'\rangle + |1'\rangle) \quad |1\rangle = \frac{1}{\sqrt{2}}(|0'\rangle - |1'\rangle)$$

we have

$$|\phi\rangle = \frac{1}{\sqrt{2}}((\alpha+\beta)|0'\rangle + (\alpha-\beta)|1'\rangle)$$

and a measurement of $|\phi\rangle$ with respect to the dual basis gives 0 (1) *with probability $\frac{1}{2}|\alpha+\beta|^2$ ($\frac{1}{2}|\alpha-\beta|^2$).*

[10] Of course, this is not the whole story. In order to illustrate the peculiarities of quantum measurement, let us consider again trines. No measurement on a trine can do better than rule out one of the three possibilities, leaving one bit of uncertainty about the original state of the trine. Thus, it is not possible to extract more than $\lg 3 - 1 \approx 0.585$ bits of classical information from a single trine. However, if we have two trines that are guaranteed to be identical, then (by Peres and Wootters, 1991), there is a measurement that can extract $\frac{\sqrt{2}}{3}(1 + \lg(17 + 12\sqrt{2})) - \frac{3}{2} \approx 1.369 > 2 \times 0.585$ bits of information from both trines. This implies that in some cases we can extract more than twice as much information from two identical qubits than from either one alone. In order to extract more information from two identical qubits than projection measurements allow, POV measurements, see Section 9.2.8, are used.

In addition, as discussed in Section 8.2.4, one can encode 3 bits into one qubit in such a way that any of them (but not all of them) can be retrieved (by a proper measurement) with success probability 0.79.

Example 2.1.5 *If the state $|0\rangle$ is measured with respect to the standard basis we get as outcome* 0 *with probability* 1 *and the state collapses into itself. On the other hand, if $|0\rangle$ is measured with respect to the dual basis we get as the outcome* 0 *or* 1*, both with probability* $\frac{1}{2}$*, and the state will collapse either into the state $|0'\rangle$ or into the state $|1'\rangle$. The result of the measurement in this case is actually a random bit.*

Remark 2.1.6 A measurement of a qubit with respect to the basis $\{|0\rangle, |1\rangle\}$ corresponds to the projection operator $\frac{1}{2}(I \pm \sigma_z)$. In the case of spin-$\frac{1}{2}$ particles this corresponds to measuring of the spin along the z-axis. Projections $\frac{1}{2}(I \pm \sigma_x)$ and $\frac{1}{2}(I \pm \sigma_y)$ correspond to measurements along the x-axis and y-axis, respectively.

Exercise 2.1.7 *Determine probabilities of the outcomes of the measurements of the qubit $|\phi\rangle = \alpha|0\rangle + \beta|1\rangle$ with respect to the bases:*
(a) $\{\frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle), \frac{1}{\sqrt{2}}(i|0\rangle - |1\rangle)\}$; *(b)* $\{\frac{1}{2}|0\rangle + \frac{\sqrt{3}}{2}|1\rangle, \frac{\sqrt{3}}{2}|0\rangle - \frac{1}{2}|1\rangle\}$.

Exercise 2.1.8 *Given $0 < p < 1$, determine the basis with respect to which the measurement of a qubit $\alpha|0\rangle + \beta|1\rangle$ gives outcome* 0 *with probability p.*

Qubit evolution

Any quantum evolution of a qubit, or any quantum operation on a qubit is described, as already mentioned, by a unitary matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

which transforms any qubit state $\alpha|0\rangle + \beta|1\rangle$ into the state $(a\alpha + b\beta)|0\rangle + (c\alpha + d\beta)|1\rangle$. [11]

For example, the evolution given by the Hadamard matrix (transformation)

$$H = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

which is called also Hadamard rotation, transforms the states $|0\rangle$, $|1\rangle$, $|0'\rangle$ and $|1'\rangle$ as follows;

$$H|0\rangle = |0'\rangle, \qquad H|0'\rangle = |0\rangle,$$
$$H|1\rangle = |1'\rangle, \qquad H|1'\rangle = |1\rangle.$$

The Hadamard transformation can also be seen as the following mapping of the basis states:

$$H|i\rangle = \frac{1}{\sqrt{2}}(|0\rangle + (-1)^i|1\rangle.$$

[11] As basic textbooks on quantum mechanics show, the basic evolution of a spin-$\frac{1}{2}$ particle can be well controlled by a clever application of magnetic field and the 2×2 unitary matrix describing time evolution of the spin in the spin-up spin-down basis, from time $t = 0$ to time $t = T$, has, up to the phase factors, in general the form:

$$U = \begin{pmatrix} e^{i\omega T/2} & 0 \\ 0 & e^{-i\omega T/2} \end{pmatrix} \begin{pmatrix} \cos(\omega T/2) & -\sin(\omega T/2) \\ \sin(\omega T/2) & \cos(\omega T/2) \end{pmatrix},$$

where the second matrix is a rotation matrix. For $\omega T = \pi$ we get, up to the phase, the matrix for the NOT operation, and for $\omega T = -\frac{\pi}{2}$, the Hadamard rotation.

The basis $\mathcal{B} = \{|0\rangle, |1\rangle\}$ is called the **standard basis**, or the **computational basis**, $\mathcal{D} = \{|0'\rangle, |1'\rangle\}$ is called the **dual basis** or the **Hadamard basis** or the **Fourier basis**. As we could see, by applying H we can switch between the standard and the dual bases. Observe also that $H^2 = I$. The so-called **circular (polarization) basis**

$$|0''\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle), \quad |1''\rangle = \frac{1}{\sqrt{2}}(|0\rangle - i|1\rangle)$$

is also of importance.

Exercise 2.1.9 *Construct matrices to transform a qubit state from: (a) standard basis to circular polarization basis and vice versa; (b) dual basis to circular polarization basis and vice versa.*

If the states $|0'\rangle$ and $|1'\rangle$ are measured with respect to the standard basis $\mathcal{B}$, we get both outcomes—0 and 1—with the same probability $\frac{1}{2}$. The evolution H applied on the states of the standard basis can therefore be seen as implementing a **fair coin tossing**.

Example 2.1.10 *If the matrix*

$$H' = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

is applied to the states of the standard basis, then $H'^2|0\rangle = -|1\rangle$, $H'^2|1\rangle = |0\rangle$, and therefore H'^2 acts as a NOT operation, up to the phase sign.

Exercise 2.1.11 *What do you get if the matrix from Figure 1.4e is applied: (a) once; (b) twice, to the standard basis states? (Transformation defined by this matrix is known as the* **square root of not**. *Explain why.)*

Three other important unitary operations on qubits are shown in Figure 2.3: rotation (by θ) $R(\theta)$; phase shift (with respect to α) $PS(\alpha)$; and scale (with respect to δ) $Scal(\delta)$.

$$R(\theta) = \begin{pmatrix} \cos\frac{\theta}{2} & \sin\frac{\theta}{2} \\ -\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix} \quad PS(\alpha) = \begin{pmatrix} e^{i\frac{\alpha}{2}} & 0 \\ 0 & e^{-i\frac{\alpha}{2}} \end{pmatrix} \quad Scal(\delta) = \begin{pmatrix} e^{i\delta} & 0 \\ 0 & e^{i\delta} \end{pmatrix}$$

Figure 2.3: Unitary matrices of degree 2

Exercise 2.1.12 *(Barenco, 1996) Show that the following properties hold for matrices R, PS, $Scal$ and the Pauli matrix σ_x.*
(a) $R(\theta_1) \cdot R(\theta_2) = R(\theta_1 + \theta_2)$;
(b) $PS(\alpha_1) \cdot PS(\alpha_2) = PS(\alpha_1 + \alpha_2)$;
(c) $Scal(\delta_1) \cdot Scal(\delta_2) = Scal(\delta_1 + \delta_2)$;
(d) $\sigma_x \cdot R(\theta) \cdot \sigma_x = R(-\theta)$; (e) $\sigma_x \cdot PS(\alpha) \cdot \sigma_x = PS(-\alpha)$.

A general form of the unitary matrices of degree 2 is well known:

Theorem 2.1.13 *Each unitary matrix U of degree 2 can be written as follows:*

$$U = e^{i\gamma} \begin{pmatrix} e^{i\alpha} & 0 \\ 0 & e^{-i\alpha} \end{pmatrix} \begin{pmatrix} \cos\theta & i\sin\theta \\ i\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} e^{i\beta} & 0 \\ 0 & e^{-i\beta} \end{pmatrix}$$

Exercise 2.1.14 *Show how to construct from the rotation, phase shift and scale matrices: (a) the Hadamard matrix; (b) the matrix* $\begin{pmatrix} \cos\theta & i\sin\theta \\ i\sin\theta & \cos\theta \end{pmatrix}$.

Exercise 2.1.15 *Prove Theorem 2.1.13 using the fact that a matrix is unitary if and only if its rows and columns are orthogonal.*

Exercise 2.1.16 *(Vazirani, 1997) Given a matrix M and a real λ, define $e^{i\lambda M} = \sum_{k=0}^{\infty} \frac{(i\lambda)^k}{k!} M^k$. (a) Compute $e^{i\sigma_x}$, $e^{i\sigma_y}$, $e^{i\sigma_z}$, where $\sigma_x, \sigma_y, \sigma_z$ are Pauli matrices; (b) show that each unitary matrix of degree 2 has the form $U = e^{i\gamma} e^{i\alpha\sigma_z} e^{i\theta\sigma_x} e^{i\beta\sigma_z}$.*

In principle, there is a continuous range of rotation, phase shift and scale matrices. However, already finitely many of them are sufficient to perform all quantum computations with an arbitrary precision.

Exercise 2.1.17 *Show that any rotation $R(\alpha)$ can be decomposed, with an arbitrarily small error, into polynomially many, with respect to the error, gates $R(\theta)$, with $\theta = 2\pi \sum_{k=0}^{\infty} \frac{1}{2^{2^k}}$.*

One of the basic tools for unitary operations in H_2 is the **beam splitter**, see Figure 2.4 for two ways a beam splitter is depicted. The beam splitter has two input and two output ports. By varying both phases of the incoming beams and the reflectivity of the beam splitter, one can realize by a beam splitter any unitary operation in H_2. A beam splitter is often depicted with only one input and with no specification of phasing and reflecting. In such a case it is assumed that the second input refers to the vacuum or other reference state and the incoming basis state is transferred to an equally weighted superposition of both bases states.

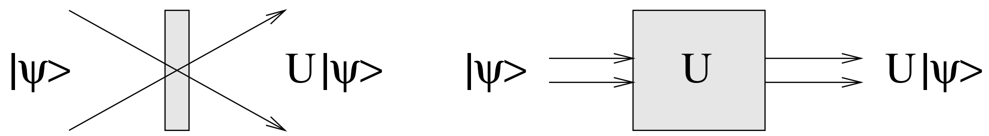


Figure 2.4: Beam splitter

Example 2.1.18 *In the case of photons a half-silver mirror (Figure 2.5b), acts as the beam splitter (see Figure 2.5a). A full-silver mirror (Figure 2.5c), reflects the photon. In the case of the half-silver mirror both detectors D1 and D2 detect the photon with the same probability. The fact that this cannot be interpreted as "photon goes one way with probability $\frac{1}{2}$ and the other way with the same probability" is well demonstrated using the so-called Mach–Zehnder* *interferometer (Figure 2.5d). In the case two half-silver mirrors and two full-silver mirrors are arranged as Figure 2.5d shows, then, as one can easily calculate, detector D1 detects the photon with probability* 1. *However, if we put an obstacle on one of the paths (Figure 2.5e), then both detectors detect with the same probability. If we put on one path an obstruction the photon can get through, but this obstruction acts as a measuring device, then both detectors detect the photon with the same probability (see Figure 2.5f). Finally, if we put on one path a glass that causes a precalculated delay (see Figure 2.5g), then only detector D2 detects the photon.*

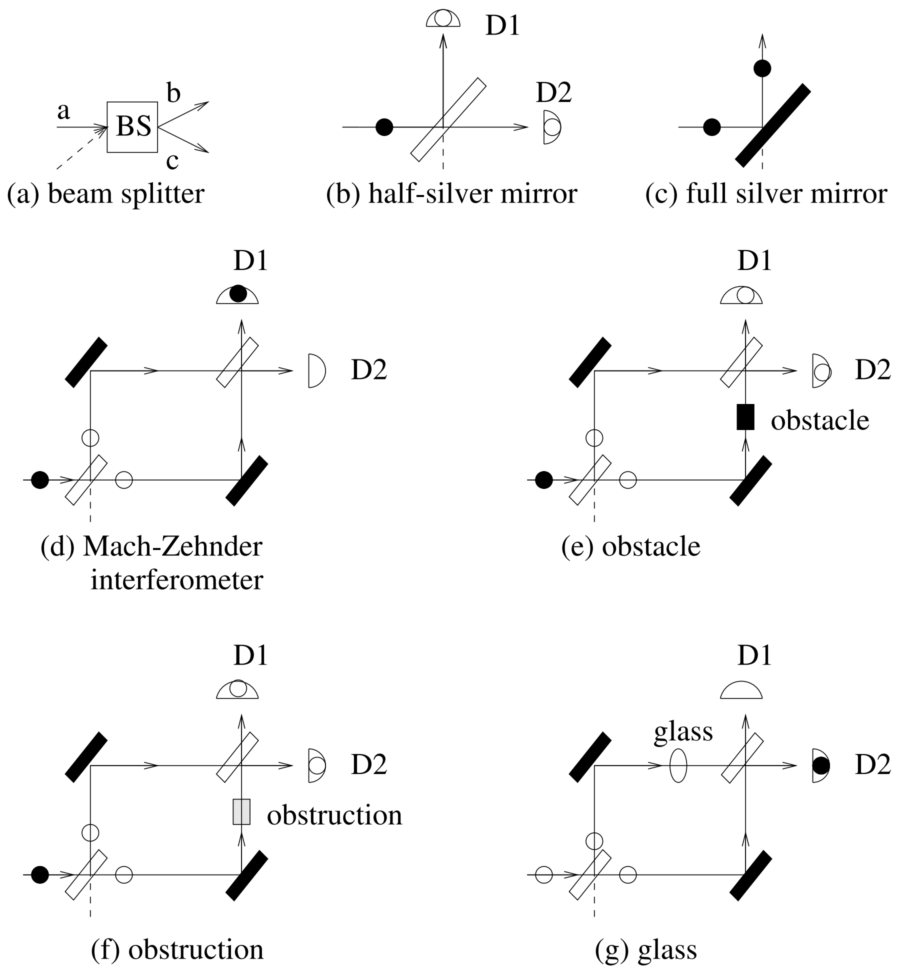


Figure 2.5: Mach–Zehnder interferometer in various situations

Remark 2.1.19 In addition to qubits as states of H_2, the special name **(qu)trit** is used for states in H_3, (for spin–1 particles). Their general form is

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle + \gamma|2\rangle,$$

where $|\alpha|^2 + |\beta|^2 + |\gamma|^2 = 1$, and $(\{|0\rangle, |1\rangle, |2\rangle\})$ is its standard basis.

One cannot do too much computation with a single qubit. Actually, it is the concept of an n-bit **quantum register** that is a proper framework for designing quantum algorithms.

2.1.2 Two-qubit registers

A tensor product of two qubits is called a 2**-qubit quantum register**. The corresponding Hilbert space is H_4. Several of its important/interesting bases are shown in Figure 2.6.

basis	basis state 1	basis state 2	basis state 3	basis state 4
standard	$\|00\rangle$	$\|01\rangle$	$\|10\rangle$	$\|11\rangle$
dual	$\|0'0'\rangle$	$\|0'1'\rangle$	$\|1'0'\rangle$	$\|1'1'\rangle$
Bell	Φ^+	Φ^-	Ψ^+	Ψ^-
	$\frac{1}{\sqrt{2}}(\|00\rangle + \|11\rangle)$	$\frac{1}{\sqrt{2}}(\|00\rangle - \|11\rangle)$	$\frac{1}{\sqrt{2}}(\|01\rangle + \|10\rangle)$	$\frac{1}{\sqrt{2}}(\|01\rangle - \|10\rangle)$
magic	$\frac{1}{\sqrt{2}}(\|00\rangle + \|11\rangle)$	$\frac{i}{\sqrt{2}}(\|00\rangle - \|11\rangle)$	$\frac{i}{\sqrt{2}}(\|01\rangle + \|10\rangle)$	$\frac{1}{\sqrt{2}}(\|01\rangle - \|10\rangle)$

Figure 2.6: Bases in H_4; Bell states are denoted by $\Phi^{\pm}$ and $\Psi^{\pm}$, and for each of these states the corresponding value is given one row below; the state Ψ^- is called **singleton**

It is usual to represent states of the standard basis in one of the following forms:

$$|0\rangle = |00\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \; |1\rangle = |01\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \; |2\rangle = |10\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \; |3\rangle = |11\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

A general state of a 2-qubit quantum register has the form

$$|\psi\rangle = \alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle, \tag{2.2}$$

where $|\alpha_{00}|^2 + |\alpha_{01}|^2 + |\alpha_{10}|^2 + |\alpha_{11}|^2 = 1$.

Exercise 2.1.20 *Express the basis states* $|00\rangle, |01\rangle, |10\rangle$ *and* $|11\rangle$ *in terms of the states* $|0'0'\rangle, |0'1'\rangle, |1'0'\rangle, |1'1'\rangle$ *and vice versa.*

Exercise 2.1.21 *(a) Design a unitary matrix that maps standard basis* $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ *into Bell's basis. (b) Design a unitary matrix that maps Bell's basis into the standard one; (c) Hadamard matrix transforms standard basis of* H_2 *into its dual basis and vice versa. Does there exist a unitary matrix that maps the standard basis of* H_4 *into Bell's basis and vice versa?*

Exercise 2.1.22 *Show that vectors of Bell's basis are eigenvectors of the unitary matrix implementing the mapping* $|x, y\rangle \rightarrow |\bar{x}, \bar{y}\rangle$.

Exercise 2.1.23 *(DiVincenzo et al. 1998a) For the following* 5 *states of two qutrits:* $|\phi_1\rangle = \frac{1}{\sqrt{2}}|0\rangle(|0\rangle - |1\rangle)$, $|\phi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)||2\rangle$, $|\phi_3\rangle = \frac{1}{\sqrt{2}}|2\rangle(|1\rangle - |2\rangle)$, $|\phi_4\rangle = \frac{1}{\sqrt{2}}(|1\rangle - |2\rangle)|0\rangle$ *and* $|\phi_5\rangle = \frac{1}{3}(|0\rangle + |1\rangle + |2\rangle)(|0\rangle + |1\rangle + |2\rangle)$ *show that they form "unextendable product base" of* H_9 *in the sense that: (a) they form an orthonormal set; (b) there is no state* $|\phi_1\rangle|\phi_2\rangle$*, where* $|\phi_1\rangle$ *and* $|\phi_2\rangle$ *are qutrit states, which is orthogonal to all* 5 *states above.*

Measurements of two-qubit registers

A measurement of the state $|\psi\rangle$, see (2.2), with respect to the standard basis produces as the outcome two bits ij with probability $|\alpha_{ij}|^2$, and causes the collapse of $|\psi\rangle$ into the state $|ij\rangle$.

Exercise 2.1.24 *Determine probabilities of possible outcomes when the state (2.2) is measured with respect to: (a) dual basis; (b) Bell's basis; (c) magic basis.*

It is often necessary to measure only one qubit. This can be done using observable

$$\begin{array}{ll} \mathcal{B}_1 = \{E_1^0, E_1^1\}, & \text{in the case of the first qubit;} \\ \mathcal{B}_2 = \{E_2^0, E_2^1\}, & \text{in the case of the second qubit.} \end{array}$$

where E_1^i, $i = 0, 1$, is the subspace spanned by the vectors $\{|i0\rangle, |i1\rangle\}$ and E_2^i, $i = 0, 1$ is the subspace spanned by the vectors $\{|0i\rangle, |1i\rangle\}$.

Hence, if the first qubit is measured, we get as the outcome bit 0 with probability $|\alpha_{00}|^2 + |\alpha_{01}|^2$, and the post-measurement state

$$|\psi'\rangle = \frac{\alpha_{00}|00\rangle + \alpha_{01}|01\rangle}{\sqrt{|\alpha_{00}|^2 + |\alpha_{01}|^2}}.$$

(Note that the state $|\psi'\rangle$ is projected and then renormalized.) In a similar way probabilities and resulting state are determined when outcome is 1 and when the second qubit is measured.

Quantum evolution of 2 qubits

Among unitary transformations on two qubits states the following transformation has a special role

$$\text{XOR} : |x, y\rangle \to |x, x \oplus y\rangle,$$

with the matrix representation

$$\text{XOR} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

XOR[12] matrix performs the same mapping as the CNOT gate considered in Section 1.7. (Observe that there is an infinite number of inequivalent quantum gates that reproduce the classical CNOT gate . Namely, all gates performing a transformation $|x, y\rangle \to e^{i\theta}|x, x \oplus y\rangle$.)

[12]For the first implementation of the XOR gate see Monroe et al. (1995) and for an introductory description of it see Berman et al. (1998).

Exercise 2.1.25 *Verify that if I denotes the unit matrix of degree 2, then*

$$XOR = |0\rangle\langle 0| \otimes \mathbf{1} + |1\rangle\langle 1| \otimes (|0\rangle\langle 1| + |1\rangle\langle 0|).$$

As discussed in detail in Section 2.3, the following unitary matrix is universal, in a reasonable sense, in the class of all unitary matrices

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & e^{i\alpha}\cos\theta & -ie^{i(\alpha-\phi)}\sin\theta \\ 0 & 0 & -ie^{i(\alpha+\phi)}\sin\theta & e^{i\alpha}\cos\theta \end{pmatrix},$$

if α, ϕ and θ are irrational multiples of π and of each other.

Exercise 2.1.26 *Show that the tensor products of matrices and quantum states have the following property:*

$$(A \otimes B)(|\psi_1\rangle \otimes |\psi_2\rangle) = (A|\psi_1\rangle) \otimes (B|\psi_2\rangle).$$

Exercise 2.1.27 *Show that XOR matrix cannot be obtained as a tensor product of two unitary matrices of degree 2.*

2.1.3 No-cloning theorem

The following simply stated and easily proved theorem (see Wootters and Zurek, 1982, and Dieks, 1982), shows what a unitary transformation cannot do on two qubit registers. No-cloning theorem has far reaching consequences for quantum information processing.

Theorem 2.1.28 (No Cloning (copying) Theorem) *An unknown quantum state cannot be cloned. (Namely, there is no unitary transformation U, such that for any one-qubit state $|\psi\rangle$, $U(|\psi, 0\rangle) = |\psi, \psi\rangle$.* [13]*) No cloning theorem holds for any Hilbert space.*

Proof 1. Assume that such a U exists and for two different orthogonal states $|\alpha\rangle$ and $|\beta\rangle$, $U(|\alpha, 0\rangle) = |\alpha, \alpha\rangle$, $U(|\beta, 0\rangle) = |\beta, \beta\rangle$. Let $|\gamma\rangle = \frac{1}{\sqrt{2}}(|\alpha\rangle + |\beta\rangle)$. Then $U(|\gamma, 0\rangle) = \frac{1}{\sqrt{2}}(|\alpha, \alpha\rangle + |\beta, \beta\rangle) \neq |\gamma, \gamma\rangle = \frac{1}{2}(|\alpha, \alpha\rangle + |\beta, \beta\rangle + |\alpha, \beta\rangle + |\beta, \alpha\rangle)$.[14] $\square$

[13] The no-cloning theorem seems to be bad news. However, this would be a very simplified view of its impacts. For example, very good news that unconditional secure quantum key generation is possible, see Section 6.2, is to a large extend due to this "bad news". In addition, new techniques have appeared that allow one to make approximate copies of qubits (Bužek and Hillery, 1996, and Bužek et al. 1997). They proposed the Universal Quantum Copy Machine that can produce two (imperfect, but equivalent in some weaker sense) copies of any qubit and the quality of copying is independent of particular qubits. It is even possible to make three imperfect copies. However, this good news has also 'bad" aspects. The copies obtained are entangled.

[14] The discovery that no general quantum copying procedure exists is a surprising and profound result of quantum mechanics. The reason behind it is that any attempt to copy a coherent superposition of states results in a state reduction, destruction of coherence, and the addition of noise. Moreover, the feasibility of cloning would have surprising consequences and would lead to paradoxes. For example, in combination with quantum teleportation, Section 6.4, it would allow faster-than-light transmission of information.

Proof 2. Assume there is a unitary operator U such that $U|\phi, 0\rangle = |\phi, \phi\rangle$, $U|\psi, 0\rangle = |\psi, \psi\rangle$, for arbitrary ϕ, ψ. Since U is unitary we have

$$\langle\psi|\phi\rangle\langle\psi|\phi\rangle = \langle\psi\psi|\phi\phi\rangle = \langle U\psi 0|U\phi 0\rangle = \langle\psi 0|\phi 0\rangle = \langle\psi|\phi\rangle\langle 0|0\rangle.$$

A contradiction is obtained if ϕ and ψ are chosen in such a way that $\langle\psi|\phi\rangle < 1$, which is always possible. ∎

Exercise 2.1.29 *Design a classical circuit to copy an unknown classical bit.*

Exercise 2.1.30 *Show that we can learn an unknown quantum state with arbitrary precision when we have an unlimited number of copies of the state and can measure it with respect to any observable we need.*

Observe also that “No-cloning theorem” only implies that there is no general unitary transformation for perfect copying of quantum information without destroying the original copy of information. As we shall see in Section 6.4, quantum information can be teleported (a copy of it can be moved to some different place) but the original copy will get destroyed.

Exercise 2.1.31 *Show that to any qubit state $|\psi\rangle$ of a qubit there is a unitary transformation $U_{|\psi\rangle}$ such that $U_{|\psi\rangle}(|\psi\rangle|0\rangle) = |\psi\rangle|\psi\rangle$.*

Exercise 2.1.32 *Show that if two different states $|\phi\rangle$ and $|\psi\rangle$ can be copied by the same circuit, then $\langle\phi|\psi\rangle = 0$.*

Exercise 2.1.33 *Show that there is no universal quantum NOT gate in the following sense—there is no unitary one-qubit transformation NOT such that $\langle NOT(\phi)|\phi\rangle = 0$ for any one qubit state $|\phi\rangle$.*[15]

2.1.4 Quantum registers

The generalization of 2-qubit registers to the case of n-qubit registers is straightforward.

Hilbert space of an n-qubit register

To deal with n-qubit registers we work within a 2^n-dimensional Hilbert space with the following set of basis vectors:

$$\mathcal{B} = \{|i\rangle \mid i \in \{0,1\}^n\}$$

or, another notation, used perhaps more often,

$$\mathcal{B} = \{|i\rangle \,|\, 0 \leq i < 2^n\},$$

which are said to form the **standard** or **computational** basis.

[15] However, as shown by Bužek et al. (1999), there is a method how to construct, to any qubit state $|\psi\rangle$ and any $\varepsilon > 0$ an (ε-orthgonal) state $|\psi_\varepsilon\rangle$ such that $\langle\psi|\psi_\varepsilon\rangle < \varepsilon$, provided a sufficient number $n_{\varepsilon,\psi}$ copies of $|\psi\rangle$ is available,

The general qubit state of the n-qubit register is

$$|\psi\rangle = \sum_{i=0}^{2^n-1} \alpha_i |i\rangle \text{ with } \sum_{i=0}^{2^n-1} |\alpha_i|^2 = 1.$$

Remark 2.1.34 n-qubit registers have several remarkable properties:

1. The number of basis states and the size of superpositions grows exponentially with the number of qubits. Note that already for $n = 200$ the total number of states of any basis is larger than estimated number of atoms in the universe.[16]
2. In order to store a number n one requires $\lceil \lg(n+1) \rceil$ qubits and therefore one has to use a Hilbert space of dimension $2^{\lceil \lg(n+1) \rceil}$, i.e. with $2^{\lceil lg(n+1) \rceil} > n$ distinguished basis states.
3. In spite of the exponential size of the set of basis states there is a fast, linear time, access from one state of the standard basis to another. Indeed, at most n NOT operation applications to individual qubits can bring any state $|i\rangle$ into any state $|j\rangle$.
4. An n-qubit register seems to be able to store exponentially more information than an n-bit register. (However, as follows from Holevo theorem (1973), n is a maximum number of bits one can "store" in an n-qubit register in such a way that one can again retrieve them faithfully.)

Evolution of n-qubit registers

An evolution of an n-qubit register state is given by a unitary operator with a $2^n \times 2^n$ matrix. In order to perform classically such an evolution step an exponential number, $2^n(2 \cdot 2^n - 1)$, arithmetical operations are needed in general.

On the other hand, in designing quantum algorithms and networks we mostly use operators that have a simple structure due to which the overall effort to simulate an evolution step is significantly smaller than in the general case.

It is mostly the case that an evolution step is performed by applying a unitary rotation operator to a single qubit or by applying a unitary operation, especially XOR, to two qubits.

If a unitary matrix U is applied to the ith qubit, then the overall unitary operation applied to the n-qubit register states has the form $(\bigotimes_{k=1}^{i-1} I) \otimes U \otimes (\bigotimes_{k=i+1}^{n} I)$. In order to simulate such an evolution step only $3 \cdot 2^n$ arithmetical operations are needed.

Exercise 2.1.35 *Let* $U = \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix}$. *Describe the structure of the following unitary matrices: (a)* $U \otimes I \otimes I$; *(b)* $U \otimes \bigotimes_{i=1}^{n-1} I$; *(c)* $\bigotimes_{i=1}^{k} I \otimes U \otimes \bigotimes_{i=1}^{k} I$; *(d)* $XOR \otimes I \otimes I$.

Example 2.1.36 *Let a unitary matrix* $U = \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix}$ *be applied to the first qubit of an n-qubit register in the state* $|\phi\rangle = \sum_{x \in \{0,1\}^n} \alpha_x |x\rangle$. *In such a case*

$$U|\phi\rangle = \sum_{y_1 \ldots y_n \in \{0,1\}^n} (\sum_{j=0}^{1} a_{y_1 j} \alpha_{j y_2 \ldots y_n}) |y_1 \ldots y_n\rangle.$$

[16] Therefore, already 100-qubit computers should have, in some sense, incredible power.

One of the aims of quantum algorithms and network design is to decompose a problem-solving unitary operator into a sequence of one and two qubit operations from a small set of such available operations. As discussed in Section 2.3, for any reasonable set of such basic operations such a decomposition always exists. However, an important problem is to determine for particular unitary operations whether such a decomposition into a polynomial number of one or two qubit operations can be done. When a considered unitary operation can be decomposed into a polynomial number $p(n)$ of one and two qubit operations, then the overall quantum computation time is $\mathcal{O}(p(n))$. Where a classical simulation is to be performed the total number of operations needed is $\mathcal{O}(p(n)2^n)$.

Measurements

If the state $|\phi\rangle$ of an n-qubit register is measured with respect to the standard basis we get as the outcome n bits, each n-tuples of bits with a precalculated probability, and the state $|\phi\rangle$, in the (potential) superposition of 2^n basis states, collapses to just one of the basis states.

Let us now consider a measurement of the jth qubit only. The corresponding observable is $\mathcal{B}_j = \{E_j^0, E_j^1\}$, where E_j^0 (E_j^1) is the subspace of the 2^n-dimensional Hilbert space spanned by all basic vectors having 0 (1) in the jth component.

The measurement of the jth qubit gives

$$0\ (1) \text{ with probability } \sum_{i \mid i_j=0} |\alpha_i|^2 \ (\sum_{i \mid i_j=1} |\alpha_i|^2),$$

where i_j denotes the jth bit in the binary representation of i.

A measurement of n qubits of an $(n+m)$-qubit register reduces the dimensionality of the Hilbert space we work with by the factor 2^n.

When we want to measure the first m qubits of an $(n+m)$-qubit quantum register, it is convenient to express its states as:

$$|\psi\rangle = \sum_{i=0}^{2^n-1} \sum_{j=0}^{2^m-1} c_{ij}|i,j\rangle \quad \text{with} \quad \sum_{i,j} |c_{ij}|^2 = 1.$$

In this case the states of the basis are seen as pairs of integers $0 \leq i < 2^n$ and $0 \leq j < 2^m$.

If we now measure the first n qubits we get each number $i \in [0, 2^n)$, with probability

$$p(i) = \sum_{j=0}^{2^m-1} |c_{ij}|^2$$

and the state $|\psi\rangle$ collapses into the state

$$|\psi_i\rangle = \frac{1}{\sqrt{p(i)}} \sum_{j=0}^{2^m-1} c_{ij}|i,j\rangle.$$

Exercise 2.1.37 *What are the results of the measurement of the first qubit of the following states with respect to the dual basis: (a)* $\alpha|000\rangle + \beta|111\rangle$*; (b)* $\frac{1}{\sqrt{2}}(|0000\rangle + |1111\rangle)$*; (c)* $\frac{1}{\sqrt{2}}(|0^{(n)}\rangle + |1^{(n)}\rangle$ *for an arbitrary* n*?*

Dual (Fourier) basis

It is the basis of the states $\{|i'\rangle \,|\, 0 \leq i < 2^n\}$. Transformation between standard and dual (or Fourier) basis is given by the so-called **Hadamard transformation** or **Walsh transformation** or **Walsh–Hadamard transformation** or **discrete Fourier transformation over $\mathbf{Z}_2^n$**.[17]

$$H_n = \bigotimes_{i=1}^{n} H.$$

It is easy to see that[18]

$$H_n|0^{(n)}\rangle = \bigoplus_{i=1}^{n} H|0\rangle = \bigoplus_{i=1}^{n} |0'\rangle = |0'^{(n)}\rangle = \frac{1}{\sqrt{2^n}} \sum_{i=0}^{2^n-1} |i\rangle$$

and, in general, for $x \in \{0,1\}^n$

$$H_n|x\rangle = \frac{1}{\sqrt{2^n}} \sum_{y \in \{0,1\}^n} (-1)^{x \cdot y} |y\rangle.^{19} \tag{2.3}$$

Example 2.1.38 (Setting of equally weighted superposition of basis states) *For many quantum algorithms the following initial state, which contains equally weighted distribution of all basis states, is of importance:*

$$|\psi\rangle = \frac{1}{\sqrt{2^n}} \sum_{i=0}^{2^n-1} |i\rangle. \tag{2.4}$$

Of key importance is the fact that it is very easy (theoretically at least) to set an "empty" register $|0^{(n)}\rangle$ into the state (2.4). One just needs to apply the Hadamard transformation H to each qubit as described above. In other words, using a linear number of operations we can transform one basis state into an exponential large and equally weighted superposition of all basis states.

Of importance in various areas of quantum information processing are the so-called **cat states** $\frac{1}{\sqrt{2}}(|0^{(n)}\rangle + |1^{(n)}\rangle)$. In the case $n = 2$ we have one of the Bell states, and in the case $n = 3$ one of the **GHZ states** (where GHZ stands for Greenberg, Horne and Zeilinger, 1998a). In general there are eight GHZ states of the form $\frac{1}{\sqrt{2}}(|abc\rangle \pm |\bar{a}\bar{b}\bar{c}\rangle)$, where $a, b, c \in \{0,1\}$ and at most one of a, b, c is one.

Exercise 2.1.39 *Describe unitary matrices that map the state $|0^{(n)}\rangle$ into the cat state $\frac{1}{\sqrt{2}}(|0^{(n)}\rangle + |1^{(n)}\rangle)$ for: (a) $n = 3$; (b) $n = 4$; (c) $n = 5$.*

Example 2.1.40 (Inversion about the average) *The unitary transformation*

$$D_n : \sum_{i=0}^{2^n-1} a_i|\phi_i\rangle \rightarrow \sum_{i=0}^{2^n-1} (2E - a_i)|\phi_i\rangle,$$

[17] Often the term "transform" is used instead of transformation .
[18] The notation $0^{(n)}$ stands for n 0s.
[19] The dot product is defined as follows: $x \cdot y = \bigoplus_{i=1}^{n} x_i y_i$.

where E is the average of $\{a_i \,|\, 0 \le i < 2^n\}$, can be performed by the matrix[20]

$$D_n = \begin{pmatrix} -1+\frac{2}{2^n} & \frac{2}{2^n} & \dots & \frac{2}{2^n} \\ \frac{2}{2^n} & -1+\frac{2}{2^n} & \ddots & \frac{2}{2^n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{2}{2^n} & \frac{2}{2^n} & \dots & -1+\frac{2}{2^n} \end{pmatrix}.$$

The matrix D_n is clearly unitary and it can be shown to have the form $D_n = -H_n R_n^1 H_n$, where $R_n^1[i,j] = 0$ if $i \ne j$, $R_n^1[1,1] = -1$ and $R_n^1[i,i] = 1$ if $1 < i \le n$.

Classical versus dual basis

In various considerations concerning Hilbert spaces H_{2^n}, especially in connection with quantum error-correcting codes, see Section 7.4, of large importance an interplay between the standard basis $\{|i\rangle \,|\, 0 \le i \le 2^n - 1\}$ and the dual basis $\{|i'\rangle \,|\, 0 \le i \le 2^n - 1\}$ is, where $|i'\rangle = |i'_{n-1}i'_{n-2}\dots i'_0\rangle$ and $i = i_{n-1}i_{n-2}\dots i_0$ in the binary representation.

Exercise 2.1.41 *(Steane, 1996) Let*

$$|\phi\rangle = \sum_{x\in\{0,1\}^n} \alpha_x|x\rangle = \sum_{x'\in\{0',1'\}^n} \alpha'_x|x'\rangle$$

and let $|\phi_j\rangle$ be obtained from $|\phi\rangle$ in the standard basis by the flipping jth bit of the basis states, i.e. $|\phi_j\rangle = \sum_{x\in\{0,1\}^n} \alpha_x|flip_j(x)\rangle$. Show that then

$$|\phi_j\rangle = \sum_{x'\in\{0',1'\}^n} \alpha'_x(-1)^{x'_j}|x'\rangle.$$

Exercise 2.1.42 *(Steane, 1996) If $|\phi_c\rangle = \sum_{x\in\{0,1\}^n} \alpha_x|flip(x)\rangle$, where $flip(x)$ flips all bits of the binary representation of x, then $|\phi_c\rangle = \sum_{x'\in\{0',1'\}^n} \alpha'_x(-1)^{parity(x')}|x'\rangle$, i.e if all basis words are flipped in the standard basis, then all odd parity words in the dual basis change their sign.*

Exercise 2.1.43 *Express the states $|1^{(n)}\rangle$ and $\frac{1}{\sqrt{2}}(|0^{(n)}\rangle + |1^{(n)}\rangle)$ in the dual basis.*

In all of the above exercises short states in one basis have long representation in another basis. This is inevitable because the following inequality holds (see Bialynicki-Birula and Mycielski, 1975, Deutsch, 1983, and Steane, 1996).

Theorem 2.1.44 *If a state $|\phi\rangle$ of a Hilbert space H_{2^n} can be written as a superposition of m_1 basis states in the standard basis and m_2 basis states in the dual basis, then*

$$m_1 m_2 \ge 2^n. \tag{2.5}$$

[20] The name of the operation comes from the fact that $2E - x = E + E - x$ and therefore the new value is as much above (below) the average as it was initially below (above) the average—which is precisely the inversion about the average.

2.2 Quantum Entanglement

One of the most specific and also most important concepts for quantum computing and quantum information theory is that of quantum entanglement—also one of the most puzzling concepts of quantum physics.

2.2.1 Entanglement of pure states

Let us start with an example. A 2-qubit register can be in the state

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle).$$

If we observe the first bit of this state using the standard observable $\mathcal{B}_1$, then we get the value 0 with probability $\frac{1}{2}$ and the value 1 also with probability $\frac{1}{2}$. (Hence the outcome is completely random.) After such an observation the state $|\psi\rangle$ collapses into the state $|00\rangle$ in the first case and into the state $|11\rangle$ in the second case.[21] If, afterwards, we measure the second qubit its value is determined uniquely, with probability 1. We see that if a quantum register is in the above state $|\psi\rangle$, then the two qubits are not independent. In addition, in such a case particular qubits of the quantum register no longer have an identity! Each of them is actually in a totally mixed state—with probability $\frac{1}{2}$ in the state $|0\rangle$ and with probability $\frac{1}{2}$ in the state $|1\rangle$; only both together form a pure state.

How specific and important is this example? We naturally expect that there should be cases where the qubits are completely independent and therefore they can be separately acted on.

This is actually the case if a state $|\psi\rangle$ of a 2-qubit register is the tensor product $|\psi_1\rangle \otimes |\psi_2\rangle$ of two 1-qubit states $|\psi_1\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ and $|\psi_2\rangle = \beta_0|0\rangle + \beta_1|1\rangle$, i.e.,

$$|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle = (\sum_{i=0}^{1} \alpha_i|i\rangle) \otimes (\sum_{j=0}^{1} \beta_j|j\rangle). \tag{2.6}$$

Exercise 2.2.1 *Determine: (a) $|\psi\rangle \otimes |\psi\rangle \otimes |\psi\rangle$ for $|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$; (b) $|\psi\rangle \otimes |\psi\rangle \otimes |\psi\rangle \otimes |\psi\rangle$ for $|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$; (c) $|\psi\rangle \otimes |\psi\rangle \otimes |\psi\rangle$ for $\psi = \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle)$.*

If we now observe the first qubit of the state $|\psi\rangle$, in (2.6), we get:

$$0 \text{ with probability } |\alpha_0\beta_0|^2 + |\alpha_0\beta_1|^2 = |\alpha_0|^2,$$

$$1 \text{ with probability } |\alpha_1\beta_0|^2 + |\alpha_1\beta_1|^2 = |\alpha_1|^2.$$

Moreover, after the observation of the first qubit the state $|\psi\rangle$ is reduced to $|\psi_2\rangle$; after the observation of the second qubit to $|\psi_1\rangle$.

If a quantum pure state of a Hilbert space H cannot be obtained as a tensor product of two quantum states from Hilberts spaces of the dimension smaller than that of H, then the state is called **entangled**. If this is a state of an n-qubit register, then we also say that all its qubits are entangled.

[21]Two qubits in the state $|\psi\rangle$ therefore behave like a pair of magic coins with the property that if they are tossed one after another then the second of them yields head if and only if the first one does so.

Entangled states arise in a natural way as a result of interactions between quantum systems. For example, if a pair of particles is created simultaneously and in such a way that some of their attributes (say spin or polarization) are preserved; for example, if one particle has spin $\frac{1}{2}$ and second has spin $-\frac{1}{2}$, or if one particle has polarization 0° and the second 90°. In addition, some quantum operations create entangled states out of unentangled states. For example, if XOR operation is applied to the state $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes |0\rangle$, the entangled state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ is produced.[22] Observe that entangled states do not exist in classical systems.

Exercise 2.2.2 *(Cleve et al. 1998) Show that the state* $\sum_{y=0}^{2^n-1} e^{\frac{2\pi i a y}{2^n}}|y\rangle$ *is unentangled if* $a \in \{0, \ldots, 2^n - 1\}$ *and can be expressed in the form* $\bigotimes_{i=1}^{n}(|0\rangle + \alpha_i|1\rangle)$*, for proper amplitudes* α_i*.*

Exercise 2.2.3 *(Cleve et al. 1998) Show that the state* $\sum_{y=0}^{2^n-1} e^{2\pi i \phi y}|y\rangle$ *is unentangled for all* ϕ*, and find its decomposition into the tensor product of one-qubit states.*

A pair of two-state particles which is in the entangled state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ or $\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$, is often said to be in the Bohm state or EPR state or Bell state or to form an EPR pair, or the EPR effect—"EPR" stands here for "Einstein, Podolsky and Rosen", see Section 9.1.5—and that it creates a so-called "EPR channel". Such a channel can be used to "teleport" quantum information as discussed in Section 6.4.

An EPR channel is created, for example, when certain types of atom or molecule decay with the emission of two photons, and consists of the fact that the two photons are always found to have opposite polarization, regardless of the basis used to measure them, provided both are measured with respect to the same basis. Figure 2.7 shows a way how an EPR-pair and channel can be generated.

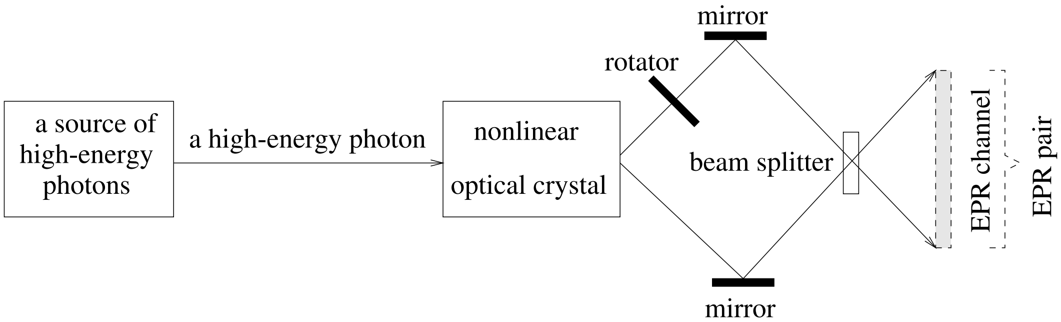


Figure 2.7: Generation of an EPR pair of polarized photons and an EPR-channel, adopted from Glanz (1995).

[22]Entangled are, for example, the states of Bell basis. Observe that XOR maps all states of Bell basis into states that are not entangled; it therefore performs **disentanglement** on the states of Bell basis.

The phenomenon of quantum entanglement was first noticed by Schrödinger, in 1935, and the English term "entanglement" is the direct translation of the German term "Verschränkung" used by Schrödinger.

Exercise 2.2.4 *Show that by an appropriate choice of the basis any entagled (pure) state of two qubits can be written as* $|\phi\rangle = \cos\phi|00\rangle + \sin\phi|11\rangle$.

Paradoxical nature of entanglement

The source of various paradoxes related to entangled states (see Section 9.1.5), is the fact that a pair of particles in an entangled state can be much physically separated. Each measurement on one particle of such an entangled pair immediately determines the state of another one, no matter how far apart they are.

A measurement of an entangled particle exhibits therefore the so-called "non-local effect" (see page 15), something that cannot happen from the point of view of classical physics without an instantaneous communication among the particles.

Remark 2.2.5 The concept of an entangled state is an excellent example of a difference between thinking in terms of Hilbert space constructs and dealing with real quantum systems. Mathematically the concept of entanglement is a trivial one. Physically the existence of entangled states of distant particles is one of the most puzzling phenomena, and the Gedanken experiment devised by Einstein and his colleagues in 1935 concerning entanglement was to illustrate how unreasonable, from Einstein's point of view, quantum mechanics is (see Section 9.1.5). It was a very challenging experimental task to show experimentally entanglement of even 2 qubits. The entanglement of two spatially separated particles was experimentally demonstrated by Wu and Shaknov (1950), Kwiat et al. (1995) and Zeilinger et al. (1997)—polarized photons—and Hagley et al. (1997), pairs of atoms. The entanglement of two photons over 10 km has been demonstrated at the University of Geneva (see Tittel et al. 1998). Recent experimental demonstrations of GHZ states are considered as important achievements. First Laflamme et al. (1997) reported, using NMR technology, that three nuclear spins within a single molecule locally exhibited GHZ state. Later, Bouwmeester et al. (1998), demonstrated polarization entanglement in GHZ state for three spatially separated photons (such that each can be addressed separately).

The following set-up further illustrates peculiarities of entanglement.

Example 2.2.6 *Let both Alice and Bob possess one particle of an entangled pair in the state* $\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$ *and assume they travel to distant places and agree to measure their particles immediately after their arrival. Let us assume that Alice gets as the result of her measurement state the* $|0\rangle$. *She knows immediately that Bob's particle has the state* $|1\rangle$. *However, this situation could happen in two ways. The first possibility is that Alice has arrived first and her measurement determined the state of Bob's particle. The other possibility is that Bob has arrived first, got* $|1\rangle$ *as the result of his measurement, and that determined* $|0\rangle$ *as the state of Alice's particle. How to find out for Alice which of these two cases did really happen? She could call Bob to ask about the time of his measurement. Interestingly enough, such a classical communication seems to be the only way to solve this problem.*

Entanglement as a computational and communicational resource

Entanglement plays a central role in quantum information theory that extends classical information theory—see Chapter 8. According to Bennett (1998a), entanglement between a pair of quantum systems in a maximally entangled state is the purest form of inherently quantum information: it is capable interconnecting two parties far apart, it cannot be copied, eavesdropped without disturbance, nor it can be used by itself to send classical messages. At the same time it can assist in speeding up both classical and quantum communication.

Quantum entanglement is also the main reason why quantum computers cannot be efficiently simulated by classical ones.

To describe fully a state of n-qubit register we need to write down in general 2^n complex coefficients. Already for a small $n = 100$ this would require $2^{100} \approx 10^{30}$ numbers, which is outside the potential of foreseeable classical computers. (In addition, to simulate a quantum computer with 100-qubit register we would need to manipulate matrices of degree 10^{30}.) Since quantum computers are probabilistic it could seem that keeping a complete state description at each stage of simulation is not the only way to simulate quantum computers on probabilistic ones. It is therefore natural to ask whether it is possible to simulate quantum computers on probabilistic ones which do not always keep a complete description of the current quantum state and only provide various outcomes with the same probability as the simulated quantum computer does. Could not we perform some "local simulations" in which each qubit has a definite value at each computation step and each quantum gate can act on the input qubits in various possible ways, only one of which is always selected as determined by a (pseudo)random generator? Such simulations could avoid a need for exponential space! However, it is a consequence of quantum entanglement, of its non-local correlations, that we cannot always divide the state of the quantum system under consideration into parts and compute them separately. Due to the quantum entanglement there is in general no local probabilistic classical algorithm simulating quantum computers.

An easy way to demonstrate directly that it is exactly because of quantum entanglement that quantum computers cannot be efficiently simulated by classical ones goes as follows.

Let us assume that an evolution of a state $|\psi\rangle$ of an n qubit register goes through a sequence of states $|\psi_i\rangle$, $i \geq 1$, and $|\psi_i\rangle = |\phi_{i1}\rangle \otimes |\phi_{i2}\rangle \otimes \ldots \otimes |\phi_{in}\rangle$, where each $|\phi_{ij}\rangle$ is a qubit state. No matter how long such an evolution is, it can be easily simulated by a classical computer because it can simulate evolution qubit by qubit and there is no exponential increase in the number of coefficients a classical computer has to store.

It often used to be emphasized that quantum superposition is the root of the extraordinary power of quantum computing. However, it is nowadays clear that this is too simplified a view of quantum computing.

Indeed, as already discussed in Section 1.5, classical waves also exhibit superpositions. As a consequence, any effect depending on quantum interference alone seems to be readily simulated by classical waves. Entanglement can be seen as a very special type of superposition that has no classical analog. Jozsa (1997) provides arguments that we cannot always simulate quantum parallelism (see page 103), by classical waves.

Quantum entanglement should be seen as a computational resource that allows qualitatively and quantitatively new types of information processing. At the same time entanglement is a resource which is very difficult to create and to preserve.

Applications of quantum entanglement: speed-up of classical computations, Sections 3.1, 3.2, 3.3; quantum key generation, Sections 6.2 and 6.2.4; teleportation, Section 6.4; superdense coding, Section 6.4.4; entanglement enhanced classical communication (Bennett, Fuchs and Smolin, 1997); quantum data compression, Section 8.2; error-correction codes, Section 7.4; fault-tolerant computing, Section 7.5, dense coding, Section 8.2.4.

2.2.2 Quantifying entanglement

The amount of entanglement $E(\phi)$ in a pure state $|\phi\rangle$ of a compound (bipartite) quantum system $A \otimes B$ is defined (see Bennett et al. 1996a), using the concept of von Neumann quantum entropy by

$$E(\phi) = -Tr\rho_A \lg \rho_A = -Tr\rho_B \lg \rho_B,$$

where $\rho = |\phi\rangle\langle\phi|$. (See Sections 1.4.5 and 9.2.3 for references to density matrices and the tracing out operation.)

Bell states and GHZ states are an important example of maximally entangled two-qubit states.[23]

A pair of maximally entangled qubits forms the so-called **ebit**—the term coined by Ch. Bennett et al. (1996)[24].

An example of weakly entangled states in H_2 are states $\frac{1}{\sqrt{k}}|00\rangle + \sqrt{\frac{k-1}{k}}|11\rangle$ for large k.

Another examples of maximally entangled states in H_n are the states

$$|\phi_n\rangle = \frac{1}{\sqrt{n}} \sum_{1=1}^{n} |i\rangle|i\rangle.$$

It is easy to see that $|\phi_{2^n}\rangle = |\Phi^+\rangle^{\otimes n}$, i.e., $|\phi_{2^n}\rangle$ is the tensor product of the Bell state $|\Phi^+\rangle$ with itself n times.

There is a variety of technical results showing that entropy of entanglement as defined above is a reasonable measure of entanglement. For example, if two parties share an entangled pair, then they cannot change the entropy of entanglement by local actions (even with the help of classical communication). In addition, two pure entangled states with the same entropy of entanglement can be transferred into each other by local actions (see Bennett, 1998a).

The concept of entanglement is defined for mixed states and density matrices similarly as for pure states.

A mixed state $[\psi\rangle$ is **separable** or **disentangled** if $[\psi\rangle = \bigoplus_{i=1}^{n}(p_i, |\phi_i\rangle \otimes |\psi_i\rangle)$, where $\sum_{i=1}^{n} p_i = 1$ and $|\phi_i\rangle$, $|\psi_i\rangle$ are pure states. $[\psi\rangle$ is entangled if it is not disentangled.

A density matrix ρ is disentangled if $\rho = \sum_{i=1}^{n} p_i\rho_i \otimes \rho'_i$, $\sum_{i=1}^{n} p_i = 1$ and it is entangled if it is not disentangled.

Several approaches to quantification of entanglement of mixed states and density matrices are dealt with in Section 8.3 where various ways of creation and manipulation of entangled states are dealt with.

Remark 2.2.7 As illustrated in the rest of the book, entanglement can be used in various ways to make quantum communication more efficient and more secure. It is also well known in quantum mechanics that to any entangled state one can find operators whose correlation violate Bell inequalities and contradict "local realism" view of quantum physics (see Section 2.7). An important role in quantum information processing have quantum error-correcting codes whose highly entangled codewords protect quantum information. As

[23] On the Bell states, that play an important role in the quantum theory of entanglement, one can also illustrate enormous difference between what is possible in quantum theory and practice. Theoretically, a single projection measurement can distinguish four Bell states. However, until now no experimental way to do that in H_4 is known! Only recently (see Kwiat and Weinfurter, 1998), a way has been demonstrated how to distinguish these four states, but only by working in a larger Hilbert space—making use of additional entanglement.

[24] The term "ebit" is also almost a trademark for "Electron-Beam Ion Trap".

discussed and illustrated by DiVincenzo and Peres(1997), entangled codewords appear to lead, under the assumption of local realism, to a reach variety of logical contradictions—"quantum paradoxis".

2.2.3 Substituting entanglement for communication

One of the surprising applications of entanglement was discovered in the area of multiparty communications. For example, as will now be shown, by sharing entangled particles communicating parties can reduce the amount of bits they need to broadcast in order to solve together a special communication/cooperation problem—to compute together a function whose input data are distributed among the parties.

There are k parties $P_1, \dots, P_k$ and a k-variable function f whose inputs are n-bit strings $x_1, \dots, x_k$. The party P_j holds only the string x_j. All parties want to compute together $f(x_1, \dots, x_k)$. In order to achieve it, the parties have to communicate. The average number of bits the parties have to broadcast if they work with classical information only, for valid inputs of length n, is called the communication complexity of f and denoted by $C(f, n, k)$. (This definition is a natural modification of the usual concepts in communication complexity.[25]) For the case where the parties can work also with quantum information, the corresponding communication complexity is denoted by $QC(f, n, k)$.

We show, following van Dam, Høyer and Tapp (1997), an example of functions $f_k, k = 1, 2, \dots$ such that $QC(f_k, n, k)$ is asymptotically smaller, with respect to k, than $C(f_k, n, k)$.

An input $(x_1, \dots, x_k) \in \{0, \dots, 2^n - 1\}^k$ is defined to be **valid** if

$$\left(\sum_{i=1}^{k} x_i \right) \bmod 2^{n-1} = 0.$$

Denote by $f_k : \{0, \dots, 2^n - 1\}^k \to \{0, 1\}$, the Boolean function defined, on valid inputs only, by

$$f_k(x_1, \dots, x_k) = \left(\frac{1}{2^{n-1}} \sum_{i=1}^{k} x_i \right) \bmod 2.$$

Clearly, $f_k(x_1, \dots, x_k) \in \{0, 1\}$ for all valid inputs.

At first we show that $QC(f_k, n, k) = k$ provided k communicating parties share k entangled qubits, each party having one. Later we show that $C(f_k, n, k) \geq k \lg k - k$ for $n \geq \lg k$. In this way we find that for functions f_k defined above communication with the help of entanglement is asymptotically better.

Let us assume that k parties share k qubits, the party P_i the qubit q_i and together these qubits are in the entangled state $|q_1 \dots q_k\rangle = \frac{1}{\sqrt{2}}(|0^{(k)}\rangle + |1^{(k)}\rangle)$. Let each party P_j independently apply on its qubit q_j the following procedure.

1. A phase-changing transformation
$$|0\rangle \to |0\rangle, \quad |1\rangle \to e^{\frac{2\pi i x_j}{2^n}} |1\rangle.$$
2. Hadamard transformation.
3. A measurement of the jth qubit q_j, with respect to the standard observable $\{|0\rangle, |1\rangle\}$ providing an output $b_j \in \{0, 1\}$.

[25] See Hromkovič (1997), Gruska (1997).

After the first step the resulting state is $\frac{1}{\sqrt{2}}\left(|0^{(k)}\rangle + (e^{\frac{2\pi i}{2^n}\sum_{j=1}^k x_j})|1^{(k)}\rangle\right)$. (Observe that for a valid input $e^{\frac{2\pi i}{2^n}\sum_{j=1}^k x_j}$ equals 1 or -1.) After the second step the resulting state is

$$\frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2^k}}\sum_{j=0}^{2^k-1}|j\rangle + e^{\frac{2\pi i}{2^n}\sum_{j=1}^k x_j}\frac{1}{\sqrt{2^k}}\sum_{j=0}^{2^k-1}(-1)^{parity(j)}|j\rangle\right),$$

where $parity(j) = 0$ if j is even and 1, otherwise. Since $e^{\frac{2\pi i}{2^n}\sum_{j=1}^k x_j} = (-1)^{f_k(x_1,\dots,x_k)}$ we see that the resulting state is

$$\frac{1}{\sqrt{2^{k-1}}}\sum_{parity(j)=f_k(x_1,\dots,x_n)}|j\rangle.$$

It is now easy to determine that

$$b_1 \oplus b_2 \oplus \dots \oplus b_k = f_k(x_1,\dots,x_k),$$

no matter in which order parties perform their measurements.

Exercise 2.2.8 *Determine in detail that $\oplus_{i=1}^k b_i = f_k(x_1,\dots,x_n)$ for: (a) $k = 2,3$; (b) for an arbitrary k.*

All parties can compute the value of f_k if bits b_i, $i = 1,\dots,k$ are broadcast. On the other hand, broadcasting of less than k bits cannot be sufficient to compute f_k because if one of the parties does not broadcast its bit, then no other party can compute the value of f_k.

Let us now consider communications between the parties in the case that only classical information is used.

The simplest way to communicate is that all but one of the parties broadcasts their inputs. The last party then computes the result and broadcast it to all other parties. This implies

$$C(f_k,n,k) \leq (k-1)n + 1.$$

Another possibility is that parties $P_1, P_2, \dots, P_{k-1}$ broadcast the d most significant bits of their inputs, i.e., $y_i = x_i - x_i \bmod 2^{n-d}$ for some $d > 1$. The last party, say P_k, then computes the sum

$$(\sum_{i=1}^k x_i) - \delta,$$

where

$$\delta = \sum_{i=1}^{k-1}(x_i \bmod 2^{n-d}).$$

If $n \geq d = 1 + \lceil \lg(k-1)\rceil$, then

$$0 < \delta \leq (k-1)(2^{n-d}-1) < 2^{n-1}.$$

This implies that P_k knows the value of the sum $\sum_{i=1}^k x_i$ up to an additional non-negative term smaller than 2^{n-1}. Since the sum $\sum_{i=1}^k x_i$ is divided by 2^{n-1} for valid inputs, P_k can determine the sum $\sum_{i=1}^k x_i$ exactly and therefore it can compute f_k.

In this case it holds, for the total number of bits to be broadcast,

$$C(f_k, n, k) \leq (k-1)(\lceil \lg(k-1) \rceil + 1) + 1.$$

$\square$

Exercise 2.2.9 (Cleve and Buhrman, 1997) *Let each of three parties A, B and C possess an n bit string w^A, w^B and w^C, such that $w_i^A \oplus w_i^B \oplus w_i^C = 1$ for $1 \leq i \leq n$. In addition, let each party P possess one of the particles q_i^P, $1 \leq i \leq n$, of n triples of particles each in the entangled state $|\psi\rangle = \frac{1}{2}(|001\rangle + |010\rangle + |100\rangle - |111\rangle)$. Show that the three parties can compute the function $f(w^A, w^B, w^C) = \bigoplus_{i=1}^n w_i^A \wedge w_i^B \wedge w_i^C$ in such a way that each party $P \in \{A, B, C\}$ performs the following protocol:*
for i **from** 1 **to** n **do**
  if $w_i^P = 0$ **then** *apply Hadamard rotation to* q_i^P;
  get the bit s_i^P *by measuring* q_i^P;
compute $s^P \leftarrow \bigoplus_{i=1}^n s_i^P$.
and then let parties B and C send two bits, namely s^B and s^C, to A such that A can compute $s^A + s^B + s^C$ which equals $f(w^A, w^B, w^C)$. (It can be shown that the classical communication complexity of this problem is 3. Using quantum entanglement only two bits for communication are necessary, as demonstrated by the above protocol.)

Remark 2.2.10 In the result presented in this section it has been demonstrated that entanglement has the potential to act as a substitute for communication during multiparty cooperation. On the other hand, the potential of entanglement for direct communication is very restricted. This is discussed in more detail in Section 6.4 on teleportation—an another important application of entanglement.

Open problem 2.2.11 *Can quantum entanglement decrease also two-party communication complexity for some communication problems?*

2.3 Quantum Circuits

In a similar way as Boolean gates and circuits are building blocks of classical computers, quantum analogues of them are natural elements from which quantum computing devices are designed. However, the related theoretical, design and implementation problems for quantum gates and circuits are of significantly larger complexity.

2.3.1 Quantum gates

Unitarity is the main new requirement quantum gates have to satisfy.

Definition 2.3.1 *A* **quantum gate** *with n inputs and n outputs is specified by a unitary operator $U : \mathcal{H}_{2^n} \to \mathcal{H}_{2^n}$, and it is represented by a unitary matrix of degree 2^n.*

Example 2.3.2 *The so-called Hadamard (rotation) gates are represented by matrices*

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad H' = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \quad H'' = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Example 2.3.3 *The following one parameter set of* **rotation gates** *(represented by matrices) is also often used:*

$$R_x(\theta) = \begin{pmatrix} \cos\theta & i\sin\theta \\ i\sin\theta & \cos\theta \end{pmatrix}, \quad R_y(\theta) = \begin{pmatrix} i\cos\theta & \sin\theta \\ \sin\theta & i\cos\theta \end{pmatrix}, \quad R_z(\theta) = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix},$$

and

As already mentioned, of central importance for quantum computing is the XOR gate (Figure 2.8). Observe that if the target qubit has the input $|0\rangle$, then this gate can be used to copy qubits $|0\rangle$ and $|1\rangle$ from the control qubit. At the same time the gate in Figure 2.8 can be seen as a classical wire. Indeed, inputs $|0\rangle$ and $|1\rangle$ on the control qubit come out on the target qubit output, but a superposition $\alpha|0\rangle + \beta|1\rangle$ on the control qubit is transferred into the entangled state $\alpha|00\rangle + \beta|11\rangle$ and if we consider as the overall output that of the target qubit then this output is a mixed state: $|0\rangle$ with probability $|\alpha|^2$ and $|1\rangle$ with probability $|\beta|^2$, and if $|\alpha| = |\beta|$ then the output is a random $|0\rangle$ or $|1\rangle$.

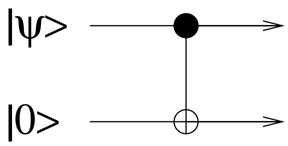


Figure 2.8: XOR gate as a real wire

Just as with classical gates, quantum gates also can be described by "truth tables" showing that the outputs for the case inputs are states $|0\rangle$ and $|1\rangle$. The point is that once such a truth table is given, linearity of quantum gate mappings allows us to determine gate outputs for all possible input states. In this way several often-used gates have been described in Figures 1.14 and 1.15: CNOT or XOR gate as well as Fredkin and Toffoli gates.

Exercise 2.3.4 *Describe unitary matrices for (a) Toffoli gate; (b) Fredkin gate.*

Exercise 2.3.5 *XOR gate can be expressed using outer and tensor products as follows:* $|0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes \sigma_x$. *Find a similar representation for: (a) Toffoli gate; (b) Fredkin gate.*

The gates for basic Boolean reversible operations NOT, CNOT and CCNOT can be described also using a notation for registers as follows:

$$\begin{aligned} \text{NOT} &: |a\rangle \rightarrow |\bar{a}\rangle \\ \text{XOR} = \text{CNOT} &: |a, b\rangle \rightarrow |a, a \oplus b\rangle \\ \text{CCNOT} &: |a, b, c\rangle \rightarrow |a, b, (a \wedge b) \oplus c\rangle \end{aligned}$$

Observe that each of these gates is its own inverse.

An important difference between the classical and quantum computing is that for any n the number of n input/output quantum gates is uncountable. This immediately gives rise to several basic questions. Which of the gates are of real importance for quantum computing? Which of them can be well implemented with current or foreseeable technology?

A direct implementation of arbitrarily large/complex quantum gates is outside the possibilities of the current and foreseeable technology. This immediately arises another fundamental question: Can all large unitary matrices be effectively decomposed, in some reasonable sense, into a small (polynomial) number of unitary matrices of very small degree?

This brings us immediately to the concept of quantum circuits as an analogue of Boolean circuits.

Definition 2.3.6 *A* **quantum (Boolean) circuit** *is a collection of quantum gates acyclicly connected (by "quantum wires").*[26] *The* **size** *and the* **depth** *of a circuit refer to the number of nodes and depth of the underlying connection graph.*

A relation between a quantum circuit and the corresponding unitary matrix is far from being very transparent even for simple circuits and some experience is needed to get proper feelings in this respect. That is why there are quite a few (very simple) examples and exercises in this section, worth paying detailed attention to.

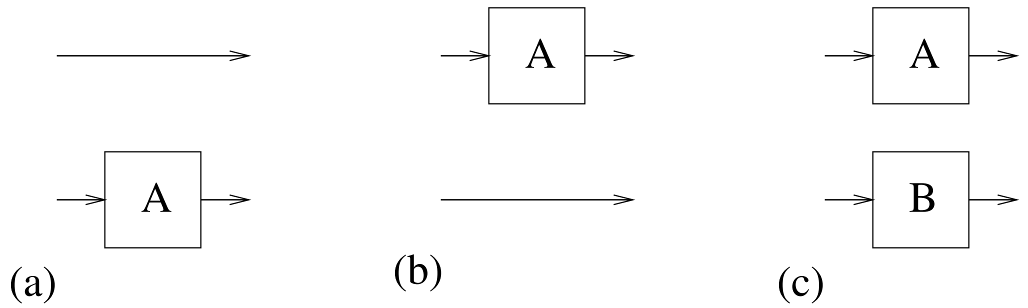


Figure 2.9: Elementary networks I

Exercise 2.3.7 *Let* $a = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, $B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$ *be unitary matrices of degree* 2*. Design unitary matrices of degree 4 that represent mappings realized by networks: (a) in Figure 2.9a; (b) in Figure 2.9b; (c) in Figure 2.9c.*

Exercise 2.3.8 *Let us assume that unitary matrices* A, B, C *of degree* 2 *are given. Design unitary matrices representing networks shown: (a) in Figure 2.10a; (b) in Figure 2.10b; (c) in Figure 2.10c; (d) for a network obtained by serial composition of networks in Figure 2.10a,b; (e) for a network obtained by serial composition of networks from Figure 2.10a,c.*

If gates G_1 and G_2 realizes the mappings described by unitary matrices A_1 and A_2, then the network in Figure 2.11a realizes the mapping described by the matrix $A_1 \otimes A_2$.

[26] We are using here similar terminology as for classical circuits, in spite of the fact that there are no real wires on the quantum level. Two ports of the gates communicate either by sharing a physical qubit or via field interactions or using other physical means.

When considering a class of quantum circuits $\{C_i\}_{i=1}^{n}$ of a certain type, it is necessary to assume, if we want to assign computations on such circuits to uniform complexity classes, that all such circuits can be designed (computed) by a single classical Turing machine in a polynomial time with respect to n.

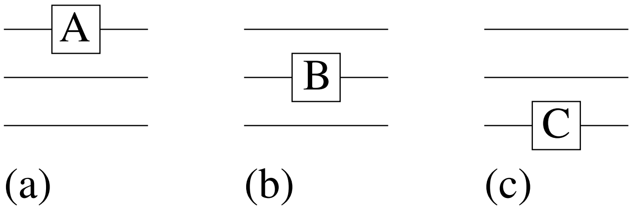


Figure 2.10: Elementary networks II

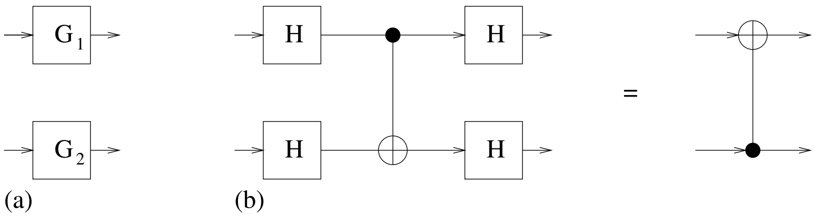


Figure 2.11: An implementation of the inverse of the XOR gate.

The processing in the network on the left side of the identity in Figure 2.11b for the input $|0\rangle|1\rangle$ can be depicted as follows:

$$
\begin{aligned}
|0\rangle|1\rangle \quad & \overset{H\text{–gates}}{\longrightarrow} \quad \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|0\rangle - |0\rangle|1\rangle - |1\rangle|1\rangle) \\
& \overset{\text{XOR gate}}{\longrightarrow} \quad \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle - |0\rangle|1\rangle - |1\rangle|0\rangle) \\
& \overset{H \text{ gates}}{\longrightarrow} \quad |1\rangle|1\rangle.
\end{aligned}
$$

Exercise 2.3.9 *Prove the equivalence of networks shown in: (a) Figure 2.11b; (b) Figure 2.12.*

Exercise 2.3.10 *Show how to design quantum circuits producing the cat state* $\frac{1}{\sqrt{2}}(|0^{(l)}\rangle + |1^{(l)}\rangle)$ *for: (a)* $l = 2$*; (b)* $l = 3$*; (c) for an arbitrary* l*.*

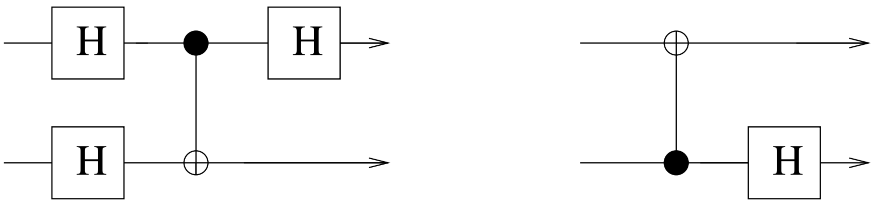
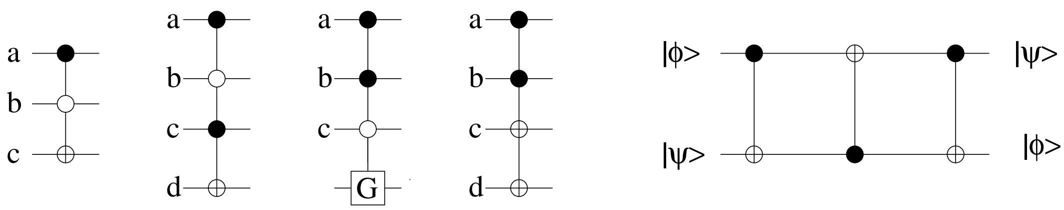


Figure 2.12: Two equivalent circuits

Example 2.3.11 *The XOR gate determined by the matrix XOR is an important example* *of a 2-qubit gate. If depicted as in Figure 1.14, then a simple circuit of three such gates, shown in Figure 2.13b, flips the qubits.*

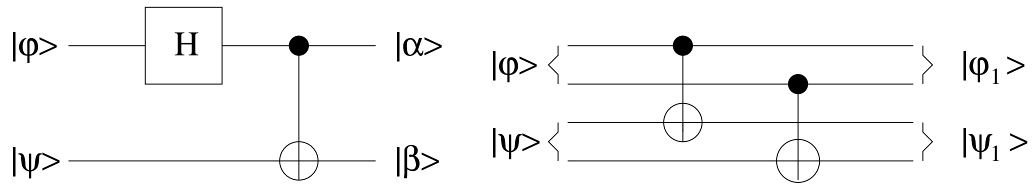


(a1) (a2) (a3) (a4) (b)

Figure 2.13: Generalized XOR gate notations and a quantum circuit to flip the qubits

Notation used for XOR and Toffoli gates in Figure 1.14 is often generalized to consider cases that target bits should flip if one of the control bits is 0 (represented by the empty circle) and for the cases there are several control and/or target bits or we have a controlled gate. For example, the circuit in Figure 2.13a1 realizes the mapping $(a, b, c) \to (a, b, (a \wedge \bar{b}) \oplus c)$; the circuit in Figure 2.13a2, the mapping $(a, b, c, d) \to (a, b, c, (a \wedge \bar{b} \wedge c) \oplus d)$. In the circuit in Figure 2.13a3 the gate G is applied if and only if $a \wedge b \wedge \bar{c} = 1$. Finally, in the circuit from Figure 2.13a4 the last two bits flip if and only if $a \wedge b = 1$ (this stands actually for two consecutive XOR gates with the same control bits but different target bits). In addition, notation $\mathrm{XOR}_{i,j}$ will be used to denote the case that an XOR gate is applied to the ith and the jth qubit as the control and the target qubits.

Example 2.3.12 *Circuit in Figure 2.14a produces the Bell states.*



(a) (b)

Figure 2.14: A circuit to produce Bell states and a circuit to map pairs of Bell states into pairs of Bell states

Exercise 2.3.13 *Show that the quantum circuit, so-called bilateral quantum XOR (BXOR), shown in Figure 2.14, transforms a pair of Bell states $|\phi\rangle$ and $|\psi\rangle$ into a pair of Bell states $|\phi_1\rangle$, $|\psi_1\rangle$.*

Exercise 2.3.14 *Show that each of the Pauli matrices maps in a one-to-one way states of the Bell basis on themselves if the matrix is applied to one of the two qubits of the given Bell state. (In other words, Bell states can be connected into one another by unilateral Pauli rotations.)*

XOR is an important example of quantum gates with two or more inputs that performs the so-called **conditional quantum dynamics**, in which one system (qubit) undergoes an evolution that depends on the quantum state of another system (of other qubits).

Exercise 2.3.15 *The circuit on Figure 2.15 (see Moore and Nilsson, 1998), performs a permutation of* 3 *qubit states using* 3 *ancilla qubits. Show: that any permutation of* n *qubit states can be performed: (a) using* 4 *layers of XOR gates with* n *ancilla qubits; (b) using* 6 *layers of XOR gates and no ancilla qubits.*

Exercise 2.3.16 *Show that any mapping* $(x_1, \dots, x_n, b) \rightarrow (x_1, \dots, x_n, b \oplus x_1 \oplus \dots \oplus x_n)$ *can be performed by a* $\mathcal{O}(\lg n)$*-depth circuit using only XOR gates.*

Exercise 2.3.17 *Design: (a) a circuit to recognize Bell states; (b) a circuit to generate GHZ states; (c) a circuit to recognize GHZ states.*

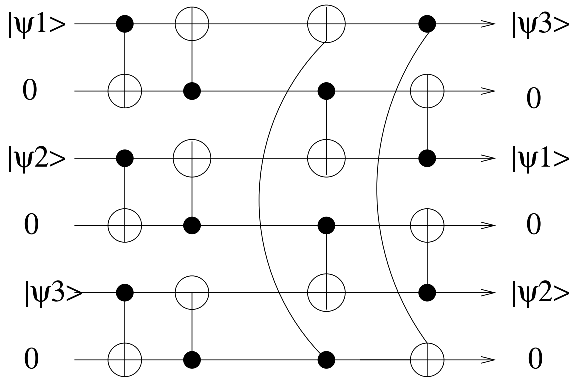


Figure 2.15: Permutation circuit

Example 2.3.18 *The Hadamard transform* H_n *is implemented by the circuit in Figure 2.16a, and Figure 2.16b contains the usual notation for the circuit for* H_n.

Exercise 2.3.19 *Apply the Hadamard transform* H_n *on the following states where* $x, y \in \{0,1\}^n$*: (a)* $\frac{1}{\sqrt{2}}(|x\rangle + |y\rangle)$*; (b)* $\frac{1}{\sqrt{2}}(|x\rangle - |y\rangle)$.

Exercise 2.3.20 *Design a quantum circuit transforming the state* $\alpha|0\rangle + \beta|1\rangle$ *into the state: (a)* $-\beta|0\rangle + \alpha|1\rangle$*; (b)* $\beta|0\rangle - \alpha|1\rangle$*; (c)* $\alpha|0\rangle - \beta|1\rangle$*; (d)* $-\alpha|0\rangle + \beta|1\rangle$.

XOR gate is also an example of so-called **measurements gates**. It is called so because if the value of the control input is 0, then the control qubit is in effect an apparatus to perform an accurate measurement of the target qubit.

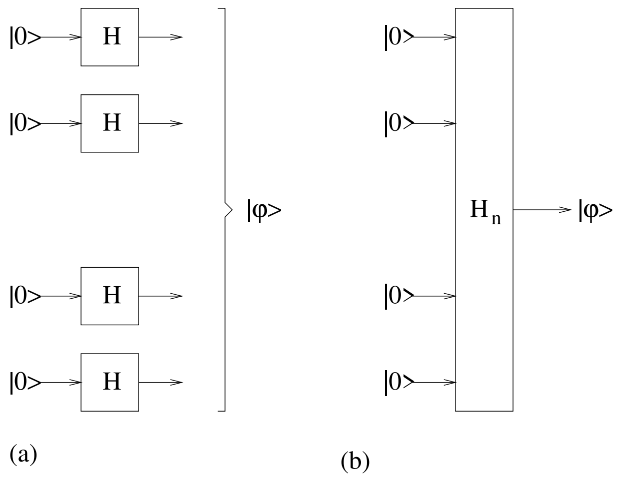


Figure 2.16: The Hadamard circuit H_n and its application to the state $|0^{(n)}\rangle$ with the outcome $|\phi\rangle = \frac{1}{\sqrt{2^n}}\sum_{i=0}^{2^n-1}|i\rangle$.

Exercise 2.3.21 *Show that the operation AND : $(x, y) \rightarrow (x, y, x \wedge y)$ can be implemented: (a) up to a phase by a quantum circuit consisting of four Hadamard gates and three XOR gates; (b) by a quantum circuit consisting of six XOR gates and eight one-bit gates.*

Exercise 2.3.22 *(Barenco et al. 1997) Design a quantum circuit to realize the unitary transformation $U_k : |0^{(k)}\rangle \rightarrow \frac{1}{\sqrt{k}}(|0^{(k)}\rangle + \sum_{i=1}^{k-1}|2^i\rangle)$. (Hint: use gates corresponding to the following unitary matrices: $A_k = \frac{1}{\sqrt{k+1}}\begin{pmatrix} 1 & -\sqrt{k} \\ \sqrt{k} & 1 \end{pmatrix}$ and*

$$T_{j,k} = \frac{1}{\sqrt{k-j+1}}\begin{pmatrix} \sqrt{k-j+1} & 0 & 0 & 0 \\ 0 & 1 & \sqrt{k-j} & 0 \\ 0 & -\sqrt{k-j} & 1 & 0 \\ 0 & 0 & 0 & \sqrt{k-j+1} \end{pmatrix}$$

for $1 \leq j < k$.)

The computational meaning of quantum circuits is defined as follows. For any quantum circuit C with input variables $x_1, \ldots, x_n$ and output variables $y_1, \ldots, y_m$, $m \leq n$ (they are to be a subset of outputs), we associate to any input $x \in \{0,1\}^n$ the probability distribution ρ_x over $\{0,1\}^m$ defined in the following way (see Yao, 1993).

For any input x the final quantum state v, corresponding to all output wires, not only to those carrying output variables, has the form

$$v = \sum_{y \in \{0,1\}^m} \alpha_y |y\rangle,$$

where α_y is the amplitude obtained by the projection of v when the output variables are set to the value y, i.e. α_y is the sum of amplitudes of these final outcomes having value y in the wires corresponding to output variables. Then $\pi_x(y) = |\alpha_y|^2$ is the corresponding probability and $\{\pi_x \mid x \in \{0,1\}^n\}$ is said to be the distribution generated by the circuit C.

One of the main results of classical computation theory says that each TM can be simulated in polynomial time by a uniform family of Boolean circuits. A similar result holds also for quantum Turing machines and quantum circuits, and due to this result one uses nowadays more quantum circuits than QTM to develop quantum algorithms. To present the corresponding result we need at first to introduce the concept of simulation of QTM by quantum circuit.

Definition 2.3.23 *A quantum circuit C with n input variables is said to (n,t)-simulate a QTM M, if for each $x \in \{0,1\}^n$ the probability distribution ρ_x generated by C is identical to the distribution of the configurations of M after t steps with x as input.*

Yao (1993) has shown the following result.

Theorem 2.3.24 *If M is a QTM and $n, t \in \mathbf{N}^{\geq 0}$, then there exists a quantum Boolean circuit C of size (number of gates) polynomial in n and t that (n,t)-simulates M.*

A **quantum program** is a quantum circuit together with some measurement (observation) gates/processes attached to some of the wires. There is a certain controversy about whether measurements can appear inside and not only at the end of quantum circuits. Arguments in favour are based on the observation that there are no physical reasons for not doing that. Arguments against are based on the point of view that measurement mostly changes a pure state into a mixed one.

Remark 2.3.25 Quantum gates are far from easy to implement. It seems safe to say that the potential computational power of quantum computers does not come from the intrinsic speed of quantum gates, but from the fact that quantum circuits to solve some problems can have exponentially less gates than their classical counterparts.

The task of designing efficient quantum algorithms can be seen as unitary matrix factorization problem: given a universal set $\mathcal{U}$ of basic unitary matrices, and an $n \times n$ unitary matrix U, how to decompose U into a product of *poly*$(\lg n)$ matrices from $\mathcal{U}$. Quantum programming—quantum algorithms and networks design and analysis—requires therefore quite a different expertise (see, for example Høyer, 1997). Development of efficient factorization methods and proper sets of basic unitary matrices is still task to deal with. In the next chapter we deal with quantum algorithm design problems in a more traditional spirit—as an art of composition of unitary transformation to perform the final unitary transformation.

2.3.2 Measurement gates

Measurement gates are not only to magnify results of quantum evolution to provide its outcomes to the classical world. They can be used also to influence, in an essential way, the whole process of quantum computation. Actually, some of the key results of the next chapter, fast quantum algorithms, are based on doing clever measurements in proper times.

Example 2.3.26 *Consider the two quantum circuits depicted in Figure 2.17. The first one consists at first of two Hadamard gates H_n and ends with the measurement gate, with respect*

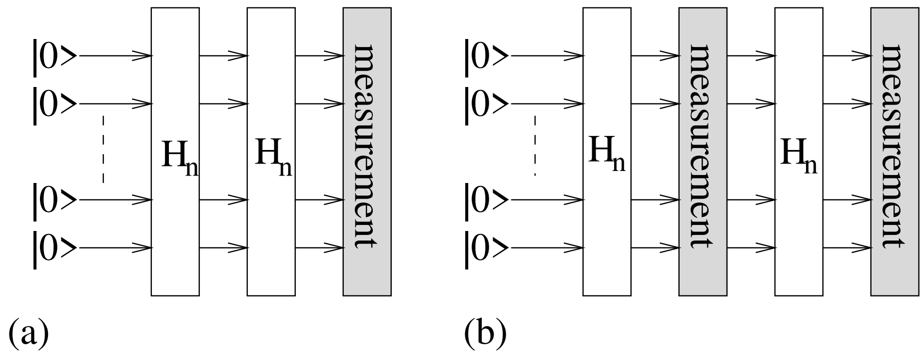


Figure 2.17: Measurement gates and their role

to the standard observable. The second circuit has in addition a measurement gate also in between two Hadamard gates.

There is an essential difference between these two circuits, and the inclusion of the measurement gate also between two Hadamard gates makes the second circuit more interesting.

Indeed, since $H_n^2 = 1$, the result of the measurement gate of the circuit in Figure 2.17a is $|0^n\rangle$ with probability 1. On the other hand, in the second circuit, the first measurement gate will observe a random n-bit string x with probability $\frac{1}{2^n}$. The output of the second gate is then $H_n|x\rangle = \frac{1}{\sqrt{2^n}}\sum_{y\in\{0,1\}^n}(-1)^{y\cdot x}|y\rangle$. The second measurement gate therefore observes any n-bit string with probability $\frac{1}{2^n}$.

Generation of a random string, as provided by the gates in Figure 2.17b, is often needed in quantum computation. It is therefore natural to ask whether the inclusion of an intermediate measurement is the only (easy) way of doing that or whether there is a simple way to avoid such intermediate measurements. As we shall see, there is—the so-called copying technique, using XOR gates. This will be illustrated in the following simple example using only H gates but the technique can be used in general. This techniques will play an important role in Sections 7.4 and 7.5.

Example 2.3.27 *As discussed in the previous example, the network in Figure 2.18a can be seen as generating a random bit. However, the same can do networks in Figure 2.18b,c.*

In the circuit in Figure 2.18b, after the XOR gate the state is $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ and the first bit is random. After the next application of the Hadamard gate on the first bit we get

$$(H \otimes I)(\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = \frac{1}{2}(|00\rangle + |10\rangle + |01\rangle - |11\rangle)$$

and the first bit is perfectly random.

In the circuit in Figure 2.18c, after the XOR gate we find the state $\frac{1}{\sqrt{2}}(|00\rangle+|11\rangle)$. After the application of Hadamard gates on both bits we get

$$(H \otimes H)(\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)) = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

and the first bit is again perfectly random.

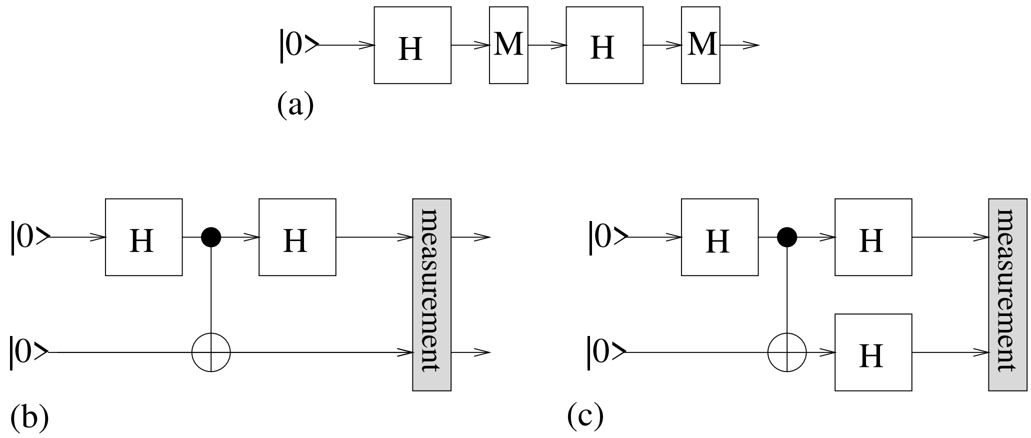


Figure 2.18: Power of copying circuits

2.3.3 Universality of quantum gates

Since quantum operations can be seen as continuous, the concept of universality for a set of quantum gates is naturally a bit more complex than in the classical case. As a consequence, the maximum we can hope for is that any unitary transformation can be approximated by a transformation realized by a circuit consisting of gates from a fixed set of gates. The concept of an ε-approximation of a unitary matrix U by another matrix M is therefore of importance. Approximations of unitary transformations are defined as follows:

Definition 2.3.28 *A matrix M is ε-close to a unitary matrix U, or it is an ε-approximation U, if $||U - M|| \leq \varepsilon$.*

It is easy to see that if a matrix M is ε-close to a unitary matrix, then $1-\varepsilon \leq ||M|| \leq 1+\varepsilon$ and its rows, M_i, have the norm close to 1 and, in addition, they are almost orthogonal.

Exercise 2.3.29 *(Bernstein and Vazirani, 1997.) Show that if a matrix M of degree d is ε-close to a unitary matrix, then the following holds for its rows M_i: (a) $1 - \varepsilon \leq ||M_i|| \leq 1 + \varepsilon$; (b) if $i \neq j$, then $||M_i M_j^*|| \leq 2\varepsilon + 3\varepsilon^2$.*

Definition 2.3.30 *A set of quantum gates is* **universal** *if any unitary transformation U on any qubit register can be performed, with arbitrary precision $\varepsilon > 0$, by a quantum circuit $C_{U,\varepsilon}$, consisting of the gates from that set. (In other words, the unitary matrix defined by $C_{U,\varepsilon}$ is ε-close to U.) A quantum gate is* **universal** *if by itself it forms a universal set when supported by constant inputs $|0\rangle$ and $|1\rangle$*[27].

[27] One also says that a set of gates is universal if the subgroup of unitary transformations generated by the unitary transformations corresponding to the gates in the set is dense in the group of all unitary transformations $U(n)$, for any n.

Exercise 2.3.31 *Let ϕ_0 be an irrational fraction of π. (a) Show that any gate*

$$U_\phi = \begin{pmatrix} e^{i\phi} & 0 \\ 0 & 1 \end{pmatrix}$$

can be implemented with an arbitrary precision using (several copies of) a single gate U_{ϕ_0}; (b) show that the number of gates required to achieve ε accuracy of approximation is $\mathcal{O}(\frac{1}{\varepsilon})$.

The first result concerning the universality of quantum gates was due to Deutsch (1989), who has shown that any 3-qubit gate with the unitary matrix

$$D(\theta) = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & i\cos\theta & \sin\theta \\ 0 & 0 & \sin\theta & i\cos\theta \end{matrix} \end{pmatrix},$$

where $\mathbf{0}$ and $\mathbf{1}$ are zero and unit matrices of degree 4, depicted in short in Figure 2.19a, with $D = R_y(\theta)$ (see page 82), is universal provided θ/π is irrational. Observe that the Deutsch gate can be seen as a quantum generalization of the Toffoli gate.

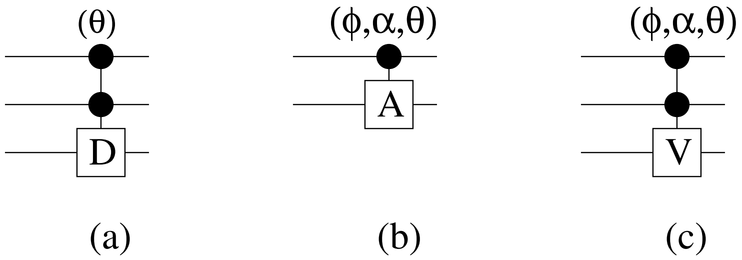


Figure 2.19: Short notation for gates $D(\theta)$, $A(\phi,\alpha,\theta)$ and $V(\phi,\alpha,\theta)$

The following example presents the first step which led to a universal 2-qubit gate.

Example 2.3.32 (Sleator and Weinfurter, 1995) *Consider the gates realizing the operators specified by the matrices:*

$$S(\tau) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & e^{\frac{i\pi}{4}}\cos\frac{\pi\tau}{2} & e^{\frac{-i\pi}{4}}\sin\frac{\pi\tau}{2} \\ 0 & 0 & e^{\frac{-i\pi}{4}}\sin\frac{\pi\tau}{2} & e^{\frac{i\pi}{4}}\cos\frac{\pi\tau}{2} \end{pmatrix}, S^{-1}(\tau) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & e^{\frac{-i\pi}{4}}\cos\frac{\pi\tau}{2} & e^{\frac{i\pi}{4}}\sin\frac{\pi\tau}{2} \\ 0 & 0 & e^{\frac{i\pi}{4}}\sin\frac{\pi\tau}{2} & e^{\frac{-i\pi}{4}}\cos\frac{\pi\tau}{2} \end{pmatrix}$$

Both gates perform transformations of the target bit conditional to the case that the control bit is 1. Clearly $S(\tau)S^{-1}(\tau) = I$. In addition, it holds

$$S^2(\frac{\tau}{2}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & i\cos\pi\frac{\tau}{2} & \sin\pi\frac{\tau}{2} \\ 0 & 0 & \sin\pi\frac{\tau}{2} & i\cos\pi\frac{\tau}{2} \end{pmatrix},$$

and therefore $S^2(\frac{1}{2}) = XOR$.[28] *Since $a \wedge b = \frac{1}{2}(a + b - a \oplus b)$, it is easy to verify that the circuit in Figure 2.20a is an implementation of a generalization of the Toffoli gate provided that the schemes in 2.20b denote the gates for the matrices $S(\frac{\tau}{2})$ and $S^{-1}(\frac{\tau}{2})$. The gate $S(\tau)$ has been constructed by Sleator and Weinfurter as a universal gate provided τ is an irrational multiple of π.*

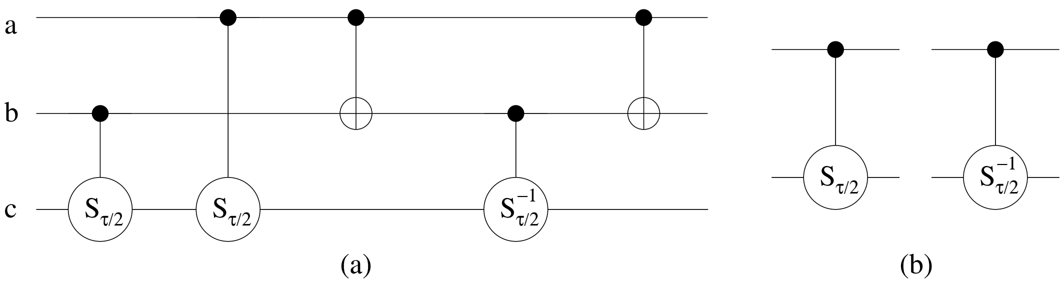


Figure 2.20: A circuit for the Toffoli gate

We show now a modified result, adopted from Barenco (1995), namely that any 2-input/output gate with the matrix

$$A(\phi, \alpha, \theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & e^{i\alpha}\cos\theta & -ie^{i(\alpha-\phi)}\sin\theta \\ 0 & 0 & -ie^{i(\alpha+\phi)}\sin\theta & e^{i\alpha}\cos\theta \end{pmatrix},$$

and with the short notation for the gate in Figure 2.19b, is universal if α, ϕ and θ are irrational multiples of π and of each other.

The proof is by providing an explicit construction of the circuit implementing the gate $D(\theta)$ via a 3-qubit gate specified by the matrix

$$V(\phi, \alpha, \theta) = \begin{pmatrix} \mathbf{1} & & & \mathbf{0} & \\ & 1 & 0 & 0 & 0 \\ \mathbf{0} & 0 & 1 & 0 & 0 \\ & 0 & 0 & e^{i\alpha}\cos\theta & -ie^{i(\alpha-\phi)}\sin\theta \\ & 0 & 0 & -ie^{i(\alpha+\phi)}\sin\theta & e^{i\alpha}\cos\theta \end{pmatrix},$$

with the short notation for the gate in Figure 2.19c.

Let us first observe that for any integer n

$$A^n(\phi, \alpha, \theta) = A(\phi, n\alpha \bmod 2\pi, n\theta \bmod 2\pi).$$

Since α and θ are irrational in the above sense, any transformation $A(\phi, \alpha_1, \theta_1)$, with $\alpha_1, \theta_1 \in [0, 2\pi]$, can be implemented with arbitrary precision using several gates A.

Exercise 2.3.33 *Show that if α_1 and θ_1 are specified with accuracy $\pm\varepsilon$, then $\mathcal{O}(\frac{1}{\varepsilon^2})$ applications of the gate A are needed to approximate $A(\phi, \alpha_1, \theta_1)$ with a given precision ε.*

[28] $S(\frac{1}{2})$ is therefore called "a square root of XOR".

Let us call "repertoire of A" the set of unitary transformations that can be approximated with an arbitrary precision by networks composed of A-gates only. In this repertoire there is clearly the inverse of the transformation A because

$$A^{-1}(\phi, \alpha, \theta) = A(\phi, 2\pi - \alpha, 2\pi - \theta).$$

Let A_{ij}, $1 \leq i \neq j \leq 3$, denote the two-qubit gate obtained from A by having the ith qubit to be its control qubit and the jth qubit to be its target qubit. All such gates are clearly in the repertoire of A. It is now easy to verify that the network from Figure 2.21, described by the matrix

$$A_{23}(\phi, \frac{\alpha}{2}, \frac{\theta}{2})A_{13}(\phi, \frac{\alpha}{2}, \frac{\theta}{2})A_{12}(\phi, \frac{\pi}{2}, \frac{\pi}{2})A_{23}^{-1}(\phi, \frac{\alpha}{2}, \frac{\theta}{2})A_{12}(\phi, \frac{\pi}{2}, \frac{\pi}{2})$$

implements the gate $V(\phi, \alpha, \theta)$—see Example 2.3.32.

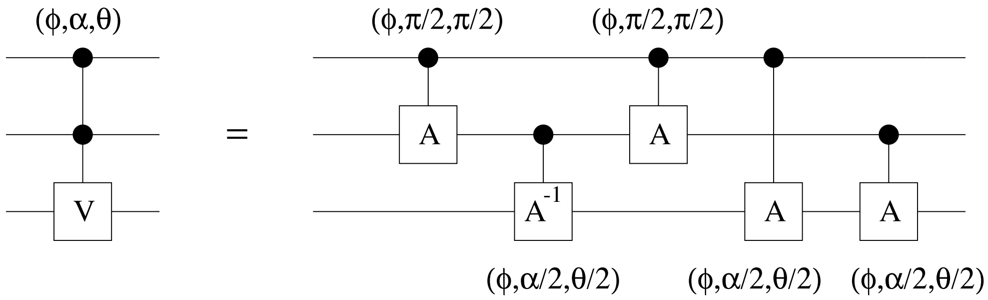


Figure 2.21: An implementation of the gate V -80%

Denote now by V' the gate obtained from V by exchanging the second and the third qubit, and let us denote by P the matrix (gate) $V'(\phi, \pi/2, \pi/2)$. In such a case

$$P = V'(\phi, \pi/2, \pi/2) = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{-i\phi} \\ 0 & 0 & 1 & 0 \\ 0 & e^{i\phi} & 0 & 0 \end{matrix} \end{pmatrix}.$$

In addition, let us denote

$$Q = V'(\phi, \frac{\pi}{2}, -\frac{\pi}{2})V(\phi, \frac{\pi}{2}, -\frac{\pi}{2})V'(\phi, \frac{\pi}{2}, -\frac{\pi}{2}) = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{matrix} \end{pmatrix}$$

and

$$T(\phi, \beta) = Q[V(\phi, 0, \beta)P]^2[V(\phi, 0, -\beta)P]^2Q.$$

For very small β, we get

$$T(\phi, \beta) = \mathbf{1} + i\beta^2 \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & ie^{-i\phi} \\ 0 & 0 & -ie^{i\phi} & 0 \end{matrix} \end{pmatrix} + \mathcal{O}(\beta^3),$$

where $\mathcal{O}(\beta^3)$ denotes a matrix the norm of which is $\mathcal{O}(\beta^3)$. Hence the transformation

$$V(\phi - \frac{\pi}{2}, 0, \beta) = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\beta & e^{-i\phi}\sin\beta \\ 0 & 0 & e^{i\phi}\sin\beta & \cos\beta \end{matrix} \end{pmatrix} = \lim_{n\to\infty} T(\phi, \sqrt{\beta/n})^n$$

can also be performed with an arbitrary precision by networks with the gate A as the only gate, and the same is therefore true for the transformation (gate).

$$\begin{aligned} R_z(\beta) &= \lim_{n\to\infty}[V\left(\phi, 0, \sqrt{\frac{\beta}{2n}}\right) V\left(\phi - \frac{\pi}{2}, 0, \sqrt{\frac{\beta}{2n}}\right) V\left(\phi, 0, -\sqrt{\frac{\beta}{2n}}\right) V\left(\phi - \frac{\pi}{2}, 0, -\sqrt{\frac{\beta}{2}n}\right)]^n \\ &= \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & e^{i\beta} & 0 \\ 0 & 0 & 0 & e^{-i\beta} \end{matrix} \end{pmatrix}. \end{aligned}$$

The universal Deutsch gate can now be constructed as follows:

$$D(\phi) = R_z(\frac{\phi}{2}) V(\phi, \frac{\pi}{2}, \phi) R_z(-\frac{\phi}{2}).$$

An additional significant step in the search for universal two qubit gates has been made by Deutsch, Barenco and Ekert (1995) and Lloyd (1995). They have shown that almost any nontrivial two-qubit gate is universal.

Exercise 2.3.34 *(Barenco, 1996) Show that: (a) there is no one-qubit universal gate; (b) no classical gate can be universal for quantum computing; (c) gate $A(\frac{\pi}{2}, \frac{\pi}{4}, \theta)$ is universal under certain conditions.*

Open problem 2.3.35 *Determine the set of all non-universal quantum gates.*

The task of finding 2-qubit universal gates is of theoretical and also practical importance. However, this is not the only way to go in searching for simple gates that can be used to design a quantum circuit.

As already mentioned, XOR gate is not universal for quantum computing. In spite of that it has its firm role in the search for universality in quantum computing.

Barenco et al. (1995) have shown that the XOR gate, when supplemented by a set of the following one-qubit gates (that perform general rotation of single qubits), is sufficient to implement any unitary transformation

$$\begin{pmatrix} e^{i(\delta+\frac{\alpha}{2}+\frac{\beta}{2})}\cos\frac{\theta}{2} & e^{i(\delta+\frac{\alpha}{2}-\frac{\beta}{2})}\sin\frac{\theta}{2} \\ -e^{i(\delta-\frac{\alpha}{2}+\frac{\beta}{2})}\sin\frac{\theta}{2} & e^{i(\delta-\frac{\alpha}{2}-\frac{\beta}{2})}\cos\frac{\theta}{2} \end{pmatrix}$$

On the base of the above results we see that while one- and two-bit operations are **classical computation primitives**, one- and two-qubit unitary operations are **quantum computation primitives**.

XOR gate has also been shown to be an important component of various decompositions of 3-qubit gates into 2-input/output gates. All these results make the XOR gate of central importance for quantum computation.

In addition, it can be shown that XOR gate and a single one-qubit gate form a universal set of gates (see Section 5.1).

Remark 2.3.36 Current quantum computing algorithms use qubits and quantum registers as the basic building blocks. This implies that current quantum mechanics systems used for quantum computing are composed of two-state quantum systems. Theoretically, we could use as basic quantum systems three or more states quantum systems. This would be a generalization of old attempts to use 3-valued or multi-valued logic for classical computing. In the classical case it has not been demonstrated that such a generalization brings some essential advantages. The quantum case has not yet been investigated and the situation may be quite different. Of interest is also the non-trivial question of universal quantum gates for 3-valued logic.

Remark 2.3.37 Presentation of quantum computations as an evolution of quantum states as presented usually in quantum computing, including this book, can be seen as corresponding to the Schrödinger view of quantum mechanics, see Section 9.1.1. An alternative approach, more corresponding to Heisenberger's matrix view of quantum mechanics, seeing a quantum computation as an evolution of operators in time, has been illustrated by Gottesman (1998) as useful in understanding an important class of quantum operators.

2.3.4 Arithmetical circuits

An efficient realization of networks for basic arithmetical and modular arithmetical operations (see Vedral, Barenco and Ekert, 1996), is of large importance for efficient implementation of various quantum algorithms, for example of those discussed in Section 3.2.

The two very basic problems are the design of networks for addition $(a, b) \to (a, a + b)$ and for modular addition $(a, b) \to (a, (a + b) \bmod N)$ of n-bit numbers a and b with $N = 2^n$.

Using an easy-to-design carry-gate (Figure 2.22a) and the summation gate (Figure 2.22b), the design of the adder is a straightforward application of the idea from Figure 1.17.

In the first phase all carry bits are computed using n carry gates. In the second phase at first $a_{n-1} \oplus b_{n-1}$ is transferred to b_{n-1}, using one XOR gate, and then, using $n - 1$ reverse carry gates and n summation gates, all $(a + b)_i$ bits are computed and all carry bits are reset to 0 (to remove all garbage), for $i = n - 2, \ldots, 0$.

Exercise 2.3.38 *(a) Design the carry gate from Figure 2.22a, using one XOR and one Toffoli gate; (b) design the summation gate from Figure 2.22b, using two XOR gates.*

Figure 2.23a shows a schematic notation for a binary adder. By reversing the order in which gates of an adder are applied we get a network, schematically shown in Figure 2.23b, to compute an ordinary subtraction $(a, b) \to (a, b - a)$ when $b \geq a$, and a "modulo 2^n subtraction" $(a, b) \to (a, 2^n + (b - a))$ if $a > b$. In the last case the most significant bit of the second register contains always 1. This will be essentially used in the next construction of the modular adder.

A modular adder for $(a + b) \bmod N$ is shown in Figure 2.24. The basic idea is very simple. Adder A1 provides the outputs $(a, a + b)$ and the subtractor S1 produces $a + b - N$ for the

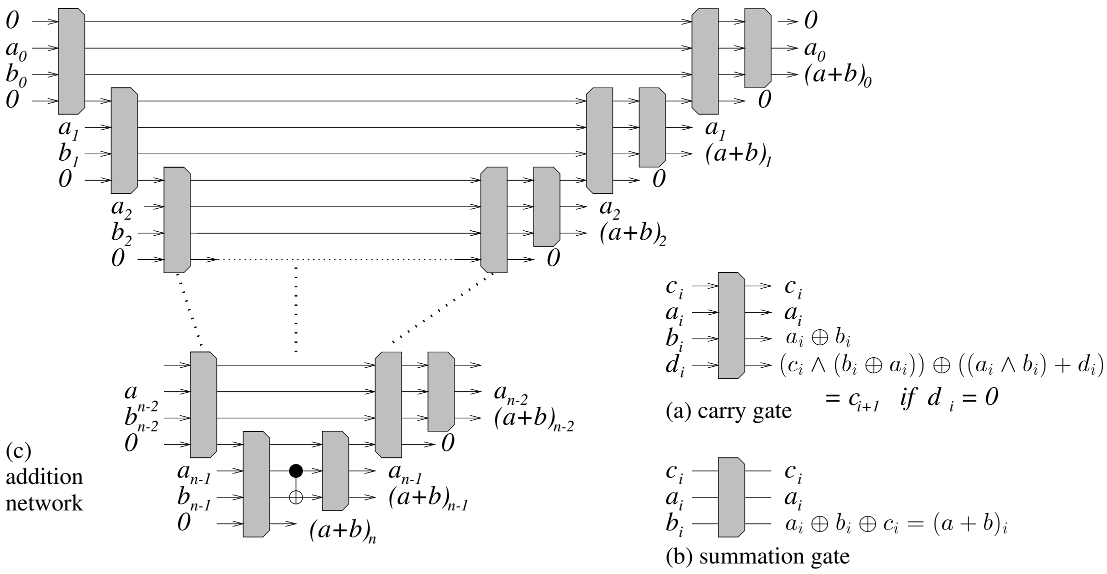


Figure 2.22: Quantum network for binary addition

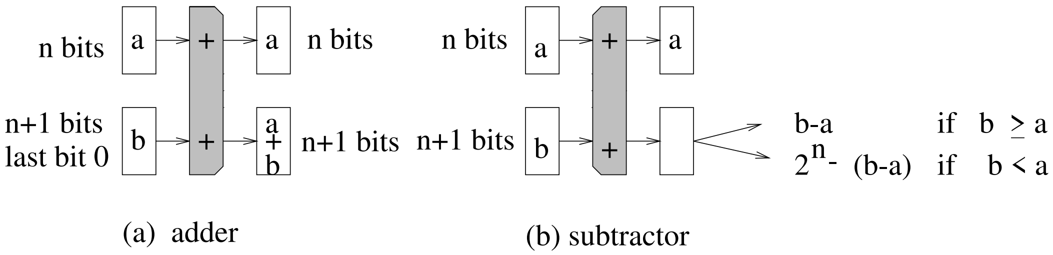


Figure 2.23: Gates for adders and subtractors

case $a + b \geq N$, which is exactly $(a + b) \bmod N$. However, in the case $a + b < N$ there is an overflow after S1 and the rest of the network is needed to handle this case. Overflow is first copied into the additional bit t set initially to 0, using the first XOR gate. Conditionally on that the gate G1 produces either N or 0. The second output of the adder A2 is therefore $(a + b) \bmod N$. The gate G2 resets the first register to N, if this is needed. The subtractor S2 then produces outputs $(a, (a + b) \bmod N - a)$. The second register has then the overflow exactly when S1 had and this is then used, with the help of the second XOR gate, to reset t to the value 0. The last adder is just to produce the output $(a, (a + b) \bmod N)$.

There are several natural ways to build networks for (modular) multiplication, and (modular) exponentiation on the base of networks for addition and modular addition. In the case of $f_{a,N}(b) = a \cdot b \bmod N$ we can use the decomposition of $f_{a,N}$:

$$\underbrace{(b + \ldots + b)}_{(a-1)-\text{plus operations}} \mod N \quad \text{or} \quad (\ldots \underbrace{((b + b) \bmod N) + b) \bmod N) + \ldots + b) \bmod N)}_{(a-1)-\text{plus operations}}$$

or as

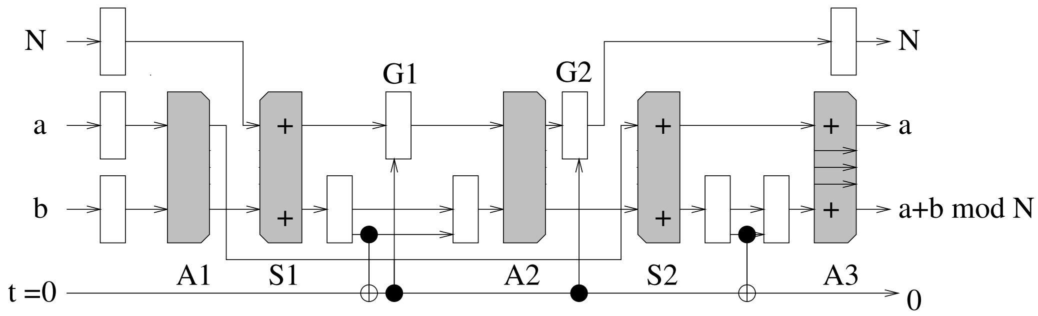


Figure 2.24: A quantum network for modular addition-75%

$$ab = (\sum_{i=0}^{n-1} a2^i b_i) \bmod N,$$

where $b = \sum_{i=0}^{n-1} b_i 2^i$, i.e. as $(n-1)$-additions, where in the ith addition $a2^i$ is added if and only if $b_i = 1$.

Exercise 2.3.39 *(a) Design a quantum network for modular multiplication; (b) design a quantum network for modular exponentiation.*

Similarly, $g_{a,N}(b) = a^b \bmod N$ can be computed using the decomposition

$$a^b = \prod_{i=0}^{n-1} (a^{2^i})^{b_i}.$$

Concerning the efficiency of networks for (modular) arithmetical operations two quantitative measures are of importance: the total number of elementary gates and the total number of qubits needed. The problem has received special attention because in Shor's algorithm efficiency of exponentiation $a^x \bmod n$ is of key importance for the overall efficiency. By Vedral, Barenco and Ekert (1996), $4n + 3$ qubits are sufficient for exponentiation in Shor's factorization algorithm, where n is the number of qubits needed to store N, $2n$ is the number of qubits needed to store x (because x can be there as large as N^2), and $n + 3$ temporary qubits are sufficient.

2.3.5 Quantum superoperator circuits

In spite of the fact that the model of quantum circuits from page 83 is fundamental for processing pure quantum states of an isolated quantum system, circuits processing density matrices with superoperators defined below (see Aharonov et al. 1998), are also of importance. They better correspond to computations in noisy environments.

Definition 2.3.40 *A superoperator gate G of type (k, l) is a completely positive map which maps density matrices on k qubits to density matrices on l qubits. Its action on a density matrix ρ will be denoted symbolically by $G \circ \rho$.*

Two important special cases of superoperator gates are unitary and measurement gates. In the case of a unitary gate U the corresponding operator is $U \cdot U^*$. For a pure state $|\phi\rangle$, U maps $\rho = |\phi\rangle\langle\phi|$ into $U \circ \rho = U\rho U^*$. A measurement gate represents a probabilistic projection into a set of mutually orthogonal subspaces, which produces a mixed state. Superoperators are in general not reversible.

There is a well-known and well-understood relation between superoperators and unitary operators.

Lemma 2.3.41 *The following conditions are equivalent for any two Hilbert spaces H_n and H_m and sets of linear operators $\mathcal{L}(H_n)$ and $\mathcal{L}(H_m)$.*

1. *A superoperator $T : \mathcal{L}(H_n) \to \mathcal{L}(H_m)$ is trace-preserving and completely positive.*

2. *There is a Hilbert space F with $\dim(F) \leq \dim(H_n)\dim(H_m)$, and a unitary embedding $E : H_n \to H_n \otimes F$ such that $T\rho = Tr_F(E\rho E^*)$ for all $\rho \in \mathcal{L}(H_n)$.*

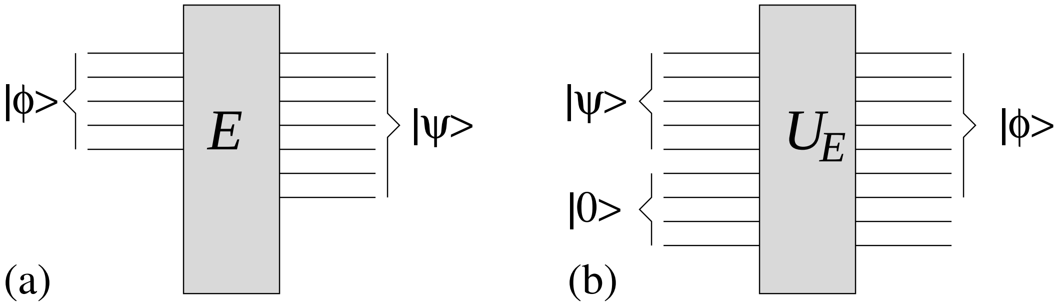


Figure 2.25: Encoder as superoperator and its unitary embedding

Example 2.3.42 *Superoperators of special interest for quantum information processing are encoders and decoders. An encoder E that maps n-qubits to m-qubits, $m \geq n$ (see Figure 2.25a), can be seen in a larger quantum space, extended by $k \geq m - n$ qubits in the initial state $0^{(k)}$, the so called* **ancilla** *qubits, as a unitary operator U_E, Figure 2.25b, some of the outputs of which are then discarded.*

Actually, any operation on quantum data can be represented as a superoperator, equivalent to a unitary interaction with ancilla qubits, followed by discarding some qubits (see Figure 2.25b).

Superoperators of any Hilbert space H can be naturally extended to superoperators operating on a composed Hilbert space $H \otimes H'$ by taking tensor product with the unit matrix of H'. This way any $T : \mathcal{L}(H) \to \mathcal{L}(H')$ can be extended to $T \otimes I : \mathcal{L}(H \otimes H') \to \mathcal{L}(H' \otimes H')$. This fact will be used below when the computation on superoperator circuits is defined.

Definition 2.3.43 *A quantum superoperator circuit over a set of superoperators $\mathcal{G}$ is a directed acyclic graph to whose nodes are assigned superoperators from $\mathcal{G}$ in such a way that if a node has indegree (outdegree) k (l), then the corresponding superoperator is of the type (k, l). A subset of inputs is labeled "blank" and a subset of outputs is labeled "results".*

Each superoperator circuit produces, for a given input density matrix, an output density matrix which is defined in the natural way as follows. If Q is a quantum superoperator circuit and $G_1, \ldots, G_t$ is a topological sort of its gates, then Q computes the density matrix

$$Q \circ \rho = G_t \circ G_{t-1} \circ \ldots \circ G_1 \circ \rho.$$

In order to show that this definition is consistent one has to show that the two different topological orderings of gates yield the same result. A step in this direction is to show that if G_1 and G_2 are superoperator gates operating on different qubits, then $G_1 \circ G_2 \circ \rho = G_2 \circ G_1 \circ \rho$, for any density matrix ρ. This can be shown easily if we consider an extension of superoperators by tensoring with unitary matrices as discussed above. The probability distribution that such a circuit computes is defined in the following way:

Definition 2.3.44 *Let Q be a quantum circuit with n inputs (blanks) and m outputs (results). The probability distribution $f_Q : \{0,1\}^n \to [0,1]^{\{0,1\}^m}$ that Q computes is defined as follows: For an input i the probability of the output j is*

$$f_Q(i)(j) = \langle j|(Q \circ |i\rangle\langle j|)_A|j\rangle,$$

where A is the set of "outputs".

As a quite straightforward corollary of Lemma 2.3.41 we have the main result concerning the computational power of superoperator gates.

Lemma 2.3.45 *If $G : \mathcal{L}(H_{2^n}) \to \mathcal{L}(H_{2^n})$ is a superoperator gate of type (n, m), then there exists a unitary quantum gate U_g on $2n + m$ qubits such that for any density matrix of the order n,*

$$G \circ \rho = (U_g \circ (\rho \otimes |0^{(n+m)}\rangle\langle 0^{(n+m)}|))|_A,$$

where A is the set of the first n qubits.

Proof. Let $G : \mathcal{L}(H_{2^n}) \to \mathcal{L}(H_{2^m})$. By Lemma 2.3.41, G has a representation of the form $G = Tr_{H_{2^{n+m}}}(V \cdot V^*)$, where $V : H_{2^n} \to H_{2^m} \otimes H_{2^{n+m}}$ is a unitary embedding. Let $\{|\phi_{ij}\rangle \,|\, 0 \le i < 2^n, 0 \le j < 2^{n+m}\}$ be an orthonormal basis on $H_{2^n} \otimes H_{2^{n+m}}$ such that $|\phi_{i,0^{n+m}}\rangle = V|i\rangle$, for any i and other basis vectors are arbitrary. Clearly, there is a unitary operator $U : |i,j\rangle \to |\phi_{ij}\rangle$ such that $V = UV_0$, where $V_0 : |x\rangle \to |x\rangle \otimes |0^{n+m}\rangle$. This implies

$$G = Tr_{H_{2^{n+m}}}(UV_0 \cdot V_0^* U^*).$$

As a corollary we get

Theorem 2.3.46 *The model of quantum circuits with mixed states is polynomially equivalent, in computational power, to the standard model of quantum circuits over pure states.*

Chapter 3

ALGORITHMS

INTRODUCTION

Quantum algorithms make use of several specific features of the quantum world, for example quantum superposition, to get from classical inputs, through entangled states, to classical outputs more efficiently than classical algorithms. A variety of quantum algorithms are presented in this chapter. They range from pioneering algorithms, simple but powerful, for several promise problems, through seminal Shor's algorithms and a variety of algorithms for various search problems and their modifications, due to Grover and others.

Design of faster-than-classical quantum algorithms for important algorithmic problems has been an interesting intellectual adventure and achievement. Their existence keeps being one of the key stimuli to those trying to overcome enormous technology problems to build (powerful) quantum computers.

Methods to design quantum algorithms and to show limitations of quantum power have also been developed gradually and will be presented and illustrated in this chapter.

LEARNING OBJECTIVES

The aim of the chapter is to learn:

1. the power of quantum superpositions, parallelism and entanglement;
2. the efficient quantum algorithms for several basic promise problems;
3. the quantum Fourier transform and its properties and implementation;
4. Shor's quantum algorithms for integer factorization and discrete logarithm computation;
5. the hidden subgroup problems and their role in quantum computing;
6. a variety of search algorithms due to Grover and others;
7. methods to design efficient quantum algorithms;
8. methods to show lower bounds and limitations of quantum computing.

> An important scientific innovation rarely makes its way by gradually winning over and converting its opponents: it rarely happens that Saul becomes Paul. What does happen is that its opponents gradually die out and that the growing generation is familiarized with the idea from the beginning.
>
> Max Planck (1936)

Quantum algorithms are methods using quantum networks and processors to solve algorithmic problems. On a more technical and prosaic level, a design of a quantum algorithm can be seen as a process of an efficient decomposition of a complex unitary transformation into products of elementary unitary operations (or gates), performing simple local changes.

Quite a few classical algorithms for practically important algorithmic problems can be asymptotically speeded-up by replacing its computationally most intensive subalgorithms by quantum algorithms, with the same classical input and output, that perform asymptotically less (quantum) operations—or quantum gates when performed by quantum circuits.

For a variety of basic computational problems quantum algorithms have already been constructed that make a clever use of various quantum phenomena to solve algorithmic problems more efficiently than classical algorithms. Quantum "tricks" of the first quantum algorithms have been gradually developed into methods of utilizing the power of quantum evolution, especially of such quantum phenomena as quantum superposition, parallelism and entanglement. Programming of quantum computers requires the use of a variety of innovative techniques. Two basic techniques are: extraction of a global property of a function to be computed and amplitude amplification to make desirable results most probable. It is not yet known whether the power of quantum parallelism can be really harnessed for a wide class of important applications. However the results obtained so far are interesting, stimulating and important.

In the area of quantum algorithms (circuits) the research has been concentrated in the following four topics.

1. The design of quantum algorithms (circuits), faster than classical ones, for important computational problems of classical computing.

2. Abstraction of general methods to design efficient quantum algorithms. They fall into two main categories: methods to mix amplitudes carefully (for example quantum Fourier transform); and methods to manipulate carefully chosen amplitudes, especially the amplitude amplification methods. Especially the first method is very important.

3. The development of special tools, inherently quantum, of designing efficient quantum algorithms, for example efficient quantum transformations (see Høyer, 1997, and Hogg et al. 1998).

4. Experimental realization of simple quantum algorithms.

3.1 Quantum Parallelism and Simple Algorithms

Quantum evolution is performed by linear (unitary) operators. An application of an operator A to a state

$$|\phi\rangle = \sum_{i=0}^{2^n-1} c_i|i\rangle \tag{3.1}$$

yields

$$A|\phi\rangle = \sum_{i=0}^{2^n-1} c_i A|i\rangle,$$

i.e. by a single application of the operator A (on a "single processor"), exponentially many, namely 2^n, operations on basis states are performed. This phenomenon is called **quantum parallelism** and it is of great importance for the design of efficient quantum algorithms. Observe that quantum parallelism is already for a modest n a really massive parallelism.

Quantum computing can therefore trade exponentiality in time for exponentiality in quantum interference. In addition, in quantum registers the amount of parallelism increases exponentially with the size of the system, and this exponential growth of parallelism requires only a linear increase in the amount of physical space needed.[1]

Example 3.1.1 *Let $f : \{0, 1, \ldots, 2^m - 1\} \rightarrow \{0, 1, \ldots, 2^n - 1\}$ be a function. The mapping $X_f : a \rightarrow (a, f(a))$ is one-to-one and therefore, as discussed in Section 1.7, there is a unitary transformation U_f such that if we take an $(n + m)$-qubit register, composed of two subregisters—an m-qubit register for $x \in \{0, 1, \ldots, 2^m - 1\}$ and an n-qubit register for $b \in \{0, 1, \ldots, 2^n - 1\}$, and binary representations for x, b and $f(x)$—then*

$$|x, b\rangle \xrightarrow{U_f} |x, b \oplus f(x)\rangle,$$

for any b. If now the operator U_f is applied to the uniform coherent superposition of all 2^m basis states $|\phi\rangle = \frac{1}{\sqrt{2^m}} \sum_{x=0}^{2^m-1} |x\rangle$ we get, in one computational step,

$$|\phi\rangle = \frac{1}{\sqrt{2^m}} \sum_{x=0}^{2^m-1} |x, 0\rangle \xrightarrow{U_f} \frac{1}{\sqrt{2^m}} \sum_{x=0}^{2^m-1} |x, f(x)\rangle = U_f|\phi\rangle = |\psi\rangle \tag{3.2}$$

and therefore a single application of U_f computes all values $f(i)$, $0 \leq i < 2^m$.

Another often used unitary operator related to functions $f : \{0, \ldots, 2^N - 1\} \rightarrow \{0, 1\}$ is the operator

$$V_f|x\rangle \rightarrow (-1)^{f(x)}|x\rangle,$$

[1] This is behind a popular saying that quantum computers have memory "exponentially larger than their physical size".

which changes the sign of the amplitude for those basis states $|x\rangle$ for which $f(x) = 1$. Using one additional qubit, in the state $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$, the operator V_f can be expressed using the operator U_f as follows:

$$\begin{aligned} U_f|x, \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\rangle &= \frac{1}{\sqrt{2}}(|x, f(x)\rangle - |x, 1 \oplus f(x)\rangle) \\ &= (-1)^{f(x)}|x, \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\rangle. \end{aligned}$$

With the exception of some trivial cases, the resulting state $|\psi\rangle$ in (3.2) is entangled. Indeed, by measuring the first m qubits (or "x-register") with respect to the standard basis, we get a value x_0 randomly chosen from the set $\{0, 1, \ldots, 2^m - 1\}$ and the state collapses to $|x_0, f(x_0)\rangle$. The following measurement of the second register gives us then $f(x_0)$. However, this way quantum algorithms provide no advantage over the classical one. Fortunately, as illustrated in the following examples, in some cases there is a more clever way to make use of quantum entanglement in (3.2) to compute efficiently some global properties of f.

Example 3.1.2 *(van Dam, 1998) Let a function $f : \{1, \ldots, n\} \to \{0, 1\}$ be given as a black box. To determine f classically, n calls of f are needed—to get the string $w_f = f(1)f(2)\ldots f(n)$. Quantumly, this can be done, with probability greater than 0.95, using $\frac{n}{2} + \sqrt{n}$ quantum calls of f. Indeed, by (2.3)*

$$w_f = H_n \sum_{x \in \{0,1\}^n} (-1)^{x \cdot w_f}|x\rangle \tag{3.3}$$

In order to compute $x \cdot w_f$ one needs $hw(x)$ calls of f, where $hw(x)$ is the Hamming weight of x—the number of 1's in x.

The basic trick is to compute the sum in (3.3) but only for x such that $hw(x) \leq k$, for a suitable k.

If F_k is such a function that for $x \in \{0, 1\}^n$, $F_k(x) = x \cdot w_f$ if $hw(x) \leq k$ and $F_k(x) = 0$ otherwise, then

$$V_{F_k}|x\rangle = (-1)^{x \cdot w_f}|x\rangle,$$

if $hw(x) \leq k$ and $V_{F_k}|x\rangle = |x\rangle$, otherwise. Therefore if V_{F_k} is applied to the (initial) state

$$|\psi_k\rangle = \frac{1}{\sqrt{M_k}} \sum_{x \in \{0,1\}^n}^{hw(x) \leq k} |x\rangle,$$

where $M_k = \sum_{i=0}^{k} \binom{n}{k}$, then

$$|\psi'_k\rangle = V_{F_k}|\psi_k\rangle = \sum_{x \in \{0,1\}^n}^{hw(x) \leq k} (-1)^{x \cdot w_f}|x\rangle.$$

In order to compute $|\psi'_k\rangle$, at most k calls of f are needed. Let us now measure all n qubits of $|\psi''_k\rangle = H_n|\psi'_k\rangle$. The probability that this way we get w_f is

$$Pr(|\psi''_k\rangle \text{ yields } w_f) = |\langle w_f|\psi''_k\rangle| = \frac{M_k}{2^n} = \frac{1}{2^n} \sum_{i=1}^{k} \binom{n}{k}$$

and, as one can calculate, this probability is less then 0.95 if $k = n + \sqrt{n}$.

3.1.1 Deutsch's problem

The next algorithmic problem for which we design a quantum algorithm is, informally, a very simple problem of guessing whether a given coin is genuine (with head on one side and tail on the other) or fake (with both sides the same). The question is how many times we need to look at the coin to find out which case it is. In the classical world twice, to both sides. In the quantum world only once, but to a quantum superposition (of both sides).

Example 3.1.3 (Deutsch's XOR problem—original randomized solution) *Given a function $f : \{0,1\} \rightarrow \{0,1\}$, as a black box*[2]*, the task is to determine whether $f(0) \oplus f(1) = 0$ or 1 (i.e. whether f is constant or balanced).*

In classical computing, where the black box for f contains a classical circuit, we need two applications of f, to 0 and to 1, to solve the problem. The first quantum algorithm presented below to solve the problem assumes that f is given as a black box with a quantum circuit and uses only one application of f, but to a superposition of basis states, to solve the problem with probability $\frac{1}{2}$ in such a way that we know whether the answer obtained is correct.

Let U_f be the unitary mapping of $|x,y\rangle$ into $|x, y \oplus f(x)\rangle$—the so-called f-controlled NOT. One application of this transformation to the state $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)|0\rangle$ yields the state $\frac{1}{\sqrt{2}}(|0, f(0)\rangle + |1, f(1)\rangle)$, which can be written in the standard and dual basis as follows: if f is constant:

$$\frac{1}{\sqrt{2}}(|0, f(0)\rangle + |1, f(1)\rangle) = \frac{1}{\sqrt{2}}(|0', 0'\rangle + (-1)^{f(0)}|0', 1'\rangle)$$

and if f is balanced:

$$\frac{1}{\sqrt{2}}(|0, f(0)\rangle + |1, f(1)\rangle) = \frac{1}{\sqrt{2}}(|0', 0'\rangle + (-1)^{f(0)}|1', 1'\rangle).$$

If the measurement of the second qubit, in the dual bases, provides 0 we have lost all information about f. However, if the measurement of the second qubit yields 1, then the measurement of the first qubit yields the correct result.

The corresponding circuit is shown in Figure 3.1a. Observe that the above algorithm (due to Deutsch, 1985), has three possible outcomes; undefined (probability $\frac{1}{2}$), and either constant or balanced (with the total probability again $\frac{1}{2}$).

Surprisingly, there is also a "deterministic" solution to the Deutsch's problem, due to Cleve, Ekert, Macchiavello and Mosca (1998): a quantum algorithm that uses the black box for f only once and provides in all cases the exact answer.

Example 3.1.4 (Deutsch's XOR problem—deterministic solution) *Apply first the Hadamard transform on both registers in the initial state $|0,1\rangle$ and then U_f to get*

$$\begin{aligned}
|0\rangle|1\rangle &\overset{H_2}{\rightarrow} \frac{1}{2}(|0\rangle + |1\rangle)(|0\rangle - |1\rangle) \\
&= \frac{1}{2}(|0\rangle(|0\rangle - |1\rangle) + |1\rangle(|0\rangle - |1\rangle)) \\
&\overset{U_f}{\rightarrow} \frac{1}{2}(|0\rangle(|0 \oplus f(0)\rangle - |1 \oplus f(0)\rangle) + |1\rangle(|0 \oplus f(1)\rangle - |1 \oplus f(1)\rangle)) \qquad (3.4)
\end{aligned}$$

[2]From now on the assumption that a function f is given as a black box, or oracle, means that it is not possible to obtain knowledge about f by any other means than by evaluating it on points of its domain.

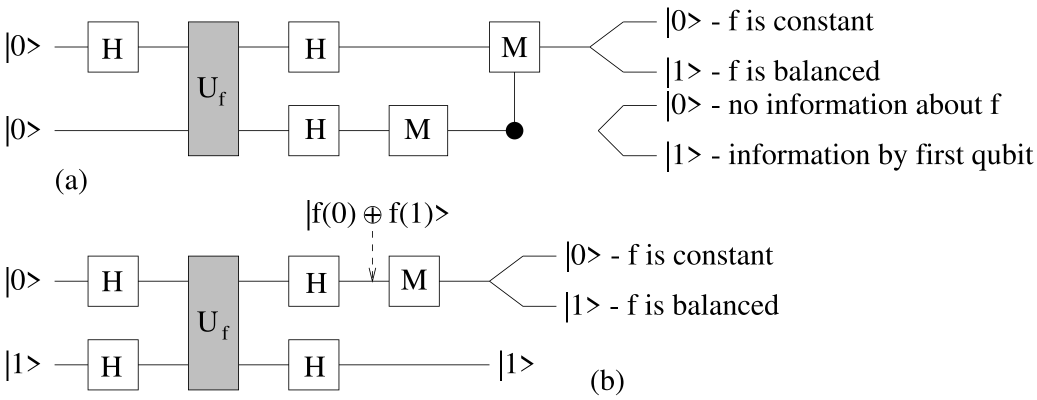


Figure 3.1: Circuits for randomized and deterministic solution of Deutsch's problem

$$
\begin{aligned}
&= \frac{1}{2}(\sum_{x=0}^{1}(-1)^{f(x)}|x\rangle)(|0\rangle - |1\rangle) \\
&= \frac{1}{2}(-1)^{f(0)}(|0\rangle + (-1)^{f(0)\oplus f(1)}|1\rangle)(|0\rangle - |1\rangle). \qquad (3.5)
\end{aligned}
$$

From the right side in (3.4), the two possibilities for f to be constant lead to the left sides in (3.6) and (3.7) and two possibilities for f to be balanced lead to the left sides in (3.8) and (3.9):

$$
\begin{aligned}
\frac{1}{2}(|0\rangle + |1\rangle)(|0\rangle - |1\rangle) &= |0'\rangle|1'\rangle \text{ if } f(0) = 0; && (3.6) \\
\frac{1}{2}(|0\rangle + |1\rangle)(|1\rangle - |0\rangle) &= -|0'\rangle|1'\rangle \text{ if } f(0) = 1; && (3.7) \\
\frac{1}{2}(|0\rangle - |1\rangle)(|0\rangle - |1\rangle) &= |1'\rangle|1'\rangle \text{ if } f(0) = 0; && (3.8) \\
\frac{1}{2}(|0\rangle - |1\rangle)(|1\rangle - |0\rangle) &= -|1'\rangle|1'\rangle \text{ if } f(0) = 1. && (3.9)
\end{aligned}
$$

By measuring the first bit, with respect to the dual basis, we can immediately see whether f is constant or balanced.

Another way, and a more straightforward one, to come to the same outcome is to transform, at the right side in (3.5), the states of both qubits to the dual basis with the outcome

$$(-1)^{f(0)}|(f(0) \oplus f(1))'\rangle|1'\rangle.$$

The circuit for this algorithm is in Figure 3.1b. It is now easy to see how we can simplify the algorithm and the corresponding circuit. Indeed, since the final measurement is on the first qubit only we can omit the second Hadamard rotation on the second qubit. In addition, we can also omit the first Hadamard rotation on the second qubit, if its initial state is $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$. The resulting circuit is in Figure 3.2a, where a special notation is used for the f-controlled NOT.

In the second algorithm we have used two simple but powerful techniques which one often encounters in the design of efficient algorithms and quantum error-correcting networks: a

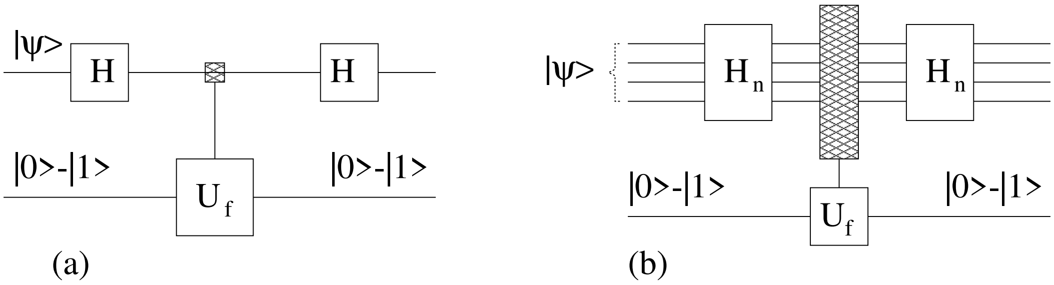


Figure 3.2: Circuit for the Deutsch's problem and the "Hadamard-twice scheme". The state $|0\rangle - |1\rangle$ should be normalized

change between the standard and the dual basis (for some qubits), and the computation scheme, called **Hadamard twice** , depicted in the general form in Figure 3.2b, which uses again f-controlled NOT . On closer examination one sees that the key point of the "Hadamard twice" scheme is again the change of the basis from standard to dual, some natural computations, and again the change of the basis back.

Exercise 3.1.5 *Given a unitary operation U_y which maps any state $|\psi\rangle$ into $(-1)^y|\psi\rangle$, for a fixed y, design a network using a conditional U_y-gate and two Hadamard gates to determine y.*

Deutsch's problem was the first one at which a separation was found between what classical and quantum computers can do. In this case a better performance of the quantum algorithm is due to the fact that a quantum algorithm can act in one step on a superposition of states $|0\rangle$ and $|1\rangle$ and in this way it can extract global information about the function.

The second algorithm for Deutsch's problem was the first quantum algorithm experimentally implemented, using NMR technology (see page 312 for more detail).

Exercise 3.1.6 *The parity* $\mathrm{par}(f)$ *of a function* $f : \{1,2\} \rightarrow \{-1,1\}$ *is defined as* $\mathrm{par}(f) = f(1)f(2)$. *Design a quantum algorithm to compute* $\mathrm{par}(f)$ *such that the black-box for f is used only once.*

3.1.2 The Deutsch–Jozsa promise problem

As already mentioned, no quantum measurement can extract exponentially large number of information which such states as in (3.1) and (3.2) contain. However, in some cases, as in the above algorithm, there are ways of obtaining from such states important information about some joint (or global) properties of all output values, computation of which on classical computers seems to require exponential time. An important example of this type—computation of the period of functions—will be described in Section 3.2. Now we present other examples of such a type.

In the first of these examples we consider a generalization of Deutsch's problem in which we deal with the following two properties of Boolean functions.

Definition 3.1.7 *A function $f : \{0,1\}^n \to \{0,1\}$ is* **balanced** *if none of the values of f has majority and it is* **constant** *if there exists no $x, y \in \{0,1\}^n$ such that $f(x) \neq f(y)$.*

Example 3.1.8 (The Deutsch–Jozsa problem) *Given a computable function $f : \{0,1\}^n \to \{0,1\}$, as a black box (circuit), that is (promised to be) either balanced or constant function, decide which of these two properties f has.*

It is easy to see that classical deterministic computers need, in the worst case, exponential time to solve the problem. Indeed, it may happen that by evaluating f for 2^{n-1} arguments it is not sufficient to determine whether f is constant or balanced—for example, if all 2^{n-1} values of f are 0.

Surprisingly, there is a quantum algorithm to solve the Deutsch-Jozsa problem by applying f only once. The algorithm can be designed using a similar technique as before, and as discussed later, but first we demonstrate a slightly different method (Berthiaume, 1997).

Let us consider one quantum register with n qubits and apply the Hadamard transformation H_n to the first register. This yields

$$|0^{(n)}\rangle \xrightarrow{H_n} |\phi\rangle = \frac{1}{\sqrt{2^n}} \sum_{i=0}^{2^n-1} |i\rangle.$$

By applying the transformation V_f on the first register we get

$$V_f|\phi\rangle = \frac{1}{\sqrt{2^n}} \sum_{i=0}^{2^n-1} (-1)^{f(i)}|i\rangle = |\phi_1\rangle.$$

What has been achieved by these operations? The values of f were transferred to the amplitudes, relative to each of the basis states. This can now be utilized, through the power of quantum superposition and a proper observable, to solve the problem through a single measurement as follows.

Let us consider the observable $\mathcal{D} = \{E_a, E_b\}$, where E_a is the one-dimensional subspace spanned by the vector

$$|\psi_a\rangle = \frac{1}{\sqrt{2^n}} \sum_{i=0}^{2^n-1} |i\rangle,$$

and $E_b = (E_a)^\perp$. The projection of $|\phi_1\rangle$ into E_a and E_b has the form

$$|\phi_1\rangle = \alpha|\psi_a\rangle + \beta|\psi_b\rangle \quad \text{with} \quad |\alpha|^2 + |\beta|^2 = 1,$$

where $|\psi_b\rangle$ is a vector in E_b such that $|\psi_a\rangle \perp |\psi_b\rangle$. A measurement through $\mathcal{D}$ will provide "the value a or b" with probability $|\alpha|^2$ or $|\beta|^2$, respectively.

It is easy to determine α, using the projection of $|\phi_1\rangle$ onto E_a by the computation

$$\begin{aligned}
\alpha &= \langle\psi_a|\phi_1\rangle = \left(\frac{1}{\sqrt{2^n}} \sum_{i=0}^{2^n-1} \langle i|\right) \left(\frac{1}{\sqrt{2^n}} \sum_{j=0}^{2^n-1} (-1)^{f(j)}|j\rangle\right) \\
&= \frac{1}{2^n} \sum_{i=0}^{2^n-1} \sum_{j=0}^{2^n-1} (-1)^{f(j)}\langle i|j\rangle = \frac{1}{2^n} \sum_{i=0}^{2^n-1} (-1)^{f(i)},
\end{aligned}$$

because $\langle i|j\rangle = 1$ if and only if $i = j$ and 0 otherwise.

If f is balanced, then the sum for α contains the same number of 1*s and* -1*s and therefore* $\alpha = 0$. *A measurement of $|\phi_1\rangle$, with respect to $\mathcal{D}$ therefore provides, for sure, the outcome b.*

If f is constant, then either $\alpha = 1$ or $\alpha = -1$ and therefore the measurement of $|\phi_1\rangle$ with respect to $\mathcal{D}$ always gives the outcome a.

A single measurement of $|\phi_1\rangle$, with respect to $\mathcal{D}$, therefore provides the solution of the problem with probability 1.

The Deutsch-Jozsa problem was the first one that was found to need only linear time on a quantum computer but exponential time on a deterministic Turing machine.

Exercise 3.1.9 *Show that the Deutsch–Jozsa problem can be solved by first applying the Hadamard transformation to the state $|\phi_1\rangle$ and then checking whether all resulting qubits are $|0\rangle$.*

Exercise 3.1.10 *(Bernstein–Vazirani (1993) problem) For $y \in \{0,1\}^n$, let $f_y : \{0,1\}^n \to \{0,1\}$ be a function defined by $f_y(x) = x \cdot y$. Show that there is a quantum algorithm to distinguish the 2^n functions f_y by evaluating f only once.*

The quantum algorithm presented above solves the Deutsch–Jozsa problem exactly in polynomial time. As shown above, the problem cannot be solved in polynomial time on a deterministic computer. However, it can be solved in polynomial time on a PTM.

Exercise 3.1.11 *Show that the Deutsch–Jozsa problem can be solved on PTM in polynomial time provided an arbitrarily small one-sided error is allowed.*

There are several variations of the Deutsch–Jozsa problem that can be solved with a small modification of the above techniques.

Exercise 3.1.12 *(Cleve et al. 1998) Given a function $f : \{0,1\}^n \to \{0,1\}^m$, $m \leq n$, that is promised to have the property that the parity of the elements in the range of f is either constant or equally balanced. Show that there is a quantum algorithm to determine which of these two properties f has. (Hint: choose an auxiliary register of m qubits all in the initial state $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$.)*

3.1.3 Simon's problems

Simon (1994) was the first to show a nice and simple problem with expected polynomial time quantum algorithm but with no polynomial time randomized algorithm.

Example 3.1.13 (Simon's XOR Problem) *Let $f : \{0,1\}^n \to \{0,1\}^n$ be a function such that either f is one-to-one or f is two-to-one and there exists a single non-zero $s \in \{0,1\}^n$ such that*

$$\forall x \neq x'(f(x) = f(x') \Leftrightarrow x' = x \oplus s).$$

The task is to determine which of the above conditions holds for f and, in the second case, to determine also s.

To solve the problem two registers are used, both with n qubits and the initial states $|0^{(n)}\rangle$, and (expected) $\mathcal{O}(n)$ repetitions of the following version of the Hadamard-twice scheme:

1. *Apply the Hadamard transformation on the first register, with the initial value $|0^{(n)}\rangle$, to produce the superposition $\frac{1}{\sqrt{2^n}}\sum_{x\in\{0,1\}^n}|x,0^{(n)}\rangle$.*
2. *Apply U_f to compute $|\psi\rangle = \frac{1}{\sqrt{2^n}}\sum_{x\in\{0,1\}^n}|x,f(x)\rangle$.*
3. *Apply Hadamard transformation on the first register to get*
$$\frac{1}{2^n}\sum_{x,y\in\{0,1\}^n}(-1)^{x\cdot y}|y,f(x)\rangle.$$
4. *Observe the resulting state to get a pair $(y,f(x))$.*

Case 1: *f is one-to-one. After performing the first three steps of the above procedure all possible states $|y,f(x)\rangle$ in the superposition are distinct and the absolute value of their amplitudes is the same, namely 2^{-n}. $n-1$ independent applications of the scheme* Hadamard-twice *therefore produce $n-1$ pairs $(y_1,f(x_1)),\ldots,(y_{n-1},f(x_{n-1}))$, distributed uniformly and independently over all pairs $(y,f(x))$.*

Case 2: *There is some $s\neq 0^{(n)}$ such that $\forall x\neq x'((f(x)=f(x')\Leftrightarrow x'=x\oplus s)$. In such a case for each y and x the states $|y,f(x)\rangle$ and $|y,f(x\oplus s)\rangle$ are identical. Their total amplitude $\alpha(x,y)$ has the value*
$$\alpha(x,y)=2^{-n}((-1)^{x\cdot y}+(-1)^{(x\oplus s)\cdot y}).$$
If $y\cdot s\equiv 0 \bmod 2$, then $x\cdot y\equiv(x\oplus s)\cdot y \bmod 2$ and therefore $\alpha(x,y)=2^{-n+1}$; otherwise $\alpha(x,y)=0$. n independent applications of the scheme Hadamard-twice *therefore yield $n-1$ independent pairs $(y_1,f(x_1)),\ldots,(y_{n-1},f(x_{n-1}))$ such that $y_i\cdot s\equiv 0 \pmod 2$, for all $1\le i\le n-1$.*

In both cases, after $n-1$ repetitions of the scheme Hadamard-twice, *$n-1$ vectors $y_i, 1\le i\le n-1$, are obtained. If these vectors are linearly independent, then the system of $n-1$ linear equations in $\mathbf{Z}_2$, $y_i\cdot s=0$ can be solved to obtain s. In Case 2, if f is two-to-one, s obtained in such a way is the one to be found. In Case 1, s obtained in such a way is a random string. To distinguish these two cases, it is enough to compute $f(0)$ and $f(s)$. If $f(0)\neq f(s)$, then f is one-to-one. If the vectors obtained by the scheme* Hadamard-twice *are not linearly independent, then the whole process has to be repeated.*

As shown in the next lemma, the vectors $y_i, 1\le i\le n$, obtained in this way are linearly independent with probability at least $\frac{1}{4}$. The total expected computation time is therefore $\mathcal{O}(nt(n)+g(n))$, where $t(n)$ is time needed to compute f on inputs of length n and $g(n)$ is time needed to solve the system of n linear equations in $\mathbf{Z}_2$. ∎

Lemma 3.1.14 *If u is a non-zero binary vector of length n, then $n-1$ randomly chosen binary vectors y of length n such that $u\cdot y\equiv 0 \bmod 2$ are linearly independent with probability at least $\frac{1}{4}$.*

Proof. Let $y_1,\ldots,y_{n-1}$ be such vectors. For any $1<i\le n-1$ at most 2^{i-1} vectors are linear combinations of vectors $y_1,\ldots,y_{i-1}$. Hence the probability that y_i is linearly independent from $y_1,y_2,\ldots,y_{i-1}$ is
$$\frac{2^{n-1}-2^{i-1}}{2^{n-1}}=1-\frac{1}{2^{n-i}}.$$

The probability that $y_1, \ldots, y_{n-1}$ are all linearly independent is then

$$(1-\frac{1}{2^{n-1}})(1-\frac{1}{2^{n-2}})(1-\frac{1}{2^{n-3}})\ldots(1-\frac{1}{2})$$

and the probability that $y_1, \ldots, y_{n-2}$ are linearly dependent is at most $\frac{1}{2^{n-1}}+\frac{1}{2^{n-2}}+\ldots+\frac{1}{4} \leq \frac{1}{2}$. Hence the probability that vectors $y_1, \ldots, y_{n-2}$ are linearly independent is at least $\frac{1}{2}$ and the probability that y_{n-1} is linearly independent from $y_1, \ldots, y_{n-2}$ is at least $\frac{1}{2}$. The total probability that vectors $y_1, \ldots, y_{n-1}$ are linearly independent is at least $\frac{1}{4}$. ∎

Finally (see Vazirani, 1997), we show that any randomized Monte Carlo algorithm needs exponential time to solve the above version of Simon's problem. The proof uses the following variant of Yao's Minimize Principle (Motwani and Raghavan, 1995):

Lemma 3.1.15 (Yao's Lemma) *Let δ be a probability distribution on possible inputs of a problem P such that no deterministic algorithm running in time $t(n)$ gives a correct answer with probability at least p when inputs are drawn according to δ. Then there is no randomized algorithm for P running in time $t(n)$ with the probability of correct answer at least p.*

For Simon's problem a "hard probability distribution" is easy to choose. $s \neq 0$ is chosen randomly and all pairs $(u, u \oplus s)$ are mapped randomly to $\{0,1\}^n$ in such a way that no two pairs are mapped into the same element.

After m steps, a deterministic algorithm computes at least m values of f: $f(x^{(1)}), \ldots, f(x^{(m)})$. These values provide for the algorithm the following information:

1. If $f(x^{(i)}) = f(x^{(j)})$ and $x^{(i)} \neq x^{(j)}$, then $s = x^{(i)} \oplus x^{(j)}$.
2. If $f(x^{(i)}) \neq f(x^{(j)})$ and $x^{(i)} \neq x^{(j)}$, then $s \neq x^{(i)} \oplus x^{(j)}$.

Assume that $x^{(1)}, \ldots, x^{(k)}$ and also $f(x^{1)}), \ldots, f(x^{(k)})$ are mutually different. Then s is none of the $\binom{k}{2}$ values $x^{(i)} \oplus x^{(k)}$. (It can be shown that all other values are equally probable.)

The probability that $x^{(i)} \oplus x^{(k+1)} = s$ for some $i \in \{1, \ldots, k\}$ is at most

$$\frac{k}{2^n - 1 - \binom{k}{2}},$$

because there are at least $2^n - 1 - \binom{k}{2}$ possible values of s. Summing over all k we get

$$\sum_{k=1}^{m} \frac{k}{2^n - 1 - \binom{k}{2}} \leq \sum_{k=1}^{m} \frac{k}{2^n - k^2} \leq \frac{m^2}{2^n - m^2}.$$

For $m = 2^{(1/2-\varepsilon)n}$ we have

$$\frac{m^2}{2^n - m^2} = \frac{2^{(1-2\varepsilon)n}}{2^n - o(2^n)} = 2^{-2\varepsilon n} - o(2^{-2\varepsilon n}).$$

This implies that, with respect to the chosen distribution, any deterministic algorithm running in exponential time $m = 2^{(1/2-\varepsilon)n}$ has exponentially low probability of a correct answer. By Yao's Lemma the same bound holds for probabilistic algorithms.

Observe that Simon's algorithm is polynomial only in the expected time and there is no upper bound to how long it can run in the worst case. However, there is another algorithm, discovered by Brassard and Høyer (1997), that always produces an exact solution in a time which is upper bounded in the worst case by a polynomial.

3.2 Shor's Algorithms

Shor's polynomial time algorithms for factorization and computation of discrete logarithms, discussed in detail in this section, represent so far perhaps the most important, stimulating, and impressive results in quantum computing.

Shor, inspired by Simon (1994), cleverly put together several old ideas, namely that

- factorization of integers can be reduced to the problem of finding the period of a function;
- Fourier transform puts the period of any periodic function into multiples of the reciprocal of the period,

with a clever use of Quantum Fourier Transform (QFT) to get approximations of the period and new clever methods to extract the exact period. In addition, he was able to show that all that can be done efficiently, which was far from obvious, especially for QFT.

Shor's result has been of the utmost importance. Modern public key cryptography and the security of digital signature methods has largely relied so far on the belief that no effective integer factorization or computation of discrete logarithm exists. A design of quantum computer capable of performing Shor's polynomial time quantum algorithms for factorization and discrete logarithm could therefore have great impact on modern cryptography and on current methods of secure information transmission and storage, on which society is more and more dependent. That is why Shor's result interested also people and institutions outside science and technology in seeing the implications of having in the near future really powerful quantum computers available. Since Shor's result, the security of modern cryptography has relied on confidence in the slowness of technological progress in the design of quantum computers.

3.2.1 Number theory basics

Two simple reduction results (Vazirani, 1997), are behind an effective quantum factorization algorithm.

Lemma 3.2.1 *If there is a polynomial time deterministic (randomized) [quantum] algorithm to find a nontrivial solution* $x \not\equiv \pm 1 \pmod{n}$ *of the equation* $x^2 \equiv 1 \pmod{n}$*, then there is a polynomial time deterministic (randomized) [quantum] algorithm to factorize integers.*

Proof. Let $a \not\equiv \pm 1 \pmod{n}$ be such that $a^2 \equiv 1 \pmod{n}$. Since $a^2 - 1 = (a+1)(a-1)$, if n is not prime, then a prime factor of n has to be a prime factor of either $a+1$ or $a-1$. By applying Euclid's algorithm to $(n, a+1)$ and $(n, a-1)$ we can find, in $\mathcal{O}(\log n)$ steps, a prime factor of n. □

The second concept to be used in the following is that of the **period** of the function $f_{n,x}(k) = x^k \bmod n$. It is the smallest integer r such that $f_{n,x}(k+r) = f_{n,x}(k)$ for any k; i.e., the smallest r such that $x^r \equiv 1 \pmod{n}$. Such an r is also called the **order** of x, in short $ord(x)$, in $\mathbf{Z}_n^*$.[3] The problem to find the period of a function is more technically named as the **order problem**.

Exercise 3.2.2 *Show that if* $gcd(x,n) \neq 1$*, then* $x^r \equiv 1 \pmod{n}$ *for no* $r > 0$.

[3] $\mathbf{Z}_n^* = \{i \mid 0 < i \leq n, gcd(i,n) = 1\}$ is a multiplication group.

Let us now consider the following algorithm:

Algorithm 3.2.3

1. Choose randomly $1 < y < n$.
2. Compute $gcd(y,n)$. If $gcd(y,n) \neq 1$ we have a factor, otherwise continue.
3. Find the period r of the function $y^k \bmod n$.
4. If r is odd or $y^{r/2} \equiv \pm 1 \pmod{n}$, then go to step 1, otherwise stop.

If this algorithm stops, then $y^{r/2}$ is a nontrivial solution of the equation $x^2 \equiv 1 \pmod{n}$.

Exercise 3.2.4 *Show the following result, which indicates why it is a good idea to exclude powers of primes in Algorithm 3.2.3. Let $n = p^e$, where p is an odd prime and $e > 1$, let y be an integer with $gcd(y,p) = 1$ and let r be the order of y. Then either r is odd or $y^{\frac{r}{2}} \equiv \pm 1 \pmod{n}$.*

Exercise 3.2.5 *Show the following result which implies that the exclusion of powers of primes in the factorization Algorithm 3.2.3 is not an essential restriction: powers of primes can be factorized in polynomial time.*

Lemma 3.2.6 *If a y such that $1 < y < n$ and $gcd(n,y) = 1$ is selected randomly and an odd n is not a power of a prime, then $Pr\{r$ is even and $y^{r/2} \not\equiv \pm 1 \pmod{n}\} \geq \frac{1}{4}$.*

Proof. Let a prime factorization of n be $n = \prod_{i=1}^{k} p_i^{e_i}$. By the Chinese remainder theorem, the groups $\mathbf{Z}_n^*$ and $\mathbf{Z}_{p_1^{e_1}}^* \times \ldots \times \mathbf{Z}_{p_k^{e_k}}^*$ are isomorphic by the following mapping $a \bmod n \leftrightarrow (a \bmod p_i^{e_i}, \ldots, a \bmod p_k^{e_k})$.

In the rest of the proof we consider the following unique decompositions $\phi(n) = 2^l m$, $\phi(p_i^{e_i}) = 2^{l_i} m_i$,[4] $1 \leq i \leq k$, where m and all m_i are odd.

Since all groups $\mathbf{Z}_{p_i^{e_i}}^*$ are cyclic (see, for example, Gruska, 1997, page 53), in each of them a generator g_i can be found and fixed. In such a case choosing randomly and independently $x_i \in \{1, \ldots, 2^{l_i} m_i\}$ and considering $a = (g_1^{x_1}, \ldots, g_k^{x_l}) \in \mathbf{Z}_{p_1^{e_1}}^* \times \ldots \times \mathbf{Z}_{p_k^{e_k}}^*$ is a way to get a random $a \in \mathbf{Z}_n^*$. The claim of the Lemma now follows from the following two sublemmas:

Lemma 3.2.7 *If $a \in \mathbf{Z}_n^*$ is chosen randomly, n is odd, then $Pr\{r = ord(a)$ is even$\} \geq \frac{1}{2}$.*

Proof. The order r of a in $\mathbf{Z}_n^*$ is the smallest integer such that all $\frac{rx_i}{2^{l_i} m_i}$ are integers. Hence r is the least common multiplier (LCM) of the set

$$\left\{ \frac{2^{l_1} m_1}{x_1}, \ldots, \frac{2^{l_k} m_k}{x_k} \right\}.$$

Since n is odd all p_i have to be odd primes and therefore all $\phi(p_i^{e_i})$ are even and, naturally, all $l_i > 0$. Thus if any of the x_i is odd, then LCM must be even. Since x_i are chosen randomly, such a probability is at least $1 - \frac{1}{2^k} \geq \frac{1}{2}$. ∎

[4] $\phi(n)$ is Euler totient function and $\phi(n)$ is the number of elements of the group $\mathbf{Z}_n^*$.

Lemma 3.2.8 *For a random* $a \in \mathbf{Z}_n^*$, $r = ord(a)$, r *even,* $Pr\{a^{\frac{r}{2}} \not\equiv \pm 1 \pmod{n}\} \geq \frac{1}{2}$.

Proof. Fix an a and let $r = ord(a)$. $a^{\frac{r}{2}}$ corresponds to $(g_1^{\frac{x_1 r}{2}}, \ldots, g_k^{\frac{x_k r}{2}})$. Since all $\mathbf{Z}^*_{p_i^{e_i}}$ are cyclic, in each $\mathbf{Z}^*_{p_i^{e_i}}$ there are only two square roots of 1, namely $+1$ and -1. This implies that square roots of 1 in $\mathbf{Z}_n^*$ are exactly those corresponding to k-tuples $(\pm1, \ldots, \pm1)$ with the correspondence $1 \leftrightarrow (1, \ldots, 1)$, $-1 \leftrightarrow (-1, \ldots, -1)$. For an $1 \leq i \leq k$, $g_i^{\frac{x_i r}{2}}$ is -1 if $\frac{x_i r}{2}$ is not a multiple of $2^{l_i} m_i$. This happens if the highest power of 2 dividing $x_i r$ is at most l_i.

It is clear that not all $g_i^{\frac{x_i r}{2}}$ are 1 because otherwise r would not be the order of a. This implies that in order to show the lemma, it is sufficient to bound the probability that all $g_i^{\frac{x_i r}{2}}$ are -1.

The only way this can happen is that for all $1 \leq i \leq k$, the highest power of 2 dividing rx_i is l_i. Suppose now that each x_i is chosen randomly. Let t be the highest power of 2 dividing x_1. In order that $g_1^{\frac{x_1 r}{2}}$ is -1, the highest power of 2 dividing r has to be $l_1 - t > 0$. The probability of choosing x_2 such that the highest power of 2 dividing it is exactly $l_2 - (l_1 - t)$ (which implies $g_2^{\frac{x_2 r}{2}} = -1$) is less than or equal to $\frac{1}{2}$. This proves the lemma. ∎

Corollary 3.2.9 *If there is a polynomial time randomized [quantum] algorithm to compute the period of the function* $f_{n,x}$ *for any* n, x*, then there is a polynomial time randomized [quantum] algorithm to find non–trivial solutions of the equation* $x^2 \equiv 1 \pmod{n}$*, (and therefore also to factorize any integer* n*).*[5]

> **Exercise 3.2.10** *Show that if* $1 < y < n$ *is selected randomly, then the probability that* $gcd(y, n) = 1$ *is: (a) greater than* $\Omega(\frac{1}{\lg n})$*; (Hint: use the Prime Number Theorem); (b) greater than* $\Omega(\frac{1}{\lg \lg n})$.
>
> **Exercise 3.2.11** *Use Lemma 3.2.1 and Algorithm 3.2.3 to factorize: (a)* 91*; (b)* 899*; (c)* 5183.

Example 3.2.12 *Let* $n = 15$ *and select* $1 < y < 15$ *such that* $gcd(y, 15) = 1$. *The set of such* y *is* $\{2, 4, 7, 8, 11, 13, 14\}$. *Let us choose* $y = 11$. *Values of* $11^x \bmod 15$ *form, for* $x = 1, 2, \ldots$, *the sequence* $11, 1, 11, 1, 11, 1, \ldots$ *with the period* $r = 2$. *Hence* $y^{r/2} = 11$ *and we have to compute* $gcd(15, 11 + 1) = 3$ *and* $gcd(15, 11 - 1) = 5$*—to get both factors of* 15. *Observe also that the corresponding periods of elements* $2, 4, 7, 8, 11, 13, 14$ *are* $4, 2, 4, 4, 2, 4, 2$ *and in this case any choice of* y *with the exception of* $y = 14$ *leads to a desirable factorization. For* $y = 14$ *we get* $r = 2$, $14^{2/2} \equiv -1 \bmod 15$ *and the method fails.*

> **Exercise 3.2.13** *Analyse the case* $n = 21$. *Find all integers* y *such that* $gcd(21, y) \equiv 1$ *and their order.*

The task now is to find out how to make use of quantum parallelism to compute the period of the function $f_{n,x}(k)$ for $n = 2^d - 1$. Let $U_{f_{n,x}}$ be the unitary operator to realize

[5] It has been shown (see, for example, Gruska, 1997, Section 1.8.1), that if n is not prime and has at least 2 different odd factors, then the equation $x^2 \equiv 1 \pmod{n}$ has at least four solutions.

the mapping $(k,0) \rightarrow (k, f_{n,x}(k))$. An application of this operator to the state

$$|\psi\rangle = \frac{1}{\sqrt{2^d}} \sum_{k=0}^{2^d-1} |k, 0^{(d)}\rangle,$$

yields

$$U_{f_{n,k}}|\psi\rangle = \frac{1}{\sqrt{2^d}} \sum_{k=0}^{2^d-1} |k, f_{n,x}(k)\rangle = |\psi_1\rangle. \tag{3.10}$$

Observe that all possible values of $f_{n,x}$ are encoded in values of the second register of the state $|\psi_1\rangle$. However, as already pointed out, in this context we are actually not interested in particular values of the function $f_{n,x}$, only in its period. It is therefore of importance and interest to locate the potentials and pitfalls of attempts to find the period from the state $|\psi_1\rangle$. To see that let us consider again the case $n = 15$ and $x = 7$. In such a case (3.10) has the form

$$\frac{1}{4}(|0\rangle|1\rangle + |1\rangle|7\rangle + |2\rangle|4\rangle + |3\rangle|13\rangle + |4\rangle|1\rangle + |5\rangle|7\rangle + \ldots + |14\rangle|4\rangle + |15\rangle|13\rangle).$$

If we measure at this point the second register, then we get as the outcome one of the numbers $1, 4, 7$ or 13, and the following table shows the corresponding post-measurement states in the second column. The corresponding sequences of values of the first register are periodic with period 4 but they have different offsets (pre-periods) listed in column 3 of the table.

result	post-measurement state	offset
1	$\frac{1}{2}(\|0\rangle + \|4\rangle + \|8\rangle + \|12\rangle)\|1\rangle$	0
4	$\frac{1}{2}(\|2\rangle + \|6\rangle + \|10\rangle + \|14\rangle)\|4\rangle$	2
7	$\frac{1}{2}(\|1\rangle + \|5\rangle + \|9\rangle + \|13\rangle)\|7\rangle$	1
13	$\frac{1}{2}(\|3\rangle + \|7\rangle + \|11\rangle + \|15\rangle)\|13\rangle$	3

One natural way to obtain the period seems to be to repeat computation (3.10) many times and each time to measure at first the second register and then the first one. If we get for some value z of the second register the values $y_1 < y_2 < y_3$ of the first register we know that the period is at most $gcd(y_2 - y_1, y_3 - y_2)$.

Unfortunately, this method is not efficient enough. Due to the fact that preperiods may be different we cannot compare values of the first register for different values of the second register. In addition, on average the period r grows exponentially with d. Therefore an exponential number of repetitions of computation 3.10 would be in general needed to get the period this way.

Fortunately, there is a method of "massaging" the state (3.10) in such a way that from the result the period can be obtained efficiently, without sampling the state. The key step is to transform the pre-period into the phase in which it has no influence on the corresponding probabilities. The key tool to use is the Quantum Fourier Transform discussed in the next subsection.

3.2.2 Quantum Fourier Transform

Fourier transforms are one of the most powerful tools of applied mathematics. In general they map from time domain to frequency domain—they map functions of period r to functions having non-zero values only at the multiples of the frequency $\frac{1}{r}$.

The quantum Fourier transform is a quantum variant of the **discrete Fourier transform** (DFT). It maps a discrete function to another discrete one with equally distant points as its domain. For example it maps a q-dimensional complex vector $\{f(0), f(1), \ldots, f(q-1)\}$, into $\{\bar{f}(0), \bar{f}(1), \ldots, \bar{f}(q-1)\}$ as follows:

$$\bar{f}(c) = \frac{1}{\sqrt{q}} \sum_{a=0}^{q-1} e^{2\pi iac/q} f(a), \tag{3.11}$$

for $c \in \{0, \ldots, q-1\}$.

DFT plays such an important role in computing that in the course of time several ingenious devices have been invented for it. All these efforts were superseded by the invention of the algorithm known as Fast Fourier Transform (FFT).

The quantum version of DFT is a quantum variant of FFT, known as Quantum Fourier Transform (QFT). QFT with the base q (or in the group $\mathbf{Z}_q$) is the unitary transformation

$$\mathrm{QFT}_q : |a\rangle \rightarrow \frac{1}{\sqrt{q}} \sum_{c=0}^{q-1} e^{2\pi iac/q} |c\rangle \tag{3.12}$$

for $0 \leq a < q$, with the unitary matrix

$$F_q = \frac{1}{\sqrt{q}} \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & \omega & \omega^2 & \ldots & \omega^{(q-1)} \\ 1 & \omega^2 & \omega^4 & \ldots & \omega^{2(q-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{(q-1)} & \omega^{2(q-1)} & \ldots & \omega^{(q-1)^2} \end{pmatrix},$$

where $\omega = e^{2\pi i/q}$ is the qth root of unity.

Exercise 3.2.14 *Demonstrate why is the nature of QFT different from DFT.*

If applied to a quantum superposition, QFT_q transforms the state $\frac{1}{\sqrt{q}} \sum_{a=0}^{q-1} f(a)|a\rangle$ as follows

$$\mathrm{QFT}_q : \sum_{a=0}^{q-1} f(a)|a\rangle \rightarrow \sum_{c=0}^{q-1} \bar{f}(c)|c\rangle,$$

where $\bar{f}(c)$ is defined by (3.11).

Observe that

$$QFT_q : |\mathbf{0}\rangle \rightarrow \frac{1}{\sqrt{q}} \sum_{i=0}^{q-1} |i\rangle,$$

and therefore the impact of QFT on $|\mathbf{0}\rangle$ is the same as of the Hadamard transformation—see Section 2.1.

Most of the known important/interesting quantum algorithms use QFT either in its full strength or its special case—the Hadamard transformation. Because of that the question of how efficiently one can compute QFT on quantum computers is of key importance.

Exercise 3.2.15 *(Mosca, 1998b) Let $x, y \in [0, 2^n - 1]$ and $W_y : |x\rangle|\psi\rangle \to \xi^{x \cdot y}|x\rangle|\psi\rangle$, $\xi_n = e^{2\pi i/2^n}$ be a unitary transformation. Design a network to determine y that uses two QFT gates.*

QFT_q is usually used with the base $q = 2^n$. In such a case the classical Fourier Transform algorithm requires time $\mathcal{O}(2^{2n})$. The classical Fast Fourier Transform algorithm requires only time $\mathcal{O}(n2^n)$—a very significant saving. With a quantum implementation time can be reduced to $\mathcal{O}(n^2)$ for some n. The fact that the QFT can be performed in polynomial time is of key importance for polynomial running time of quantum algorithms using it.

We first prove a partial result, namely that if

$$q = p_1^{e_1} p_2^{e_2} \dots p_k^{e_k} \tag{3.13}$$

is a prime factorization of q, then QFT_q can be performed by successively applying $\mathrm{QFT}_{p_i^{e_i}}$, $i = 1, \dots, k$, and therefore QFT_q can be performed efficiently on a quantum computer if q is **smooth**, i.e. if all prime factors $p_i^{e_i}$ of q are of size $\mathcal{O}((\lg q)^c)$ for some constant c. This result is due to Cleve (1994).

We show how to compute $\mathrm{QFT}_q|a\rangle$ by applying successively $\mathrm{QFT}_{p_i^{e_i}}$ to proper subcircuits.

As already discussed on page 113, the correspondence

$$a \leftrightarrow (a_1, a_2, \dots, a_k),$$

where $a_i = a \bmod p_i^{e_i}$ is one-to-one (and therefore it can be performed by a unitary transformation).

For any $0 \le c < q$ let $c_i = c \bmod p_i^{e_i}$. Then $ac \equiv a_i c_i \bmod p_i^{e_i}$ and, therefore, by the Chinese remainder theorem,

$$ac \equiv \sum_{i=1}^{k} a_i c_i r_i \prod_{j \ne i} p_j^{e_j} \pmod{q}, \tag{3.14}$$

where

$$r_i = (\prod_{j \ne i} p_j^{e_j})^{-1} \bmod p_i^{e_i} \tag{3.15}$$

and all r_i can be computed easily using the extended Euclid's algorithm.

The mapping

$$|a_1, \dots, a_k\rangle \to |a_1 r_1, \dots, a_k r_k\rangle.$$

is also one-to-one because each r_i is invertible modulo $p_i^{e_i}$ and therefore, if we apply $\mathrm{QFT}_{p_i^{e_i}}$ on the ith register of $|a_1 r_1, \dots, a_k r_k\rangle$, then from (3.11) the following cumulative result

$$\frac{1}{\sqrt{q_1^{e_1}} \cdots \sqrt{q_k^{e_k}}} \sum_{c_1=0,\dots,c_k=0}^{q_1^{e_1},\dots,q_k^{e_k}} e^{2\pi i \left(\frac{a_1 r_1 c_1}{p_1^{e_1}} + \dots + \frac{a_k r_k c_k}{p_k^{e_k}} \right)} |c_1, \dots, c_k\rangle$$

follows. This expression can be simplified using the relations (3.14) and (3.15) to the form

$$\frac{1}{\sqrt{q}} \sum_{c=0}^{q-1} e^{2\pi i ac/q} |c_1, \dots, c_k\rangle$$

and by relabeling $|c_1, \ldots, c_k\rangle$ with $|c\rangle$ we get exactly the same expression as in (3.12).

A simple implementation of QFT_q was discovered by Coppersmith (1994) and Deutsch (see Ekert and Jozsa, 1996) for the case that $q = 2^n$. The circuit implementing QFT_q uses the Hadamard gate H and conditional phase shift on second qubit provided the first qubit is in the state $|1\rangle$. Phase shift by $e^{\frac{2\pi i}{2^j}}$ is represented by the matrices X_j, $j = 0, \ldots, n-1$, where

$$X_j = \begin{pmatrix} 1 & 0 \\ 0 & e^{2\pi i/2^j} \end{pmatrix}.$$

Let us denote by H_j the gate H operating on the jth qubit and by $S_{j,k}$ the conditional X_{k-j} gate operating on the jth and kth qubit, $j < k$.

The algorithm is based on the fact that if $q = 2^n$, $a = \sum_{i=0}^{n-1} a_i 2^{n-i-1}$, then the $\mathrm{QFT}_q|a\rangle$ is not entangled and it holds

$$QFT_q|a\rangle = \frac{1}{\sqrt{2^n}}(|0\rangle + e^{2\pi i 0.a_{n-1}}|1\rangle)(|0\rangle + e^{2\pi i 0.a_{n-2}a_{n-1}}|1\rangle)\ldots(|0\rangle + e^{2\pi i 0.a_0\ldots a_{n-1}}|1\rangle)$$

and it can be realized by the network shown in Figure 3.3.

Exercise 3.2.16 *Show that in the network in Figure 3.3 the output value of the qubit with input $|a_j\rangle$ is $|0\rangle + e^{2\pi i 0.a_j\ldots a_{n-1}}|1\rangle$.*

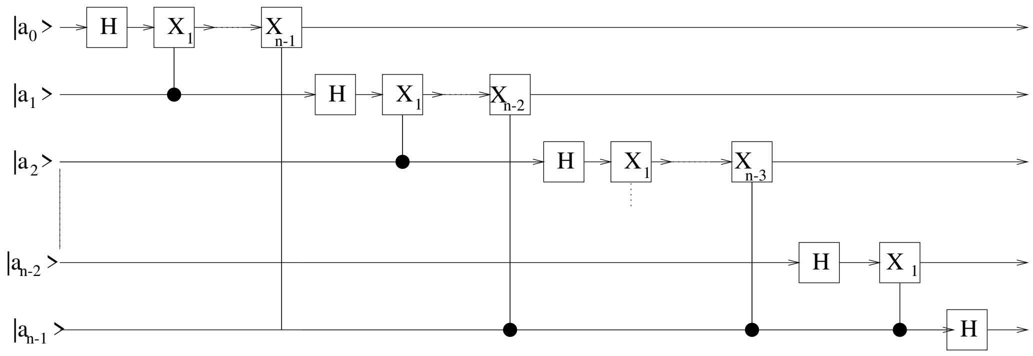


Figure 3.3: An implementation of QFT

The number of gates and, consequently, the computation time of this network is $\theta(n^2)$. For details see Coppersmith (1994) and Cleve et al. (1998). These results were generalized by Kitaev (1997), who showed how to design a polynomial time approximate quantum algorithm for Fourier transform on any finite Abelian group presented as a product of cyclic groups. Beals (1997) showed how to compute quantum Fourier transform over symmetric groups. For an analysis of the role Fourier transform has in design of quantum algorithms, and for the general construction of the Fourier transform on Abelian groups see Jozsa (1997a).

Exercise 3.2.17 *Make a formal proof that the above quantum computes QFT.*

Exercise 3.2.18 *Show that if $\mathcal{B} = \{\beta_i\}_{i=1}^n$ is a basis, then $\mathcal{G} = \{\gamma_i\}_{i=1}^n$ is also a basis if $\gamma_k = \frac{1}{\sqrt{n}}\sum_{j=1}^n \beta_j e^{2\pi ijk/n}$ and these two bases are mutually unbiased. (Quantum measurements corresponding to two mutually unbiased bases (see page 367), are called* **complementary***.)*

Quantum Fourier transform has been so far the key tool in designing efficient quantum algorithms. It is therefore of importance to analyze the performance of the QFT in the presence of decoherence. Barenco et al. (1996) have shown that so-called approximate QFT can provide better results concerning the period estimation than (exact) QFT.

The main problem with using QFT to extract the period, as we shall see, is that it works only approximately in general and a special effort is needed to derive from the approximation the exact period.

3.2.3 Shor's factorization algorithm

The best-known classical factorization heuristic for m-bit numbers has complexity $\mathcal{O}(e^{c(\ln m)^{1/3}(\ln\ln m)^{2/3}})$, where $c = (\frac{64}{9})^{1/3} \approx 1.9$. It is therefore an exponential time algorithm. This fact is of crucial importance for current digital communication and signatures because the security of some key cryptographical systems, for example RSA[6], is based on the belief that no polynomial time factorization algorithm exists.

Shor's factorization algorithm, a special and the most interesting case of which is presented in this section, requires $\mathcal{O}(m^2 \lg m \lg\lg m)$ steps on a quantum computer plus a polynomial number of steps on classical computers (which could also be done in polynomial time on a quantum computer).

We present here the original Shor's algorithm as the historically very important contribution to quantum computing. For a conceptually simpler version (see Cleve et al. 1998).

The general scheme of Shor's algorithm is depicted in Figure 3.4. The key part is the period finding quantum subroutine which will now be shown. As already discussed in Section 3.2.1, the problem is to determine the period, or the order, of the function $f_{n,x}(k) = x^k \bmod n$.

Remark 3.2.19 If we can factorize an integer n we can break any RSA cryptosystem with the public key n, e. In order to do factorization, as we can see from the flow diagram in Figure 3.4, we need in general to do order computing several times. However, to break RSA we actually do not need to factorize n. There is a simple method of breaking RSA, as pointed out by Ekert (1997) and Cleve et al. (1998), at which it is sufficient to compute the order of the cryptotext, and only once.

Indeed, given a cryptotext $c = w^e \bmod n$ for an integer plaintext w we have, since e is relative prime to $\phi(n)$, $order(c) = order(w)$. Let now d be such that $ed \equiv 1 \bmod (order(c))$, i.e. $ed = k \cdot order(w) + 1$ for some k. In such a case $c^d \equiv w^{ed} = w^{order(w)k+1} \equiv w \bmod n$ and in this way we can get the plaintext from the cryptotext.

[6] To design an RSA cryptosystem two large primes (512–1024 bits) are first chosen and $n = pq$, $\phi(n) = (p-1)(q-1)$ are computed. d is then chosen such that $gcd(d, \phi(n)) = 1$ and e is computed such that $ed \equiv 1 \bmod (\phi(n))$. n and e form the public key; p, q, d form the secret key. Encoding of a plaintext w: $c = w^e \bmod n$; decoding of the cryptotext c; $w = c^d \bmod n$. Encryption seems to be secure provided it is not feasible to get p and q from n—though it is not known if breaking RSA is as hard as integer factorization.

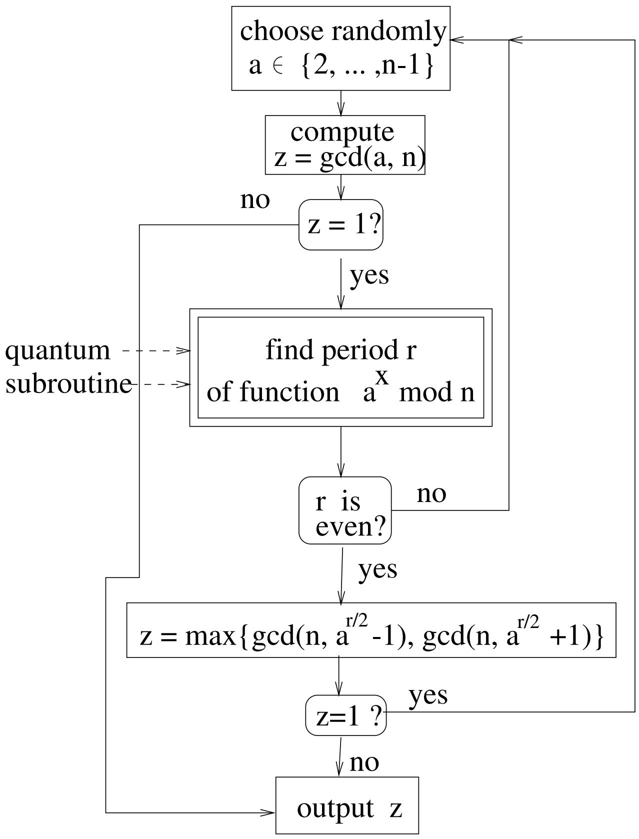


Figure 3.4: A general scheme of the Shor's factorization algorithm

Period computation algorithm

The basic idea of the algorithm is: (i) to create a state with the period we need to determine; (ii) to apply QFT to get rid of the offset; (iii) to extract the period by a computation.

Given an m bit integer n we choose first a $q \in \mathcal{O}(n^2)$ and start with five registers in states $|n, x, q, \mathbf{0}, \mathbf{0}\rangle$, where the last two registers have $\lceil \lg n \rceil$ qubits.

Phase I: construction of a state whose amplitudes have the same periods as the function $f_{n,x}$. An application of the Hadamard transformation to the fourth register yields the state

$$\frac{1}{\sqrt{q}} \sum_{a=0}^{q-1} |n, x, q, a, \mathbf{0}\rangle.$$

Using quantum parallelism we compute $x^a \bmod n$ for all a in one step and store the result in the fifth register. This yields

$$\frac{1}{\sqrt{q}} \sum_{a=0}^{q-1} |n, x, q, a, x^a \bmod n\rangle.^7$$

[7]It can be shown (see Shor, 1997), that modular exponentiation $x^a \bmod n$ can be done on a quantum computer in time $\mathcal{O}(\lg^2 n \lg \lg n \lg \lg \lg n)$ and space $\mathcal{O}(\lg n \lg \lg n \lg \lg \lg n)$.

As the next step we perform a measurement on the last register. Let y be the value obtained, i.e. $y = x^l \bmod n$ for the smallest l with this property. If r is the period of $f_{n,x}$, then $x^l \equiv x^{jr+l} \pmod{n}$ for all j. Therefore, the measurement actually selects the following sequence of a's values (in the fourth register), $l, l+r, l+2r, \ldots, l+Ar$, where A is the largest integer such that $l + Ar \leq q$, and $l \leq r$ has been chosen essentially randomly by the measurement. Since $l \leq r < n$ and $q = \Theta(n^2)$, we get $A \approx \frac{q}{r}$. The post-measurement state is then

$$|\phi_l\rangle = \frac{1}{\sqrt{A+1}} \sum_{j=0}^{A} |n, x, q, jr + l, y\rangle. \tag{3.16}$$

Since n, x, q and y will be fixed from now on, we will no longer write them down explicitly and therefore the previous state can be considered as having the form

$$|\phi_l\rangle = \frac{1}{\sqrt{A+1}} \sum_{j=0}^{A} |jr + l\rangle. \tag{3.17}$$

Phase II: amplitude amplification by QFT. From now on we consider in detail only a special case. Namely that $A = \frac{q}{r} - 1$. In such a case the state (3.17) can be written in the form

$$|\phi_l\rangle = \sqrt{\frac{r}{q}} \sum_{j=0}^{\frac{q}{r}-1} |jr + l\rangle$$

and after QFT_q is applied on $|\phi_l\rangle$ we get, by (3.12):

$$\begin{aligned} \mathrm{QFT}_q|\phi_l\rangle &= \frac{1}{\sqrt{q}} \sum_{c=0}^{q-1} \sqrt{\frac{r}{q}} \sum_{j=0}^{\frac{q}{r}-1} e^{2\pi i c(jr+l)/q} |c\rangle & (3.18) \\ &= \frac{\sqrt{r}}{q} \sum_{c=0}^{q-1} e^{2\pi i lc/q} \left(\sum_{j=0}^{\frac{q}{r}-1} e^{2\pi i jcr/q} \right) |c\rangle = \sum_{c=0}^{q-1} \alpha_c |c\rangle. & (3.19) \end{aligned}$$

If c is a multiple of $\frac{q}{r}$, then $e^{2\pi i jcr/q} = 1$ and if c is not a multiple of $\frac{q}{r}$, then

$$\sum_{j=0}^{\frac{q}{r}-1} e^{2\pi i jcr/q} = 0,$$

because the above sum is over a set of $\frac{q}{r}$ roots of unity equally spaced around the unit circle. Thus

$$\alpha_c = \begin{cases} \frac{1}{\sqrt{r}} e^{2\pi i lc/q}, & \text{if } c \text{ is a multiple of } \frac{q}{r}; \\ 0, & \text{otherwise;} \end{cases}$$

and therefore

$$|\phi_{out}\rangle = \mathrm{QFT}_q|\phi_l\rangle = \frac{1}{\sqrt{r}} \sum_{j=0}^{r-1} e^{2\pi i lc/q} |j\frac{q}{r}\rangle.$$

The key point is that the trouble-making offset l appears now in the phase factor $e^{2\pi i lc/q}$ and has no influence either on the probabilities or on the values in the register.

Phase III: period extraction. Each measurement of the state $|\phi_{out}\rangle$ therefore yields one of the multiples $c = \lambda\frac{q}{r}$, $\lambda \in \{0, 1, \ldots r-1\}$, where each λ is chosen with the same probability $\frac{1}{r}$. Observe also that in this case the QFT transforms a function with the period r (and an offset l) to a function with the period $\frac{q}{r}$. After each measurement we therefore know c and q and

$$\frac{c}{q} = \frac{\lambda}{r},$$

where λ is randomly chosen. If $gcd(\lambda, r) = 1$, then from q we can determine r by dividing q with $gcd(c, q)$. Since λ is chosen randomly, the probability that $gcd(\lambda, r) = 1$ is greater than $\Omega(\frac{1}{\lg\lg r})$. If the above computation is repeated $\mathcal{O}(\lg\lg r)$ times, then the success probability can be as close to 1 as desired and therefore r can be determined efficiently.[8]

In the general case, i.e., if $A \neq \frac{q}{r} - 1$, there is only a more sophisticated computation of the resulting probabilities and a more sophisticated way to determine r (using a continuous fraction method to extract the period from its approximation). No new "quantum-computing" ideas are involved. For details see Shor (1997).

Analysis of Shor's factorization algorithm

It is natural and important to analyze how many gates Shor's algorithm requires and how practical it could be. The first subproblem is how to implement quickly such key operators of Shor's algorithm as Fourier transform and modular exponentiation.

An efficient implementation of QFT has already been discussed. Quantum networks for basic arithmetical operations, including modular exponentiation, were designed by Vedral, Barenco and Ekert (1996)—see Section 2.3.4. A detailed analysis of the number of gates needed to implement Shor's algorithm was done by Beckman et al. (1996) and shows that about $300\lg n$ elementary quantum gates are needed. By Steane (1997), this implies that to factor a 130-digit number, which is within the limits of the current classical technology, would require two weeks on an ideal quantum computer with switching frequency 1 MHz. However, for 260-digit numbers, which are completely beyond the current technology, this would require only 16 times more time.

As already indicated in the flow diagram in Figure 3.4, it may happen that one has to use the period-finding subroutine several times when attempting to factorize an integer using the above Shor's quantum algorithm.

Shor's bounded-error algorithm has been shown, by Shor, to be polynomial but only in the expected time. No polynomial upper bound is known for the worst case.

Let us now look more schematically at particular steps of the Shor's order-computing subalgorithm in order to reveal the essence and some physical background of its particular steps.

Starting point is one basis state, $|0^{(n)}\rangle$, of the register, Figure 3.5a. After the first Hadamard transformation is applied we get an equally weighted superposition of all basis states, Figure 3.5b. After the parallel application of the transformation U_f we get a curve as sketched in Figure 3.5c (but discrete). Fourier transform makes out of it a nice tooth function, but with a pre-period, see Figure 3.5d. The last step, Figure 3.5e, removes the offset.

The curve in Figure 3.5e, as observed by Bennett (1998b), reminds us of what we get when depicting the amount of incoming particles in an n-slit experiment (for n going to

[8] As observed by Shor (1994) and shown by Cleve et al. (1998), the expected number of trials can be put down to a constant.

infinity), a natural generalization of the basic interference two-slit experiment presented in Section 1.5.

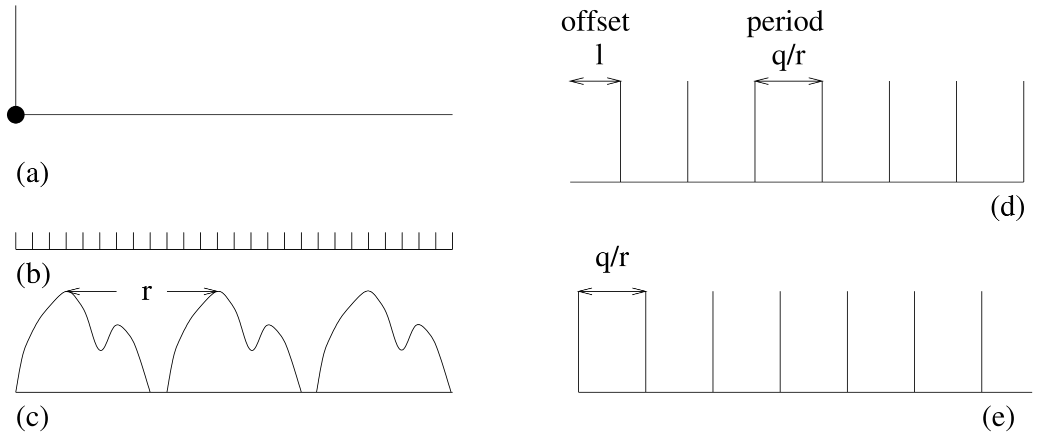


Figure 3.5: Representation of particular steps of Shor's order-finding algorithm—adapted from Bennett (1998b)

Remark 3.2.20 An interesting/important task is also to get realistic estimations of the amount of physical resources needed to implement Shor's factorization algorithm.

Beckman et al. (1996) explored possibilities to optimize networks for factorization with Shor's algorithm. Their analysis shows that an m-bit integer could be factorized in $\Theta(m^2 \lg m \lg\lg m)$ time using a quantum computer capable storing $5m + 1$ qubits. A network for modular exponentiation, which is the bottleneck of Shor's algorithm, can be implemented using $72\, m^3$ elementary quantum gates and an implementation using a linear ion trap technology would require about $400m^3$ laser pulses. A "proof-of-principle" network to factorize number 15 could be performed with only 6 trapped ions and 38 laser pulses.

On the other hand, Plenio and Knight (1996) gave much more pessimistic estimations. They found bounds $\mathcal{O}(m^8)$ for linear ion trap approach and $\mathcal{O}(m^6)$ for the cavity QED implementations—see page 310.

In practice it may be important to reduce both the number of qubits needed to solve a problem and the time for which qubits have to be isolated. This problem is discussed in detail by Mosca and Ekert (1998). For example, if certain hardware requirements are satisfied (one can measure and resume control qubits during the computation), then the number of qubits for modular exponentiation can be reduced to $3m$ (see Mosca and Ekert, 1998 and Zalka, 1998).

Perspectives of factorization

Hughes (1997) has analyzed the perspectives of factoring using on one side currently the most powerful factorization technique (Number Field Sieve method), on state-of-the-art workstations (assuming that the power of processors keeps increasing by Moore law), and on the other side a potential quantum computer with minimal clock speed of 100 MHz.

Table 3.6 shows estimations of factoring times on networks of 1000 workstations. Table 3.7 provides estimations for the number of qubits, gates and factoring time for a (potential) quantum computer.

Number of bits	1024	2048	4096
Factoring in 2006	10^5 years	5×10^{15} years	3×10^{29} years
Factoring in 2024	38 years	10^{12} years	7×10^{25} years
Factoring in 2042	3 days	3×10^8 years	2×10^{22} years

Figure 3.6: Future factoring times on networks of classical computers

size in bits	1024	2048	4096
number of qubits	5124	10244	20484
number of gates	3×10^9	2×10^{11}	2×10^{12}
factoring time	4.5 min.	36 min.	4.8 hours

Figure 3.7: Factoring on quantum computers

The above analyses show that using 2048-bit numbers seems to be safe for the next 50 years for classical computers. However, this is not so even for 4096-bit numbers if sufficiently powerful quantum computers will be available.

3.2.4 Shor's discrete logarithm algorithm

Just as for factorization, we present a polynomial time quantum algorithm for computing the discrete logarithm only for a special case. The general case again requires only more sophisticated classical computations.

The task is to determine an r such that $g^r \equiv x \pmod p$ given a prime p, a generator g of the multiplicative group $\mathbf{Z}_p^*$ and a $0 < x < p$. The special case we consider is that $p-1$ is smooth.

Using QFT_{p-1} twice, on the third and fourth subregister of the register $|x, g, \mathbf{0}, \mathbf{0}, \mathbf{0}\rangle$, we get

$$|\phi\rangle = \frac{1}{p-1} \sum_{a=0}^{p-2} \sum_{b=0}^{p-2} |x, g, a, b, \mathbf{0}\rangle,$$

a uniform distribution of all pairs (a, b), $0 \leq a, b \leq p-2$. By applying to $|\phi\rangle$ a uniform transformation mapping $(x, g, a, b, \mathbf{0}) \to (x, g, a, b, g^a x^{-b} \bmod p)$ we have

$$|\phi'\rangle = \frac{1}{p-1} \sum_{a=0}^{p-2} \sum_{b=0}^{p-2} |x, g, a, b, g^a x^{-b} \bmod p\rangle.$$

Since x, g will not be changed in the following computations we will not write them explicitly any longer.

As the next step we apply QFT_{p-1} on $|\phi'\rangle$ twice, once to map $a \to c$ with amplitude $\frac{1}{\sqrt{p-1}} e^{2\pi i ac/(p-1)}$ and once to map $b \to d$ with amplitude $\frac{1}{\sqrt{p-1}} e^{2\pi i bd/(p-1)}$. Since $p-1$ is smooth, this can be done in polynomial time. The resulting state is

$$|\phi_1\rangle = \frac{1}{(p-1)^2} \sum_{a,b,c,d=0}^{p-2} e^{\frac{2\pi i}{p-1}(ac+bd)} |c, d, g^a x^{-b} \bmod p\rangle.$$

Let us now measure the last register and let us determine the probability that what we get is the state $y \equiv g^k \bmod p$ for some k.

The probability equals the square of the absolute value of the sum of all amplitudes of all states having y in the last register, i.e., the sum of amplitudes over all a and b satisfying the equality $a - rb \equiv k \pmod{p-1}$ for some k. This is due to the fact that the computational paths interfere only if $y \equiv g^a(g^r)^{-b} = g^{a-rb} \equiv g^k \bmod p$.[9] Indeed, $g^{a-rb} = g^{j(p-1)+k} \equiv g^k$ (mod p)). The probability is therefore

$$\left| \frac{1}{(p-1)^2} \sum_{a,b,c,d=0}^{p-2} \{ e^{\frac{2\pi i}{p-1}(ac+bd)} \, | a - rb = k \} \right|^2 .$$

By substituting $a = k + rb$ in the above expression we get the probability in the form

$$\left| \frac{1}{(p-1)^2} \sum_{b,c,d=0}^{p-2} e^{\frac{2\pi i}{p-1}(kc+b(d+rc))} \right|^2 .$$

This probability is 0 if $d + rc \not\equiv 0 \bmod (p-1)$ because, as in the previous algorithm, in such a case the sum in the above expression is over a set of $(p-1)$st roots of unity equally spaced around the unit circle. On the other hand, if $d \equiv -rc \pmod{(p-1)}$, then the above sum does not depend on b and it is equal to $(p-1)^{-1}e^{(2\pi ikc/(p-1))}$. The square of its absolute value, the probability, is therefore $\frac{1}{(p-1)^2}$.

The measurements on the first and second register provide a (random) $c < p-1$ such that $d \equiv rc \pmod{p-1}$. If $gcd(c, p-1) = 1$, r can be obtained by division. As already mentioned, the probability that $gcd(c, p-1) = 1$ is $\Omega(\frac{1}{\lg p})$. Therefore, the number of computations needed to perform in order to get the probability close to 1 for finding r is polynomial in $\lg p$. ∎

3.2.5 The hidden subgroup problems

Shor provided bounded-error

polynomial time algorithms for two problems of great practical importance. This naturally initiated intensive search in two directions.

1. To design bounded-error polynomial time algorithms for other problems, especially for those problems of which factorization and discrete logarithm are special cases.

2. To determine for which of the above problems there exist also the exact polynomial time algorithms.

An important outcome in this direction has been an observation (for example, see Høyer, 1997), that all currently known quantum algorithms which run superpolynomially faster than their most efficient probabilistic classical counterparts solve a **hidden subgroup problem**.

The first results along these lines were due to Simon (1994), Shor (1994) and Kitaev (1995), who discovered a bounded-error polynomial time algorithm for the so-called Abelian subgroup stabilizer problem to which both integer factorization and discrete logarithm problem can be reduced. This problem is also a special case of the following problem:

[9] It follows from Fermat's theorem that if p is a prime and $a \equiv b \pmod{p-1}$, then $g^a \equiv g^b \pmod{p}$, because $g^{p-1} \equiv 1 \pmod{p}$.

Hidden subgroup problem

Given: An (efficiently computable) function $f : G \to R$, where G is a finite group and R a finite set.

Promise: There exists a subgroup $G_0 \leq G$ such that f is constant and distinct on the cosets of G_0.

Task: Find a generating set for G_0 (in polynomial time (in $\lg |G|$) number of calls to the oracle for f and in the overall polynomial time).[10]

Examples (Mosca and Ekert, 1998a)

Deutsch's problem, Section 3.1.1. $G = \mathbf{Z}_2$, $f : \{0,1\} \to \{0,1\}$, $x - y \in G_0 \Leftrightarrow f(x) = f(y)$. Decide whether $G_0 = \{0\}$ (and f is balanced) or $G_0 = \{0,1\}$ (and f is constant).

Simon's problem, Section 3.1.3. $G = \mathbf{Z}_2^n$, $f : G \to R$. $x - y \in G_0 \Leftrightarrow f(x) = f(y)$, $G_0 = \{0^{(n)}, s\}$, $s \in \mathbf{Z}_2^n$. Decide whether $G_0 = \{0^{(n)}\}$ or $G_0 = \{0^{(n)}, s\}$, with an $s \neq 0^{(n)}$ (and in the second case find s).

Order-finding problem, Section 3.2.1. $G = \mathbf{Z}$, $a \in \mathbf{N}$, $f(x) = a^x$, $x - y \in G_0 \Leftrightarrow f(x) = f(y)$, $G_0 = rk \,|\, k \in \mathbf{Z}$ for the smallest r such that $a^r = 1$. Find r.

Discrete logarithm problem, Section 3.2.4. $G = \mathbf{Z}_r \times \mathbf{Z}_r$, $a^r = 1$, $b = a^m$, $a, b \in \mathbf{N}$, $f(x, y) = a^x b^y$, $f(x_1, y_1) = f(x_2, y_2) \Leftrightarrow (x_1, y_1) - (x_2, y_2) \in G_0$. $G_0 = \{(k, -km) \,|\, k \in \mathbf{Z}_r\}$. Find G_0 (or m).

Abelian group stabilizer problem. Let G be a group of elements acting on a set R; that is if $a \in G$, then $a : R \to R$ and if $a, b \in G$, then $a(b(x)) = (ab)(x)$ for each $x \in R$. For $x \in R$ let $St_x = \{a \,|\, a \in G, a(x) = x\}$. St_x is a subgroup—the so-called stabilizer for x. For each $x \in R$ let $f_x : G \to R$ be such that $f_x(a) = a(x)$. The hidden subgroup corresponding to f_x is $G_0 = St_x$.

It is still not known whether the hidden subgroup problem has a bounded-errorl polynomial time algorithm also in the general case of non-abelian groups. This problem is of interest for various reasons. One of them is that the graph isomorphism problem is of such a type. (Graph isomorphism problem is reducible to finding a hidden subgroup of the symmetric group S_n.)

[11] Polynomial time algorithms for the hidden subgroup problem for certain types of non-abelian groups have been designed by Rötteler and Beth (1998) and Ettinger and Høyer (1998, 1999). In the first paper the problem is solved for certain semi-direct (namely wreath) group products; in the second paper for the so-called dihedral groups but the algorithm is

[10] A way to solve the problem is to show that in polynomial number of oracle calls (or time) the states corresponding to different candidate subgroups have exponentially small inner product and are therefore distinguishable.

[11] Indeed, let G be the disjoint union graph of connected graphs G_1 and G_2. The automorphism group H of G is a subgroup of the graph $S_n \wr S_2$ (where S_i is the symmetric group of the "order" i) and $\wr$ stands for the wreath product of groups). Knowledge of a set of generators for H is sufficient to decide isomorphism of G_1 and G_2. Ettinger and Høyer (1999) defined an observable on $l_2((S_n \wr S_2)^m)$, for any m, through a projection P such that if $|\psi\rangle$ is the tensor product of the coset states of H, then if G_1 and G_2 are (are not) isomorphic, then $\langle\psi|P|\psi\rangle = 1$ ($\geq 1 - \frac{n!}{2^m}$). It remains, as an open problem, to determine whether this observable is efficiently implementable.

In this connection of importance seems to be to determine for which non-abelian groups there are efficient QFT algorithms. For permutation non-commutative groups the existence of such an efficient algorithm was shown by Beals (1997).

polynomial only with respect to the number of quantum oracle calls—the classical postprocessing requires exponential time. In both cases the key subresult is an efficient mplementation of the Fourier transform for some non-abelian groups (for that see also Püschel et al. 1998). Ettinger et al. (1999) showed that the hidden subgroup problem can be solved in linear ($\mathcal{O}(\lg |G|)$) number of calls for any finite group G. However, their algorithm requires again exponential time for classical postprocessing. They have actually shown that there is a POV measurement that can distinguish among the possible states corresponding to different subgroups. An open problem remains whether there is such an POVM which can do the same, is efficiently implementable and also the postprocessing can be done efficiently.

There are two basic methods for solving the hidden subgroup problems. The first one is presented in Section 3.2 and follows an already familiar scheme: a Fourier transform, a function evaluation, again a Fourier transform and a sampling of the resulting superposition distribution.[12] ,[13],[14] The second approach, introduced by Kitaev (1995), is based on an estimation of eigenvalues of certain unitary operators. (For a detailed exposition of Kitaev's algorithm see Aharonov (1998).) These two approaches have been shown equivalent (see Mosca and Ekert, 1998). Shor's and Kitaev's algorithms are bounded-error polynomial time algorithms. An important question is for which hidden subgroup problems there exist also exact polynomial time algorithms. So far only one partial result is known. Brassard and Høyer (1997) have shown the existence of such an algorithm for a generalization of Simon's problem. By that, see also Brassard and Høyer, (1996), they established an exponential gap between the power of exact quantum computation and that of classical bounded-error randomized computation for decision problems.

The hidden subgroup problem for finitely generated Abelian subgroups is dealt with in detail by Mosca and Ekert (1998).

3.3 Quantum Searching and Counting

Shor's efficient quantum algorithms created large expectations concerning the power of quantum computing and it was expected that more of such algorithms would soon appear. However, despite the large effort of the quantum computing community the number of really useful quantum algorithms discovered remained for quite a while small.[15]

A significant breakthrough was achieved by Grover (1996). He developed a new quantum method that can be applied to the whole class of problems for which it is hard to find a solution but it is easy to check a to-be solution, as for a problem in **NP**. Grover's method, and its generalizations, can be used to show that quantum computers are faster than classical ones for a whole class of problems for which the lower bound on the efficiency of classical algorithms is known.

In the previous two sections the power of quantum algorithms was in their capability to extract efficiently a global property of a function, such as a period. In this section we deal

[12] This approach works when the group is represented as a product of a fixed number of cyclic groups.

[13] One can also say that finding hidden subgroups has been so far the main application of QFT.

[14] In both Shor's algorithms it was easy to make a use of such samples in the case QFT_p was performed for "nice" p, and we have skipped dealing with the hard cases with less nice, but more probable p. This technique has been improved by Hales and Hallgren (1998). They showed that distribution sampled after a QFT_p can be efficiently approximated by sampling after QFT_q is performed for a large range of q.

[15] Ozhigov (1998a) has shown that the set of classical algorithms for which a quantum speed-up exists has probability measure 0. However, it is far from clear whether this negative result has a significant implication concerning algorithmic problems one is really interested in.

with algorithms whose power is in a clever amplitude magnification in such a way that the desirable outcome has far the largest probability to come up at the measurement.

3.3.1 Grover's search algorithm

Grover discovered a fast quantum algorithm for the following search problem:

> In an unsorted database (list) of N items there is one satisfying a given condition; retrieve it.[16]

In addition, it is assumed that it is easy to verify whether a chosen element satisfies the condition. (This requirement is sometimes captured by the assumption that elements satisfying the condition are **marked**.)

Many algorithmic problems can be phrased as the search problems of the above type. For example, sorting of a sequence $\{a_i\}_{i=1}^n$ can be seen as searching for a permutation π of the set $\{1, \ldots, n\}$ such that $a_{\pi(i)} \leq a_{\pi(i+1)}$ for all $1 \leq i < n$. The Hamiltonian cycle problem is also of such a search type.

In order to present a slight modification of Grover's algorithm let us assume that a function $f : \{0,1\}^n \rightarrow \{0,1\}$ is given as a black box such that $f(x_0) = 1$ for a single x_0.

Grover's search algorithm

1. Using Hadamard transformation H_n create the state

$$|\phi\rangle = \frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} |x\rangle.$$

2. Apply the sign-changing operator V_f to $|\phi\rangle$ to provide

$$|\psi\rangle = \frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} (-1)^{f(x)} |x\rangle.$$

3. Apply the inversion about average operator $D_n = -H_n R_n^1 H_n$, page 72, to the state received in the previous step;

4. Iterate $\lceil \frac{\pi}{4}\sqrt{2^n} \rceil$ times steps 2 and 3, i.e., the transformation $G_n = -H_n R_n^1 H_n V_f$ (the so-called Grover's iterate, or G-iteration).

5. Measure the x-register.

The heart of Grover's algorithm is the "amplitude amplification process" performed in steps 2 and 3. The goal is to boost the probability of finding x_0.

The basic idea of Grover's algorithm is simple and it is illustrated in Figure 3.8. Let x_0 be the single string that meets the condition. To start with we create the quantum superposition of all strings with the same amplitude (see Figure 3.8a). The next task is to "massage" this superposition in such a way that the probability for x_0 increases and that of all other strings decreases (see Figure 3.8c). As the first step the sign at x_0 is inverted, using V_f transformation, see Figure 3.8b and then all terms are inverted about the average.

[16] For example, find in a telephone directory a person if her/his phone number is known.

By iterating these steps the absolute value of the amplitude at x_0 keeps increasing and that of other x keeps decreasing. After $\frac{\pi}{4}\sqrt{n}$ iteration of these steps the absolute value of the amplitude at x_0 is 1. The "miracle tool" to do that is again a double Hadamard transformation. In a proper moment, see Figure 3.8c, a measurement is performed. (However, one has to be very careful. If the proper moment to measure x is missed the probability at x_0 starts to decrease and at other strings to increase (see Figure 3.8d). Fortunately, there is a way to calculate quite well when the measurement should be performed.)[17]

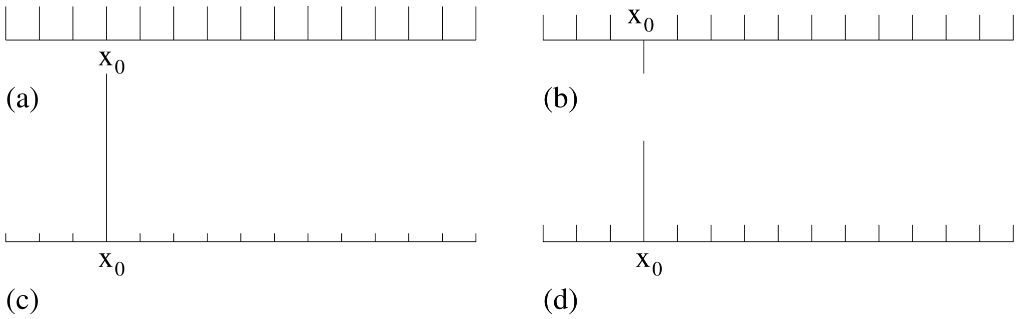


Figure 3.8: "Cooking" the solution with Grover's algorithm - 70%

Grover's algorithm is very simple and it is easy to verify that it works. Less trivial is to get a proper insight why it works. For that see a detailed analysis by Jozsa (1999). For example, as also discussed later, instead of the transformation H_n, essentially any unitary operator can be used.

It is clear that unless some structure of the problem is given, any classical algorithm to solve the unsorted databases search problem has to try to check all elements until the one with the desired property is found and for that $\frac{N}{2}$ checks are needed in the average. Grover's quantum algorithm presented above can do that in $\mathcal{O}(\sqrt{N})$ steps. [18]

An analysis by Bøyer et al. (1996) showed that after $\frac{\pi}{8}\sqrt{2^n}$ iterations the failure rate is $\frac{1}{2}$. Zalka (1997) has shown that Grover's algorithm is optimal, for finding a solution with probability at least $\frac{1}{2}$. Grover (1998a) gives a simple proof that at least $0.707\sqrt{2^n}$ queries are needed by any quantum search algorithm.[19]

Exercise 3.3.1 *Show that Grover's algorithm could, in principle, be used to break such cryptosystems as DES.*

[17] Steane (1997) refers to K. Fuchs for the following view of Grover's techniques: "It is like cooking a soufflé. The state is placed in the "quantum oven" and the desired answer rises slowly. You must open the oven at the right time, neither too soon nor too late, to guarantee success. Otherwise the soufflé will fall—the state collapses to the wrong answer."

[18] Grover's result does not put the unsorted database search problem into another complexity class. In spite of that it is remarkable that for such a surprisingly simple problem such an improvement can be obtained. In addition, such an algorithm could also be of large importance for cryptanalysis—to find a plaintext to a given cryptotext.

[19] Under special conditions (for example, if queries about multiple items are allowed or highly structured search problems are considered), see Grover (1997), Terhal and Smolin (1997) and Hogg (1998) a single database query is sufficient.

Remark 3.3.2 Experimental realization, using NMR technology, of the simplest interesting case of $n = 4$ and $f : \{0, 1, 2, 3\} \to \{-1, 1\}$ was reported by Chuang et al. (1998). In the case x_0 has the same probability to be $0, 1, 2$ or 3, the average number of classical queries is 2.25 and Grover's algorithm reduces to a single query.

Various generalizations and modifications of the original search problem started to be investigated soon. At first the requirement to find a unique solution was lifted. Then the problem of getting an exact and approximate counting of potential solutions has been considered.

Let us now formulate, following Mosca (1998b), the corresponding decision, search, counting and approximation problems and results formally. Let $f : X \to \{0, 1\}$ be a function. Define $X_1 = \{x \mid f(x) = 1\}$, $X_0 = \{x \mid f(x) = 0\}$ and $x_1 = |X_1|$, $x_0 = |X_0|$.

The **decision problem** associated with f is the problem of deciding whether $|X_1| = 0$. The **search problem**, or the **generation problem** is the problem to find an $x \in X_1$. The **counting problem** is to determine $|X_1|$. To **approximately count** X_1 with accuracy ε means to get an x such that

$$(1 - \varepsilon)x_1 \leq x \leq (1 + \varepsilon)x_1. \tag{3.20}$$

A **randomized approximation scheme** for f is a randomized algorithm that for any real parameter $\varepsilon > 0$ outputs an x such that (3.20) holds with probability $\frac{2}{3}$.

Figure 3.9 summarizes the main complexity results for the above algorithmic problems for classical and quantum computing, where $N = |X|$ and $t = |X_1|$.

problem	classical algorithms	quantum algorithms
decision	$\Theta(\frac{N}{t})$	$\Theta(\sqrt{\frac{N}{t}})$
search	$\Theta(\frac{N}{t})$	$\Theta(\sqrt{\frac{N}{t}}$
counting	$\Theta(N)$	$\Theta(\sqrt{t(N-t)}))$
approximation	$\Theta(\frac{N}{\varepsilon^2 t})$	$\Theta(\frac{1}{\varepsilon}\sqrt{\frac{N}{t}})$

Figure 3.9: Complexity of search and counting problems

For simplicity it is sufficient to consider the above problems for

$$f : \{0, 1\}^n \to \{0, 1\}$$

and this will be done in the following.

Remark 3.3.3 *Grover's search problem and his quantum algorithm allow a variety of stimulating interpretations. For example, Grover's algorithm can be seen as distinguishing, using* $\mathcal{O}(\sqrt{n})$ *oracle calls, between* n *functions* $g_j : \{1, 2, \ldots, n\} \to \{-1, 1\}$, $j = 1, 2, \ldots, n$, *such that*

$$g_j(j) = -1 \quad \textit{and} \quad g_j(k) = 1 \textit{ if } j \neq k.$$

A natural generalization of this problem was explored by Farhi et al. (1999). They showed, using methods of Sections 3.5.1 and 3.5.2, that if N *functions from* $\{1, 2, \ldots, n\}$ *to* $\{-1, 1\}$ *can be distinguished with* k *quantum oracle calls, then*

$$N \leq 1 + \binom{n}{1} + \ldots + \binom{n}{k}. \tag{3.21}$$

In addition, from the results of Example 3.1.2 it follows that all 2^n functions from $\{1,2,\ldots,n\} \rightarrow \{-1,1\}$ can be distinguished by k oracle calls with probability $(1+\binom{n}{1}+\ldots+\binom{n}{k})/2^n$.

Exercise 3.3.4 *Show that if both Alice and Bob have a subset of $\{1,2,\ldots,n\}$, then it is sufficient for them to communicate $\sqrt{n}$ qubits to determine whether their sets are disjoint or not.*

3.3.2 G-BBHT search algorithm

We shall discuss now a generalized search problem, namely that more than one x exists such that $f(x)=1$ and at first the special case of this problem—that the number $t=|\{x \mid f(x)=1\}|$ is known—and its solution, another modification of Grover's algorithm, and its analysis, due to Bøyer, Brassard, Høyer and Tapp (1996), called G-BBHT algorithm.

Algorithm 3.3.5 (G-BBHT search algorithm)

Step 1: Consider one n-qubit register in the initial state $|0^{(n)}\rangle$ and apply the Hadamard transformation to get

$$|\psi\rangle = \frac{1}{\sqrt{2^n}}\sum_{x=0}^{2^n-1}|x\rangle.$$

Step 2 Apply $\lceil \frac{\pi}{4}\sqrt{\frac{2^n}{t}}\rceil$ times the G-iteration $G=-H_nR_n^1H_nV_f$.
Step 3: Make a measurement of the x-register to get an x_0.
Step 4: Check if $f(x_0)=1$; if not go to Step 1.

Analysis of correctness and efficiency

Let $X_1=\{x \mid f(x)=1\}$, $X_0=\{x \mid f(x)=0\}$, and let, after the jth iteration of Step 2:

$$|\psi_j\rangle = k_j\sum_{x\in X_1}|x\rangle + l_j\sum_{x\in X_0}|x\rangle,$$

for suitable l_j, k_j with $k_0=\frac{1}{\sqrt{2^n}}$, $l_0=\frac{1}{\sqrt{2^n}}$.

In Step 2 the unitary operation, the inversion about the average, is performed that corresponds to the matrix

$$-H_n\cdot R_n^1\cdot H_n = \begin{pmatrix} -1+\frac{2}{2^n} & \frac{2}{2^n} & \ldots & \frac{2}{2^n} \\ \frac{2}{2^n} & -1+\frac{2}{2^n} & \ldots & \frac{2}{2^n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{2}{2^n} & \frac{2}{2^n} & \ldots & -1+\frac{2}{2^n} \end{pmatrix}.$$

For k_j and l_j the following recursive relations have been derived:

$$k_{j+1} = \frac{2^n-2t}{2^n}k_j + \frac{2(2^n-t)}{2^n}l_j, \quad (3.22)$$

$$l_{j+1} = \frac{2^n-2t}{2^n}l_j - \frac{2t}{2^n}k_j \quad (3.23)$$

with the solution

$$k_j = \frac{1}{\sqrt{t}}\sin((2j+1)\theta), \tag{3.24}$$

$$l_j = \frac{1}{\sqrt{2^n - t}}\cos((2j+1)\theta), \tag{3.25}$$

where $\sin^2\theta = \frac{t}{2^n}$.

> **Exercise 3.3.6** *(a) Derive recurrences* (3.22) *and* (3.23)*; (b) Show that k_j and l_j from* (3.24) *and* (3.25) *satisfy recurrences* (3.22) *and* (3.23)*.*

The aim and the art is to make such a number of steps that maximize k_j and minimize l_j. Let us therefore take j such that $\cos((2j+1)\theta) = 0$. This yields

$$j = \frac{\pi}{4\theta} - \frac{1}{2} + m\frac{\pi}{\theta} \qquad \text{for some } m \in \mathbf{Z}.$$

Since j has to be an integer, choose

$$j_0 = \left\lfloor \frac{\pi}{4\theta} \right\rfloor.$$

Because $\sin^2\theta = \frac{t}{2^n}$ we have $\theta \geq \sin\theta = \sqrt{\frac{t}{2^n}}$ and therefore $j_0 \leq \frac{\phi}{4}\sqrt{\frac{2^n}{t}} = \mathcal{O}(\sqrt{\frac{2^n}{t}})$. The stop condition, i.e., the value of j_0, can be determined classically.

The case $t = \frac{2^n}{4} = 2^{n-2}$ is of special interest because then $\sin^2\theta = \frac{t}{2^n} = \frac{1}{4}$ and therefore $\theta = \frac{\pi}{6}$ and $l_1 = \frac{1}{\sqrt{2^n - 2^{n-2}}}\cos(3\theta) = 0$. This implies that a solution is found after just one iteration step.

General case

Let us now consider the general case in which the number t of solutions is not known. Without loss of generality we assume that $1 \leq t \leq \frac{3}{4}2^n$. Indeed, if $t > \frac{3}{4}2^n$, then a simple algorithm, namely

Choose an x randomly and test if $f(x) = 1$

has a success probability $\geq \frac{3}{4}$, and by a repeated application of this algorithm we can easily get an exponentially small error probability.

Algorithm 3.3.7 (Quantum Search Algorithm II) *for* $t \leq \frac{3}{4}2^n$.

Step 0: Take $m = 1$ and $\lambda = \frac{6}{5}$.[20]
Step 1: Choose $j_0 \in [1, \ldots, m]$ randomly.
Step 2: Apply j_0 iterations of Grover's iterate to the state $\frac{1}{\sqrt{2^n}}\sum_{i=0}^{2^n-1}|i\rangle$.
Step 3: Make a measurement to obtain a value x_0.
Step 4: Test if $f(x_0) = 1$. If yes, stop. The problem is solved.

[20] Any $\lambda \in \{1, \frac{4}{3})$ could be used.

Step 5: Take $\min\{\lambda m, \sqrt{2^n}\}$ as a new value of m and go to Step 1.

An analysis shows that the expected computation time of the algorithm is $\mathcal{O}(\sqrt{\frac{2^n}{t}})$.

It has been shown, by Bøyer et al. (1998), that the optimal number of iterations is close to $0.58\sqrt{\frac{2^n}{t}}$, if $t \ll 2^n$ and the probability of success is close to 0.84.

Analysis of search algorithms

Concerning the analysis of the algorithms for searching an unordered list of n elements, of special interest is the relation between the error ε and the number T of queries. Let us summarize the main results (Buhrman and de Wolf, 1998):

1. ε can be an arbitrarily small constant using $\mathcal{O}(\sqrt{n})$ queries (Grover, 1996) but not using $o(\sqrt{n})$ queries (Zalka, 1997, Grover, 1998, and others);
2. ε can be at most $\frac{1}{2^{n^\alpha}}$ using $\mathcal{O}(n^{0.5+\alpha})$ queries (see Buhrman and de Wolf, 1998);
3. To achieve no error, i.e., $\varepsilon = 0$, n queries are needed (Buhrman et al. 1998, see also Section 3.5);
4. The following lower bound for ε is due to Buhrman and de Wolf (1998), where b is a fixed constant and $T < n$:
$$\varepsilon \in \Omega(e^{-\frac{4bT^2}{n} - \frac{8T}{\sqrt{n}}}).$$

Remark 3.3.8 It has been shown that for a closely related problem of **searching in an ordered list** we can gain a bit, but not too much, by using quantum algorithms. The best current lower bound, in terms of the number of comparisons, is $\frac{\lg n}{2\lg\lg n}$—for exact and bounded error computation—due to Farhi et al. (1998a). The upper bound of $\frac{3}{4}\lg n + \mathcal{O}(1)$ quantum queries on average, with probability $\frac{1}{2}$, is due to Röhrig (1998). Interesting and stimulating on the last result is that not Fourier but Haar transform (see Haar, 1910, and Høyer, 1997), was used. Farhi et al. (1999) presented an $0.55\lg n$ algorithm.

3.3.3 Minimum-finding algorithm

There is a variety of applications of Grover's algorithm. One of them, actually of quite a general character, is shown in the following example.

Example 3.3.9 (Minimum finding—Dürr and Høyer 1996)) *Let $s = s_1, \ldots, s_n$ be an unsorted sequence of distinct numbers. The task is to find an m such that s_m is the minimum of the sequence s.*

Classical probabilistic algorithms require $\mathcal{O}(n)$ comparisons. The "binary search" quantum algorithm presented below, which makes use of the G-BBHT algorithm, requires $\mathcal{O}(\sqrt{n})$ comparisons and iterations. Due to the general lower bound (Bennett et al. 1997), this is asymptotically optimal.

Algorithm 3.3.10 *1. Choose as the first threshold a random $y \in \{1, \ldots, n\}$.*

2. Repeat the following three steps until the total running time is more than $22.5\sqrt{n} + 1.4\lg^2 n$, then interupt.

(a) Initialize $|\psi_0\rangle = \frac{1}{\sqrt{m}}\sum_{i=1}^{m}|i\rangle|y\rangle$, and consider an index i as **marked** *if $s_i < s_y$.*

(b) *Apply the G-BBHT search algorithm to the first register to find a marked element.*

(c) *Measure the first register. If y' is the outcome and $s_{y'} < s_y$ take as the new threshold index y'.*

3. *Return y.*

Theorem 3.3.11 *The minimum-finding algorithm presented in the example above finds the minimum with probability at least $1 - \frac{1}{2}$ if the measurement is done after a total number of $\mathcal{O}(\sqrt{n})$ iterations.*

Proof. Let us denote by $p(M, t)$ the probability that when the above algorithm searches the minimum among M items, one of the threshold indices chosen will be that of the element s of the rank[21] $t \in \{1, \dots, M\}$. On the basis of the identity

$$Pr(k+1, t) = \frac{1}{k+1} + \frac{1}{k+1} \sum_{r>t}^{k+1} Pr(r-1, t)$$

one can show by induction that $Pr(M, t) = \frac{1}{t}$, if $t \leq M$ and $Pr(M, t) = 0$, otherwise.

As shown above, by Bøyer et al. (1996), the expected number of iterations by the G-BBHT algorithm to search for one of $t-1$ marked items among n items is at most $4.5\sqrt{\frac{n}{t-1}}$. The expected total number of time steps of the above algorithm during the execution of Step 2b is therefore

$$\sum_{t=1}^{n} \frac{1}{t} 4.5 \sqrt{\frac{n}{t-1}} \leq 4.5\sqrt{n} \left(\frac{1}{2} + \sum_{t=2}^{n-1} t^{-\frac{3}{2}} \right) \leq \frac{45}{4}\sqrt{n}.$$

The expected total number of time steps of the Phase 2a before the minimum is found is therefore at most

$$\sum_{t=2}^{n} Pr(n, t) \lg n = (H_n - 1) \lg n \leq \ln n \lg n \leq \frac{7}{10} lg^2 n,$$

where H_n denotes the nth Harmonic number.

The expected total time used by the algorithm is therefore at most

$$m = \frac{45}{4}\sqrt{n} + \frac{7}{10} \lg^2 n. \tag{3.26}$$

Of course, by using algorithm c times the probability of success can be improved to at least $1 - \frac{1}{2^c}$. Using Markov inequality we have that after $2m$ steps the minimum is found with probability at least $\frac{1}{2}$ and therefore Theorem 3.3.11 holds.

Exercise 3.3.12 *(a) Design a minimum-search algorithm for the case that not all elements of the sequence are distinct; (b) reformulate the above algorithm to find extreme points of functions.*

[21] The rank of an element is its position if the sequence is sorted.

3.3.4 Generalizations and modifications of search problems

Grover's quantum search idea has been used to design fast quantum algorithms for a variety of problems. Some of them will be briefly discussed in this section.

G-BBHT algorithm for arbitrary initial amplitude distribution

In G-BBHT algorithm the initial distribution is uniform—all marked and unmarked states have the same amplitude. Generalization of this algorithm to the case of arbitrary initial distribution is given by Biron et al. (1998). The need to deal with such a case arises when the search is used as a subroutine of larger quantum computations. They derived an expression for the optimal measurement time, $\mathcal{O}(\sqrt{\frac{n}{t}})$, where n is the total number of elements and t the number of marked ones, which depends on the initial amplitudes of the marked and unmarked states. (However, for an arbitrary initial state the probability of success even after the optimal number of iterations might be very low (see also the analysis in Mosca, 1998).)

Quantum speeding of heuristics

For many computationally hard problems there are heuristics that work in practice quite well. We show now, following Brassard, Høyer and Tapp (1998a), how one can use Grover's technique to speed-up heuristics for black box searching. The term **heuristic** is used here in a very general sense, as a randomized polynomial time algorithm that provides what it is searched for with some probability. More formally:

Let $\mathcal{F}$ be a family of functions, $F : X \to \{0,1\}$ for a set X. A heuristic is a function $H : \mathcal{F} \times R \to X$ for a finite set R. For every function $F \in \mathcal{F}$ let $t_F = |F^{-1}(1)|$ and $h_F = |\{r \in R \,|\, F(H(F,r)) = 1\}|$. The heuristic H is called efficient for an F if $\frac{h_F}{|R|} > \frac{t_F}{|X|}$.

Theorem 3.3.13 *Let F be a search problem chosen from $\mathcal{F}$ according to some probability distribution. If, using a heuristic H, a solution to F can be found in expected time t, then there is a quantum algorithm to find a solution in expected time $\mathcal{O}(\sqrt{t})$.*

Proof. We combine the G-BBHT search algorithm with the heuristic H. Define $H'(r) = F(H(F,r))$ and $x = H(F, \text{G-BBHT}(H'))$. Hence $F(x) = 1$. By the results of Section 3.3.2, for each $F \in \mathcal{F}$ we have an expected running time $\theta(\sqrt{\frac{|R|}{h_F}})$. If p_F is the probability that f is chosen, then $\sum_{F\in\mathcal{F}} p_F = 1$. The expected computation time is then of the order

$$\sum_{F\in\mathcal{F}} \sqrt{\frac{|R|}{h_F}} p_F = \sum_{F\in\mathcal{F}} \sqrt{\frac{|R|}{h_F} p_F}\sqrt{p_F} \leq \left(\sum_{F\in\mathcal{F}} \frac{|R|}{h_F} p_F\right)^{1/2} \left(\sum_{F\in\mathcal{F}} p_F\right)^{1/2} = \left(\sum_{F\in\mathcal{F}} \frac{|R|}{h_F} p_F\right)^{1/2}$$

by the Cauchy–Schwarz inequality.

Quantum counting

We present a way to perform an approximate counting of the number of solutions of the equation $f(x) = 1$, where $f : \{0,1\}^n \to \{0,1\}$, due to Brassard, Høyer and Tapp (1998a)[22], in which one combines basic ideas of Grover's and Shor's a algorithms.

[22] Another approach, due to Mosca (1998) is based on the analysis of eigenvalues of Grover's iterate.

Basic idea: in the G-BBHT algorithm the amplitudes of the sets A_0 and A_1 vary with the number of iterations, according to a periodic function. We know from Section 3.3.2 that this period is directly related to the size of these sets. An estimation of the common period, using Fourier analysis, provides an approximation of the size of sets A_0 and A_1.

The quantum algorithm COUNT presented below to provide the approximate counting has two parameters: the function f, given as a black box, and an integer $p = 2^k$, for some k, to determine the precision of the approximation. The algorithm uses two transformations

$$\begin{aligned} C_f &: |m, \psi\rangle \rightarrow |m, G_f^{(m)}\psi\rangle, \\ F_p &: |k\rangle \rightarrow \frac{1}{\sqrt{p}} \sum_{l=0}^{p-1} e^{2\pi i kl/p} |l\rangle, \end{aligned}$$

where $G_f = -H_n R_n^1 H_n V_f$ is the Grover's iterate for f, and $G_f^{(m)}$ denotes its mth iteration.

Algorithm 3.3.14 *COUNT*(f, p)

1. $|\psi_0\rangle \leftarrow (H_n \otimes H_n)|0^{(n)}, 0^{(n)}\rangle$;
2. $|\psi_1\rangle \leftarrow C_f|\psi_0\rangle$;
3. $|\psi_2\rangle \leftarrow F_p \otimes I|\psi_1\rangle$;
4. $f \leftarrow$ **if** measure of $|\psi_2\rangle > \frac{p}{2}$ **then** $p - \mathcal{M}(|\psi_2\rangle)$ **else** $\mathcal{M}(|\psi_2\rangle)$;
5. output $\leftarrow 2^n \sin^2(\frac{f\pi}{p})$.

An upper bound on m is needed in order to be able to apply C_f even if the first argument is in a quantum superposition—this is the aim of the parameter p. $\mathcal{M}(|\psi_2\rangle)$ denotes the result of the measurement of $|\psi_2\rangle$.

The following theorem states the limits on the efficiency of the algorithm COUNT.

Theorem 3.3.15 *Let* $f : \{0, 1, \ldots, 2^n - 1\} \rightarrow \{0, 1\}$ *be a Boolean function* $t = |f^{-1}(1)| \leq 2^{n-1}$ *and* $\bar{t}$ *be the output of COUNT*(f, p), *where* $p \geq 4$, *then with probability at least* $\frac{8}{\pi^2}$, $|t - \bar{t}| < \frac{2\pi}{p}\sqrt{t2^n} + \frac{\pi^2}{p^2}2^n$.

The collision problem

A collision problem for a function $f : X \rightarrow Y$ is to find two different $x \neq y$ such that $f(x) = f(y)$ under the promise that such a pair exists. A function f is defined to be r-to-one if every element in its image has exactly r pre-images. Brassard, Høyer and Tapp (1998) found a fast quantum algorithm for the collision problem for r-to-one functions working in time $\mathcal{O}((n/r)^{1/3})$, where n is the cardinality of the domain.

Verification of the predicate calculus formulas

The quantum algorithm presented in Section 3.3.1 can be seen as a method to verify formulas $\exists x P(x)$, where $x \in \{0, 1\}^n$ and P is a predicate determined by an oracle providing the answer **true** or **false** for a given input x.

Once the search problem is interpreted this way it is natural to ask whether there is a fast quantum algorithm to verify formulas of the more general type:

$$\forall x_1 \exists y_1 \forall x_2 \exists y_2 \ldots \forall x_k \exists y_k P(x_1, y_1, x_2, y_2, \ldots, x_k, y_k). \tag{3.27}$$

This can be done with bounded error probability, in time $\mathcal{O}(\sqrt{N})$, $N = 2^{2k}$, where $\mathcal{O}(N)$ is time any classical algorithm neeeds. The result was obtained by Buhrman et al. (1998) in a more general setting and by Ozhigov (1998). Namely, they showed that one can compute in time $\mathcal{O}(\sqrt{2^n}n^{d-1})$ the function

$$\mathrm{SIGMA}_d(f) = \bigvee_{x^{(1)}\in\{0,1\}^{m_1}} \left(\cdots\left(\bigwedge_{x^{(d)}\in\{0,1\}^{m_d}} f(x^{(1)},\ldots,x^{(d)})\right)\cdots\right),$$

where $m_1,\ldots,m_d$ are implicit parameters and $m_1+\ldots+m_d = n$.

3.4 Methodologies to Design Quantum Algorithms

A general method, usually referred to as **amplitude amplification**, to boost the success probability of quantum algorithms has already emerged. We present here two approaches. Their common feature is: (a) to use a generalization of Grover's iterate in which the Hadamard transformation is replaced by arbitrary unitary transformation and, (b) the fact that iterative applications of these transformations to vectors of a certain small-dimensional subspace preserve this subspace.

3.4.1 Amplitude amplification–boosting search probabilities

The first approach to amplitude amplification presented here is due to Brassard et al. (1998a).

Let $H = H_1 \otimes H_2$ be a composed Hilbert space. A pure state $|\phi\rangle$ of H can be written in the form

$$|\phi\rangle = \sum_{i\in Z} \alpha_i|i,\psi_i\rangle,$$

where $\{|i\rangle \mid i \in Z\}$ forms a basis of H_1, and ψ_i are states of H_2. Every $f : Z \to \{0,1\}$ induces two orthogonal subspaces of H spanned by the vectors

$$|\phi_1\rangle = \sum_{i\in f^{-1}(1)} \alpha_i|i,\psi_i\rangle, \qquad |\phi_0\rangle = \sum_{i\in f^{-1}(0)} \alpha_i|i,\psi_i\rangle.$$

If we call states $|i,\cdot\rangle$ good provided $i \in f^{-1}(1)$, and bad otherwise, then $|\phi_1\rangle$ ($|\phi_0\rangle$) is a projection of $|\phi\rangle$ into the subspaces spanned by good (bad) state. Moreover, the probability that by measuring $|\phi\rangle$ we get a good (bad) state is $\langle\phi_1|\phi_1\rangle$ ($\langle\phi_0|\phi_0\rangle$).

Let U be any unitary transformation on H. The key tool for the amplitude amplification procedure to be presented below is the unitary operator

$$Q = Q(U,f,p,q) = -US_0^pU^{-1}S_f^q,$$

where p, q are complex numbers such that $|p| = |q| = 1$. The operator S_f^q conditionally changes the phase by the phase factor q as follows

$$\begin{aligned} S_f^q|i,\cdot\rangle &\to q|i,\cdot\rangle \text{ if } f(i)=1; \\ S_f^q|i,\cdot\rangle &\to |i,\cdot\rangle \text{ if } f(i)=0. \end{aligned}$$

The operator S_0^p changes the phase of a state by a factor p if and only if the first register holds 0. Using this notation the original Grover's iterate has the form $Q(H_n, f, -1, -1)$.

The following properties of the operators $US_0^pU^{-1}$ and Q are easy to verify.

Lemma 3.4.1 *If $|\phi\rangle$ is any state of H, then*

$$\begin{aligned} US_0^pU^{-1}|\phi\rangle &= |\phi\rangle - (1-p)\langle\phi|U|0\rangle^* U|0\rangle; \\ Q|\phi_1\rangle &= -q|\phi_1\rangle + q(1-p)\langle\phi_1|U|0\rangle^* U|0\rangle; \\ Q|\phi_0\rangle &= -|\phi_0\rangle + (1-p)\langle\phi_0|U|0\rangle^* U|0\rangle. \end{aligned}$$

In particular, if $|\psi\rangle = U|0\rangle$ and $|\psi\rangle$ is decomposed, similarly as above, into $|\psi_1\rangle$ and $|\psi_0\rangle$, then it holds

Lemma 3.4.2

$$\begin{aligned} Q|\psi_1\rangle &= q((1-p)a-1)|\psi_1\rangle + q(1-p)a|\psi_0\rangle \\ Q|\psi_0\rangle &= -((1-p)a+p)|\psi_0\rangle + (1-p)(1-a)|\psi_1\rangle, \end{aligned}$$

where $a = \langle\psi_0|\psi_0\rangle$.

Exercise 3.4.3 *Prove in detail Lemmas 3.4.1 and 3.4.2.*

As a consequence of the above lemma, for any vector $|\psi\rangle$ the subspace spanned by vectors $|\phi_0\rangle$, $|\phi_1\rangle$, $|\psi_0\rangle$ and $|\psi_1\rangle$ is invariant under the transformation Q. For the special case $p = q = -1$ simpler relations are obtained:

Lemma 3.4.4 *Let $U|0\rangle = |\psi\rangle = |\psi_0\rangle + |\psi_1\rangle$, $Q = Q(U, f, -1, -1)$, then*

$$Q|\psi_1\rangle = (1-2a)|\psi_1\rangle - 2a|\psi_0\rangle, \tag{3.28}$$

$$Q|\psi_0\rangle = (1-2a)|\psi_0\rangle + 2b|\psi_1\rangle, \tag{3.29}$$

where $a = \langle\psi_1|\psi_1\rangle$, $b = 1 - a = \langle\psi_0|\psi_0\rangle$.

The recurrences (3.28) and (3.29) have actually been solved in Section 3.3.2 with the result

Theorem 3.4.5 *Let $U|0\rangle = |\psi\rangle = |\psi_0\rangle + |\psi_1\rangle$ and $Q = Q(U, f, -1, -1)$. Then, for $j \geq 0$,*

$$Q^jU|0\rangle = k_j|\psi_1\rangle + l_j|\psi_0\rangle,$$

where

$$k_j = \frac{1}{\sqrt{a}}\sin((2j+1)\theta) \quad \textit{and} \quad l_j = \frac{1}{\sqrt{1-a}}\cos((2j+1)\theta)$$

and θ is such that $\sin^2\theta = a = \langle\psi_1|\psi_1\rangle$, $0 \leq \theta \leq \frac{\phi}{2}$.

In Theorem 3.4.5 U can be any quantum algorithm that uses no measurement. This way we therefore get a general method to increase probability of the success at searching and let us therefore analyze this situation in more detail.

If $\mathcal{A}|0\rangle$ is computed, then a is the success probability to get into a good state. On the other side, if the transformation $Q^j\mathcal{A}$ is applied, then the success probability is $ak_j^2 = \sin^2((2j+1)\theta)$. One can achieve high probability of success by chosing j such that $\sin^2((2j+1)\theta) \approx 1$. However, for that one needs to know θ which depends in turn on a. In case $a > 0$ and $j = \sqrt{\frac{\pi}{4\theta}}$, we have $ak_j^2 \geq 1 - a$ and therefore it holds.

Theorem 3.4.6 (Quadratic speed–up theorem) *Let $\mathcal{A}$ be a quantum algorithm that uses no measurement, and let $f : Z \to \{0,1\}$ be a Boolean function. Let the initial success probability a and the angle θ be defined as in Theorem 3.4.5. Let $a > 0$ and $m = \lceil \frac{\pi}{4\theta} \rceil$. If $Q^m \mathcal{A}|0\rangle$ is computed and the resulting state is measured, then the outcome is good with probability at least* $\max\{0, 1-a\}$.

The case that no knowledge of a is known can be treated in a way similar to that discussed in Section 3.3.2.

3.4.2 Amplitude amplification—speeding of the states searching

A slightly different framework for the speed-up of quantum algorithms by amplitude amplification, developed Grover (1998, 1998b).

Let $|\phi_1\rangle$ and $|\phi_2\rangle$ be two basis states and U an arbitrary unitary transformation of a Hilbert space H. If a^U_{st} is the amplitude of reaching the state $|\phi_t\rangle$ when U is applied to the state $|\phi_s\rangle$, then we need $\Omega(\frac{1}{|a^U_{st}|^2})$ repetitions of such steps (to apply U and to make a measurement immediately) before a single success. This can be improved to $\mathcal{O}(\frac{1}{|a^U_{st}|})$ steps, as shown below, using similar amplitude amplification technique as in Section 3.4.1.

Let us denote by I_x the unitary operator that switches the sign of the amplitude of the state $|\phi_x\rangle$, and consider the transformation $Q = -I_s U^{-1} I_t U$ and two states, $|\phi_s\rangle$ and $U^{-1}|\phi_t\rangle$. It is easy to verify that it holds

$$\begin{aligned} Q|\phi_s\rangle &= (1 - 4|a^U_{st}|^2)|\phi_s\rangle + 2a^U_{st} U^{-1}|\phi_t\rangle, && (3.30)\\ QU^{-1}|\phi_t\rangle &= -2a^{U^*}_{ts}|\phi_s\rangle + U^{-1}|\phi_t\rangle, && (3.31)\end{aligned}$$

and therefore we have basically the same recurrences (with different parameters) as in Sections 3.3.2 and 3.4.1.

Exercise 3.4.7 *Derive in detail recurrences 3.30 and 3.31.*

By applying a similar procedure as in Section 3.3.2 one can show that after T iterations of Q we get the superposition $a_s|\phi_s\rangle + a_t U^{-1}|\phi_t\rangle$, where $\alpha_s = \cos(2T|\alpha^U_{st}|)$, $|\alpha_t| = |\sin(2T|a^U_{st}|)|$. If $T = \frac{\pi}{4|a^U_{st}|}$, then we get superposition $U^{-1}|\phi_t\rangle$ and by an application of U we reach the target state $|\phi_t\rangle$. Therefore in $\mathcal{O}(\frac{1}{|a^U_{st}|})$ steps we reach $|\phi_t\rangle$ when starting in $|\phi_s\rangle$.

3.4.3 Case studies

The methodology presented in Section 3.4.2 will now be used (see Grover, 1998b), to sketch the design of fast quantum algorithms for several problems.

Search of any basis state from $|0^{(n)}\rangle$ in H_{2^n}

Let the state to reach from the initial state $|0^{(n)}\rangle$ be an n-qubit basis state $|t\rangle$, $1 \le t < 2^n$, as the target state.

If the Hadamard transformation H_n is used, then the amplitude $a^{H_n}_{\mathbf{0}t}$ is $\frac{1}{\sqrt{2^n}}$. The method presented in Section 3.4.2 provides an algorithm to reach the target state in $\mathcal{O}(\frac{1}{|a^{H_n}_{0t}|}) = \mathcal{O}(\sqrt{2^n})$ steps. In this case the operator $-H_n I_0 H_n$ is simply the inversion about the average operator as introduced on page 72.

Search of any basis state from any basis state

If Hadamard transformation H_n is again used, then $a_{st}^{H_n} = \frac{1}{\sqrt{2^n}}$ and the methodology of Section 3.4.2 provides us with an $\mathcal{O}(\sqrt{2^n})$ algorithm. The iterative operator used is $-I_s H_n I_t H_n$ and the main difference with respect to the previous case is that we no longer have the inversion about the average interpretation of the iteration step.

Search when a basis state near the target basis state is given

The aim is to reach in H_{2^n} a basis state $|t\rangle$ starting from the initial state $|s\rangle$ provided a state $|q\rangle$ is given such that q and t differ in k bits and k is known.

This time instead of the Hadamard transformation, the transformation $W_\alpha^n = \bigoplus_{i=1}^n W_\alpha$ is applied to each of n qubits, where

$$W_\alpha = \begin{pmatrix} \sqrt{1-\frac{1}{\alpha}} & \frac{1}{\sqrt{\alpha}} \\ \frac{1}{\sqrt{\alpha}} & -\sqrt{1-\frac{1}{\alpha}} \end{pmatrix}.$$

In this case $|a_{st}^{W_\alpha^n}| = (1-\frac{1}{\alpha})^{\frac{n-k}{2}}(\frac{1}{\alpha})^{\frac{k}{2}}$ and this value is maximal if $\alpha = \frac{n}{k}$. The algorithm presented in Section 3.4.2 provides the solution in time $\mathcal{O}(\frac{1}{|a_{st}^{W_\alpha^n}|})$.

It remains to find out how good this solution is with respect to the classical case of doing exhaustive search. The size of the search space is now $\binom{n}{k}$. Using Stirling' approximation for factorial we have $\lg\binom{n}{k} \approx n \lg \frac{n}{n-k} - k \lg \frac{k}{n-k}$. On the other hand, for $\alpha = \frac{n}{k}$, $\lg|a_{st}^{W_n^\alpha}| = \frac{n}{2}\lg\frac{n-k}{n} - \frac{k}{2}\lg\frac{n-k}{k}$ and therefore the number of steps of the quantum algorithm just derived is about the square root of the number of steps of the classical exhaustive search algorithm.

Estimation of the median and mean

Grover (1998b) has derived using the methodology of Section 3.4.2 new and significantly simpler algorithms than previously (Grover, 1996a, 1997b), to estimate, to a given precision, the median and the mean of a sequence.

3.5 Limitations of Quantum Algorithms

After having demonstrated several cases in which quantum algorithms outperform asymptotically classical ones it is of interest and importance to explore circumstances and cases in which no asymptotic quantum speed-up is possible. In this section we discuss such limitations in two ways. At first we give a direct proof that for a specific classical computation task no quantum speed-up is possible. Afterwards we present a general methodology to prove lower bounds which implies, for example, that in the area of black box computations ony promise problems may have larger than polynomial quantum speed-up.

3.5.1 No quantum speed-up for the parity function

Grover's result suggested the possibility that any problem involving computation of a function with n inputs could be solved in $\mathcal{O}(\sqrt{n})$ time. In this section we show, following Farhi et al. (1998), that this is not so for the following version of the **parity problem**.

Let $\mathcal{F}_n = \{f \mid f : \{1, \ldots, n\} \to \{-1, 1\}\}$. The parity of a function $f \in \mathcal{F}_n$ is defined by

$$par(f) = \prod_{x=1}^{n} f(x).$$

If $f \in \mathcal{F}_n$ is given as a black box, then any classical deterministic or randomized algorithm to compute $par(f)$ needs n function calls of f, because all values of f matter.

Let A_f be the unitary transformation such that for $x \in \{1, \ldots, n\}$, $y \in \{-1, 1\}$,

$$A_f|x, y\rangle = |x, yf(x)\rangle,$$

where the second register is considered as a qubit taking value -1 or 1. If we now define

$$|x, 0'\rangle = \frac{1}{\sqrt{2}}(|x, 1\rangle + |x, -1\rangle), \quad |x, 1'\rangle = \frac{1}{\sqrt{2}}(|x, -1\rangle - |x, 1\rangle),$$

then for $y \in \{0, 1\}$, if we denote $f^y(x) = (f(x))^y$,

$$A_f|x, y'\rangle = f^y(x)|x, y'\rangle \tag{3.32}$$

and therefore, in the basis $\{|x, 0'\rangle, |x, 1'\rangle\}$, the quantum operator A_f is simply a multiplication by $f^y(x)$.

Our aim is to show that any quantum algorithm computing $par(f)$ has to use at least $\frac{n}{2}$ calls of f. In order to do that we need to consider also the cases where a quantum algorithm computing $par(f)$ works in a larger Hilbert space than $(2n)$-dimensional Hilbert space spanned by vectors $\{|x, y\rangle \mid 1 \le x \le n, y \in \{0, 1\}\}$. We shall therefore assume that a quantum algorithm $\mathcal{A}$ to compute $par(f)$ is given which works in a Hilbert space $H_{n,Z}$ spanned by the vectors $\{|x, y, z\rangle \mid 1 \le x \le n, y \in \{0, 1\}, 1 \le z \le Z\}$, for some Z.

$\mathcal{A}$ can be seen as a sequence of unitary operators that act on an initial vector $|\psi_0\rangle$ and produce a final vector $|\psi_f\rangle$ in such a way that there is a projection operator (observable) P corresponding to a decomposition of the underlying Hilbert space into two orthogonal subspaces, such that if P is applied to $|\psi_f\rangle$, then either the value 0 is obtained, corresponding to the case $par(f) = -1$ or the value 1, corresponding to $par(f) = 1$.

The algorithm $\mathcal{A}$ will be considered as computing $par(f)$ with error ε, if for the expectation value of P, with respect to $|\psi_f\rangle$, it holds

$$\langle\psi_f|P|\psi_f\rangle \begin{cases} \ge \frac{1}{2} + \varepsilon, & \text{if } par(f) = 1; \\ \le \frac{1}{2} - \varepsilon, & \text{if } par(f) = -1. \end{cases} \tag{3.33}$$

The operator A_f can clearly be expressed (as a generalization of (3.32)), in the form

$$A_f = \sum_{x=1}^{n} \sum_{y=0}^{1} f^y(x) P_x P_y,$$

where, for any $1 \le x \le n$, P_x is the projection operator into the space spanned by the vectors $\{|x, y', z\rangle \mid y \in \{0, 1\}, 1 \le z \le Z\}$ and P_y is the projection into the subspace with the basis $\{|x, y', z\rangle \mid 1 \le x \le n, 1 \le z \le Z\}$. (This is actually the spectral representation of the query operator A_f.)

If the algorithm $\mathcal{A}$ contains k applications of A_f, then it can be seen as having the form

$$\mathcal{A} = U_k A_f U_{k-1} A_f \ldots A_f U_1 A_f, \tag{3.34}$$

where U_i are unitary operators which do not depend on f. Therefore,

$$|\psi_f\rangle = \mathcal{A}|\psi_0\rangle = U_k A_f U_{k-1} A_f \dots A_f U_1 A_f |\psi_0\rangle.$$

In such a case

$$\langle\psi_f|P|\psi_f\rangle = \sum_{x_1=1}^{n}\sum_{q_1=0}^{1}\sum_{x_2=1}^{n}\sum_{q_2=0}^{1}\dots\sum_{x_{2k}=1}^{n}\sum_{q_{2k}=0}^{1}\Phi(x_1,q_1,x_2,q_2,\dots,x_{2k},q_{2k})\prod_{i=1}^{2k} f^{q_i}(x_i),$$

where

$$\Phi(x_1,q_1,\dots,x_{2k},q_{2k}) = \langle\psi_0|P_{x_1}P_{q_1}U_1^*\dots U_k^* P_f U_k \dots U_1 P_{x_{2k}} P_{q_{2k}}|\psi_0\rangle$$

does not depend on f.

There are 2^n functions $f \in \mathcal{F}_n$. The summation over all such functions yields

$$\sum_{f\in\mathcal{F}_n} par(f)\langle\psi_f|P|\psi_f\rangle = \sum_{f\in\mathcal{F}_n}\sum_{x_1=1}^{n}\sum_{q_1=0}^{1}\dots\sum_{x_{2k}=1}^{n}\sum_{q_{2k}=0}^{1}\Phi(x_1,q_1,\dots,x_{2k},q_{2k})\prod_{i=1}^{2k} f^{q_i}(x_i)\prod_{y=1}^{n} f(y). \tag{3.35}$$

Since the summation is over all functions $f \in \mathcal{F}_n$, it holds

$$\sum_{f\in\mathcal{F}_n} f(z) = 0 \quad \text{for any } z \in \{1,\dots,n\},$$

because to each $f \in \mathcal{F}_n$ with $f(z) = 1$ there is an $f' \in \mathcal{F}_n$ such that $f'(z) = -1$ and $f'(i) = f(i)$ otherwise—and vice versa. Using a similar argument we can show that if $z_1,\dots,z_n$ are distinct, then

$$\sum_{f\in\mathcal{F}_n}\prod_{i=1}^{n} f(z_i) = 0. \tag{3.36}$$

Consider now the key computation in the sum (3.35), namely

$$S = \sum_{f\in\mathcal{F}_n}\prod_{i=1}^{2k} f^{q_i}(x_i)\prod_{y=1}^{n} f(y),$$

where $x_1,\dots,x_{2k}$ and also $q_1,\dots,q_{2k}$ are fixed. Since $f^0(x_i) = 1$, we have

$$S = \sum_{f\in\mathcal{F}_n}\prod_{\{i\,|\,q_i=1\}} f(x_i)\prod_{y=1}^{n} f(y). \tag{3.37}$$

Observe that $f^2(z) = 1$ for any f and any z. From (3.36) and (3.37) it follows that $S = 0$ unless each term in the second product in (3.37) can be matched by a different term of the first product.

The first product has at most $2k$ terms, and the second product has n terms. This immediately implies that $S = 0$ if $2k < n$. In addition, in such a case

$$\sum_{f\in\mathcal{F}} par(f)\langle\psi_f|P|\psi_f\rangle = 0$$

and therefore, for $2k < n$,

$$\sum_{\{f\,|\,par(f)=1\}}\langle\psi_f|P|\psi_f\rangle = \sum_{\{f\,|\,par(f)=-1\}}\langle\psi_f|P|\psi_f\rangle. \tag{3.38}$$

The ε-error condition (3.33) is therefore not satisfied.

The last equality (3.38) implies that the bounds (3.33) hold if we weaken the success criterion to $\varepsilon = 0$. Indeed, the condition (3.38) forces any algorithm which computes $par(f)$ for every f with less than $\frac{n}{2}$ applications of A_f and which claims the probability of success at least $\frac{1}{2}$ for every f, to have this probability to be exactly $\frac{1}{2}$.

We have shown that $par(f)$ cannot be computed with less than $\frac{n}{2}$ oracle calls. To show that the bound $\frac{n}{2}$ is optimal we need to show that there is a way to compute $par(f)$ with exactly $\frac{n}{2}$ f-oracle calls for n even. However, this follows easily from the fact that Deutsch's problem requires only one oracle call. Indeed, n values of $f(n)$ can be grouped into $\frac{n}{2}$ pairs, and for each one we compute parity using one oracle call. Once this is done the rest of the computation can be done without any additional f-oracle call.

Exercise 3.5.1 *Show that the above bound holds also for n odd. Namely, that with $\frac{n+1}{2}$ applications of A_f one can determine parity of f.*

Exercise 3.5.2 (Farhi et al. 1998) *Consider computation of $f^n(x)$ for functions that map a set of size $2n$ to itself. Show that no quantum algorithm which uses f as a black box can solve the problem with fewer than $\frac{n}{2}$ applications of the unitary operator corresponding to the given function.*

3.5.2 Framework for proving lower bounds

To prove lower bounds for quantum solutions of algorithmic problems is in general hard. One methodology to do that (due to Beals et al. 1998), has already emerged and will now be presented. It reduces proving lower bounds of quantum algorithms for Boolean functions to a hard classical problem of computing degrees of polynomials that represent or approximate Boolean functions. Using deep results from that classical complexity area it is possible to show with this methodology, see the next section, that without promises quantum computing can achieve only polynomial speed-up for black box computations .

The reduction of quantum algorithms to polynomials for a network with T oracle calls is done by showing that the acceptance probability of such a network after T oracle calls can be written as n-variate multilinear polynomial, of the input variables, of the degree $2T$.

Basic concepts and models

Three types of algorithm to compute a Boolean function $f : \{0,1\}^n \to \{0,1\}$ are to consider: **exact algorithms** providing $f(x)$ for any $x \in \{0,1\}^n$; **Las Vegas** or **zero-error** algorithms providing a result with probability at least $\frac{1}{2}$ (and if they deliver a result, then surely a correct one); **Monte Carlo (2–sided error)** or **bounded-error algorithms** providing a result that is correct with probability $\frac{2}{3}$.

Representation and approximation of Boolean functions by polynomials is defined as follows:

Definition 3.5.3 *A multilinear polynomial $p : \mathbf{R}^n \to \mathbf{R}$ is said to represent (to approximate) a Boolean function $f : \{0,1\}^n \to \{0,1\}$ if $p(x) = f(x)$ for all $x \in \{0,1\}^n$ (if $|p(x) - f(x)| \leq \frac{1}{3}$ for all $x \in \{0,1\}^n$).*

Observe that $x^n = x$ for any integer n whenever $x \in \{0,1\}$. It is therefore sufficient to consider multilinear polynomials when representation or approximation of Boolean functions by a polynomials is the task.

Example 3.5.4 *(1) The polynomials x_1x_2 and $1-(1-x_1)(1-x_2)$ represent Boolean functions $x_1 \wedge x_2$ and $x_1 \vee x_2$; (2) the polynomial $x_1 + x_3 - x_1x_3$ represents the Boolean function $(x_1 \vee x_2 \vee x_3) \wedge (x_1 \vee \overline{x_2} \vee x_3)$.*

Definition 3.5.5 *The* **degree** *of a multilinear polynomial is the maximum number of variables occurring in a term of the polynomial. For a Boolean function f let $\deg(f)$ ($\widehat{\deg}(f)$) be the the degree of a minimum degree polynomial representing (approximating) f.*

Remark 3.5.6 It is well known that for any Boolean function there is exactly one polynomial representing this function. Since n-ary OR function, $x_1 \vee x_2 \vee \ldots \vee x_n$, can be represented by the polynomial $1 - \prod_{i=1}^{n}(1-x_i)$, we have that *deg*(OR) $= n$. Clearly, the same holds for n-ary AND.

Several general lower bounds on the degree of Boolean functions were shown by Nisan and Szegedy (1994), von zur Gathen and Rucke (1997) and by Paturi (1992):

Theorem 3.5.7 *If $f : \{0,1\}^n \rightarrow \{0,1\}$ is a Boolean function that depends on all its variables, then*
(1) $\deg(f) \geq \lg n - \mathcal{O}(\lg\lg n)$;
(2) $\deg(f) \geq n - \mathcal{O}(n^{0.548})$, *if f is non-constant and symmetric;*
(3) $\deg(f) = n$, *if f is non-constant and symmetric and $n+1$ is prime.*
(4) $\widehat{\deg}(f) = \theta(\sqrt{n(n-\Gamma(f))})$, *where*

$$\Gamma(f) = \min\{|2k-n+1| \mid f_k \neq f_{k+1} \text{ and } 0 \leq k \leq n-1\}$$

and $f_k = f(X)$, for X with k 1s if f is non-constant and symmetric.

Exercise 3.5.8 *(a) Show for functions f equal to OR and AND that $\Gamma(f) = n-1$ (and therefore $\widehat{\deg}(f) = \theta(\sqrt{n})$); (b) show for functions f equal to PARITY or MAJORITY that $\Gamma(f) = 1$, and therefore $\widehat{\deg}(f) = \theta(n)$.*

Model of computation

We shall consider the following oracle setting for computing Boolean functions. There is a vector of N Boolean variables $X = (x_0, \ldots, x_{N-1})$ given by an oracle that produces x_i on the input i and we want to compute a Boolean function $f : X \rightarrow \{0,1\}$. (If $f : \{0,1\}^n \rightarrow \{0,1\}$, then $X = (x_0, \ldots, x_{2^n-1})$.)

A quantum network with T oracle calls to an oracle X will be represented by a sequence of unitary transformations

$$U_0, O, U_1, O, U_2, O, \ldots\ldots, O, U_{T-1}, O, U_T,$$

where U_i are arbitrary unitary transformations and O is a special unitary transformation defined below.

It will be assumed that all unitary transformations U_i, O operate on three registers $|i,b,y\rangle$, where the first register has n qubits, second one has one qubit and the last one some m' qubits. Let $m = m' + n + 1$ be the total number of qubits in the registers. The unitary transformation O—the oracle transformation—is defined to map $|i,b,y\rangle \to |i, b \oplus x_i, y\rangle$. The result of the algorithm is defined to be obtained at the end of the computation by measuring the last qubit of the third register.

Basic results

Lemma 3.5.9 *Let $\mathcal{N}$ be a quantum circuit that makes T calls to an oracle X. Then there exist complex-valued multilinear polynomials p_i, $0 \leq i < 2^m$, each of the degree at most T, such that the final state of the network is*

$$|\phi_X\rangle = \sum_{k=0}^{2^m-1} p_k(X)|k\rangle.$$

for any oracle X.

Proof. Let $|\phi_i\rangle$ be the state of the network just before the ith oracle call; i.e. $|\phi_{i+1}\rangle = U_i O|\phi_i\rangle$. The amplitudes in $|\phi_0\rangle$ depend on the initial state and on U_0, but not on the oracle, and therefore they are polynomials of the degree 0. An oracle call maps a basis state $|i,b,z\rangle$ to $|i, b \oplus x_i, z\rangle$.[23] If therefore the amplitude of $|i,0,z\rangle$ ($|i,1,z\rangle$) is α (β), then the amplitude of $|i,0,z\rangle$ ($|i,1,z\rangle$), after the oracle call becomes $(1-x_i)\alpha + x_i\beta$ ($x_i\alpha + (1-x_i)\beta$), which are polynomials of degree 1. In the same way we can show that if the amplitudes of the state before an oracle call are polynomials of the degree $\leq j$, then after the oracle call they are polynomials of the degree $\leq j+1$. Moreover, no unitary transformation U_i increases the maximal degree of the amplitude polynomials because such transformations only create linear combinations of the already existing polynomials. By induction we can now prove that all amplitude polynomials have degree at most T. ∎

Lemma 3.5.10 *Let $\mathcal{N}$ be a quantum circuit that makes T calls to an oracle X and $\mathcal{B}$ be a set of basis states. Then there exist a real-valued multilinear polynomial $p(X)$ of degree at most $2T$ which equals the probability that observing the final state of $\mathcal{N}$ with oracle X yields a state from $\mathcal{B}$.*

Proof. By Lemma 3.5.9, the final state of the network can be written as

$$\sum_{j=0}^{2^m-1} p_j(X)|j\rangle,$$

where p_j are complex valued polynomials of degree at most T. The probability of observing a state from $\mathcal{B}$ is

$$P_{\mathcal{B}}(X) = \sum_{j \in \mathcal{B}} |p_j(X)|^2.$$

If $p_j(X) = Re_j(X) + iIm_j(X)$, where $Re_j(X)$ and $Im_j(X)$ are real and imaginary parts—real valued polynomials of degree at most T—then $|p_j(X)|^2 = (Re_j(X))^2 + (Im_j(X))^2$ and therefore the resulting polynomials have degree at most $2T$. ∎

[23] The mapping $O : |i,b,y\rangle \to |i, b \oplus x_i, y\rangle$ has the following interpretation: i is information for the oracle and x_i is the output of the oracle.

Now we are in position to formulate the main result:

Theorem 3.5.11 *Let f be a Boolean function and $\mathcal{N}$ a quantum circuit computing f with T oracle calls. Then $T \geq \deg(f)/2$.*

Proof. Let B be the set of those basis states $|i\rangle$ that have 1 as the rightmost bit (of the binary representation of i). By Lemma 3.5.10, there is a real-valued polynomial p of degree at most $2T$ and such that, for all X, $p(X)$ equals the probability that the result of measurement of the last qubit is 1. Since $\mathcal{N}$ computes f exactly it must hold that $p(X) = 1$ if and only if $f(X) = 1$, and $p(X) = 0$ if and only if $f(X) = 0$. Hence $p(X) = f(X)$ for all X and therefore the degree of p has to be at least $deg(f)$. □

As a corollary of Theorems 3.5.7 and 3.5.11 we have

Theorem 3.5.12 *Let f be a Boolean function that depends on n variables and $\mathcal{N}$ a quantum network that computes f exactly using T oracle calls. Then $T \geq \frac{1}{2}\lg n - \mathcal{O}(\lg\lg n)$.*

Better bounds can be obtained if more can be assumed about f. By Theorems 3.5.7 and 3.5.11 we have

Corollary 3.5.13 *Let f be a non-constant symmetric Boolean function on n variables and $\mathcal{N}$ a quantum circuit that computes f exactly using T oracle calls. Then $T \geq \frac{n}{2} - \mathcal{O}(n^{0.548})$, and $T \geq \frac{n}{2}$ if $n+1$ is prime.*

For bounded-error computations the lower bound is similar.

Theorem 3.5.14 *Let f be a Boolean function and $\mathcal{N}$ be a quantum circuit that computes f with 2-sided error probability at most $\frac{1}{3}$ using T oracle calls. Then $T \geq \widehat{\deg}(f)/2$.*

Proof. By Lemma 3.5.10 there is a real valued multilinear polynomial p of degree at most $2T$ such that, for all X, $p(X)$ is the probability that the output of the circuit is 1. Since $\mathcal{N}$ computes f with error probability $\leq \frac{1}{3}$, we have $p(X) \geq \frac{2}{3}$ if and only if $f(X) = 0$. As a consequence $|f(X) - p(X)| \leq \frac{1}{3}$ for all X. This implies that p approximates f and so its degree has to be at least $\widehat{deg}(f)$. □

Quantum zero-error search

As shown in Section 3.3.2, we can perform an n-item unsorted database search with high probability of success in $\mathcal{O}(\sqrt{n})$ queries. Since an n-ary OR function can be seen as a special case of the unsorted database search, the upper bound $\mathcal{O}(\sqrt{n})$ holds for the OR function also.

The next task is to investigate the quantum potential for finding zero-error solutions for the unsorted database search and for the OR function.

Where it is promised that the number of solutions is 0 or t, then by Beals et al. (1998)—due to Høyer and Mosca—the exact solution can be found in $\mathcal{O}(\sqrt{\frac{n}{t}})$ queries and if a promise is made that the number of solutions is either 0 or odd, then the search needs $\frac{n}{2}$ queries and $\frac{n}{2} + 1$ queries are sufficient. However, without promises quantum search does not provide an exact solution more cheaply.

Lemma 3.5.15 *Any quantum circuit that computes n-variable OR with zero-error requires n oracle calls.*

Proof. By Lemma 3.5.9, there are complex-valued multilinear polynomials p_k, of degree at most T, such that the final state of the circuit is

$$|\phi_X\rangle = \sum_{k=0}^{2^m-1} p_k(X)|k\rangle.$$

Let B be the set of all basis states ending with 0—with the output 0. For every $k \in B$ there has to be $p_k(X) = 0$ if $X \neq \mathbf{0}$; otherwise the probability of getting the incorrect answer on the input $\mathbf{0}$ would be non-zero. At the same time there has to be at least one $k_0 \in B$ such that $p_{k_0}(\mathbf{0}) \neq 0$, since the probability of getting the correct answer 0 on $|\phi_{\mathbf{0}}\rangle$ must be non-zero. Let $p(X)$ be the real part of the polynomial $1 - p_{k_0}(X)/p_{k_0}(\mathbf{0})$. This polynomial has to have degree at least t and can be seen as representing the OR function. Hence p has to have degree at least $deg(\text{OR}) = n$. Therefore, $t \geq n$. ∎

As a consequence we have

Corollary 3.5.16 *A quantum circuit for exact or zero-error n-element database search requires n oracle calls.*

Finally, let us discuss again the problem of computing parity function of n arguments. Since $deg(\text{par}) = n$, it follows from Theorem 3.5.11 that at least $\frac{n}{2}$ oracle calls are needed, and as discussed at the end of Section 3.5.2, the fact that the Deutsch XOR problem can be solved with one oracle call implies that parity computation can be done for n even with $\frac{n}{2}$ oracle calls.

3.5.3 Oracle calls limitation of quantum computing

Strong limitations of quantum computing have been shown by Beals et al. (1998) for black-box computations. The results imply that quantum algorithms can achieve no more than polynomial time speed-up comparing to classical algorithms with respect to the number of black-box calls or, in other words, with respect to the number of accesses to input data. As a consequence exponential speed-up can be hoped for only for "computationally intensive" problems, such as factoring, with small input, but not for "input-intensive" problems such as a database search.

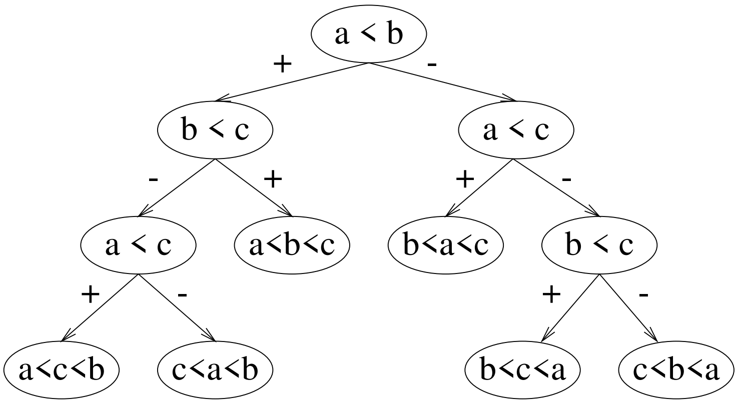


Figure 3.10: A decision tree-80%

Decision trees are perhaps the simplest computational model convenient to deal with the above problems. In **deterministic decision trees** (see Figure 3.10 for a decision tree to sort three different numbers), to each node corresponds a query concerning the input data (a use of the black box), and computation then proceeds according to the result of the query. In order to simplify the matter we will consider, without loss of generality concerning the complexity, decision trees to compute Boolean functions. The depth-cost of such a tree is the length of the shortest path and let $D(f)$ be the minimum of depth-costs over all deterministic decision trees for f.

Randomized decision trees are the corresponding model for randomized computation. They can be seen as a probability distribution on the set of deterministic decision trees. The depth-cost of such a decision tree is the expected number of calls on the worst case input. Depending on the type of error allowed several types of randomized decision trees are considered: zero-error, one-sided error and bounded-error. Randomized bounded-error complexity $R(f)$ of a Boolean function f is the minimum cost of randomized decision tree that computes f with bounded-error probability at most $\frac{1}{3}$ for all inputs.

Quantum complexity $Q(f)$ of a Boolean function f is defined as the number of the black box calls in the best network that computes f with error probability at most $\frac{1}{3}$.

Relation $D(f) = \mathcal{O}(R(f)^3)$ is due to Nisan (1991), and the relation $D(f) = \mathcal{O}(Q(f)^6)$ is duet to Beals et al. (1998). The result implies that if some Boolean function can be computed quantumly with bounded-error probability, then it can be computed by a deterministic decision tree with only polynomial increase of black-box queries.

Open problem 3.5.17 *Can the upper bound $D(f) = \mathcal{O}(Q(f)^6)$ be improved?*

In the case of monotone (symmetric) Boolean functions it holds $D(f) = \mathcal{O}(Q(f)^4)$ $(D(f) = \mathcal{O}(f)^2)$, and the best separation known is $D(f) = n$ and $Q(f) = \mathcal{O}(\sqrt{n})$ for the XOR function.

Chapter 4

AUTOMATA

INTRODUCTION

In addition to the study of problems of the design and analysis of algorithms and circuits, as well as of the computational complexity of algorithmic problems, another main method of theoretical computing to get an insight into the power of computational resources is to study models of quantum computing devices and the corresponding complexity classes. This will be done in this chapter for three most basic models of quantum automata: quantum versions of finite automata, Turing machines and cellular automata.

Quantum finite automata are perhaps the most elementary model of quantum automata. They are in addition the only model so far for which it has been fully proved that they have larger power than their classical counterparts.

Quantum Turing machines are the main model to study the most fundamental questions concerning the power of quantum computing itself and the power of quantum versus classical computing. Quantum cellular automata are of a special interest. They seem to be a model much closer to the physical reality than quantum Turing machines. In addition, it is still a major open problem whether quantum cellular automata are more powerful than quantum Turing machines.

LEARNING OBJECTIVES

The aim of the chapter is to learn:

1. the basic models of one-way and two-way quantum finite automata as well as their computational power and space efficiency in comparison with classical finite automata;

2. the basic models of quantum Turing machines and methods of implementing the main classical programming primitives in a reversible way;

3. the types and precision of amplitudes needed to ensure sufficient accuracy of computation on quantum Turing machines;

4. the basic models of quantum cellular automata and partitioned quantum cellular automata, as well as simulations between quantum cellular automata and quantum Turing machines.

By convention there is color, by convention sweetness, by convention bitterness, but in reality there are atoms and space.

Democritus (460–400 BC)

Three basic models of quantum automata are described in detail in this chapter. Each of them plays a special role in the study of the power and limitations of quantum computing.

One-way quantum finite automata can be seen as the simplest model of quantum automata that can also quite well reflect the simplest quantum computers we can expect first. In spite of their simplicity their descriptional efficiency can be remarkable. Two-way quantum finite automata are of interest because they are the first model of finite automata more powerful in polynomial time than classical deterministic finite automata.

With quantum Turing machines the most fundamental questions concerning the ultimate power and limitations of quantum computing can be studied. For example, the concept of universality and the question whether quantum Turing machine should be seen as analogue or discrete devices.

A suitable definition of more dimensional quantum cellular automata is still an important open problem. For one-dimensional cellular automata the problem has been resolved fully satisfactorily and it is with this type of quantum cellular automata we deal in this chapter. We could define in a similar way more dimensional quantum cellular automata but in such a case there are problems to decide whether a given specification of to-be quantum automaton really defines a quantum automaton.

There are two main reasons why to come with an easy to deal definition of two and more dimensional quantum cellular automata is an important and interesting task. Intuitively, two- and three-dimensional quantum cellular automata could much better reflect the nature of the quantum world than quantum Turing machines. In addition, there is still a chance that they could be computationally more powerful than quantum Turing machines.

The way quantum analogues of classical automata models are defined is already pretty standard, at least its basic steps. A quantum automaton is described formally in a similar way to a probabilistic one, only instead of the probabilities of transitions we use probability amplitudes. Evolution takes place then in the inner-product space over the set of finite configurations, which then represent the states of the (standard) basis. The main problem now is that evolution of each quantum automaton has to be unitary. Given a description of a to-be-quantum automaton, it can be far from trivial to decide whether its evolution satisfies the unitarity condition. One way out is to discover some more easy-to-check conditions, usually called "well-formedness conditions", satisfiability of which implies unitarity of the evolution.

4.1 Quantum Finite Automata

Just as classical finite automata serve as the very basic model of classical finite-size machines, so can quantum finite automata be seen as basic quantum models of finite state quantum machines.

4.1.1 Models of classical finite automata

A variety of models of finite state machines and automata have been developed and investigated in the classical theory of computation. Each of these models captures some mode of actions, or its natural abstraction, of the intuitive concept of computing with finite-size machines.

The very basic model of finite automata is that of the input-driven **deterministic finite automaton** (DFA) $\mathcal{A} = \langle Q, \Sigma, q_0, Q_a, \delta\rangle$, where Q is the finite set of states, Σ is the input alphabet, q_0 is the initial state, Q_a is the set of final (accepting) states and $\delta : Q \times \Sigma \rightarrow Q$ is the **transition function** such that for a given input symbol σ and the state q, $\delta(q, \sigma)$ is the new state of the control unit—see Figure 4.1a. Another view of DFA is that of a **one-way finite automaton** (1FA, see Figure 4.1b), with the finite control unit connected to the input tape by a read-only head moving only in a left-to-right way. A natural generalization is that of a **two-way finite automaton** (see Figure 4.1c), with the finite-control unit attached to the input tape by a read-only head which can move in both directions. The transition function of such a 2FA has the form $\delta : Q \times \Sigma \rightarrow Q \times \{\leftarrow, \downarrow, \rightarrow\}$.

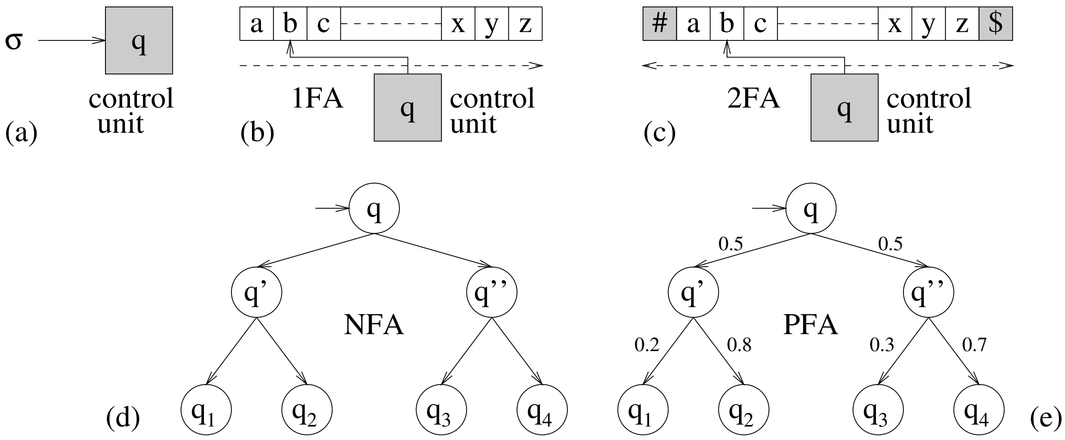


Figure 4.1: Models of finite automata-70%

Two other generalizations of FA are of importance. In a **nondeterministic finite automaton**, (NFA, see Figure 4.1d for a transition tree), there is in general a nondeterministic choice of the next state, and the transition function has the form $\delta : Q \times \Sigma \times Q \rightarrow \{0, 1\}$ (or $\delta : Q \times \Sigma \rightarrow \{0, 1\}^Q$). In **probabilistic finite automata** (PFA or 1PFA, see Figure 4.1e for a transition tree), to each transition a probability is assigned; i.e. the transition function has the form $\delta : Q \times \Sigma \times Q \rightarrow [0, 1]$ (or $\delta \in Q \times \Sigma \rightarrow [0, 1]^Q$), such that for each $q \in Q$ and $\sigma \in \Sigma$ the following **local probability condition** is satisfied: $\sum_{q' \in Q} \delta(q, \sigma, q') = 1$.

One way to compare computational power of various types of finite automata is to consider them as **acceptors** of the set of those input words that make the automaton reach a final state after reading the whole input. One of the main results of the automata theory is that one-way and two-way deterministic and nondeterministic finite automata accept the same family of languages, namely regular languages. Probabilistic finite automata are a special case. With respect to the bounded-error acceptance 1PFA can accept only regular languages and so do 2PFA, a probabilistic version of 2FA, if required to work in linear time with respect to the size of input. (On the other hand, with respect to the unbounded-error acceptance and arbitrary real probabilities 1PFA can accept also non-regular languages.)

4.1.2 One-way quantum finite automata

Two main models of quantum finite automata (QFA) will be presented in this section. One-way quantum automata (1QFA) as a quantum analogue of 1FA and two-way quantum automata (2QFA) as an analogue of 2FA.

One-way quantum finite automata seem to model very well the way very simple quantum processors work, (see Ambainis and Freivalds, 1998), and also the way simple classical/quantum processors are expected to work: the classical part reads an input, picks up the corresponding quantum operator (a transition mapping) and performs it on a quantum memory of fixed size, independent of the size of input. 1QFA are very simple but provably less powerful than classical 1FA.

Two-way quantum finite automata, on the other hand, can accept more than regular languages. The price to pay is the "size of quantum memory" needed. A 2QFA can be in a large superposition of its basis states each of which corresponds to a position of the head on a different symbol of the input. The size of quantum memory can therefore be proportional to the size of input.

Basic concepts

One-way quantum finite automata are the first type of quantum automata we consider and therefore we shall deal with them in more detail.

Definition 4.1.1 *A* **one-way quantum finite automaton** $\mathcal{A}$ *is specified by the finite (input) alphabet* Σ*, the finite set of states* Q*, the initial state* q_0*, the sets* $Q_a \subseteq Q$ *and* $Q_r \subseteq Q$ *of accepting and rejecting states, respectively, with* $Q_a \cap Q_r = \emptyset$*, and the transition function*

$$\delta : Q \times \Gamma \times Q \to \mathbf{C}_{[0,1]},$$

where $\Gamma = \Sigma \cup \{\#, \$\}$ *is the tape alphabet of* $\mathcal{A}$*, and* $\#$ *and* $\$$ *are endmarkers not in* Σ*.*

The evolution (computation) of $\mathcal{A}$ *is performed in the inner-product space* $l_2(Q)$*, i.e., with the basis* $\{|q\rangle \mid q \in Q\}$*, using the linear operators* V_σ*,* $\sigma \in \Gamma$*, defined by*

$$V_\sigma(|q\rangle) = \sum_{q' \in Q} \delta(q, \sigma, q')|q'\rangle,$$

which are required to be unitary.

In computation of $\mathcal{A}$ the so-called computational observable $\mathcal{O}$ is used that corresponds to the orthogonal decomposition

$$l_2(Q) = E_a \oplus E_r \oplus E_n,$$

where $E_a = span\{|q\rangle \mid q \in Q_a\}$, $E_r = span\{|q\rangle \mid q \in Q_r\}$ and E_n is the orthogonal complement of $E_a \oplus E_r$. Denote by P_p, $p \in \{a, r, n\}$ the projection operator into the subspace E_p.

A **computation of $\mathcal{A}$ on an input** $\#\sigma_1 \ldots \sigma_n\$$ proceeds as follows. The operator $V_\#$ is first applied to the starting configuration $|q_0\rangle$ and then the observable $\mathcal{O}$ is applied to the resulting configuration $V_\#|q_0\rangle$. This observable projects $V_\#|q_0\rangle$ into a vector $|\psi'\rangle$ of one of the subspaces E_a, E_r, E_n, with the probability equal to the square of the norm of $|\psi'\rangle$. If $|\psi'\rangle \in E_a$ the input is accepted; if $|\psi'\rangle \in E_r$ the input is rejected. If $|\psi'\rangle \in E_n$, then, after the normalization of $|\psi'\rangle$, the operator V_{σ_1} is applied to $|\psi'\rangle$ and after that the observable $\mathcal{O}$ to the resulting vector. This process goes on. Operators $V_{\sigma_1}, V_{\sigma_2}, \ldots, V_{\sigma_n}$ are applied one after another, and after each such application the measurement by the observable $\mathcal{O}$ is performed. In all cases the computation continues only if a projection into E_n occurs.

When no termination occurs the computation can be seen as an application of the composed operator

$$V'_{\sigma_n} V'_{\sigma_{n-1}} \ldots V'_{\sigma_1}|q_0\rangle,$$

where $V'_{\sigma_i} = P_n V_{\sigma_i}$.

Language recognition by 1QFA

In order to define formally the overall probability with which an input is accepted (rejected) by a 1QFA $\mathcal{A}$, we define the set $V_\mathcal{A} = l_2(Q) \times \mathbf{C} \times \mathbf{C}$ of so-called "total states" of $\mathcal{A}$, that will be used only with the following interpretation. $\mathcal{A}$ is at any time during the computation in the state (ψ, p_a, p_r) if so far in its computation $\mathcal{A}$ accepted the input with probability p_a, rejected with probability p_r and neither with probability $1 - p_a - p_r = ||\psi||^2$, and $|\psi\rangle$ is its current, unnormalized state. For each $\sigma \in \Gamma$ the evolution of $\mathcal{A}$, with respect to the total state, on an input σ is given by the operator T_σ defined by

$$T_\sigma(\psi, p_a, p_r) \to (P_n V_\sigma \psi, p_a + ||P_a V_\sigma \psi||^2, p_r + ||P_r V_\sigma \psi||^2).$$

For $x = \sigma_1\sigma_2 \ldots \sigma_n \in \Gamma^*$ let $T_{\#x\$} = T_\$ T_{\sigma_n} T_{\sigma_{n-1}} \ldots T_{\sigma_1} T_\#$. If $T_{\#x\$}(|q_0\rangle, 0, 0) = (\psi, p_a, p_r)$, then we say that $\mathcal{A}$ accepts x with probability p_a and rejects with probability p_r.

A 1QFA $\mathcal{A}$ is said to accept a language L with probability $\frac{1}{2} + \varepsilon$, $\varepsilon > 0$, if it accepts any $x \in L$ with probability at least $\frac{1}{2} + \varepsilon$ and rejects any $x \notin L$ with probability at least $\frac{1}{2} + \varepsilon$. If there is an $\varepsilon > 0$ such that $\mathcal{A}$ accepts L with probability at least $\frac{1}{2} + \varepsilon$, then L is said to be accepted by $\mathcal{A}$ with bounded-error probability. L is said to be accepted with unbounded error probability if $x \in L$ is accepted with probability at least $\frac{1}{2}$ and $x \notin L$ rejected with probability at least $\frac{1}{2}$.

Example 4.1.2 *(Ambainis and Freivalds, 1998) We show that the language $L_0 = \{0^i 1^j \mid i, j \geq 0\}$ can be recognized by a 1QFA $\mathcal{A}_0$ with probability p such that $p = 1 - p^3$ ($p \approx 0.68$).*

$\mathcal{A}_0$ has the set of states $Q = \{q_0, q_1, q_2, q_a, q_r\}$, $Q_a = \{q_a\}$, $Q_r = \{q_r\}$. q_0 is the initial state. The transition function σ is specified by transitions:

$$V_\#|q_0\rangle = \sqrt{1-p}|q_1\rangle + \sqrt{p}|q_2\rangle,$$

$$V_0|q_1\rangle = (1-p)|q_1\rangle + \sqrt{p(1-p)}|q_2\rangle + \sqrt{p}|q_r\rangle,$$

$$V_0|q_2\rangle = \sqrt{p(1-p)}|q_1\rangle + p|q_2\rangle - \sqrt{1-p}|q_r\rangle,$$

$$V_1|q_1\rangle = |q_r\rangle, V_1|q_2\rangle = |q_2\rangle,$$
$$V_\$|q_1\rangle = |q_r\rangle, V_\$|q_2\rangle = |q_a\rangle,$$

and the remaining transitions are defined arbitrarily so that unitarity requirements are satisfied.

Case 1 $x = 0^i$. *Since* $V_0(\sqrt{1-p}|q_1\rangle + \sqrt{p}|q_2\rangle) = \sqrt{1-p}|q_1\rangle + \sqrt{p}|q_2\rangle$, $\mathcal{A}_0$ *remains in the state* $\sqrt{1-p}|q_1\rangle + \sqrt{p}|q_2\rangle$ *while reading* 0^i. *At the right endmarker the operator* $V_\$$ *is applied to yield the state* $\sqrt{1-p}|q_r\rangle + \sqrt{p}|q_a\rangle$. *Therefore* $\mathcal{A}_0$ *accepts with probability* p.

Case 2. $x = 0^i1^j$, $i \geq 0, j > 0$. *As in Case 1,* $\mathcal{A}_0$ *is in the state* $\sqrt{1-p}|q_1\rangle + \sqrt{p}|q_2\rangle$ *after reading* 0^i. *The first* 1 *on the input changes the state into* $\sqrt{1-p}|q_r\rangle + \sqrt{p}|q_2\rangle$. *The nonhalting part is* $\sqrt{p}|q_2\rangle$. *It is unchanged by the following* 1*s until the right endmarker changes it to* $\sqrt{p}|q_a\rangle$. $\mathcal{A}_0$ *therefore again accepts with probability* p.

Case 3. x *is not in* $\{0^i1^j \,|\, i \geq 0, j \geq 0\}$. *Hence there is an initial substring of* x *of the form* $0^i1^j0^k$, *where* $i \geq 0$, $j > 0$, $k > 0$. *After reading the first* 1, $\mathcal{A}_0$ *is in the state* $\sqrt{1-p}|q_r\rangle + \sqrt{p}|q_2\rangle$ *and the automaton rejects with probability* $1-p$. *The nonhalting part,* $\sqrt{p}|q_2\rangle$, *is not changed by the following* 1*s and the first* 0 *changes it into* $p\sqrt{1-p}|q_1\rangle + p\sqrt{p}|q_2\rangle - \sqrt{p(1-p)}|q_r\rangle$. *At this moment* $\mathcal{A}_0$ *rejects with probability* $p(1-p)$. *The total probability that* x *is rejected is therefore at least*

$$(1-p) + p(1-p) + p^2(1-p) = 1 - p^3 = p.$$

Exercise 4.1.3 *What is the largest probability with which a 1QFA can accept the language* $\{0^i1^j0^k \,|\, i,j,k \geq 0\}$*?*

On $V_\mathcal{A}$ we define a “norm” $||\cdot||_u$ as follows

$$||(\psi, p_a, p_r)||_u = \frac{1}{2}(||\psi|| + |p_a| + |p_r|)$$

and let $\mathcal{D} = \{v \in V_\mathcal{A} \,|\, ||v||_u \leq 1\}$. $\mathcal{D}$ contains all global states of $\mathcal{A}$.

Exercise 4.1.4 *Show: (a)* $||\cdot||_u$ *is a norm on* $V_\mathcal{A}$*; (b) there is a constant* c *such that* $||T_xv - T_xv'||_u \leq c||v - v'||_u$ *for any* $v, v' \in V_\mathcal{A}$, $x \in \Gamma^*$*; (c) if a set* $A \subset \mathcal{D}$ *is such that there exists an* $\varepsilon > 0$ *such that for all* $v, v' \in \mathcal{A}$ *it holds* $||v - v'||_u > \varepsilon$, *then there can be only finitely many elements in* A.

Remark 4.1.5 Another possibility to define acceptance by one-way QFA was considered by Moore and Crutchfield (1997)—to measure only at the right endmarker. Let us denote one-way QFA with such acceptance by mc-1QFA. Even if this definition seems to be at first sight more natural, technical results do not confirm that. It holds mc-QFA $\subsetneq 1QFA$.

Exercise 4.1.6 *Show that if a language is accepted by a mc-1QFA, it is accepted by a 1QFA, but that the opposite statement does not have to be true.*

Exercise 4.1.7 *Explore closure properties of languages accepted by mc-QFA.*

4.1.3 1QFA versus 1FA

Two basic problems concerning one-way quantum automata are dealt with in this section: their power and space efficiency.

Language recognition

The basic question is whether 1QFA have larger recognition power than 1FA. The answer is negative (Kondacs and Watrous, 1997).

Theorem 4.1.8 *If a language L is accepted by a 1QFA with bounded-error probability, then L is regular.*

Proof. The proof is a small modification of the proof that probabilistic FA accept only regular languages, see Rabin (1976). Here is the basic idea:

Let $\mathcal{A} = \langle Q, \Sigma, q_0, Q_a, Q_r, \delta\rangle$ be a 1QFA recognizing the language L with probability $\frac{1}{2} + \varepsilon$.

For $w, w' \in \Sigma^*$ we define that the prefix relation $w \equiv_L w'$ holds if for all $y \in \Sigma^*$, $wy \in L$ if and only if $w'y \in L$. It is well known (see, for example Gruska, 1997), that a language L is regular if and only if there are only finitely many equivalence classes with respect to its prefix equivalence.

Let $W \subseteq \Sigma^*$ be *any* set of strings that are mutually inequivalent with respect to the equivalence $\equiv_L$. If we prove that W is finite, the theorem will be proved. This can be done as follows.

If $w, w' \in W$, $w \not\equiv_L w'$, then there must exist a y such that only one of the strings wy and $w'y$ is in L. Therefore, for $v = T_{\#w}(|q_0\rangle, 0, 0)$ and $v' = T_{\#w'}(|q_0\rangle, 0, 0\rangle$ it has to hold:

$$||T_{y\$}v - T_{y\$}v'||_u > 2\varepsilon,$$

because $\mathcal{A}$ accepts L with bounded-error probability ε. Consequently, by Exercise 4.1.4, $||v-v'||_u \geq \frac{2\varepsilon}{c}$ for some constant c that does not depend on ε. This implies, by Exercise 4.1.4 again, that the set W has to be finite. ∎

The next basic question is whether 1QFA have the same recognition power as 1FA. The answer is again negative (Kondacs and Watrous, 1997).

Theorem 4.1.9 *The regular language $L_0 = \{0,1\}^*0$ cannot be recognized by a 1QFA with bounded-error probability.*

Proof. The proof is by contradiction. Let $\mathcal{A} = \langle Q, \Sigma, \delta, q_0, Q_a, Q_r\rangle$ be a 1QFA recognizing the language L_0. To each $x = \sigma_1 \ldots \sigma_n \in \Gamma^*$ we assign the state

$$|\psi_x\rangle = V'_{\sigma_n} \ldots V'_{\sigma_1}|q_0\rangle$$

and let $\mu = \inf_{w\in\{0,1\}^*}\{||\psi_{\#w}||\}$. For each $w \in \{0,1\}^*$, $w0 \in L_0$ and $w1 \notin L_0$. If $\mu = 0$, then clearly $\mathcal{A}$ cannot recognize L_0 with bounded-error probability $\frac{1}{2} + \varepsilon$. Let us therefore assume that $\mu > 0$.

For any $\varepsilon > 0$ there is a w such that $||\psi_{\#w}|| < \mu + \varepsilon$, and also $||\psi_{\#wy}|| \in [\mu, \mu + \varepsilon]$ for any $y \in \{0,1\}^*$. In particular, for any $m > 0$

$$||V_1'^m|\psi_{\#w0}\rangle|| \in [\mu, \mu + \varepsilon). \tag{4.1}$$

This implies that the sequence $\{V_1'^i \mid \psi_{\#w0}\rangle\}_{i=0}^{\infty}$ is bounded in the finite dimensional inner-product space and must have a limit point. Therefore there have to exist j and k such that

$$||V_1'^j(|\psi_{\#w0}\rangle - V_1'^k|\psi_{\#w0}\rangle)|| < \varepsilon.$$

The last inequality together with (4.1) imply (see Lemma 4.1.10), that there is a constant c which does not depend on ε and such that

$$\left|\left||\psi_{\#w0}\rangle - V_1'^k|\psi_{\#w0}\rangle\right|\right| < c\varepsilon^{1/4}.$$

This implies that

$$||T_{\#w0\$}(|q_0\rangle, 0, 0) - T_{\#w01^k\$}(|q_0\rangle, 0, 0)||_u < c'\varepsilon^{1/4}$$

for fixed c'. However, this has to be valid for an arbitrarily small ε. This is not possible if $\mathcal{A}$ accepts L_0 because $\mathcal{A}$ should accept the string $w0$ and reject $w01^k$. Hence $\mathcal{A}$ cannot accept L_0 with bounded-error probability. ∎

Lemma 4.1.10 *(J. Watrous) If $|u\rangle$ and $|v\rangle$ are vectors such that for a linear operator A, reals $0 < \varepsilon < 1$ and $\mu > 0$, $||A(u-v)|| < \varepsilon$, and $||v||$, $||u||$, $||Au||$, $||Av||$ are in $[\mu, \mu+\varepsilon]$, then there is a constant c, that does not depend on ε, such that $||u-v|| < c\varepsilon^{1/4}$.*

Proof. First observe that if we can show

$$||u-v||^2 < ||A(u-v)||^2 + c'\sqrt{\varepsilon} \tag{4.2}$$

then we have our proof because the square root of the right-hand side in (4.2) is smaller than $\sqrt{\varepsilon^2 + c'\sqrt{\varepsilon}} \leq \sqrt{c'+1}\,\varepsilon^{1/4}$, since $\varepsilon < 1$. To show (4.2) let us compute

$$\begin{aligned} ||u-v||^2 - ||A(u-v)||^2 &= \langle u-v|u-v\rangle - \langle Au-Av|Au-Av\rangle \\ &= ||u||^2 + ||v||^2 - \langle u|v\rangle - \langle v|u\rangle \\ &\quad -||Au||^2 - ||Av||^2 + \langle Au|Av\rangle + \langle Av|Au\rangle \\ &= (I) + (II) + (III) + (IV), \end{aligned}$$

where

$$(I) = ||u||^2 - ||Au||^2, \qquad (II) = ||v||^2 - ||Av||^2,$$
$$(III) = \langle u|A^*Av - v\rangle, \qquad (IV) = \langle v|A^*Au - u\rangle.$$

Since $||u||, ||v|| \leq [\mu, \mu+\varepsilon]$, we have

$$||u||^2 - ||Au||^2 \leq 2\mu\varepsilon + \varepsilon^2, \qquad ||v||^2 - ||Av||^2 \leq 2\mu\varepsilon + \varepsilon^2.$$

In order to get an estimate for (III) and (IV) we proceed as follows:

$$\begin{aligned} ||A^*Au - u||^2 &\leq ||A^*Au||^2 + ||u||^2 - 2||Au||^2 \\ &\leq ||u||^2 - ||Au||^2 \leq 2\mu\varepsilon + \varepsilon^2, \end{aligned}$$

because $||A^*|| = ||A|| \leq 1$ and therefore

$$\langle v|A^*Au - u\rangle \leq ||v||\,||A^*Au - u|| \leq (\mu+\varepsilon)\sqrt{2\mu\varepsilon + \varepsilon^2}.$$

A similar estimation holds for (III). Altogether we have

$$||u-v||^2 - ||A(u-v)||^2 \leq \sqrt{\varepsilon}[2(2\mu\sqrt{\varepsilon} + \varepsilon^{3/2}) + 2(\mu+\varepsilon)\sqrt{2\mu+\varepsilon})].$$

By taking $c' > 2(2\mu\sqrt{\varepsilon} + \varepsilon^{3/2}) + 2(\mu+\varepsilon)\sqrt{2\mu+\varepsilon}$, for example $c' = 10$, we get the inequality (4.2). ∎

Another natural type of automata one should try to compare with 1QFA are one-way reversible finite automata (1RFA). They can be defined as 1QFA having transition amplitudes either 0 or 1. The following result, due to Ambainis and Freivalds (1998), shows that if 1QFA are required to give the correct answer with high probability, then they have the same recognition power as 1RFA.

Theorem 4.1.11 *A language can be recognized by a 1QFA with probability $\frac{7}{9}+\varepsilon$ if and only if it can be recognized by 1RFA.*

$\{0^i1^j \mid i,j \geq 0\}$ is an example of a language which can be recognized by a 1QFA with probability 0.68 (see Example 4.1.2), but not with probability $\frac{7}{9}+\varepsilon$.

Size-space efficiency

The number of states is a natural space measure for both 1FA and 1QFA. A natural related problem to investigate is the following one: if a language L is accepted by a 1QFA, then how many states does a minimal 1QFA recognizing L have in comparison with minimal 1FA recognizing L? The results obtained by Ambainis and Freivalds (1998) and Ambainis et al. (1998) show that sometimes a 1QFA is exponentially smaller than any 1FA recognizing the same language, but sometimes a 1FA is almost exponentially smaller than any 1QFA recognizing the same language. It holds:

Theorem 4.1.12 *(1) Let p be prime. Any 1FA recognizing the language $L_p = \{0^i \mid i$ is divisible by $p\}$ has to have p states, but for any $\varepsilon > 0$ there is a 1QFA $\mathcal{A}_{p,\varepsilon}$, with $\mathcal{O}(\lg p)$ states accepting L_p with probability $\frac{1}{2}+\varepsilon$.*

(2) Let $n \geq 1$ be an integer. The language $L_n = \{w0 \mid w \in \{0,1\}^*, |w| \leq n\}$ can be recognized by a 1FA with $2n+3$ states but for any $\varepsilon > 0$ any 1QFA recognizing L_n with probability $\frac{1}{2}+\varepsilon$ has to have $2^{\Omega(n/\lg n)}$ states.

It is easy to design a $(2n+3)$-state 1FA accepting the language L_n. Concerning the proof of the lower bound on the number of states of any 1QFA for L_n, the very basic idea is simple. Since a 1QFA can read each input symbol only once, any 1QFA for L_n which is necessary reversible is forced to "remember" all symbols read until it is clear whether the input symbol is in the language. Consequently, the state the automaton reaches after reading n input symbols has to be an encoding of these n symbols. Since this encoding has to be such that any input n-bit word can be recovered, the number of states has to be at least 2^n. Unfortunately, the above idea is not fully valid because a 1QFA can terminate before it reads all inputs. The problem is therefore more complex and consequently the proof is more involved. It makes use of the ideas of dense coding discussed in Section 8.2.4.

It follows from the above theorem that in some cases the requirement of unitarity makes 1QFA much larger than the minimal equivalent 1FA.

4.1.4 Two-way quantum finite automata

The model of two-way quantum finite automata (2QFA) due to Watrous (1997c), presented in this section is a natural quantum analogue of the classical two-way finite automata. 2QFA are provably more powerful than their classical counterparts.

Basic concepts

Definition of two-way QFA is significantly more complex than of a 1QFA, especially due to the need to make sure their evolutions are unitary.

Definition 4.1.13 *A two-way quantum finite automaton $\mathcal{A}$ is specified by the finite (input) alphabet Σ, the finite set of states Q, the initial state q_0, the sets $Q_a \subset Q$ and $Q_r \subset Q$ of accepting and rejecting states, respectively, with $Q_a \cap Q_r = \emptyset$, and the transition function*

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

where $\Gamma = \Sigma \cup \{\#, \$\}$ is the tape alphabet of $\mathcal{A}$ and $\#$ and $\$$ are endmarkers not in Σ, which satisfies the following conditions (of well-formedness) for any $q_1, q_2 \in Q$, $\sigma, \sigma_1, \sigma_2 \in \Gamma$, $d \in \{\leftarrow, \downarrow, \rightarrow\}$:

1. **Local probability and orthogonality condition.** $\sum_{q',d} \delta^*(q_1, \sigma, q', d)\delta(q_2, \sigma, q', d) = \begin{cases} 1, & \text{if } q_1 = q_2; \\ 0, & \text{otherwise.} \end{cases}$
2. **Separability condition I.** $\sum_{q'} \delta^*(q_1, \sigma_1, q', \rightarrow)\delta(q_2, \sigma_2, q', \downarrow) = 0.$
3. **Separability condition II.** $\sum_{q'} \delta^*(q_1, \sigma_1, q', \downarrow)\delta(q_2, \sigma_2, q', \leftarrow) = 0.$
4. **Separability condition III.** $\sum_{q'} \delta^*(q_1, \sigma_1, q', \rightarrow)\delta(q_2, \sigma_2, q', \leftarrow) = 0.$

(These conditions are equivalent, as shown below, to the requirement that evolution of $\mathcal{A}$ is unitary.) Formally, $\mathcal{A} = \langle \Sigma, Q, q_0, Q_a, Q_r, \delta \rangle$.

States from $Q_a \cup Q_r$ are called halting states and states from $Q_n = Q - (Q_a \cup Q_r)$ are called non-halting states.

In order to process an input word $x \in \Sigma^*$ by $\mathcal{A}$ we assume that the input is written on the tape with the endmarkers in the form $w_x = \#x\$$ and that such a tape, of length $|x| + 2$, is circular, i.e. the symbol to the right of $\$$ is $\#$.[1]

For an integer n let C_n be the set (of size $(n + 2)|Q|$) of all possible configurations of $\mathcal{A}$ for inputs of length n. For each specific input x such a configuration is uniquely determined by a pair (q, k), where q is the state of the configuration and $0 \leq k \leq |x| + 1$ denotes the position of the head.

To a computation of $\mathcal{A}$ on an input x of length n corresponds a unitary evolution in the underlying inner-product space $\mathcal{H}_{\mathcal{A},n} = l_2(C_n)$. For each configuration $c \in C_n$, $|c\rangle$ denotes the basis vector in $l_2(C_n)$. Each state in $\mathcal{H}_{\mathcal{A},n}$ will therefore have a form

$$\sum_{c \in C_n} \alpha_c |c\rangle \quad \textit{where} \quad \sum_{c \in C_n} |\alpha_c|^2 = 1.$$

The automaton $\mathcal{A}$ induces for any input $x \in \Sigma^n$ a linear operator U_x^δ on $\mathcal{H}_{\mathcal{A},n}$ defined for a configuration $(q, k) \in C_n$ by

$$U_x^\delta |q, k\rangle = \sum_{q',d} \delta(q, w_x(k), q', d)\,|\,q', (k + \mu(d)) \bmod (n + 2)\rangle,^2$$

where $w_x(k)$ denotes the kth symbol of $w_x = \#x\$$ for $0 \leq k \leq |x| + 1$. By linearity U_x^δ is extended to map any superposition of basis states.

[1] The requirement of circularity of the tape is not essential but slightly simplifies the treatment of 2QFA.

[2] By definition, $\mu(d) = -1(0)[1]$ if $d = \leftarrow (\downarrow)[\rightarrow]$.

Lemma 4.1.14 *For any nonempty input string x the mapping U_x^δ is unitary if and only if the conditions (1) to (4) of Definition 4.1.13 are satisfied.*

Proof. To prove the lemma, it is sufficient to investigate the orthogonality of the vectors $U_x^\delta|q,k\rangle$ for $q \in Q$, $0 \leq k \leq |x|+1$. The condition (1) is equivalent to the statement that for every x, $|x| \geq 0$, $||U_x^\delta|q,k\rangle|| = 1$ for all q and k and that $U_x^\delta|q_1,k\rangle \perp U_x^\delta|q_2,k\rangle$ for $q_1 \neq q_2$. Conditions (2) to (4) are equivalent to the statement that $U_x^\delta|q_1,k_1\rangle \perp U_x^\delta|q_2,k_2\rangle$ if k_1 and k_2 differ at most by 2. Finally, it is trivially true that $U_x^\delta|q_1,k_1\rangle \perp U_x^\delta|q_2,k_2\rangle$ if the head positions are more than two cells away because the heads can move only one cell per step.
□

Verification of well-formedness condition of a 2QFA can be in general a quite cumbersome task, but it is easy for so-called simple 2QFA.

Informally, a 2QFA is simple if its transitions (amplitudes) are uniquely determined by transitions (amplitudes) between current states and the states the automaton comes into and, in addition, the move of the head is always uniquely determined by the new state only.

Definition 4.1.15 *A 2QFA $\mathcal{A} = \langle \Sigma, Q, q_0, Q_a, Q_r, \delta \rangle$ is* **simple**, *or* **unidirectional**, *if for each $\sigma \in \Gamma$ there is a linear unitary operator V_σ on the inner-product space $l_2(Q)$ and a function $D : Q \to \{\leftarrow, \downarrow, \rightarrow\}$ such that for each $q \in Q, \sigma \in \Gamma$,*

$$\delta(q, \sigma, q', d) = \begin{cases} \langle q'|V_\sigma|q\rangle, & \text{if } D(q') = d; \\ 0, & \text{otherwise,} \end{cases} \tag{4.3}$$

(where $\langle q'|V_\sigma|q\rangle$ denotes the coefficient at $|q'\rangle$ in $V_\sigma|q\rangle$).

It is straightforward to verify that if we rewrite the well-formedness conditions using the relation 4.3, then we get:

Lemma 4.1.16 *A simple 2QFA $\mathcal{A}$ satisfies the well-formedness condition if and only if*

$$\sum_{q'} \langle q'|V_\sigma|q_1\rangle^* \langle q'|V_\sigma|q_2\rangle = \begin{cases} 1, & \text{if } q_1 = q_2; \\ 0, & \text{otherwise,} \end{cases}$$

for each $\sigma \in \Gamma$, which holds if and only if every operator V_σ is unitary.

Exercise 4.1.17 *Prove in detail Lemma 4.1.16.*

Example 4.1.18 *Let $\mathcal{A} = \langle \Sigma, Q, q_0, Q_a, Q_r, \delta \rangle$ be a simple 2QFA with $Q = \{q_0, q_1, q_2, q_3, q_4\}$, $\Sigma = \{0, 1\}$, $Q_a = \{q_3\}$, $Q_r = \{q_4\}$ and*

$$\begin{array}{lllll}
V_\#|q_0\rangle = |q_0\rangle, & V_0|q_0\rangle = |q_0\rangle, & V_1|q_0\rangle = |q_1\rangle, & V_\$|q_0\rangle = |q_1\rangle, & D(q_0) = +1, \\
V_\#|q_1\rangle = |q_2\rangle, & V_0|q_1\rangle = |q_2\rangle, & V_1|q_1\rangle = |q_0\rangle, & V_\$|q_1\rangle = |q_0\rangle, & D(q_1) = -1, \\
V_\#|q_2\rangle = |q_4\rangle, & V_0|q_2\rangle = |q_4\rangle, & V_1|q_2\rangle = |q_2\rangle, & V_\$|q_2\rangle = |q_3\rangle. & D(q_2) = +1, \\
V_\#|q_3\rangle = |q_3\rangle, & V_0|q_3\rangle = |q_3\rangle, & V_1|q_3\rangle = |q_3\rangle, & V_\$|q_3\rangle = |q_2\rangle, & D(q_3) = \ \ 0, \\
V_\#|q_4\rangle = |q_1\rangle, & V_0|q_4\rangle = |q_1\rangle, & V_1|q_4\rangle = |q_4\rangle, & V_\$|q_4\rangle = |q_4\rangle, & D(q_4) = \ \ 0.
\end{array}$$

By inspection one sees that all V_σ operators are unitary and therefore if δ is defined as in 4.3, then $\mathcal{A}$ is well-formed. Below, Example 4.1.19, we will see the way $\mathcal{A}$ works.

4.1.5 2QFA versus 1FA

Both acceptance and rejection can be defined for QFA in a natural and straightforward way. We only need to form a natural quantum version of the observation used in the classical FA.

For each input x with $n = |x|$ and a QFA $\mathcal{A} = \langle \Sigma, Q, q_0, Q_a, Q_r, \delta \rangle$ let $C_n^a = \{(q,k) \,|\, (q,k) \in C_n, q \in Q_a\}$, $C_n^r = \{(q,k) \,|\, (q,k) \in C_n, q \in Q_r\}$ and $C_n^- = C_n - C_n^a - C_n^r$. Let E_a, E_r and E_n be the subspaces of $l_2(C_n)$ spanned by C_n^a, C_n^r and C_n^-, respectively.

In order to define language recognition by $\mathcal{A}$ we use the "computational observable" $\mathcal{O}$ that corresponds to the orthogonal decomposition $l_2(C_n) = E_a \oplus E_r \oplus E_n$. The outcome of any observation will be either "accept" or "reject" or "non-terminating".

For example, if the amplitudes of all accepting and rejecting configurations in a superposition are 0, then the result of an observation, according to the observable $\mathcal{O}$ will be "non-terminating" with probability 1 and the state "collapses" into itself (and it will therefore not be disturbed by the observation).

The language recognition by $\mathcal{A}$ is now defined as follows; For an $x \in \Sigma^*$ we consider as the input the string $w_x = \#x\$$, and we assume that the computation starts with $\mathcal{A}$ being in the state $|q_0, 0\rangle$, i.e. with the head on the leftmarker $\#$. Each step of the computation consists of two parts. At first the operator $U_{w_x}^{\delta}$ is applied to the current state and then the resulting state is observed using the computational observable $\mathcal{O}$ defined above. The computation continues this way until the result of an observation is "accept" or "reject", and in such a case the computation halts.

The language $L(\mathcal{A})$ accepted by $\mathcal{A}$ is defined in a similar way as for 1QFA. This way we can talk about bounded-error acceptance and unbounded-error acceptance.

Example 4.1.19 *We show that $L(\mathcal{A}) = 0^*1^*$ for the automaton from the Example 4.1.18. More exactly, each $x \in L$ is accepted with probability 1 and each $x \in \{0,1\}^* - L$ is rejected with probability 1.*

Let us start with an illustration of computations on $\mathcal{A}$ for two different input strings. For the input $0^3 1^2$ we get the following sequence of states:

$$|q_0, 0\rangle \xrightarrow{\#} |q_0, 1\rangle \xrightarrow{0} |q_0, 2\rangle \xrightarrow{0} |q_0, 3\rangle \xrightarrow{0} |q_0, 4\rangle \xrightarrow{1} |q_1, 3\rangle \xrightarrow{0} |q_2, 4\rangle \xrightarrow{1} |q_2, 5\rangle \xrightarrow{1} |q_2, 6\rangle \xrightarrow{\$} |q_3, 6\rangle,$$

and for the input $0^2 10$:

$$|q_0, 0\rangle \xrightarrow{\#} |q_0, 1\rangle \xrightarrow{0} |q_0, 2\rangle \xrightarrow{0} |q_0, 3\rangle \xrightarrow{1} |q_1, 2\rangle \xrightarrow{0} |q_2, 3\rangle \xrightarrow{1} |q_2, 4\rangle \xrightarrow{0} |q_4, 4\rangle.$$

It is easy to verify that for an input $0^i 1^j$, $i, j \geq 0$, the automaton $\mathcal{A}$ enters the state $|q_3, i + j + 1\rangle$ after $i + j + 4$ steps and at that time it will be for the first time that $\mathcal{A}$ gets into a halting state. An input $1^i 0x$ with $i > 0$ is rejected after $i + 3$ steps and the input of the form $0^i 1^j 0x$ with $i > 0$ and $j > 0$ after $i + j + 4$ steps.

Exercise 4.1.20 *Design a simple 2QFA accepting the languages: (a) $\{0^*1^*0^*\}$; (b) $\{(00)^*(11)^*\}$.*

As the next step we show that the family of languages recognized by one-sided bounded-error simple 2QFA in linear time contains also non-regular languages, namely the language $L = \{0^i 1^i \,|\, i > 0\}$.

For each integer n let $\mathcal{A}^{(n)} = \langle\{0,1\}, Q^{(n)}, q_0, Q_a^{(n)} Q_r^{(n)}, \delta^{(n)}\rangle$ be a simple 2QFA with

$$\begin{aligned} Q^{(n)} &= \{q_0, q_1, q_2, q_3\} \cup \{r_{j,k} \,|\, 1 \le j \le n, 0 \le k \le \max\{j, n-j+1\}\} \\ &\quad \cup \{s_j \,|\, 1 \le j \le n\}, \end{aligned}$$

$Q_a^{(n)} = \{s_n\}$, $Q_r^{(n)} = \{q_3\} \cup \{s_j, 1 \le j < n\}$. The transition function $\delta^{(n)}$ is defined as in (4.3) where $V_\sigma^{(n)}, \sigma \in \{\#, 0, 1, \$\}$ and D are given in Figure 4.2. An extension of $V_\sigma^{(n)}$ to other basis states in such a way that all $V_\sigma^{(n)}$ are unitary is straightforward.

$V_\#\|q_0\rangle = \|q_0\rangle$,	$V_\$\|q_0\rangle = \|q_3\rangle$,
$V_\#\|q_1\rangle = \|q_3\rangle$,	$V_\$\|q_2\rangle = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\|r_{j,0}\rangle$,
$V_\#\|r_{j,0}\rangle = \frac{1}{\sqrt{n}}\sum_{l=1}^{n} e^{\frac{2\pi i}{n}jl}\|s_l\rangle, 1 \le j \le n$,	
$V_0\|q_0\rangle = \|q_0\rangle$,	$D(q_0) = \rightarrow$,
$V_0\|q_1\rangle = \|q_2\rangle$,	$D(q_1) = \leftarrow$,
$V_0\|q_2\rangle = \|q_3\rangle$,	$D(q_2) = \rightarrow$,
$V_0\|r_{j,0}\rangle = \|r_{j,j}\rangle, 1 \le j \le n$,	$D(q_3) = \downarrow$,
$V_0\|r_{j,k}\rangle = \|r_{j,k-1}\rangle, 1 \le k \le j, 1 \le j \le n$,	
$V_1\|q_0\rangle = \|q_1\rangle$,	$D(r_{j,0}) = \leftarrow, 1 \le j \le n$,
$V_1\|q_2\rangle = \|q_2\rangle$,	$D(r_{j,k}) = \downarrow, 1 \le j \le n, k \ne 0$,
$V_1\|r_{j,0}\rangle = \|r_{j,n-j+1}\rangle, 1 \le j \le n$,	$D(s_j) = \downarrow, 1 \le j \le n$,
$V_1\|r_{j,k}\rangle = \|r_{j,k-1}\rangle, 1 \le k \le j \le n$.	

Figure 4.2: Transition function of $\mathcal{A}^{(n)}$.

Lemma 4.1.21 *Let $x \in \{0,1\}^*, n \in \mathbf{N}$. If $x \in \{0^i1^i \,|\, i \ge 1\}$, then the 2QFA $\mathcal{A}^{(n)}$ accepts x with the probability 1; otherwise $\mathcal{A}^{(n)}$ rejects x with the probability at least $1 - \frac{1}{n}$. In either case $\mathcal{A}^{(n)}$ halts after $\mathcal{O}(n|x|)$ steps.*

Proof. Figure 4.3 illustrates the basic trick of a 2QFA described in Figure 4.2, that accepts strings from the language $\{0^i1^i \,|\, i > 0\}$. Each computation of $\mathcal{A}^{(n)}$ consists of three phases. In the first phase, in which only states q_0, q_1, q_2 and q_3 are involved, any input word not of the form 0^i1^j is rejected, in a similar way as in Example 4.1.19. For words of the type 0^i1^j the phase ends in the state $|q_2\rangle$ with the head on the rightmost endmarker \$. As the first step of the second phase the operator $V_\$$ is applied and a superposition of n states is formed. This way computation branches into n parallel paths starting in the states $|r_{1,0}\rangle, \ldots, |r_{n,0}\rangle$, each with the amplitude $\frac{1}{\sqrt{n}}$.

In the jth of the paths, starting in the state $|r_{j,0}\rangle$ the head moves, deterministically, to the left endmarker according to the following rules. Each time the head is on a new cell and reads 0 (1) it remains stationary for j $(n-j+1)$ steps and then moves one cell left. Therefore, for an input of the form 0^u1^v the jth head requires exactly $(j+1)u + (n-j+2)v + 1$ steps to reach the left endmarker. If $j \ne j'$, then

$$(j+1)u + (n-j+2)v + 1 = (j'+1)u + (n-j'+2)v + 1 \;\text{ if and only if }\; u = v.$$

This implies that any two heads of all n different computational paths reach the left endmarker at the same time if and only if $u = v$.[3]

[3]This means that in the second phase the automaton is always in a superposition of n basis states (configurations) each corresponding to a different branch, and therefore one can see the computation as being performed by n different heads which can be on different cells.

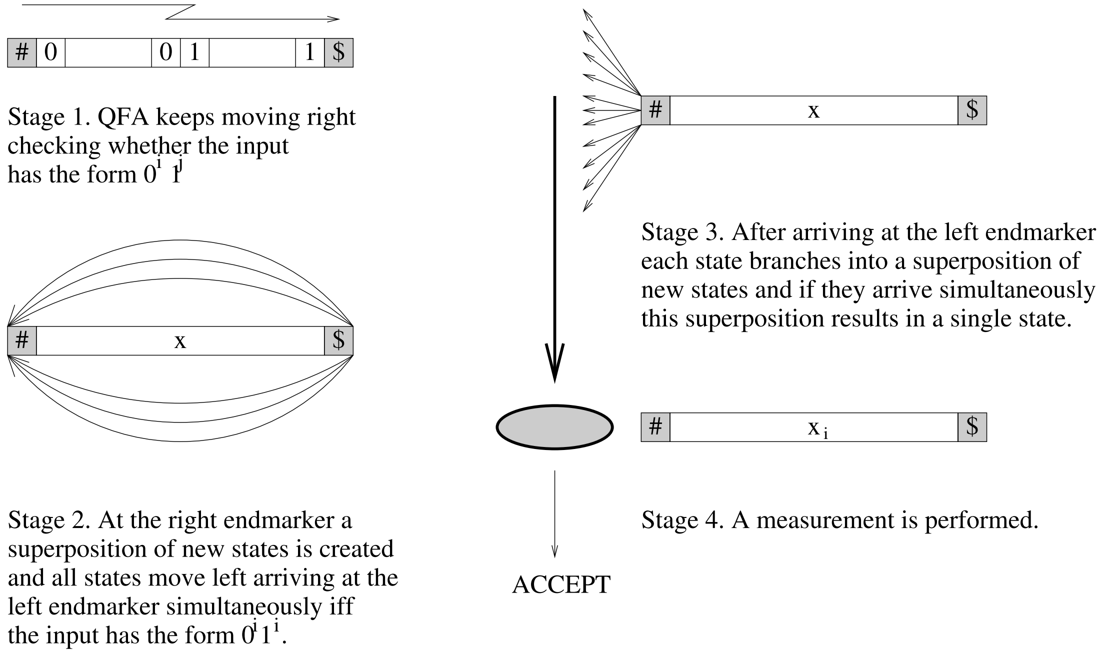


Figure 4.3: QFA recognizing the language $\{0^i 1^i \,|\, i \geq 1\}$ - 60%

In the third phase, consisting of only one step, each computation path splits again, this time the resulting superposition is obtained by an application of the QFT.

In the case $u = v$ all these splittings occur simultaneously and the resulting superposition has the form

$$\frac{1}{n}\sum_{j=1}^{n}\sum_{l=1}^{n} e^{\frac{2\pi i}{n} jl}|s_j\rangle, \tag{4.4}$$

and equals exactly to $|s_n\rangle$ on the basis of the same reasoning as in Sections 3.1 and 3.2. In the last phase, in addition, an observation is performed using the observable $\mathcal{O}$. In the case $u = v$ the result of such an observation is "accept" with probability 1.

Finally, in the case $u \neq v$, no two different computational paths come to the left endmarker at the same time and therefore after Phase 3, when the first observation is made, we get "accept" with probability only $\frac{1}{n}$ and "reject" with probability $1 - \frac{1}{n}$. ▯

The above method can be used to show that some non-context-free languages can also be accepted in linear time by bounded error 2QFA.

Exercise 4.1.22 *Which of the following languages can be accepted by bounded-error linear time QFA: (a)* $\{a^n b^n a^n \,|\, n \geq 1\}$*; (b)* $\{a^n b^{2n} \,|\, n \geq 1\}$*; (c)* $\{a^i b^j c^j \,|\, i, j \geq 1\} \cup \{a^i b^i c^j \,|i, j \geq 1\}$*?*

Simple 2QFA can be shown to accept all regular languages. Actually a stronger statement has been shown. Namely, that any deterministic (one-way) FA can be simulated by a two-way reversible finite automaton (which can be defined as a 2QFA, amplitudes of which are only 0 or 1). As a corollary we have

Theorem 4.1.23 *Every regular language is accepted by a 2QFA.*

This theorem has been shown by Kondacs and Watrous (1997) directly and it also follows from a more general result due to Lange at al. (1997).

Exercise 4.1.24 *Show that any 1FA can be simulated by a reversible two-way finite automaton.*

Exercise 4.1.25 *Show that the language of words over the alphabet* $\{0, 1\}$ *that contain the same number of* 0*'s and* 1*'s can be accepted by a 2QFA*

Exercise 4.1.26 *Let us call a 2QFA* $\mathcal{A}$ **one-directional** *if it never moves left. Show that one-directional 2QFA can accept non-regular languages.*

Open problem 4.1.27

1. *Is it true that one-directional 2QFA accept all regular languages?*
2. *Is it true that unidirectional 2QFA are as powerful as 2QFA?*
3. *Can we increase the power of 2QFA by considering some other observables?*
4. *Can exponential time, bounded-error, 2QFA recognize languages not recognizable by linear time, bounded-error 2QFA?*
5. *How powerful are 2QFA compared to multihead FA?*
6. *Find an elegant characterization of the family of languages acceptable by 2QFA?*

Remark 4.1.28 One-directional 2QFA seem to be very powerful. Amano and Iwama (1999), who denote them as 1.5QFA, showed that one of the basic decision problem, the emptiness problem, which is decidable even for pushdown automata, is undecidable for 1.5QFA. Perhaps it would have been more proper to call them 1QFA and those denoted that way so far as “real-time QFA”,

There are several variants of the models of quantum finite automata discussed here and they need to be explored in order to get a better insight into the power of quantum computation on the “finite state” level. Some of these models are presented in the following exercises (Ambainis and Freivalds (1998)).

Exercise 4.1.29 *Let us consider special two-way QFA in which the head keeps making left-to-right and right-to-left passes between endmarkers. (a) Show that such a quantum automaton can recognize the language* $\{0^i1^i \mid i \geq 0\}$ *in the following sense. If* $x \notin L$ *the automaton stops with probability* 1 *after* $\mathcal{O}(|x|)$ *scans of the tape; if* $x \in L$*, then it never stops; (b) explore more potentials of such quantum automata.*

Exercise 4.1.30 *Show that for any $\varepsilon > 0$ there is a 2-way probabilistic finite automaton $\mathcal{A}_1$ and an 1QFA $\mathcal{A}_2$ such that with probability at least $1 - \varepsilon$: (1) $\mathcal{A}_1$ stops in time quadratic in the length of the input; (2) $\mathcal{A}_2$ accepts the output of $\mathcal{A}_1$ if and only if the input x of $\mathcal{A}_1$ is in $\{0^i1^i \mid i > 0\}$.*

The result of Exercise 4.1.30 is of interest for the following reason: neither 1QFA nor 2PFA can recognize non-regular languages (see Freivalds, 1981, and Kaneps and Freivalds, 1991), but their composition can.

4.2 Quantum Turing Machines

Quantum versions of Turing machines play a similar fundamental role in the theory of quantum computing as classical Turing machines play in the classical complexity theory.

4.2.1 One-tape quantum Turing machines

One-tape quantum Turing machines are the main type of quantum Turing machines to be investigated in the following.

Basic model and its evolution

Definition 4.2.1 *A (one-tape)* **quantum Turing machine** $\mathcal{M} = \langle \Sigma, Q, q_0, q_f, \delta \rangle$, *QTM in short, is defined analogously to a PTM, with the initial state q_0 and the final state q_f, but the transition amplitude mapping*

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

is required to be such that quantum evolution of $\mathcal{M}$, to be defined later, is unitary. If the unitarity condition is not required to be valid we talk about a quasi-QTM (in short qQTM). Moreover, when neither the initial nor the final state is of importance a QTM is specified in the form $\langle \Sigma, Q, \delta \rangle$.[4]

Exercise 4.2.2 *(A random walk QTM—Watrous, 1998) Let us consider a QTM $\mathcal{M}$ with two states alphabet $\{0, 1\}$, $q_0 = 0$, and any tape alphabet Σ, such that at each step $\mathcal{M}$ applies Hadamard transform to its state and moves its head left or right, depending on whether new state is 0 or 1. ($\mathcal{M}$ therefore never changes the content of the tape.) Describe the superposition of $\mathcal{M}$ after i steps, $i = 1, 2, \ldots, 10$. What can you say about the behaviour of $\mathcal{M}$ in general?*

[4]Quite a different approach to QTM is pursued by Benioff, see page 261 and Benioff (1998) for the recent paper. It is the so-called "physical QTM", and the main difference is in problems on which investigation concentrates. Benioff (1998) is concerned with the process defined by a step operator that is used to construct Hamiltonian according to Feynman's ideas—see page 261. It is interesting to compare this "physical" approach to QTM with the "automata theoretic" approach as presented in this section. In spite of the fact that Benioff's approach may be seen, from an implementation point of view, as better reflecting quantum physics laws; so far the automata-theoretic approach seems to be far more insightful, stimulating, and productive.

The concept of configuration is the basic one for the description of the quantum evolution of QTM. A configuration of $\mathcal{M}$ is determined by the content τ of the tape, $\tau \in \Sigma^{\mathbf{Z}}$, by an $i \in \mathbf{Z}$ which specifies the position of the head, and by a $q \in Q$, the current state of the tape. Let $C_{\mathcal{M}}$ denote the set of all configurations of $\mathcal{M}$. Computation (evolution) of $\mathcal{M}$ is performed in the inner-product space $H_{\mathcal{M}} = l_2(C_{\mathcal{M}})$ with the basis $\{|c\rangle \mid c \in C_{\mathcal{M}}\}$.

The transition function δ uniquely determines a mapping $a : C_{\mathcal{M}} \times C_{\mathcal{M}} \to \mathbf{C}$ such that for $c_1, c_2 \in C_{\mathcal{M}}$, $a(c_1, c_2)$ is the amplitude of the transition of $\mathcal{M}$ from the basis state $|c_1\rangle$ to $|c_2\rangle$. The time evolution mapping $U_{\mathcal{M}} : H_{\mathcal{M}} \to H_{\mathcal{M}}$ is then defined as follows.

1. If $|c\rangle$ is a basis state, then
$$U_{\mathcal{M}}|c\rangle = \sum_{c' \in C_{\mathcal{M}}} a(c, c')|c'\rangle.$$

2. If $|\phi\rangle = \sum_{c \in C_{\mathcal{M}}} \alpha_c |c\rangle$ is a superposition, then
$$U_{\mathcal{M}}|\phi\rangle = \sum_{c \in C_{\mathcal{M}}} \alpha_c U_{\mathcal{M}}|c\rangle.$$

Remark 4.2.3 1. Sometimes it is convenient to see the basis states of a QTM as being tensor products of the form $|q\rangle|\tau\rangle|i\rangle$, where $q \in Q$, $\tau \in \Sigma^{\mathbf{Z}}$, $i \in \mathbf{Z}$, with i representing the position of the head.

2. Observe that the subspace of $H_{\mathcal{M}} = l_2(C_{\mathcal{M}})$ consisting of finite sums of configurations is not a Hilbert space but a dense subspace of $l_2(C_{\mathcal{M}})$.

Measurements

In addition to a standard measurement with respect to the basis $\{|c\rangle \mid c \in C_{\mathcal{M}}\}$, which, when applied to a state $|\phi\rangle = \sum_{c \in C_{\mathcal{M}}} \alpha_c |c\rangle$, provides a configuration c with probability $|\alpha_c|^2$, measurements of certain cells of the tape are also of importance. The concept of an equivalence of configurations is needed to introduce the corresponding observables.

If c is a configuration, then let $c(i)$ be the ith symbol of its tape. The equivalence relation on configurations, denoted $\sim_I$ and associated to a set $I = \{-N, \dots, N\}$, $N \in \mathbf{N}$, is defined by
$$c_i \sim_I c_j \Longleftrightarrow c_i(k) = c_j(k) \quad \textit{whenever} \quad k \in I.$$

(In other words, two configurations are equivalent over I if they have the same symbols in the positions from I.) Let us now denote by $[c]$ the equivalence class of configurations containing the configuration c. In the case of the state $|\psi\rangle = \sum_{c \in C_{\mathcal{M}}} \alpha_c |c\rangle$, the probability of such equivalence class is defined by
$$p([c]) = \sum_{c_i \in [c]} |\alpha_{c_i}|^2.$$

Definition 4.2.4 *The standard observable of the cells I of the tape yields the equivalence class $[c]$ with probability $p([c])$ and the post-observation superposition is*
$$\frac{1}{\sqrt{\sum_{c_j \in [c]} |\alpha_{c_j}|^2}} \sum_{c_j \in [c]} \alpha_{c_j} c_j,$$
provided the class $[c]$ was observed.

Well-formedness conditions

As shown in Theorem 4.2.6, unitarity of a qQTM $\mathcal{M}$ is ensured if $\mathcal{M}$ satisfies the following so-called strong well-formedness conditions.

Definition 4.2.5 *A qQTM $\mathcal{M} = \langle \Sigma, Q, q_0, q_f, \delta \rangle$ with the transition mapping*

$$\delta : Q \times \Sigma \times \Sigma \times Q \times \{\leftarrow, \downarrow, \rightarrow\} \longrightarrow \mathbf{C}$$

is said to be strongly well-formed if the following conditions are satisfied.

1. **Local probability condition.** *For any $(q_1, \sigma_1) \in Q \times \Sigma$;*
$$\sum_{(\sigma,q,d)\in\Sigma\times Q\times\{\leftarrow,\downarrow,\rightarrow\}} |\delta(q_1, \sigma_1, \sigma, q, d)|^2 = 1.$$

2. **Separability conditon I.** *For any two different pairs $(q_1, \sigma_1), (q_2, \sigma_2)$ from the set $Q \times \Sigma$:*
$$\sum_{(q,\sigma,d)\in Q\times\Sigma\times\{\leftarrow,\downarrow,\rightarrow\}} \delta^*(q_1, \sigma_1, \sigma, q, d)\delta(q_2, \sigma_2, \sigma, q, d) = 0.$$

3. **Separability condition II.** *For any $(q, \sigma, d), (q', \sigma', d')$ from the set $Q\times\Sigma\times\{\leftarrow, \downarrow, \rightarrow\}$ such that $(q, \sigma, d) \neq (q', \sigma', d')$:*
$$\sum_{(q_1,\sigma_1)\in Q\times\Sigma} \delta^*(q_1, \sigma_1, \sigma, q, d)\delta(q_1, \sigma_1, \sigma', q', d') = 0.$$

4. **Separability condition III.** *For any $(q_1, \sigma_1, \sigma_1'), (q_2, \sigma_2, \sigma_2') \in Q \times \Sigma \times \Sigma$ and $d_1 \neq d_2 \in \{\leftarrow, \downarrow, \rightarrow\}$:*
$$\sum_{q\in Q} \delta^*(q_1, \sigma_1, \sigma_1', q, d_1)\delta(q_2, \sigma_2, \sigma_2', q, d_2) = 0.$$

A **computation** of a QTM $\mathcal{M}$ is a sequence of superpositions $c_0, c_1, c_2, \ldots$, where c_0 is the initial configuration and for $i \geq 1$ $c_i = U_{\mathcal{M}}(c_{i-1})$ or c_i is obtained from c_{i-1} via an observation procedure.

Theorem 4.2.6 *(Hirvensalo[5], 1997) If a qQTM $\mathcal{M}$ satisfies the above strongly well-formedness conditions, then its evolution is unitary.*

Proof. Let us assume a fixed enumeration of all configurations. For any configuration c_i then

$$U_{\mathcal{M}}(c_i) = \sum_{l=1}^{\infty} \alpha_{li} c_l, \tag{4.5}$$

where α_{li} is the amplitude of reaching c_l from c_i. Observe that for each i the sum (4.5) is actually finite, because there are only finitely many configurations reachable from any c_i in one step. $U_{\mathcal{M}}$ can now be seen as a finite or infinite matrix with $U_{\mathcal{M}}(l, i) = \alpha_{li}$. Each column of the matrix has therefore only finitely many nonzero elements.

[5]The proof makes use of the ideas of Bernstein and Vazirani (1993).

If now $U^*_{\mathcal{M}}$ is the adjoint matrix to $U_{\mathcal{M}}$, then

$$U^*_{\mathcal{M}}(c_i) = \sum_{l=1}^{\infty} \alpha^*_{il} c_l.$$

We first show that $U^*_{\mathcal{M}} U_{\mathcal{M}} = I$. This will imply that the mapping $U_{\mathcal{M}}$ is one-to-one. Indeed,

$$U^*_{\mathcal{M}}(U_{\mathcal{M}}(c_i)) = \sum_{l=1}^{\infty} \left(\sum_{k=1}^{\infty} \alpha_{ki}\alpha^*_{kl} \right) c_l.$$

The right-hand side is exactly c_i if

$$\sum_{k=1}^{\infty} \alpha_{ki}\alpha^*_{kl} = 0, \tag{4.6}$$

whenever $l \neq i$ and

$$\sum_{k=1}^{\infty} \alpha_{ki}\alpha^*_{ki} = \sum_{k=1}^{\infty} |\alpha_{ki}|^2 = 1.$$

The last equalities follow from the condition (1) in the definition of well-formedness.

Let us now consider condition (4.6), which actually requires orthogonality of the sequence of amplitudes $\{\alpha_k \,|\, c_k$ is the successor configuration of $c_i, k \in \mathbf{N}\}$ and $\{\beta_k \,|\, c_k$ is the successor configuration of $c_l, k \in \mathbf{N}\}$.

Observe that from different configurations c_i and c_l it is possible to obtain c_k in one step only if the head positions in c_i and c_l differ at most by two positions.

Now if the head is in the same position in c_i and c_l, the orthogonality follows from condition 3; if the heads are in different positions, the orthogonality follows from condition 4 of the definition of well-formedness.

To finish the proof of the theorem it is now sufficient to show that the mapping $U_{\mathcal{M}}$ is surjective. This will imply that $U^*_{\mathcal{M}} = U^{-1}_{\mathcal{M}}$ and the theorem is proved.

A configuration c_i will be called reachable (by $\mathcal{M}$) if c_i occurs with a non-zero coefficient in some superpositions $U_{\mathcal{M}}(c_k) = \sum_{j=1}^{\infty} \alpha_j c_j$—which happens if the ith row of $U_{\mathcal{M}}$ is not empty.

We first show that each configuration is reachable. This will imply that $U_{\mathcal{M}}$ has no empty row.

Let us assume, on the contrary, that there is a configuration c' not reachable from any other configuration. Let $\sigma_1 q \sigma_2 \sigma_3$ be a "local subconfiguration" of c' containing the state q, symbols in the cell with the head on σ_2, and in two neighbouring cells σ_1, σ_3. If c' is not reachable, then the same has to be true for all configurations having $\sigma_1 q \sigma_2 \sigma_3$ as the local subconfiguration. All these configurations will be said to be locally like c'. Let us now take an $n \geq 4$ and consider all configurations $c_{i_1}, \dots, c_{i_K}$ such that in these configurations all cells outside cells $0, 1, \dots, n$ have blanks and the state symbol is not in the nth cell. For the subspace $V \subset C_{\mathcal{M}}$ generated by configurations $c_{i_1}, \dots, c_{i_K}$, we have $dim(U_{\mathcal{M}}(V)) = K = n|Q|\,|\Sigma|^n$ because $U_{\mathcal{M}}$ is injective.

In one step from any c_{i_j} one can either reach another c_{i_k}, or exit the chosen $n+1$ cells, and there are at most $2|Q||\Sigma|^n$ configurations to exit into. In total this gives $K + 2|Q|\,|\Sigma|^n$ possible successor configurations. However, at least $(n-2)|\Sigma|^{n-3}$ of them look locally like c', and therefore they cannot be reached from any configuration. Consequently we get

$$U_{\mathcal{M}}(c_{i_k}) = \sum_{j_l \in J} \alpha_l c_{j_l}$$

with the index set J of configurations of the cardinality at most $K+2|Q|\,|\Sigma|^n-(n-2)|\Sigma|^{n-3}$. Therefore,

$$K = dim(U_{\mathcal{M}}(V)) \leq K + 2|Q|\,|\Sigma|^n - (n-2)|\Sigma|^{n-3},$$

which yields $n \leq 2|Q|\,|\Sigma|^3 + 2$ and contradicts the fact that n could be chosen arbitrarily large.

Without loss of generality let us now assume that the configuration c_1 is not in the range of $U_{\mathcal{M}}$. As shown above, the first row of $U_{\mathcal{M}}$ is not empty. Moreover, it follows from the condition 2 of the definition of well-formedness of QTM that any two rows of $U_{\mathcal{M}}$ are orthogonal.

Let us now choose N such that $\alpha_{ij} = 0$ for $j > N$. From the orthogonality of rows of $U_{\mathcal{M}}$ it then follows that

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1N} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2N} \\ \vdots & \vdots & \dots & \vdots \end{pmatrix} \begin{pmatrix} \alpha_{11}^* \\ \alpha_{12}^* \\ \vdots \\ \alpha_{1N}^* \end{pmatrix} = \begin{pmatrix} A \\ 0 \\ 0 \\ \vdots \end{pmatrix},$$

where $A = |\alpha_{11}|^2 + \ldots + |\alpha_{1N}|^2 > 0$. Hence

$$U_{\mathcal{M}}(\alpha_{11}^* c_1 + \ldots + \alpha_{1N}^* c_N) = Ac_1$$

and by dividing both sides by A we get that c_1 is in the range of $U_{\mathcal{M}}$. □

Exercise 4.2.7 *Show that any reversible TM is QTM.*

Exercise 4.2.8 *Let $\delta : Q \times \Sigma \to \Sigma \times Q \times \{\leftarrow, \downarrow, \rightarrow\}$ be a partial mapping satisfying the following conditions.*

1. *If $\delta(q_1, \sigma_1) = (\sigma_1', q, d_1)$, $\delta(q_2, \sigma_2) = (\sigma_2', q, d_2)$, then $d_1 = d_2$;*
2. *δ is one-to-one if the direction components ($\{\leftarrow, \downarrow, \rightarrow\}$) are ignored.*

Show that δ can be completed to be a total mapping in such a way that $\mathcal{M} = \langle Q, \Sigma, \delta \rangle$ is an RTM.

Hirvensalo was the first to come with some conditions for a general qQTM to have unitary evolution. However, conditions presented in Definition 4.2.5 are only sufficient conditions. As shown by Ozawa, they are not necessary. The first set of sufficient and necessary well-formedness conditions is due to Ozawa (1998). Instead of separability conditions II and III, he considers the following two conditions:

3. **Separability condition II':** For any $(q_1, \sigma_1, \sigma_1')$, $(q_2, \sigma_2, \sigma_2') \in Q \times \Sigma \times \Sigma$

$$\sum_{p \in Q} \delta^*(q_1, \sigma_1, \sigma_1', p, \rightarrow)\delta(q_2, \sigma_2, \sigma_2', p, \leftarrow) = 0.$$

4. **Separability condition III':** For any $(q_1, \sigma_1, \sigma_1')$, $(q_2, \sigma_2, \sigma_2') \in Q \times \Sigma \times \Sigma$

$$\sum_{p \in Q, d \in \{\downarrow, \rightarrow\}} \delta^*(q_1, \sigma_1, \sigma_1', p, d-1)\delta(q_2, \sigma_2, \sigma_2', p, d) = 0,$$

where $d - 1 = \leftarrow$ if $d = \downarrow$ and $d - 1 = \downarrow$ if $d = \rightarrow$.

The proof that these conditions are really sufficient and necessary has been given in Ozawa and Nishimura (1998) and it is too technical and lengthy to be presented here. The following QTM, due to Ozawa, 1998, is an example of a QTM that does not satisfies conditions of Definition 4.2.5:

Exercise 4.2.9 *Verify that QTM with the set of states* $Q = \{0,1\}$, $\Sigma = \{0\}$ *and the transition function defined for simplicity as the mapping* $\delta : Q \times Q \times \{\leftarrow, \downarrow, \rightarrow\} \rightarrow \mathbf{C}$ *defined for* $a = \delta(p,q,c)$ *as follows*

q	p	d	a	q	p	d	a	q	p	d	a	q	p	d	a
0	0	$\leftarrow$	0	0	1	$\leftarrow$	$\frac{1}{2}$	1	0	$\leftarrow$	0	1	1	$\leftarrow$	$\frac{1}{2}$
0	0	$\downarrow$	$\frac{1}{2}$	0	1	$\downarrow$	$\frac{1}{2}$	1	0	$\downarrow$	$\frac{1}{2}$	1	1	$\downarrow$	$-\frac{1}{2}$
0	0	$\rightarrow$	$-\frac{1}{2}$	0	1	$\rightarrow$	0	1	0	$\rightarrow$	$\frac{1}{2}$	1	1	$\rightarrow$	0

satisfies Ozawa's but not Hirvensalo's conditions,

Input–output conventions

The initial configuration of a QTM $\mathcal{M} = \langle \Sigma, Q, q_0, q_f, \delta \rangle$ has the form q_0x, $x \in (\Sigma - \{\lambda\})^*$, where λ stands for the blank symbol, and x is written on cells numbered $0, 1, 2, \ldots, |x| - 1$, with all other cells filled with the blank and the head is on the cell number 0. $\mathcal{M}$ halts on the input x when it finally enters the final state q_f. The number of steps needed to reach q_f is the computation time of $\mathcal{M}$ on x. After $\mathcal{M}$ halts the standard measurement is performed and its output is the string on the tape of the resulting configuration consisting of the tape contents between the leftmost and rightmost non-blank symbol. Each output is therefore produced with certain probability. For each input $\mathcal{M}$ produces a sample from a probability distribution on its outputs.

Remark 4.2.10 It has been shown by Bernstein (1997) that it is sufficient to consider a single final state and a single measurement, after halting, if time efficiency of quantum Turing machines is considered. On the other hand, when space efficiency of QTM is considered, then, as in the case of QFA, a measurement after each step seems to be more appropriate.

Two QTM will be considered as equivalent if for each input their output probability distributions are close to each other; or more exactly, if their total variation distance is small.[6]

Exercise 4.2.11 *Let* $|\psi\rangle$, $|\phi\rangle$ *be states of an inner product space* H *such that* $||\phi|| = ||\psi|| = 1$, $||\phi - \psi|| \leq \varepsilon$. *Show that the total variation distance between probability distributions corresponding to measurements of* $|\phi\rangle$ *and* $|\psi\rangle$, *with respect to the standard basis, is at most* 4ε.

4.2.2 Variations on the basic model

Quantum Turing machines are powerful, but far from easy to deal with, models of computation. The requirement of unitarity imposed on their evolutions is very severe and makes

[6] Total variation distance of two probability distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ on the domain I is $\frac{1}{2}\sum_{i \in I} |\mathcal{D}_1(i) - \mathcal{D}_2(i)|$.

it far from trivial to perform even small modifications of the transition function of a QTM. For example, even a simple classical design step, to add a transition from the final state to the initial state, may not work—unitarity can disappear—and the new qQTM is not really quantum.

There are several other models of QTM that are easier to deal with because of the restrictive nature of the movement of the head, and their computational power, or even efficiency, is not worse than that of the most general model presented in Section 4.2.1.

DBV–model of quantum Turing machine

The basic restriction, due to Deutsch (1985), and worked out by Bernstein and Vazirani (1993, 1997), is to consider only such QTM the head of which in each step must move either right or left. If the transition function δ of a qQTM $\mathcal{M}$ satisfies this condition, then the evolution of $\mathcal{M}$ is unitary if the first two conditions of Definition 4.2.5 and the first condition of Remark 4.2.1 are satisfied, as shown in detail by Bernstein and Vazirani (1997).

The DBV-model of quantum Turing machine will be the basic one to consider in the rest of this section and in Chapter 5.

Unidirectional quantum Turing machines

In various situations, for example when designing a universal QTM, it is very handy to work with so-called unidirectional QTM, in which each state can be entered only by a head moving in the direction uniquely specified by the state itself.

Definition 4.2.12 *A QTM $\mathcal{M} = \langle \Sigma, Q, q_0, q_f, \delta \rangle$ is called* **unidirectional** *(uQTM) if $d_1 = d_2$ whenever $\delta(q_1, \sigma_1, \sigma_1', q, d_1) \neq 0 \neq \delta(q_2, \sigma_2, \sigma_2', q, d_2)$.*

It is not obvious whether a restriction to uQTM does not represent an essential reduction of the power or efficiency of QTM. In order to deal with this problem the concept of simulation of QTM is needed.

Definition 4.2.13 *A QTM $\mathcal{M}'$ simulates a QTM $\mathcal{M}$, with slowdown $f : \mathbf{N} \rightarrow \mathbf{N}$ and accuracy ε, if the following condition holds. If $\mathcal{D}$ is a distribution which is produced by $\mathcal{M}$ on input x after t steps and $\mathcal{D}'$ is the distribution produced by $\mathcal{M}'$ on the same input after $f(t)$ steps, then $|\mathcal{D} - \mathcal{D}'| \leq \varepsilon$.*

Concerning the power and efficiency of uQTM the following basic result has been shown by Bernstein and Vazirani (1997).

Theorem 4.2.14 *Any QTM can be simulated by a uQTM with slowdown factor at most* 5.

Exercise 4.2.15 *Show that if $\mathcal{M}$ is an RTM, then there is a DBV-RTM $\mathcal{M}'$ that simulates $\mathcal{M}$ with slowdown at most* 2.

Exercise 4.2.16 *Show that if $\mathcal{M}$ is a QTM, then there is a DBV-QTM $\mathcal{M}'$ that simulates $\mathcal{M}$ with constant slowdown.*

Multitape Turing machines

A natural generalization of one-tape QTM are, as in the classical case, multitape QTM. They are defined in a similar way to one-tape QTM. Concerning time efficiency, it is far from easy to see what kind of increase in the computational power such a model brings. However, as Bernstein (1997) shows, such an increase is not essential; namely, each multitape QTM can be simulated by a one-tape QTM with only polynomial time slowdown and the degree of the polynomial does not depend on the number of tapes. Multitape quantum Turing machines are a basic model to study the space efficiency of quantum Turing machines. They are discussed in more detail in Section 5.2.

The first attempt to formulate necessary and sufficient local well-formedness conditions for multitape qQTM to be QTM, similar to those for one-tape QTM, page 166, are due to Ozawa and Nishimura (1998).

4.2.3 Are quantum Turing machines analogue or discrete?

It is intuitively clear that in order to be able to investigate the computational power of QTM and to consider the idea of a universal QTM some restrictions have to be made on the amplitudes of the transitions of QTM. Otherwise we could encode sometimes into amplitudes what is to be computed. Moreover, the existence of a universal QTM would be out of the question because for that we need a clear and finite way of specifying other Turing machines.

The restriction we will make is that to each amplitude α there is a deterministic algorithm that can compute both the real and imaginary parts of α to within 2^{-n} in time polynomial in n (see Adleman et al. 1997).

Three problems concerning amplitudes will be considered in this section. (1) Do we really need such a wide range of amplitudes to be able to fully utilize the power of QTM? (2) How precisely amplitudes need to be specified in order to get outcomes with a desirable precision? (In reality to each amplitude a physical quantity corresponds, for example the angle of a polarizer, and therefore practically there are strong limitations on precision to be achieved.) (3) Are QTM analogue or discrete models of computation?

Types of amplitudes needed

The first result presented below (due to Bernstein and Vazirani, 1993), says that as far as the computational power of QTM is concerned, it is sufficient to consider only QTM with real amplitudes.

Theorem 4.2.17 *Any QTM $\mathcal{M}$ can be simulated, with constant time slowdown, by a QTM $\mathcal{M}'$ all transitions of which have real amplitudes.*

Proof sketch. The basic idea is simple: in any superposition of configurations replace each basis configuration $\alpha|q\rangle|\tau\rangle|j\rangle$ with a superposition $\alpha_i|q_i\rangle|\tau\rangle|j\rangle + \alpha_r|q_r\rangle|\tau\rangle|j\rangle$, where α_i and α_r are the imaginary and real parts of the amplitude α and q_i and q_r are new states. In correspondence with this the transition function δ' of $\mathcal{M}'$ is defined using the transition function δ of $\mathcal{M}$ as follows:

$$\begin{aligned}
\delta'(\sigma, p_r, \sigma', q_r, d) &= Re(\delta(\sigma, p, \sigma', q, d)),\\
\delta'(\sigma, p_r, \sigma', q_i, d) &= Im(\delta(\sigma, p, \sigma', q, d)),\\
\delta'(\sigma, p_i, \sigma', q_r, d) &= -Im(\delta(\sigma, p, \sigma', q, d)),\\
\delta'(\sigma, p_i, \sigma', q_i, d) &= Re(\delta(\sigma, p, \sigma', q, d)).
\end{aligned}$$

$\square$

Exercise 4.2.18 *Show that if a transition function of a QTM is modified as shown in Theorem 4.2.17, then the evolution remains unitary.*

Exercise 4.2.19 *Show that to any quantum circuit over k qubits we can design an equivalent (in a reasonable sense) quantum circuit over $k+1$ qubits all gates of which are represented by unitary matrices over reals.*

Observe that the QTM $\mathcal{M}'$ constructed in the proof of Theorem 4.2.17, which uses only real amplitudes and simulates $\mathcal{M}$, works in the Hilbert space of twice as large dimension as $\mathcal{M}$.

Theorem 4.2.17 was strengthened by Adleman et al. (1997). They showed that with respect to the computational power it is sufficient to consider only rational amplitudes, and that the set of amplitudes $\{0, \pm\frac{3}{5}, \pm\frac{4}{5}, 1\}$ is sufficient to construct a universal QTM.

Remark 4.2.20 It can be shown (see Watrous, 1995), quite surprisingly, that any QTM $\mathcal{M}$ can be simulated with a constant slowdown by a QTM $\mathcal{M}'$ with the "deterministic head position". In other words, if $\mathcal{M}'$ is observed during a computation, then the probability that the head will be observed at any given tape cell will be either 1 or 0. This implies that the position of the head of $\mathcal{M}'$ can be observed at every time step without affecting its computation.

Precision required in QTM

Two issues will now be considered (see Bernstein and Vazirani, 1997). How imprecisions, either in the initial state, or during evolution steps, influence the overall imprecision of the computation; and how precisely amplitudes need to be given if the overall precision of computation is not to be compromised. To both questions there are very encouraging answers.

The first result shows that imprecisions during computations only add (and they do not increase exponentially as in the case of classical computations).

Theorem 4.2.21 *Let $U_{\mathcal{M}}$ be the evolution operator of a QTM. If $|\phi_i\rangle$ and $|\phi_i'\rangle$, $i = 0, 1, \ldots, t$ are superpositions from $H_{\mathcal{M}}$, such that $||\phi_i - \phi_i'|| \leq \varepsilon$, $|\phi_i\rangle = U_{\mathcal{M}}|\phi_{i-1}'\rangle$, then $||\phi_t' - U_{\mathcal{M}}^t \phi_0|| \leq t\varepsilon$.*

Proof. Denote $|\psi_i\rangle = |\phi_i'\rangle - |\phi_i\rangle$. It holds.

$$\begin{aligned} |\phi_t'\rangle &= |\phi_t\rangle + |\psi_t\rangle = U_{\mathcal{M}}|\phi_{t-1}'\rangle + |\psi_t\rangle = \ldots = \\ &= U_{\mathcal{M}}^t|\phi_0\rangle + U_{\mathcal{M}}^t|\psi_0\rangle + U_{\mathcal{M}}^{t-1}|\psi_1\rangle + \ldots + U_{\mathcal{M}}|\psi_{t-1}\rangle + |\psi_t\rangle \end{aligned}$$

□

To formulate the second result we need the concept of "closeness" for QTM.

Definition 4.2.22 *Two QTM $\mathcal{M}$ and $\mathcal{M}'$ are ε-close, $\varepsilon > 0$, if they have the same sets of states and symbols and if the difference between pairs of the corresponding amplitudes has magnitude at most ε.*

Two close QTM produce two close evolutions in the following sense.

Theorem 4.2.23 *(Bernstein, Vazirani, 1998) If two QTM, $\mathcal{M}$ and $\mathcal{M}'$, with the set of states Q and alphabet Σ are ε-close, then the difference of their time evolutions has the norm bounded by $2|Q|\,|\Sigma|\varepsilon$. (The statement holds also in the case $\mathcal{M}'$ is a qQTM.)*

Proof. Let $\mathcal{M}$ and $\mathcal{M}'$ be QTM that are ε-close and let Σ and Q be their tape symbols and states. Both $H_{\mathcal{M}}$ and $H_{\mathcal{M}'}$ have the same basis $\{|c\rangle \,|\, c \in C_{\mathcal{M}} = C_{\mathcal{M}'}\}$. The difference in their evolution from a state $|\phi\rangle = \sum_{c\in C_{\mathcal{M}}} \alpha_c|c\rangle$ can be expressed as follows

$$U|\phi\rangle - U'|\phi\rangle = \sum_{c_j\in C_{\mathcal{M}}} \left(\sum_{c_i\in C_{\mathcal{M}'}} (\varepsilon_{i,j} - \varepsilon'_{i,j})\alpha_{c_i} \right) |c_j\rangle,$$

where $\varepsilon_{i,j}$ and $\varepsilon'_{i,j}$ are amplitudes of the transition from c_i to c_j in $\mathcal{M}$ and $\mathcal{M}'$. Since each configuration has at most $2|\Sigma|\,|Q|$ predecessor configurations and $|\sum_{i=1}^n a_i|^2 \le n\sum_{i=1}^n a_i^2$ for any real a_is and any n, we get

$$\begin{aligned}
||(U-U')\phi||^2 &= \sum_{c_j\in C_{\mathcal{M}}} \left| \sum_{c_i\in C_{\mathcal{M}}} (\varepsilon_{i,j} - \varepsilon'_{i,j})\alpha_{c_i} \right|^2 \\
&\le \sum_{c_j\in C_{\mathcal{M}}} 2|\Sigma|\,|Q| \sum_{c_i\in C_{\mathcal{M}}} |(\varepsilon_{i,j} - \varepsilon_{i,j})\alpha_{c_i}|^2 \\
&\le 2|\Sigma|\,|Q|\varepsilon^2 \sum_{c_i\in C_{\mathcal{M}}} \sum_{c_j\in C_{\mathcal{M}}} |\alpha_{c_i}|^2 \\
&\le 4|\Sigma|^2|Q|^2\varepsilon^2.
\end{aligned}$$

□

Corollary 4.2.24 *Let $\mathcal{M}$ be a QTM, and let $\mathcal{M}'$ be a QTM which is $\frac{\varepsilon}{24|\Sigma|\,|Q|\,t}$-close to $\mathcal{M}$, where $\varepsilon > 0$. Then $\mathcal{M}'$ simulates $\mathcal{M}$ for t steps with accuracy ε. (The statement holds also in the case when $\mathcal{M}'$ is a qQTM.)*

Proof. Let $\varepsilon > 0$. If unitary operators $U_{\mathcal{M}}$ and $U_{\mathcal{M}'}$ are applied to the same state, then, by Theorem 4.2.23, the norm of the difference of the resulting states is bounded from above by $\delta = \frac{\varepsilon}{12t}$.

An application of $U_{\mathcal{M}'}$ can also be seen as an application of $U_{\mathcal{M}}$ and then an addition of perturbations of length at most δ times the length of current superposition. Therefore the length of the superpositions of $U_{\mathcal{M}'}$ after t steps is increased by $(1+\delta)^t \le e$, because $\delta \le \frac{1}{t}$. By Theorem 4.2.21, the difference between superpositions of $\mathcal{M}$ and $\mathcal{M}'$ after t steps is a superposition of the norm at most $3\delta t \le \frac{\varepsilon}{4}$. Now the Corollary follows from Exercise 4.2.11.
□

It follows from the last corollary that $\mathcal{O}(\lg t)$ bits of precision in transition amplitudes are sufficient to support t steps of a QTM with the resulting precision ε. QTM can therefore be considered as discrete models of computation.

Remark 4.2.25 As discussed later, especially in Chapter 5, a variety of fundamental results have been obtained concerning the computational power of QTM. In spite of that many basic problems are still far from being well understood. Two of such problems are termination and measurement issues for QTM.

4.2.4 Programming techniques for quantum Turing machines

Programming of QTM is significantly more difficult than of classical TM. On a general level difficulties are caused by the necessity to produce unitary, and consequently reversible, evolutions. This makes it nontrivial to implement even such simple classical programming primitives as composition, branching and looping. In addition, one has to manage the implementation of such inherently quantum primitives as an efficient implementation of arbitrary unitary transformations.

One general technique, often very useful, is first to design a deterministic TM (DTM) to do the task and then to use Bennett's construction to produce an equivalent reversible, and consequently also quantum, TM to perform the given task. The basic techniques presented bellow are due to Bernstein and Vazirani (1997).

Standard forms for QTM

One specific source of difficulties in programming of QTM is the need to time different computational paths correctly, because configurations interfere only if they are obtained after the same number of evolution steps. In addition, only identical configurations interfere—in other words only configurations with the same tape content, state and position of the head can interfere.

In order to overcome the timing problem and other problems that need to be solved in order to combine various QTM into more complex ones, it is convenient to work only with QTM that have standard input/output behaviour.

Definition 4.2.26 *(1) A QTM $\mathcal{M}$ is called* **well-behaved** *if it halts for all inputs in a state all configurations of which are in the final state and have the head on the same cell. If this cell is always the starting cell, then $\mathcal{M}$ is called* **stationary**.

(2) A QTM $\mathcal{M}$ is said to be in the **normal form** *if each transition from the final state is into its initial state.*

Remark 4.2.27 If a QTM is in the normal form, then there can be no other transition into the initial state than those from the final state. This allows one to redirect transitions from the final state and then add transitions to the initial state.

Example 4.2.28 *We design a stationary, normal-form QTM which maps any state $|\psi\rangle$ of the n-qubit register into the state $H_n|\psi\rangle$, where H_n is the Hadamard transformation, and which halts in time $2n+4$ with its head back on the starting cell. The machine has tape alphabet $\{0,1,\lambda\}$, states $\{q_0, q_a, q_b, q_c, q_f\}$ and the transition function shown in Table 4.1:*

	λ	0	1
q_0		$\|0, q_a, \leftarrow\rangle$	$\|1, q_a, \leftarrow\rangle$
q_a	$\|\lambda, q_b, \rightarrow\rangle$		
q_b	$\|\lambda, q_c, \leftarrow\rangle$	$\frac{1}{\sqrt{2}}(\|0, q_b, \rightarrow\rangle + \|1, q_b, \rightarrow,\rangle)$	$\frac{1}{\sqrt{2}}(\|0, q_b, \rightarrow\rangle - \|1, q_b, \rightarrow\rangle)$
q_c	$\|\lambda, q_f, \rightarrow\rangle$	$\|0, q_c, \leftarrow\rangle$	$\|1, q_c, \leftarrow\rangle$
q_f	$\|\lambda, q_0, \rightarrow\rangle$	$\|0, q_0, \rightarrow\rangle$	$\|1, q_0, \rightarrow\rangle$

Table 4.1: A Turing machine realizing the Hadamard transformation

The machine processes a string x from $\{0,1\}^*$ as follows. Starting in the state q_0 on the leftmost square of x the machine moves left and right and enters the state q_b. In this state *it keeps moving right, replacing each 0 by* $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ *and each 1 by* $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$, *until the first blank is reached. Afterwards the machine keeps moving left until the first blank left from x is reached. As the last step it moves right into the final state.*

Concepts of well-behavedness, stationarity, and normal form have their meaning also for deterministic TM. They also play the key role in the following result, due to Bernstein and Vazirani (1997), that modifies Bennett's result on reversibility.

Theorem 4.2.29 (Synchronization Theorem) *If f is a string-to-string function which can be computed by a DTM in polynomial time and such that the length of $f(x)$ depends only on the length of x, then there is a polynomial time, stationary, normal-form and reversible TM which on any input x produces $x, f(x)$ and whose running time depends only on the length of x.*

Exercise 4.2.30 *Prove synchronization theorem.*

Programming primitives

Just as in the classical case, one of the techniques that makes the design and behaviour of QTM more transparent is to consider cells (tapes) as having several tracks. An equivalent way is to use tape alphabets that are cartesian products of subalphabets. Some of the very basic techniques of manipulation with several tracks that can be used in the case of quantum Turing machines are presented in the following exercises.

Exercise 4.2.31 *Show that given a QTM (RTM) $\mathcal{M} = \langle \Sigma, Q, \delta \rangle$ and a set Σ', there is a QTM (RTM) $\mathcal{M}' = \langle \Sigma \times \Sigma', Q, \delta' \rangle$ such that $\mathcal{M}'$ behaves as $\mathcal{M}$ while leaving its second track unchanged.*

Exercise 4.2.32 *Show that given any QTM (RTM) $\mathcal{M} = \langle \Sigma_1 \times \ldots \times \Sigma_k, Q, \delta \rangle$ and a permutation π on $\{1, \ldots, k\}$, there exists a QTM (RTM) $\mathcal{M}' = \langle \Sigma_{\pi(1)} \times \ldots \times \Sigma_{\pi(k)}, Q, \delta' \rangle$ such that $\mathcal{M}'$ behaves exactly as $\mathcal{M}$ except that its tracks are permuted according to the permutation π.*

The results of the following two easy exercises will be used below in showing how to compose programs in several ways. The first exercise deals with swapping of transitions. We talk about swapping of outgoing transitions of states q_1 and q_2 in a QTM with transition function δ, if δ is replaced by δ' defined as follows; $\delta'(q_1, \sigma) = \delta(q_2, \sigma)$, $\delta'(q_2, \sigma) = \delta(q_1, \sigma)$ for all tape symbols σ and $\delta'(q, \sigma) = \delta(q, \sigma)$ for all $q \notin \{q_1, q_2\}$. In a similar way swapping of incoming transition of two states is defined.

Exercise 4.2.33 *Show that if $\mathcal{M}$ is a well-formed QTM (RTM), then swapping of the incoming or outgoing transitions between a pair of states in $\mathcal{M}$ yields another well-formed QTM (RTM).*

Exercise 4.2.34 *Show that if $\mathcal{M}_1 = \langle \Sigma, Q_1, \delta_1 \rangle$ and $\mathcal{M}_2 = \langle \Sigma, Q_2, \delta_2 \rangle$ are two well-formed QTM (RTM) with disjoint sets of states, then the "union" of these two QTM (RTM) $\mathcal{M} = \langle \Sigma, Q_1 \cup Q_2, \delta_1 \cup \delta_2 \rangle$ is also a well-formed QTM (RTM).*

Lemma 4.2.35 *If $\mathcal{M}_1$ and $\mathcal{M}_2$ are normal-form QTM (RTM) with the same alphabet and q is a state of $\mathcal{M}_1$, then there is a normal-form QTM $\mathcal{M}$ which acts as $\mathcal{M}_1$ except that each time $\mathcal{M}_1$ would enter the state q it runs $\mathcal{M}_2$ instead of that.*

Proof. Let $q_0^{(1)}$, $q_0^{(2)}$, $q_f^{(1)}$ and $q_f^{(2)}$ be the initial and final states of $\mathcal{M}_1$ and $\mathcal{M}_2$. $\mathcal{M}$ will be designed as union of $\mathcal{M}_1$ and $\mathcal{M}_2$ (see Exercise 4.2.34), with the initial state $q_0^{(1)}$ if $q \neq q_0^{(1)}$ and $q_0^{(2)}$ otherwise. In addition, the incoming transitions of q and $q_0^{(2)}$ and also the outgoing transitions of q and $q_f^{(2)}$will be swapped (see Exercise 4.2.33). According to the above two exercises, the resulting QTM $\mathcal{M}$ is well-formed. Since $\mathcal{M}_1$ is in the normal form the final state of $\mathcal{M}$ leads back to its initial state (no matter whether q is the initial or the final state of $\mathcal{M}_1$ or neither of them). ∎

Now we can formulate the first two basic QTM design techniques.

Lemma 4.2.36 (Composition Lemma) *If $\mathcal{M}_1$ and $\mathcal{M}_2$ are well-formed, normal-form QTM (RTM) with the same alphabet, then there is a normal-form QTM (RTM) $\mathcal{M}$ which performs at first computation of $\mathcal{M}_1$, and after $\mathcal{M}_1$ halts, it performs computation of $\mathcal{M}_2$. (That is $\mathcal{M}$ acts as a composition of $\mathcal{M}_1$ and $\mathcal{M}_2$.)*

Proof. Let again $q_0^{(1)}$, $q_0^{(2)}$, $q_f^{(1)}$ and $q_f^{(2)}$ be the initial and final states of $\mathcal{M}_1$ and $\mathcal{M}_2$. To design $\mathcal{M}$ we simply insert $\mathcal{M}_2$ into $\mathcal{M}_1$ for the final state as shown in the proof of Lemma 4.2.35. To show that the resulting machine satisfies the claim of the lemma we proceed as follows.

Since both $\mathcal{M}_1$ and $\mathcal{M}_2$ are in the normal form, the only transition to $q_0^{(1)}$ is from $q_f^{(1)}$ and to $q_0^{(2)}$ from $q_f^{(2)}$ in $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively. Therefore no transition in $\mathcal{M}_1$ is changed when inserting $\mathcal{M}_2$ into $\mathcal{M}_1$ except for those into and out of the state $q_f^{(1)}$. $\mathcal{M}_1$ is well-behaved, and therefore when $\mathcal{M}$ starts in $q_0^{(1)}$, it computes exactly as $\mathcal{M}_1$ until $\mathcal{M}_1$ would have entered the state $q_f^{(1)}$. However, in such a case $\mathcal{M}$ will not reach the state $q_f^{(1)}$ but instead of that the state $q_0^{(2)}$. Moreover, since no other transition of $\mathcal{M}$ has been changed in the union of $\mathcal{M}_1$ and $\mathcal{M}_2$ than those into and out of $q_f^{(1)}$, $\mathcal{M}$ will proceed exactly as if $\mathcal{M}_2$ had been started in the superposition computed by $\mathcal{M}_1$. ∎

Lemma 4.2.37 (Branching Lemma) *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be well-behaved normal-form QTM (RTM) with the same alphabet. Then there is a two-track, well-behaved, normal form QTM (RTM) $\mathcal{M}$ such that if the second track is empty, then $\mathcal{M}$ runs $\mathcal{M}_1$ on the first track and leaves its second track empty, and if the second track has a 1 in the start cell (and all other cells blank), then $\mathcal{M}$ runs $\mathcal{M}_2$ on its first track and leaves 1 where its head ends up. In either case $\mathcal{M}$ makes at most 4 steps more than the corresponding machine ($\mathcal{M}_1$ or $\mathcal{M}_2$).*

Proof. $\mathcal{M}$ is constructed as follows. At first a stationary, normal-form RTM R is designed which (i) always makes four steps and leaves its input unchanged, (ii) always is in a single configuration superposition, and (iii) has two distinguished states q_1 and q_2 with the following properties. If R is run with 1 in the starting cell and blanks elsewhere, then R visits q_1 with the blank tape and with the head on the starting cell and does not visit q_2

at all. On the other hand, if R is run with the blank tape, then R visits q_2 once with the blank tape and does not visit q_1 at all. R will now be used to design $\mathcal{M}$.

At first $\mathcal{M}_1$ and $\mathcal{M}_2$ are extended to have also the second track with the alphabet of R—see Exercise 4.2.31. Afterwards R is extended to have a first track with the common alphabet of $\mathcal{M}_1$ and $\mathcal{M}_2$. Finally, $\mathcal{M}_1$ and $\mathcal{M}_2$ are inserted for the states q_1 and q_2, respectively.

R will have the alphabet $\{\lambda, 1\}$, the set of states $\{q_0, q_1, q_1', q_2, q_2', q_3, q_f\}$ and transitions:

$$\begin{array}{lll}
(q_0, \lambda, \lambda, q_2', \leftarrow) & (q_0, 1, \lambda, q_1', \leftarrow) & (q_1', \lambda, \lambda, q_1, \rightarrow) \\
(q_1, \lambda, 1, q_3, \leftarrow) & (q_2', \lambda, \lambda, q_2, \rightarrow) & (q_2, \lambda, \lambda, q_3, \leftarrow) \\
(q_3, \lambda, \lambda, q_f, \rightarrow) & (q_f, \lambda, \lambda, q_0, \rightarrow) & (q_f, 1, 1, q_0, \rightarrow)
\end{array}$$

□

Loops are the third main classical design tool. Building an RTM that loops indefinitely is easy. Finite loops can be realized in the following way.

Lemma 4.2.38 (Looping Lemma) *There is a stationary, normal-form RTM $\mathcal{M}$ and a constant c with the following property. Given as input an integer k, written in binary, $\mathcal{M}$ makes $\mathcal{O}(k \lg^c k)$ steps and halts with its tape unchanged. Moreover, $\mathcal{M}$ has a special state q^* which $\mathcal{M}$ visits exactly k times, each time with its head on the starting cell.*

Exercise 4.2.39 *Prove the looping lemma.*

Remark 4.2.40 Observe that all constructions of this section that support main classical algorithm design primitives—composition, branching, and looping—can be implemented without any inherently quantum step.

4.3 Quantum Cellular Automata

Cellular automata (CA) have been invented to model the biological world. Using this model von Neumann gave in the 1950s a positive answer to one of the intriguing questions of those times, namely whether machines can reproduce themselves.

During the last four decades, cellular automata have been studied from very many points of view and used to model a surprisingly broad variety of phenomena. The attempts to consider cellular automata as a model of the physical world (see Fredkin, 1990) naturally lead to an intensive study of the reversibility issues of cellular automata. Cellular automata are a model of computation so closely related to the physical world that some of the very first attempts to create quantum versions of the classical models of computation have been oriented on cellular automata. Surprisingly, it took quite a while for a proper model to emerge—see Watrous (1995) and Dürr, LêThanh, and Santha (1996)—for quantum one-dimensional cellular automata.

4.3.1 Classical cellular automata

Informally, a d-dimensional cellular automaton $\mathcal{A}$ with a finite set of **states** Q, a finite **neighbourhood** $N \subseteq \mathbf{Z}^d$, and a **local transition function** $\delta : Q^{|N|} \to Q$, is a d-dimensional array network of identical finite automata with Q as their set of states. For each node $n \in \mathbf{Z}^d$, the neighbourhood N determines the set $\{n\} + N$ of $|N|$ "neighbours" of the node n. Formally, $\mathcal{A} = \langle d, Q, N, \delta \rangle$ and elements (nodes) of $\mathbf{Z}^d$ are regarded as representing those finite automata which $\mathcal{A}$ consists of.

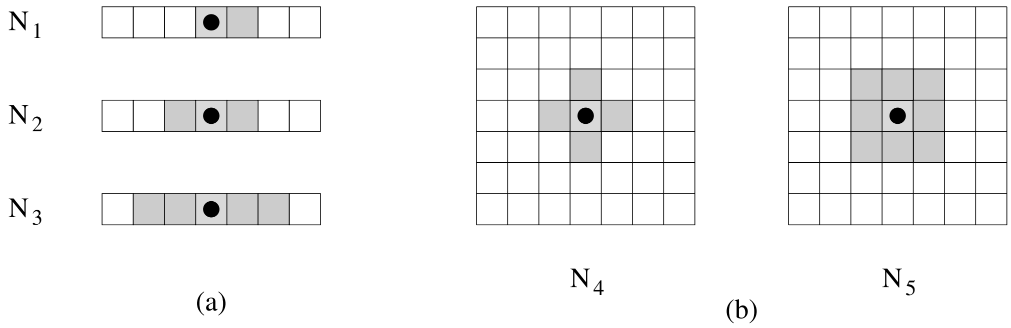


Figure 4.4: Neighbourhoods for one- and two-dimensional cellular automata-65%

Figure 4.5a depicts three neighbourhoods of the dotted nodes for one-dimensional cellular automata: $N_1 = \{0,1\}$, $N_2 = \{-1,0,1\}$, $N_3 = \{-2,-1,0,1,2\}$. Figure 4.5b illustrates two often-used neighbourhoods (again for the dotted nodes) for two-dimensional cellular automata: **von Neumann neighbourhood** $N_4 = \{(-1,0),(0,-1),(0,0),(0,1),(1,0)\}$, and **Moore neighbourhood** $N_5 = \{(-1,-1),(-1,0),(-1,1),(0,-1),(0,0),(0,1),(1,-1),(1,0),(1,1)\}$.

All finite automata of a cellular automaton work concurrently, synchronized, and in discrete time steps. At each time moment the new state of each finite automaton is defined to be the value of the local transition function applied to the current states of a cell and all its neighbours.

Figure 4.5a depicts three neighbourhoods of the dotted nodes for one-dimensional cellular automata: $N_1 = \{0,1\}$, $N_2 = \{-1,0,1\}$, $N_3 = \{-2,-1,0,1,2\}$. Figure 4.5b illustrates two often-used neighbourhoods (again for the dotted nodes) for two-dimensional cellular automata: **von Neumann neighbourhood** $N_4 = \{(-1,0),(0,-1),(0,0),(0,1),(1,0)\}$, and **Moore neighbourhood** $N_5 = \{(-1,-1),(-1,0),(-1,1),(0,-1),(0,0),(0,1),(1,-1),(1,0),(1,1)\}$. In order to describe more formally the overall behaviour of a cellular automaton $\mathcal{A} = \langle d, Q, N, \delta \rangle$ the concept of configuration

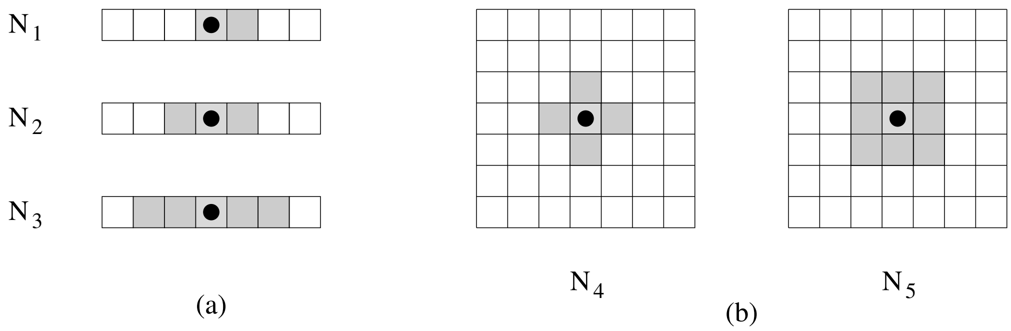


Figure 4.5: Neighbourhoods for one- and two-dimensional cellular automata-65%

is again the basic one.

A **configuration** of $\mathcal{A}$ is an element of $c \in Q^{\mathbf{Z}^d}$ assigning to each finite automaton of $\mathcal{A}$ (that is to each node of $\mathbf{Z}^d$) a state. The **global transition function**

$$G_\delta : Q^{\mathbf{Z}^d} \to Q^{\mathbf{Z}^d},$$

$$G_\delta(c)(w) = \delta(c(\{w\} + N)),$$

is used to define a **computation (behaviour)** of $\mathcal{A}$ to be a sequence of configurations $c_0, c_1, c_2, \ldots$ such that $c_i = G_\delta(c_{i-1})$, for $i \geq 1$.

For example, if $d = 1$, $N = \{-1, 0, 1\}$, then $G_\delta(c)(i) = \delta(c(i-1), c(i), c(i+1))$, if $N = \{-k, -k+1, \ldots, -1, 0, 1, \ldots, k-1, k\}$, then

$$G_\delta(c)(i) = \delta(c(i-k), c(i-k+1), \ldots, c(i+k-1), c(i+k));$$

if $d = 2$ and the von Neumann neighbourhood is considered, then

$$G_\delta(c)(i,j) = \delta(c(i-1,j), c(i,j-1), c(i,j), c(i,j+1), c(i+1,j)).$$

Of special interest for us are reversible cellular automata.

Definition 4.3.1 *A cellular automaton $\mathcal{A} = \langle d, Q, N, \delta \rangle$ is reversible, if there is another cellular automaton $\mathcal{A}' = \langle d, Q, N', \delta' \rangle$ such that for each configuration c of $\mathcal{A}$ it holds*

$$G_\delta(c) = c_1 \qquad \textit{if and only if} \qquad G_{\delta'}(c_1) = c.$$

In other words, a cellular automaton $\mathcal{A}$ is reversible if there is another cellular automaton $\mathcal{A}'$ such that for any sequence of configurations $c_1, c_2, \ldots, c_{n-1}, c_n$ of $\mathcal{A}$, where $G_\delta(c_i) = c_{i+1}$, for $1 \leq i < n$, $\mathcal{A}'$ can reverse this computation to get the following sequence of configurations $c_n, c_{n-1}, \ldots, c_2, c_1, i.e., G_{\delta'}(c_{i+1}) = c_i$. (Observe that the reverse cellular automaton $\mathcal{A}'$ may use a much smaller or larger neighbourhood than $\mathcal{A}$ does.)

Example 4.3.2 *A simple four-state cellular automaton, due to I. Korec, with the neighbourhood $\{0, 1\}$ is depicted in Figure 4.6a and its reversible counterpart, with the neighbourhood $\{-1, 0\}$ is shown in Figure 4.6b.*

$*$	0	1	2	3
0	0	1	1	0
1	2	3	3	2
2	0	1	1	0
3	2	3	3	2

(a)

$*$	0	1	2	3
0	0	0	3	3
1	2	2	1	1
2	0	0	3	3
3	2	2	1	1

(b)

Figure 4.6: A cellular automaton and its reversible counterpart

There do not seem to be many reversible cellular automata. For two-state automata with a neighbourhood N where $|N| = 2$ or $|N| = 3$ there are none. For the neighbourhood $N = \{-1, 0, 1, 2\}$ there are 65 536 cellular two-state automata but only 8 of them are reversible and all of them are insignificant modifications of the same one.[7] The following theorem, due to Toffoli (1977), Dubacq (1985) and Kari (1990), of importance for cellular automata applications, is therefore quite a surprise.

Theorem 4.3.3 *(1) Any k-dimensional CA can be simulated in real time by a $(k+1)$-dimensional reversible CA. (2) There is a universal cellular automaton that is reversible. (3) It is decidable whether a one-dimensional cellular automaton is reversible but it is undecidable whether a two-dimensional cellular automaton is reversible.*

8

Exercise 4.3.4 *Show that the one-dimensional cellular automaton with the neighbourhood $N = \{0,1\}$, states $\{0,1,\ldots,9\}$ and the transition function $\delta(x,y) = (5x + \lceil \frac{5y}{10} \rceil) \bmod 10$ is reversible.*

4.3.2 One-dimensional quantum cellular automata

The most basic model of cellular automata is that of the one-dimensional cellular automaton, and we deal in this section with its quantum version, due to Watrous (1995).

A **one-dimensional quantum cellular automaton**, in short 1QCA, $\mathcal{A} = \langle Q, \lambda, N, \delta\rangle$ is determined by a finite set Q of states that includes one so-called *quiescent state* λ, a neighbourhood $N = \{n_1, \ldots, n_r\} \subseteq \mathbf{Z}$ with $n_1 < n_2 < \ldots < n_r$ and a local transition function

$$\delta : Q^{r+1} \to \mathbf{C}_{[0,1]}$$

satisfying the following three conditions:

1. **Local probability condition**: for any $(q_1, \ldots, q_r) \in Q^r$,

$$\sum_{q\in Q} |\delta(q_1, \ldots, q_r, q)|^2 = 1,$$

2. **Stability of the quiescent state condition:** if $q \in Q$, then

$$\delta(\lambda, \ldots, \lambda, q) = \begin{cases} 1, & \text{if } q = \lambda; \\ 0, & \text{otherwise.} \end{cases}$$

(Sometimes it is more convenient to consider an equivalent definition of δ, as a mapping from Q^r to $\mathbf{C}^Q_{[0,1]}$.) In order to define the third condition several concepts have to be introduced.

A configuration $c : \mathbf{Z} \to Q$ is a mapping such that $c(i) \neq \lambda$ only for finitely many i. Let $C(\mathcal{A})$ denote the set of all configurations. Computation of $\mathcal{A}$ is then done in the inner-product space $H_{\mathcal{A}} = l_2(C(\mathcal{A}))$ with the basis $\{|c\rangle \mid c \in C(\mathcal{A})\}$.

[7] These 8 reversible CA have been found by a brute force search. For details see (http://alife.santafe.edu/alife/topics/cas/ca-faq/properties/properties.html). Amoroso and Patt (1972) found some 4-neighbour reversible CA and presented a general construction to get an infinite family of reversible CA—(Ch. Dürr).

[8] For more about cellular automata and their relations to othe models of computation see, for example, Gruska (1997).

In one step $\mathcal{A}$ transfers from one basis state $|c_1\rangle$ to another $|c_2\rangle$. The amplitude of such a transition, $\alpha(c_1, c_2)$, is defined as follows:

$$\alpha(c_1, c_2) = \prod_{i \in \mathbf{Z}} \delta(c_1(i + n_1), c_1(i + n_2) \ldots c_1(i + n_r), c_2(i)).^9$$

A state in $H_{\mathcal{A}}$ has therefore in general the form

$$|\phi\rangle = \sum_{c \in C(\mathcal{A})} \alpha_c |c\rangle, \text{ where } \sum_{c \in C(\mathcal{A})} |\alpha_c|^2 = 1.$$

The evolution operator $E_{\mathcal{A}}$ of $\mathcal{A}$ maps any state $|\phi\rangle \in l_2(C(\mathcal{A}))$ into the state $|\psi\rangle = E_{\mathcal{A}}|\phi\rangle$ such that

$$|\psi\rangle = E_{\mathcal{A}}|\phi\rangle = \sum_{c \in C(\mathcal{A})} \beta_c |c\rangle, \text{ where } \beta_c = \sum_{c' \in C(\mathcal{A})} \alpha_{c'} \alpha(c', c).$$

Now we can formulate the third condition δ has to satisfy:

3. **Unitarity of evolution condition:** the mapping $E_{\mathcal{A}}$ has to be unitary.

If $\mathcal{A}$ is such that the first two conditions are satisfied, then we say that $\mathcal{A}$ is a quasi-1QCA (in short q1QCA).

Remark 4.3.5 As mentioned at the beginning of this section, in classical cellular automata a configuration is usually defined as a mapping $c : \mathbf{Z} \to Q$ without the restriction that $c(i) \neq \lambda$ only for finitely many i, and therefore one also has so-called infinite configurations. Such an approach is normally used when the behaviour of cellular automata is studied as that of complex systems. If the computational aspects of classical cellular automata are investigated, only finite configurations are considered.

A quantum cellular automaton $\mathcal{A}$ is called **trivial** if its neighbourhood is $N = \{0\}$, i.e. each cell has only itself as the "neighbour". $\mathcal{A}$ is called **simple** if its neighbourhood is an interval of integers, i.e. $n_i = n_1 + i - 1$ for $1 < i \leq r$. $\mathcal{A}$ is called **symmetric** if it is simple and $n_r = -n_1$. A special case of a symmetric 1QCA is the one with the neighbourhood $N = \{-1, 0, 1\}$. For such a basic 1QCA the notation $\mathcal{A} = \langle Q, \lambda, \delta\rangle$ and QCA will be used.

Given a q1QCA a nontrivial problem is to decide whether the unitarity condition is satisfied. If we take as the size of a 1QCA $\mathit{size}(\mathcal{A}) = |Q|^{|N|+1}$, because the description of $\mathcal{A}$ is dominated by the description of its transition function, then for simple 1QCA the unitarity condition can be verified (by Dürr and Santha, 1996), in time $\mathcal{O}(n^{\frac{3r+1}{r+1}}) = \mathcal{O}(n^3)$, where n is the size of $\mathcal{A}$ and r is the size of the neighbourhood.

Exercise 4.3.6 *What is a good upper bound for deciding unitarity of arbitrary 1QCA?*

Unitarity is easy to verify for trivial 1QCA; we can make use of this in Section 4.3.3.

Lemma 4.3.7 *Evolution of a trivial 1QCA is unitary if and only if the following condition holds for any $q_1, q_2 \in Q$.*

$$\sum_{q \in Q} \delta^*(q_1, q)\delta(q_2, q) = \begin{cases} 1, & \text{if } q_1 = q_2 \\ 0, & \text{otherwise.} \end{cases} \tag{4.7}$$

[9] It follows from the stability of the quiescent state condition that this infinite product is a well-defined complex number.

Proof. In this case δ is a mapping from $Q \times Q$ to $\mathbf{C}_{[0,1]}$. Each cell of $\mathcal{A}$ evolves independently and actually as a 1QFA with one-letter inputs. Condition (4.7) is a necessary and sufficient condition for unitarity of the evolution of such a 1QFA (if input symbols are discarded). The evolution matrix of such a 1QFA is therefore given by the matrix U of dimension $|Q|$ such that $U(q_1, q_2) = \delta(q_2, q_1)$. The evolution matrix for any initial configuration of length n is then given by the matrix $\bigotimes_{i=1}^{n} U$. ∎

The transition function of a q1QCA is said to satisfy the **well-formedness condition** if for every state $|\phi\rangle \in H_{\mathcal{A}}$, it holds

$$||\phi|| = 1 \Leftrightarrow ||E_{\mathcal{A}}|\phi\rangle|| = 1. \tag{4.8}$$

Satisfiability of such a condition can be verified in time $\mathcal{O}(n^2)$ by Dürr, LêThanh and Santha (1996). The satisfiability of the above well-formedness condition is not sufficient for the unitarity condition to hold, as the following example shows:

Example 4.3.8 *Each classical one-dimensional CA $\mathcal{A}$ can be seen as a q1QCA. Its evolution operator is a matrix with the property that in every column indexed by a configuration there is a single non-zero entry with value 1 (for the unique next configuration). This matrix has all columns of norm 1. Its columns are all pairwise orthogonal if and only if the automaton mapping is injective. Its evolution is unitary if and only if its global mapping is bijective. Any classical 1CA whose global evolution is injective but not surjective is well-formed but not unitary.*

For example, due to Ch. Dürr, in the one-dimensional classical CA $\mathcal{A}$, called "controlled not", with states $\{0,1\}$, the quiescent state 0, the neighbourhood $N = \{0,1\}$, and the transition function $\delta(x,y) = x \oplus y$, no configuration of the form 0^*10^* has a pre-image (among configurations in $C_{\mathcal{A}}$). The global transition function of the automaton is injective but not surjective and the automaton satisfies the well-formedness condition.

Remark 4.3.9 The history of attempts to introduce the concept of quantum cellular automata goes back to Grössing and Zeilinger (1988) and a series of subsequent papers of the same authors. However, their model has little in common with the one discussed above. This history well illustrates the merits and limits of methodologies developed in physics and theoretical computing to deal with such basic problems and to come up with proper concepts.

Remark 4.3.10 In order to study more in depth computability problems for QCA and to consider the problem of universality, some restrictions have to be made on the types of amplitudes allowed. For example, for each amplitude α there has to be an algorithm to compute α to 2^{-n} in time polynomial in n; or that all amplitudes are rational.

Even for 1QCA several basic problems are still open. One of them concerns universality. Is there a single 1QCA which would be universal in a reasonable sense for the whole class of 1QCA and could efficiently simulate any 1QCA?[10]

[10] van Dam (1996) solved positively the universality problem for a special "circular" model of 1QCA, in which cells may form cycles of various length. Or, in other words, for space-periodic configurations only. This interesting result did not seem to contribute much to the solution of the basic universality problem for 1QCA. However, as shown by Dürr (1997), if a q1QCA is QCA with respect to van Dam model, then it is also 1QCA with respect to our model. This implies that a subset of 1QFA has nice properties from van Dam's model. For example, this subset has a universal instance and can be simulated by a QTM.

Remark 4.3.11 (CH. Dürr). A suitable definition of two- and more- dimensional quantum cellular automata is an untrivial issue. Formally, it can be done by a similar modification of the definition of the classical cellular automata—to assign complex amplitudes to transitions—as in the case of one-dimensional cellular automata. The difficulty with such a straightforward approach is that in such a case one cannot decide in polynomial time whether a qQCA is really a QCA. This follows from the result of Kari (1990) that reversibility of two-dimensional cellular automata is undecidable.

Another basic problem not yet fully solved is the problem of mutual efficient simulation of quantum cellular automata and quantum Turing machines. There is so far only a partial solution of this problem for a special class of 1QCA, for the so-called **partitioned quantum cellular automata** (PQCA). They are a natural quantum version of the model introduced by Morita and Harao (1989) and have been shown to be important in the classical case because they are much more easy to deal with. For PQCA it is also easy to verify whether their evolution is unitary. Results presented in the following section follow Watrous (1995).

4.3.3 Partitioned quantum one-dimensional cellular automata

For the very basic version of the partitioned one-dimensional cellular automaton (P1QCA) with which we deal in the next section, a PQCA is defined similarly to a QCA but each state (cell) is considered as being partitioned into three substates (subcells): left one, middle one and right one. The next state of any cell depends on the right substate of the left neighbour, the middle substate of the current cell and on the left substate of the right neighbour (see Figure 4.7a).

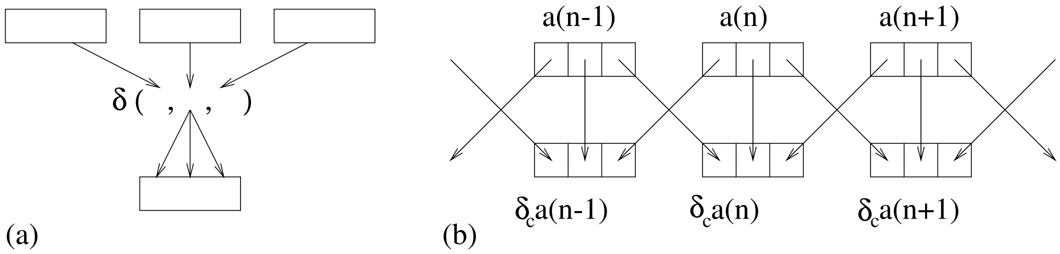


Figure 4.7: Evolution in a partial QCA-70%

In the full generality a partitioned one-dimensional quantum cellular automaton is a 1QCA $\mathcal{A} = \langle Q, \lambda, N, \delta \rangle$ which satisfies the following restrictions.

1. The set of states Q is the Cartesian product $Q = Q_1 \times \ldots \times Q_r$ of $r = |N|$ nonempty sets. (In the case of the neighbourhood $N = \{-1, 0, 1\}$ we use notation $Q = Q_l \times Q_m \times Q_r$.)
2. The local transition function $\delta : Q^r \to \mathbf{C}^Q$ is the composition of two functions:
   (a) $\delta_c : Q^r \to Q$; a classical mapping.
   (b) $\delta_q : Q \to \mathbf{C}^Q$; a quantum mapping,

   where

$$\delta_c((q_{1,1}, \ldots, q_{1,r}), (q_{2,1}, \ldots, q_{2,r}), \ldots, (q_{r,1}, \ldots, q_{r,r})) = (q_{1,1}, q_{2,2}, \ldots, q_{r,r}).$$

The function δ_c defines a 1CA (one-way CA) $\mathcal{A}_c = \langle Q, \lambda, N, \delta_c \rangle$ and the global transition function G_{δ_c} is a permutation on configurations of $\mathcal{A}$ such that for any $a \in C_{\mathcal{A}}$, $i \in N$, $[G_{\delta_c}(a)](i) = \delta_c(a(i+N))$. The case $N = \{-1, 0, 1\}$ is illustrated in Figure 4.7b.

$\mathcal{A}_c$ can also be seen as a 1QCA whose evolution operator $U_{\mathcal{A}_c}$ is defined by

$$U_{\mathcal{A}_c}(q_2, q_1) = \begin{cases} 1, & G_{\delta_c}(q_1) = q_2; \\ 0, & \text{otherwise,} \end{cases}$$

and therefore $U_{\mathcal{A}_c}$ is unitary.

The function δ_q defines a trivial 1QCA, $\mathcal{A}_q = \langle Q, \lambda, \{0\}, \delta_q \rangle$. Its evolution operator $U_{\mathcal{A}_q}$ is defined by

$$U_{\mathcal{A}_q}(q_2, q_1) = [\delta_q(q_1)](q_2).$$

$U_{\mathcal{A}_q}$ is also called the local transition matrix of $\mathcal{A}$. Unitarity of P1QCA is now easy to verify. The following theorem is due to Watrous (1995) and its proof is due to Dürr et al. (1996).

Theorem 4.3.12 *Evolution of a P1QCA $\mathcal{A}$ is unitary if and only if the evolution of the corresponding trivial 1CA is unitary, and this holds if and only if the local transition matrix of $\mathcal{A}$ is unitary.*

Proof. Since the local transition function δ of $\mathcal{A}$ is a composition of local transition functions δ_c and δ_q, the evolution operator $U_{\mathcal{A}}$ is the following product of the evolution operators $U_{\mathcal{A}} = U_{\mathcal{A}_q} U_{\mathcal{A}_c}$. Since $U_{\mathcal{A}_c}$ is trivially unitary we can state that evolution of $\mathcal{A}$ is unitary if and only if the local transition matrix $U_{\mathcal{A}_q}$ is unitary. ∎

As a consequence $\mathcal{A}$ can be specified using its local transition matrix U, i.e. in the form $\mathcal{A} = \langle Q, \lambda, N, U \rangle$ or, when the neighbourhood N is known, as $\mathcal{A} = \langle Q, \lambda, U \rangle$.

Example 4.3.13 (Watrous, 1997) *We describe a simple PQCA $\mathcal{A} = \langle Q, \lambda, U \rangle$ that simulates, in a sense, the EPR phenomenon. Let*

$$Q = Q_l \times Q_m \times Q_r, \text{ where } Q_l = Q_r = \{0, +, -\}, Q_m = \{0\}, \lambda = (0, 0, 0).$$

The matrix U of degree $|Q|$ is defined by

$$U(q', q) = \begin{cases} 1, & \text{if } q = q' \notin S = \{(-, 0, +), (+, 0, -)\}; \\ 0, & \text{if } q \neq q' \text{ and } \{q, q'\} \neq S, \end{cases}$$

and for $q, q' \in S$,

$$U(q', q) = \begin{cases} -\frac{1}{\sqrt{2}}, & \text{if } q = q' = (-, 0, +); \\ \frac{1}{\sqrt{2}}, & \text{otherwise.} \end{cases}$$

If $\mathcal{A}$ evolves from the initial basis state $|c\rangle$, where c is defined by

$$c(n) = \begin{cases} (0, 0, -), & \text{if } n = -1; \\ (+, 0, 0), & \text{if } n = 1; \\ \lambda, & \text{otherwise,} \end{cases}$$

then, after one step, $\mathcal{A}$ is in the superposition

$$\frac{1}{\sqrt{2}}(|c_0\rangle + |d_0\rangle),$$

where

$$c_0(n) = \begin{cases} (+,0,-), & \textit{if } n = 0; \\ (0,0,0), & \textit{otherwise}, \end{cases} \qquad d_0(n) = \begin{cases} (-,0,+), & \textit{if } n = 0; \\ (0,0,0), & \textit{otherwise}, \end{cases}$$

By an induction one can show that after $t > 1$ steps $\mathcal{A}$ is in the superposition

$$\frac{1}{\sqrt{2}}(|c_{t-1}\rangle + |d_{t-1}\rangle),$$

where

$$c_t(n) = \begin{cases} (+,0,0), & \textit{if } n = -t; \\ (0,0,-), & \textit{if } n = t; \\ (0,0,0), & \textit{otherwise}, \end{cases} \qquad d_t(n) = \begin{cases} (-,0,0), & \textit{if } n = -t; \\ (0,0,+), & \textit{if } n = t; \\ (0,0,0), & \textit{otherwise}. \end{cases}$$

If the states $(0,0,-)$ and $(-,0,0)$ are interpreted as negative particles and $(0,0,+),(+,0,0)$ as positive particles, then the configuration $|c_t\rangle$ models the situation in which the positive particle moves to the left and the negative one to the right; the configuration $|d_t\rangle$ models the situation when particles are reversed.

4.3.4 Quantum cellular automata versus quantum Turing machines

The problem of effective mutual simulations of 1QCA and one-tape quantum Turing machines has been solved so far only partially.

As shown below, each QTM the head of which moves always either to left or to right, can be efficiently simulated by a PQCA, and therefore by a QCA. On the other hand, only PQCA have been shown, so far, to be efficiently simulated by QTM.

One way to discuss mutual simulation of QTM and QCA is to consider both QTM and QCA as acceptors and to define when one machine model simulates the other one.

The problem of simulation of 1QCA by QTM does not seem to be easy. One reason for that may be the fact that QTM and 1QCA differ in the following property. While to each QTM $\mathcal{M}$ there exists a QTM $\mathcal{M}'$ that runs $\mathcal{M}$ backwards, this is not always true for 1QCA as the following example, due to Ch. Dürr, shows.

Example 4.3.14 *Let us consider 1QCA $\mathcal{A} = \langle\{0,1\},\{0,1\},\delta\rangle$, where $\delta(x,y) = |x\rangle$ if $y = 0$ and $\delta(x,y) = H'|x\rangle$ if $y = 1$ (with $H' = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$. Denote by $U_{\mathcal{A}}$ the global transition function of $\mathcal{A}$ and let $|c_k\rangle$ be the configuration of $\mathcal{A}$ with 1 in the cells with index $\{-k,-1\}$ and 0 otherwise.*

It can be shown, using algorithm due to Dürr and Santha (1996), that $U_{\mathcal{A}}$ is unitary and that

$$|\phi\rangle = U_{\mathcal{A}}^{-1}|c_1\rangle = \sum_{k\geq 1}\frac{1}{\sqrt{2^k}}(-1)^{k-1}|c_k\rangle.$$

$U_{\mathcal{A}}^{-1}$ is not the global transition function of a 1QCA. This follows from any of the following two observations. (1) In $|\phi\rangle$ "all cells with negative index are entangled". At the same time in any 1QCA the image of a "classical configuration" has only "unentangled states". (2) If $|c\rangle$ is the all-zero-configuration, then $U_{\mathcal{A}}^{-1}|c\rangle = |c\rangle$. This implies that $U_{\mathcal{A}}^{-1}|c_1\rangle$ has an infinite range influence. Hence all cells with negative index depend on whether there is 0 or 1 in the cell with index -1.

Exercise 4.3.15 *(a) Replace in the definition of the 1QCA from Example 4.3.14 H' with the square-root-of-not matrix, page 63, and show that the inverse of the global evolution of the resulting 1QCA is not evolution of any 1QCA; (b) find necessary and sufficient condition that if the matrix H' from Example 4.3.14 is replaced by a matrix U, then the resulting 1QCA has such a global evolution function the inverse of which is not an evolution function of a 1QCA.*

In order to consider QTM as acceptors two additional components of QTM will be needed: an integer k denoting the distinguished acceptance tape square (cell) and a set Σ_a of accepting tape symbols. An input $x \in \Sigma^*$ is accepted by a QTM $\mathcal{M} = \langle \Sigma, Q, q_0, \delta, k, \Sigma_a \rangle$, if a symbol from Σ_a is written into the tape square k, when $\mathcal{M}$ process the input x.

In order to consider 1QCA as acceptors also two additional components have to be considered: an integer k denoting the distinguished acceptance cell and a set Q_a of accepting states. An input $x \in Q^*$ will be accepted by a 1QCA $\mathcal{A} = \langle Q, \lambda, \delta, k, Q_a \rangle$ if the cell number k gets into a state from Q_a during the computation with x as the initial configuration.

Definition 4.3.16 *We say that a PQCA $\mathcal{A} = \langle Q, \lambda, U, k_{\mathcal{A}}, Q_a \rangle$ simulates a QTM $\mathcal{M} = \langle \Sigma, Q_{\mathcal{M}}, \delta, k_{\mathcal{M}}, \Sigma_a \rangle$ if and only if there are linear-time computable functions $T : C_{\mathcal{M}} \rightarrow C_{\mathcal{A}}$ and $f : \mathbf{N} \times \mathbf{N} \rightarrow \mathbf{N}$ such that for any configuration $c \in C_{\mathcal{M}}$ the probability that $\mathcal{M}$ accepts c after t steps, i.e., that $\mathcal{M}$ comes to a configuration with a terminating state, is equal to the probability that $\mathcal{A}$ accepts $T(c)$ after $f(t, |T(c)|)$ steps.*

The first problem we will deal with is the simulation of a uQTM (see Definition 4.2.12), by a PQCA. By Theorem 4.2.14, the restriction to unidirectional QTM is not essential. Let a uQTM $\mathcal{M} = \langle \Sigma, K, \delta, k, \Sigma_a \rangle$ be given. In such a case K can be partitioned into the subsets K_l and K_r. K_l (K_r) contains those states into which $\mathcal{M}$ comes only when the head moves left (right).

A PQCA $\mathcal{A} = \langle Q, \lambda, U, k_{\mathcal{A}}, Q_a \rangle$ which simulates $\mathcal{M}$ is defined as follows: $Q = Q_l \times Q_m \times Q_r$ with

$$Q_l = K_l \cup \{\#\}, \quad Q_m = \mathbf{\Sigma}, \quad Q_r = K_r \cup \{\#\},$$

where $\# \notin K$, $\lambda = (\#, b, \#)$, where b is the blank symbol of $\mathcal{M}$. The transition matrix U is defined as follows:

1. For each $(s_1, \tau_1), (s_2, \tau_2) \in K_l \times \Sigma$ let
$$U((s_2, \tau_2, \#), (s_1, \tau_1, \#)) = \delta(s_1, \tau_1, \tau_2, s_2, \leftarrow).$$

2. For each $(s_1, \tau_1) \in K_l \times \Sigma$, $(s_2, \tau_2) \in K_r \times \Sigma$, let
$$U((\#, \tau_2, s_2), (s_1, \tau_1, \#)) = \delta(s_1, \tau_1, \tau_2, s_2, \rightarrow),$$
$$U((s_1, \tau_1, \#), (\#, \tau_2, s_2)) = \delta(s_2, \tau_2, \tau_1, s_1, \leftarrow).$$

3. For each $(s_1, \tau_1), (s_2, \tau_2) \in K_r \times \Sigma$, let
$$U((\#, \tau_2, s_2), (\#, \tau_1, s_1)) = \delta(s_1, \tau_1, \tau_2, s_2, \rightarrow).$$

4. For any $q_1, q_2 \in Q$ for which $U(q_1, q_2)$ has not been defined in (1) to (3), let
$$U(q_1, q_2) = \begin{cases} 1, & \text{if } q_1 = q_2; \\ 0, & \text{otherwise.} \end{cases}$$

Let $k_{\mathcal{A}} = k$ and $Q_a = \{(q_l, q_m, q_r) \in Q \,|\, q_m \in \Sigma_a\}$. It can be shown in a straightforward way that if $\mathcal{M}$ is a QTM the transition function of which satisfies conditions on page 166, then $\mathcal{A}$ is a PQCA with the unitary matrix U and therefore the evolution of $\mathcal{A}$ is unitary.

In order to consider a simulation of $\mathcal{M}$ by $\mathcal{A}$ we need to define the mappings $T : \mathcal{C}_{\mathcal{M}} \to \mathcal{C}_{\mathcal{A}}$ and $f : \mathbf{N} \times \mathbf{N} \to \mathbf{N}$.

A configuration $c \in C_{\mathcal{M}}$ has the form xqy, where $x, y \in \Sigma^*, q \in K$. The mapping T is then defined as follows

$$T(xqy)(n) = \begin{cases} (\#, xy_n, \#), & \text{if } q \in K_l, n \neq |x| + 2; \\ (q, xy_n, \#), & \text{if } q \in K_l, n = |x| + 2; \\ (\#, xy_n, \#), & \text{if } q \in K_r, n \neq |x|; \\ (\#, xy_n, q), & \text{if } q \in K_r, n = |x|, \end{cases}$$

where xy_n denotes the nth symbol of the string xy, or λ if out of the range.

Exercise 4.3.17 *Show, for any $n \in \mathbf{N}$, that the probability that $\mathcal{M}$ accepts the initial configuration q_0w after n steps is equal to the probability that $\mathcal{A}$ accepts $T(q_0w)$ after n steps.*

To summarize:

Theorem 4.3.18 *(Watrous, 1995) To any QTM $\mathcal{M}$ there is a PQCA simulating $\mathcal{M}$ with a constant slowdown.*

The concept of simulation of PQCA by QTM is defined in an analogous way as in Definition 4.3.16.

We show now how to construct, given a PQCA $\mathcal{A} = \langle Q, \lambda, U, k, Q_a \rangle$, a QTM $\mathcal{M} = \langle \Sigma, K, q_0, \delta, k', K_a \rangle$ which simulates $\mathcal{A}$. In order to do that we assume that tape cells of $\mathcal{A}$ and of $\mathcal{M}$ are indexed. Each tape cell of $\mathcal{M}$ will represent the cell of $\mathcal{A}$ of the same index. Let $Q = Q_l \times Q_m \times Q_r$. We define

$$\Sigma = Q \cup Q' \cup \{b\},$$

where Q' is a "copy" of Q in the sense that $q' = (q'_l, q'_m, q'_r) \in Q'$ if and only if $q = (q_l, q_m, q_r) \in Q$, and b is the blank symbol of $\mathcal{M}$. The states from Q' will be used as endmarkers to limit the non-blank portion of the tape.

Moreover, let

$$K = Q_l \times \{s_0, s_1, s'_1, s_2, s'_2, s_3\} \times Q_r,$$

where $s_0, s_1, s'_1, s_2, s'_2$ and s_3 are new states not used in Q_m. Hence each state of $\mathcal{M}$ consists of three components: one element of Q_l, one from the set $\{s_0, s_1, s'_1, s_2, s'_2, s_3\}$ and one from Q_r. The elements from Q_l and Q_r are needed in order to "move" substates from one tape square to another when performing the permutation δ_c described on page 183.

The simulation of one step of $\mathcal{A}$ is done in three phases. The task of the first phase is to perform reversibly the permutation δ_c on the given configuration of $\mathcal{A}$ represented on the tape of $\mathcal{M}$. This is done in one left-to-right and one right-to-left pass of the head between the cells with endmarkers. In the left-to-right move the leftmost components of the triple states are shifted one cell (triple) to the right. In the right-to-left move the rightmost components of triple states are shifted one triple to the left. All that is done by the moves defined for each $(l,r) \in Q_l \times Q_r$, $(q_l,q_m,q_r),(q_l',q_m',q_r') \in Q_l \times Q_m \times Q_r$:

$$\begin{aligned}
\delta((l,s_0,r),(q_l,q_m,q_r)',(q_l,q_m,r)',(l,s_1,q_r),\rightarrow) &= 1,\\
\delta((l,s_1,r),(q_l,q_m,q_r),(q_l,q_m,r),(l,s_1,q_r),\rightarrow) &= 1,\\
\delta((l,s_1,r),(q_l,q_m,q_r)',(q_l,q_m,r),(l,s_1',q_r),\rightarrow) &= 1,\\
\delta((l,s_1',r),(q_l,q_m,q_r)',(q_l,q_m,r)',(l,s_1',q_r),\rightarrow) &= 1;
\end{aligned}$$

$$\begin{aligned}
\delta((l,s_1',r),(q_l,q_m,q_r),(q_l,q_m,r)',(q_l,s_2,q_r),\leftarrow) &= 1,\\
\delta((l,s_2,r),(q_l,q_m,q_r),(l,q_m,q_r),(q_l,s_2,r),\leftarrow) &= 1,\\
\delta((l,s_2,r),(q_l,q_m,q_r)',(l,q_m,q_r),(q_l,s_2',r),\leftarrow) &= 1,\\
\delta((l,s_2',r),(q_l,q_m,q_r)',(l,q_m,q_r)',(q_l,s_2',r),\leftarrow) &= 1.
\end{aligned}$$

Not all of the above transitions are really needed for simulation. Some of them are just to make evolution unitary.

The aim of the first phase was to put all the information needed for the main second phase in which transitions of $\mathcal{A}$ defined by U are simulated by transitions of $\mathcal{M}$ according to the rules:

$$\begin{aligned}
\delta((l,s_2',r),(q_l,q_m,q_r),(q_l',q_m',q_r')',(q_l,s_3,r),\rightarrow) &= U((q_l,q_m,q_r),(q_l',q_m',q_r')),\\
\delta((l,s_3,r),(q_l,q_m,q_r),(q_l',q_m',q_r'),(l,s_3,r),\rightarrow) &= U((q_l,q_m,q_r),(q_l',q_m',q_r')),\\
\delta((l,s_3,r),(q_l,q_m,q_r)',(q_l',q_m',q_r')',(l,s_0,r),\leftarrow) &= U((q_l,q_m,q_r),(q_l',q_m',q_r')).
\end{aligned}$$

In the third phase $\mathcal{M}$ just moves its head to come to the cell representing the current leftmost cell of the configuration of $\mathcal{A}$. This is achieved by the transitions:

$$\delta((l,s_0,r),(q_l,q_m,q_r),(q_l,q_m,q_r),(l,s_0,r),\leftarrow) = 1.$$

Finally, let δ take the value 0 everywhere not defined above. In addition, let $k' = k$ and $\Sigma_a = \{q, q' \mid q \in Q_a\}$.

On the base of Theorem 4.3.12 it is now straightforward to show:

Lemma 4.3.19 *If $\mathcal{A}$ and $\mathcal{M}$ are defined as above, then if $\mathcal{A}$ is a PQCA, then $\mathcal{M}$ is a QTM.*

In order to formulate the new main simulation result we need to define two mappings: $T : C_{\mathcal{A}} \rightarrow C_{\mathcal{M}}$ and $f : \mathbf{N} \times \mathbf{N} \rightarrow \mathbf{N}$.

For $a \in C_{\mathcal{A}}$ let n_l^a and n_r^a be the indices of the leftmost and rightmost non-quiescent cells of a. Define

$$T(a) = (\lambda, s_0, \lambda)c,$$

where

$$c(i) = \begin{cases} a(i), & \text{if } i \neq n_l^a, i \neq n_r^a;\\ a^*(i), & \text{otherwise,}\end{cases}$$

and let $f(t,|c|) = 4t^2 + 4|c|t - t$ be the number of steps needed by $\mathcal{M}$ to simulate t steps of $\mathcal{A}$. One can now prove (Watrous, 1995):

Lemma 4.3.20 *For any $t \in \mathbf{N}$ and $c \in C_{\mathcal{A}}$ the probability that $\mathcal{A}$ accepts c after t steps is equal to the probability that $\mathcal{M}$ accepts $T(c)$ after $f(t,|T(c)|)$ steps.*

To summarize:

Theorem 4.3.21 *To any PQCA $\mathcal{A}$ there exists a QTM simulating $\mathcal{A}$ with quadratic slowdown.*

Chapter 5

COMPLEXITY

INTRODUCTION

The study of complexity questions and of complexity classes, computational and communicational, has proved to be very enlightening and important for classical computation. It has developed a firm theoretical basis for our understanding of the potentials and limitations of computational resources, models, and modes. There is reason to expect the same for the complexity investigations in quantum computation and communication.

It is of utmost importance to determine whether quantum classification of inherent computational complexity is indeed different from the classical one. Would this prove to be the case the very basic foundations of computing would be shaken.

Quantum computational complexity theory is characterized, as its classical counterpart, by a number of fundamental open problems concerning the proper inclusions of complexity classes. In order to get a better insight into these problems, and to test potential methods to solve them, the relativized quantum complexity theory is of interest and importance.

It is also of importance to find out how much quantum features can speed up computations, shorten communications and achieve efficiency in size or space.

Investigations of the potential impacts on the power of computing of the existence of slightly non-linear evolutions in quantum physics are also of interest.

LEARNING OBJECTIVES

The aim of the chapter is to learn:

1. the way universal quantum Turing machines can be constructed;
2. the basic quantum complexity classes and their properties;
3. the basic relations between classical and quantum complexity classes;
4. the basic results concerning relativized quantum complexity;
5. the basic concepts of quantum communication complexity;
6. a reduction of quantum communication protocols to quantum computation problems;
7. the potential impacts of non-linearity on the power of quantum computing.

Every man takes the limits of his own
field of vision for the limits of the world.

Arthur Schopenhauer (1788-1860)

Quantum complexity theory is surely the most basic one to develop our complexity oriented understanding of the potentials and limitations of information processing and communication.

The classical computational complexity theory has been based on feasibility-oriented strengthening of the Church–Turing thesis that itself has formed foundations for the computability theory. The former thesis, called also the "Modern Church-Turing computational thesis", says that any "reasonable" (practically realizable) model of computation can be efficiently simulated (in polynomial time and linear space), on probabilistic Turing machines. In the development of the fundamental concepts and models of classical complexity theory an abstraction of human computational customs and means is used, and they actually correspond to the view of the physical world classical physics offers. Once the idea of quantum computing was born and shown to be successful, at least theoretically so far, it became clear that one has to base the computational complexity theory on the laws of quantum physics. The same is true for the communication complexity problems, and that is why quantum communication complexity is also of interest and importance.

5.1 Universal Quantum Turing Machines

We present the basic ideas how to construct a QTM U which is universal in the following sense. Given an arbitrary DBV-QTM $\mathcal{M}$, its input string x, time bound T, and accuracy $\varepsilon > 0$, U produces a state whose distance from the state produced by $\mathcal{M}$, after T steps, is at most ε. In addition, the time of the simulation of $\mathcal{M}$ on $\mathcal{U}$ is bounded by a polynomial in T, $|x|$ and $\frac{1}{\varepsilon}$.

The design of the universal QTM $\mathcal{U}$ presented in the following is much more complex than those designs known for the classical universal TM and therefore only the basic ideas will be given here—for details see Bernstein (1997) and Bernstein and Vazirani (1997). At the same time the structure of $\mathcal{U}$ is simple. In addition, $\mathcal{U}$ uses a single quantum operation, a one-qubit rotation. The rest of $\mathcal{U}$ can be seen as a classical reversible TM.

5.1.1 Efficient implementation of unitary transformations

The key component of the universal QTM is a QTM which can realize efficiently and with required accuracy, an arbitrary finite unitary transformation, even in the case that this unitary transformation is given only by a "sufficiently good approximation". The design of such a QTM is the subject of this section.

The main problem with having a single QTM performing an arbitrary unitary transformation is that a fixed QTM maps in a single step any configuration to a superposition of a fixed number of configurations and therefore it cannot simulate in one step a step of any QTM. The way out is first to decompose, efficiently, any given unitary transformation U (or its approximation) to a "small number" of elementary unitary transformations and then to carry out these elementary transformations, in order to realize U.

Decomposition of unitary transformations

Any unitary matrix of degree d can be decomposed by a polynomial time, with respect to d, deterministic algorithm into a product of the so-called near-trivial unitary matrices. In doing that we assume that all entries of matrices are given within accuracy bound 2^{-r} for some r.

Definition 5.1.1 *A unitary matrix M of degree d is* **near-trivial** *if M satisfies one of the following conditions.*

1. *M is like a unit matrix except for one diagonal element which has the form $e^{i\theta}$, $\theta \in [0, 2\pi]$. (Such a transformation is called a* **near-trivial shift** *and denoted by (j, j, θ) if the element $e^{i\theta}$ is in the jth row.)*
2. *M is like the unit matrix except for the elements in the intersections of a jth and a kth row and column which form the matrix*

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

for some $\theta \in [0, 2\pi]$. (Such a transformation is called a **near-trivial rotation** *and it is denoted by (j, k, θ).)*

In the proof of the decomposition theorem presented below we make use of the following result.

Lemma 5.1.2 *There is a deterministic algorithm $\mathcal{A}$ such that given a d-dimensional column vector v of complex numbers, and an $\varepsilon > 0$, $\mathcal{A}$ computes near-trivial matrices $U_1, \ldots, U_{2d-1}$ such that*

$$||U_1 \ldots U_{2d-1}v - ||v||\, e_1|| \leq \varepsilon,$$

where e_1 is the unit vector, with one in the first component, and $\mathcal{A}$ runs in polynomial time with respect to d, $\lg \frac{1}{\varepsilon}$ and length of the input.

Proof. The first step is to use, for $j = 1, \ldots, d$, phase shifts $P_j = (j, j, \phi_j)$, where $\phi_j = 2\pi - \arccos \frac{Re(v_j)}{|v_j|}$ or $\phi_j = \arccos \frac{Re(v_j)}{|v_j|}$ (depending on whether $Im(v_j)$ is positive or negative), if $v_j \neq 0$ and $\phi_j = 0$, otherwise—to replace all components v_j by $|v_j|$.

As the next step $d-1$ rotations R_j, $j = d-1, \ldots, 1$, are used to move all weights of the vector into its first component—all other components will have at the end of this procedure the value 0. This can be achieved with $R_j = (j, j+1, \psi_j)$ and $\psi_j = \arccos \frac{|v_j|}{\sqrt{\sum_{j=1}^{n} |v_j|^2}}$ if the sum of the elements is not 0 and $\psi_i = 0$ otherwise. As the result we have

$$R_1 \ldots R_{d-1} P_1 \ldots P_d v = ||v|| e_1. \tag{5.1}$$

The angles ϕ_j and ψ_j can be computed in polynomial time, with respect to d, $\lg\frac{1}{\varepsilon}$ and the length of input, with the precision $\delta = \frac{\varepsilon}{(2d-1)||v||}$. Let us denote by P_j', R_j' the near-trivial matrices corresponding to P_j, R_j but with elements determined with the precision δ. Since $|\phi_j - \phi_j'| \leq \delta$ we have $||P_j - P_j'|| \leq \delta$ and since $|\psi_j - \psi_j'| \leq \delta$ we get $|R_j - R_j'| \leq \delta$, and therefore also the inequality

$$||R_1' \dots R_{d-1}' P_1' \dots P_d' - R_1 \dots R_{d-1} P_1 \dots P_d|| \leq (2d-1)\delta,$$

and consequently,

$$||R_1' \dots R_{d-1}' P_1' \dots P_d' v - ||v|| e_1|| \leq (2d-1)\delta ||v|| = \varepsilon.$$

(The last inequality is due to the fact that if the distance between angles is at most δ, then so it is between the corresponding points on the unit circle in both real and complex plane.)
□

Theorem 5.1.3 (Decomposition Theorem) *There is a deterministic algorithm $\mathcal{A}$, which runs in polynomial time, with respect to d, $\lg\frac{1}{\varepsilon}$ and the length of input, such that given as input an $\varepsilon > 0$, and a complex matrix U which is $\frac{\varepsilon}{2(10\sqrt{d})^d}$-close to a unitary matrix, $\mathcal{A}$ computes a sequence of near-trivial matrices of degree d, $U_1, \dots, U_n$, with n polynomial in d, such that $||U - U_n \dots U_1|| \leq \varepsilon$.*

Proof. Let us say that a matrix of the degree d is k-simple if its first k rows and columns are identical with those of the unit matrix of degree d. The basic idea of the proof is to show how to reduce the problem of the decomposition of an i-simple matrix of degree d, close to a unitary matrix, to the problem of decomposition of an $(i+1)$-simple matrix of degree d, for $i = 0, 1, \dots, d-1$. (Observe that the product of two k-simple matrices is again a k-simple matrix.)

Suppose that we have started to approximate a k-simple matrix U by near-trivial matrices whose product is V. In order to produce the desirable reduction, we need to create another sequence of near-trivial matrices whose product is $W = UV^*$. In order to reduce the problem, V should be such that UV^* is close to being $(k+1)$-simple. To achieve that Lemma 5.1.2 will be used.

Let U be k-simple and δ-close to a unitary matrix. Let Z be the lower right $(d-k)\times(d-k)$ submatrix of U and Z_1 its first row. By Lemma 5.1.2, one can construct a sequence of near-trivial matrices $V_1, \dots, V_{2(d-k)-1}$ of degree $d-k$ whose product $V = V_1 \dots V_{2(d-k)-1}$ is such that $||VZ_1^T - ||Z_1|| e_1|| \leq \delta$. Finally, let us extend the matrices V and all V_i into the matrices of degree d that are k-simple.

Denote $W = UV^*$. V is unitary and W is k-simple and δ-close to a unitary matrix (because U is unitary and V is δ-close to a unitary matrix). In addition, W is close to being $(k+1)$-simple in the following sense. For the $(k+1)$th row W_{k+1} of W it holds, as discussed below,

$$||W_{k+1} - e_{k+1}|| \leq 2\delta \quad \text{and all} \quad ||W_{j,k+1}|| \leq 6\delta \quad \text{for all } j \geq k+1. \tag{5.2}$$

Let X be a $(k+1)$-simple matrix of degree d such that $X_{i,j} = W_{i,j}$ if $i, j > k+1$. From (5.2) and Exercise 1.4.21 it follows that $||W - X|| \leq 2\delta + 6\sqrt{d}\delta$. Since W is δ-close to a unitary matrix, X is $(3\delta + 6\sqrt{d}\delta)$-close to a unitary matrix.

The problem with the above idea is that it may happen that not all entries of $W = UV^*$ can be computed exactly. However, there is a way out. By Exercise 1.4.21, it is sufficient to compute them with the accuracy $\frac{\delta}{d}$ to obtain a matrix W_1 such that $||W - W_1|| \leq \delta$. Let us now use W_1 to construct a new matrix X_1 in a similar way as above. On the base of the triangle inequality we can then derive that $||W - X_1|| \leq 3\delta + 6\sqrt{d}\delta$ and that X_1 is $(4\delta + 6\sqrt{d}\delta)$-close to a unitary matrix.

The problem of decomposing a k-simple matrix U was reduced in this way to the problem of the decomposition of a $(k+1)$-simple matrix X_1 provided we are willing to accept two errors:

1. An error $||W - X_1|| \leq 3\delta + 6\sqrt{d}\delta$, because we are going to decompose in the next step X_1 and not W.
2. X_1 is only $(3\delta + 6\sqrt{d}\delta)$-close to a unitary matrix.

If $\delta' = 10\sqrt{d}\delta$, then δ' is the upper bound on both of the above errors. The total error, after d steps, is therefore $\sum_{j=1}^{d}(10\sqrt{d})^j\delta \leq 2(10\sqrt{d})^d\delta$, since $10\sqrt{d} \leq 2$. Since U is δ-close to a unitary matrix with $\delta = \frac{\varepsilon}{2(10\sqrt{d})^d}$, the total error of the approximation is, as required, at most ε.

To finish the proof, it remains to show (5.2). This requires to make several careful estimations and to use the inequality from Exercise 2.3.29b. This is left to the reader (see also Bernstein, 1997, and Bernstein and Vazirani 1997). $\Box$

Implementation of unitary transformations

The existence of a single QTM capable to implement, efficiently and with a sufficient accuracy, any unitary transformation is based on the decomposition result presented above and on the existence of a single QTM that can carry out any rotation. This is in turn based on the following lemma, which implies that a fixed rotation can be used to approximate efficiently any other rotation.

Lemma 5.1.4 *If $R = 2\pi\sum_{i=1}^{\infty} 2^{-2^i}$, then there is a deterministic polynomial time algorithm, with respect to $\lg\frac{1}{\varepsilon}$ and the length of the input, which for given $\varepsilon > 0$, and $\theta \in [0, 2\pi]$, produces an integer k, bounded by a polynomial in $\frac{1}{\varepsilon}$ such that $|kR - \theta| \bmod 2\pi \leq \varepsilon$.*

Proof. The algorithm is simple: it first computes an n, a power of 2, such that $\varepsilon > \frac{\pi}{2^{n-1}}$, and then an $m \in [1, 2^n]$ such that $|\frac{\theta}{2\pi} - \frac{m}{2^n}| \leq \frac{1}{2^n}$. The outcome is $k = m2^n$.

Indeed,

$$\begin{aligned}(m2^nR) \bmod 2\pi &= (2\pi m \sum_{i=\lg n+1}^{\infty} 2^{n-2^i}) \bmod 2\pi \\ &= (\frac{2\pi m}{2^n} + 2\pi m \sum_{i=\lg n+2}^{\infty} 2^{n-2^i}) \bmod 2\pi\end{aligned}$$

and since

$$m \sum_{i=\lg n+2}^{\infty} 2^{n-2^i} \leq m2^{n-4n+1} \leq 2^{-2n+1},$$

we have

$$\begin{aligned}|m2^nR - \theta| \bmod 2\pi &\leq |m2^nR - \frac{2\pi m}{2^n}| \bmod 2\pi + |\frac{2\pi m}{2^n} - \theta| \\ &\leq \frac{2\pi}{2^{2n-1}} + \frac{2\pi}{2^n} < \frac{2\pi}{2^{n-1}} < \varepsilon.\end{aligned}$$

□

The following theorem yields the key result used in Section 5.1.2 to design a universal QTM.

Theorem 5.1.5 (Unitarity Transformation Theorem) *There is a stationary and normal form QTM $\mathcal{M}$, with the first track alphabet $\{0, 1, \lambda\}$, that can perform, given an $\varepsilon > 0$, T steps of any unitary transformation U_0, of the degree d, on its first track, in polynomial time, with respect to $\frac{1}{\varepsilon}, d, T$ and the length of input, provided U_0 is given by an $\frac{\varepsilon}{4(10\sqrt{d})^d}$-close matrix.*

Sketch of the proof. Given an $\varepsilon > 0$ and a transformation U of degree $d = 2^k$ which is $\frac{\varepsilon}{2(10\sqrt{d})^d}$-close to a unitary transformation, the machine $\mathcal{U}$ to carry out U with accuracy ε, on the first k cells of the first track will be designed with the help of the following QTM $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$.

1. $\mathcal{M}_1$ decomposes U into a sequence $U_1, \ldots, U_n$ of near-trivial transformations such that $||U - U_n \ldots U_1|| \leq \frac{\varepsilon}{2}$, where n is polynomial in d, and writes on an empty track $\frac{\varepsilon}{2n}$. The existence of $\mathcal{M}_1$ follows from Theorem 5.1.3.

2. $\mathcal{M}_2$ applies, one after another, transformations $U_1, \ldots, U_n$, each with accuracy $\frac{\varepsilon}{2n}$. The stationary, normal-form QTM $\mathcal{M}_2$ to perform Step 2, can be designed by composition of the following quantum Turing machines $\mathcal{T}_1$ and $\mathcal{T}_2$:

$\mathcal{T}_1$ applies any near-trivial transformation specified by a triple (i, j, θ), which is a shift if $i = j$ and a rotation if $i \neq j$. Such a QTM can be designed using the branching lemma 4.2.37 and two QTM, $\mathcal{T}_s$ for shifts and $\mathcal{T}_r$ for rotations. To design $\mathcal{T}_2$ a DTM $\mathcal{T}_d$ is used that can permute a given sequence of near-trivial transformations, each given by a triple (i, j, θ)—by moving the rightmost one to the left end of the sequence. This can be clearly done by a DTM and therefore, by synchronization theorem, we can design a QTM $\mathcal{T}'$ to perform such a permutation.

3. $\mathcal{M}_3$ erases matrices $U_1, \ldots, U_n$ and $\frac{\varepsilon}{2n}$. (This can again be done by a DTM and therefore by a stationary, normal-form QTM.)

It remains to put together QTM for Steps 1 to 3. This can be done by combining methods presented in Exercises 4.2.31 and 4.2.32 and the composition lemma.

The running time of QTM $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$ is independent of the contents of the first track (which contains the description of the configuration), and therefore so is the running time of the resulting QTM $\mathcal{M}$.

If $\mathcal{M}$ is run on the first k cells of the first track, then a unitary transformation U' is applied such that

$$||U' - U|| \leq ||U' - U_n \ldots U_1|| + ||U_n \ldots U_1 - U|| \leq n\frac{\varepsilon}{2n} + \frac{\varepsilon}{2} = \varepsilon.$$

5.1.2 Design of a universal quantum Turing machine

It is in principle easy to design a multitape or multitrack classical universal Turing machine $\mathcal{U}$ (UTM) capable of simulating an arbitrary one-tape TM $\mathcal{M}$. Given a representation of $\mathcal{M}$ in a fixed-size alphabet, $\mathcal{U}$ simulates one step of $\mathcal{M}$ after another. To simulate one step of $\mathcal{M}$, $\mathcal{U}$ first determines, from the current configuration c of $\mathcal{M}$, the current state and the tape symbol under the head at c. Afterwards $\mathcal{U}$ searches, in the finite description of $\mathcal{M}$, for the transition $\mathcal{M}$ should perform and implements the required change of the state, of the tape symbol, and of the position of the head. In order to achieve polynomial time efficiency,

$\mathcal{U}$ keeps shifting the description of $\mathcal{M}$ so it is always close to the position of the head (on another track) of $\mathcal{M}$.

There are surprisingly small classical universal TM (see Roghozin 1996, Gruska 1997). Currently the smallest universal TM has 5 states and 5 tape symbols. There is also a universal TM with 2 states (and 18 tape symbols) and 2 tape symbols (and 24 states). The design of a universal QTM is much more complex.

It is natural to try to design a UQTM on the basis of similar ideas as the classical UTM. However, there are some fundamental difficulties that need to be overcome.

1. Unitary transformations specifying evolutions of QTM have in principle infinite dimension. A way out is, given a QTM, first to design an equivalent unidirectional QTM, evolution of which can be specified by a finite unitary matrix.

2. A natural idea that a UQTM $\mathcal{U}$ simulates the evolution of any given QTM $\mathcal{M}$ step by step is not so easy to realize. The problem lies in the following. Each evolution step of $\mathcal{M}$ is an application of its unitary matrix. If this is to be simulated by $\mathcal{U}$ then it seems that $\mathcal{U}$ has to implement one step of $\mathcal{M}$ using several steps of $\mathcal{U}$. Therefore $\mathcal{U}$ will not map a state of $\mathcal{M}$ immediately to the proper superposition, but it has to create this superposition during several steps. This causes difficulties because it is necessary that $\mathcal{U}$ is reversible. The way out has been found in Section 5.1.1. A complex unitary transformation can be decomposed into a sequence of near-trivial unitary matrices.

Encodings of QTM

Any universal QTM $\mathcal{U}$ gets as input an encoding of a QTM $\mathcal{M}$ and of its input x. Both encodings have to be words in the alphabet of $\mathcal{U}$. The number of tape symbols and states can be encoded in a similar way as for classical TM (see, for example, page 224 in Gruska (1997) or Appendix on web). The transition function $\delta_{\mathcal{M}}$ is encoded by an algorithm that for given arguments $q, \sigma, \sigma', q', d$ computes the amplitude $\delta_{\mathcal{M}}(q, \sigma, \sigma', q', d)$ with precision 2^{-n} in time polynomial in n.

Design of UQTM

Theorem 5.1.6 *There is a normal form QTM $\mathcal{U}$ such that for any QTM $\mathcal{M}$, its input x, $\varepsilon > 0$, and $T \in \mathbf{N}$, $\mathcal{U}$ can simulate $\mathcal{M}$ with accuracy ε for T steps in time polynomial in T and $\frac{1}{\varepsilon}$.*

Proof. The universal QTM $\mathcal{U}$, the existence of which we are going to demonstrate, gets as input a description of $\mathcal{M}$, x, T, and the simulation accuracy ε. $\mathcal{U}$ works in two phases.

Preprocessing. Using the algorithm of the proof of Theorem 4.2.14, $\mathcal{M}$ is transformed into an equivalent unidirectional QTM $\mathcal{M}' = \langle \Sigma, Q, \delta \rangle$. During this preprocessing the following outputs are produced:

1. A representation of the initial configuration of $\mathcal{M}'$ for x of length $2T + 1$. The reason for such length is the fact that T steps of $\mathcal{M}'$ are to be simulated. Each configuration of $\mathcal{M}'$ will be represented on the working tape of $\mathcal{U}$ with each cell of $\mathcal{M}'$ represented by a "supercell" consisting of $|\Sigma|\,|Q|$ cells—enough to store representation of a symbol from Σ and of a state from Q, if the head of $\mathcal{M}'$ is to scan a particular cell, and 0, otherwise.

2. The description of the unitary transformation $U_{\mathcal{M}'}$ for the evolution of $\mathcal{M}'$, written with the accuracy $\frac{\varepsilon}{40T(10\sqrt{d})^{d+2}}$, where $d = |\Sigma|\,|Q|$.

3. The binary string s of length $|Q|$ of the directions of the head moves of $\mathcal{M}'$, for $|Q|$ states of $\mathcal{M}'$.

4. The desired number of the simulation steps, $5T$, and the desired accuracy ε.

All that can be computed by a DTM in time polynomial in T, $\frac{1}{\varepsilon}$ and the length of the input and therefore, by the synchronization theorem, with the asymptotically same efficiency also by a QTM.

Simulation of the computation steps of $\mathcal{M}'$. Let us first describe a QTM called STEP that simulates one step of $\mathcal{M}'$. STEP operates as follows:

1. STEP transfers the current state q and the tape symbol σ to an empty work space near the starting cell, leaving a special marker in their places.

2. STEP applies $U_{\mathcal{M}'}$ to (q, σ) to produce, with accuracy ε, a superposition of new states and symbols (q', σ').

3. STEP reverses Step 1, to replace the marker with (q', σ') and to empty the working space.

4. STEP transfers the state on the tape either one cell left or right according to the value of q'th bit of the string s.

Steps 1, 3 and 4 can be done easily by DTM and therefore, by the synchronization theorem, there exists a stationary, normal form QTM that realizes the above steps that works, for a fixed $\mathcal{M}$, in time polynomial in T. The fact that Step 2 can be done by a QTM in time polynomial in T, d and $\frac{1}{\varepsilon}$ follows from Theorem 5.1.5. Moreover, using the results of Exercises 4.2.31 and 4.2.32, QTM for Steps 1 to 4 can be composed to get the resulting QTM STEP to perform simulation of one step of $\mathcal{M}'$. By inserting STEP for the special state in the RTM constructed according to the looping lemma 4.2.38, together with an additional input T, the resulting QTM simulating $\mathcal{M}'$ will compute in time polynomial with respect to T and $\frac{1}{\varepsilon}$ and with the accuracy $T\varepsilon$.

The resulting universal QTM is obtained by composing QTM of both phases.

Analysis. If, in the preprocessing, the transformation $U_{\mathcal{M}'}$ is computed to the specified accuracy, then the transformation provided for the simulation is within $\frac{\varepsilon}{40T(10\sqrt{d})^{d}}$ of the desired unitary transformation $U_{\mathcal{M}'}$ and therefore it will be $\frac{\varepsilon}{40T(10\sqrt{d})^{d}}$-close to a unitary transformation as required for the output of STEP. This implies that if each time STEP works with accuracy $\frac{\varepsilon}{40T}$, then it applies a unitary transformation which is within $\frac{\varepsilon}{20T}$ of $U_{\mathcal{M}'}$. After $5T$ runs of STEP the transformation is applied which is within $\frac{\varepsilon}{4}$ of the transformation of $\mathcal{M}'$. This implies that an observation of the simulation track of $\mathcal{M}'$, after it has completed its run, yields a sample from a distribution which is within the total variance distance ε of the distribution sampled by observing $\mathcal{M}$ on input x at time T (see Exercise 4.2.110. $\square$

Exercise 5.1.7 *(Bernstein and Vazirani, 1997) Let $\mathcal{M}$ be a QTM such that there is a polynomial time algorithm $\mathcal{A}$ such that, for any input x and any $t > 0$, all measurements of $\mathcal{M}$ at time t are determined by $\mathcal{A}$. Show that there is a QTM $\mathcal{M}'$ such that for any $t > 0$, any input x and any $\varepsilon > 0$, a measurement of $\mathcal{M}'$ at time t, determined by a fixed algorithm polynomial time in t and $\frac{1}{\varepsilon}$, allows sampling from a probability distribution of $\mathcal{M}'$, which is within the total variation distance ε of the distribution samples from $\mathcal{M}$, with respect to the outcomes of $\mathcal{A}$ over t steps.*

5.2 Quantum Computational Complexity

To ask the hard question is simple.

Wystan Hugh Auden (1907-1973)

Quantum variations of the main time and space computational complexity classes of classical computing are other important theoretical concepts to use in order to get a deeper insight into the power of quantum computing.

For quantum time complexity classes, and for their relations to classical computational complexity classes, several inclusion results are already known and will be discussed below. Unfortunately, as is usual in complexity theory, in many cases it is not known whether inclusions obtained are proper. However, it is known that many major new open problems cannot be resolved without making a breakthrough concerning the separation of the classical computational complexity classes.

In the case of quantum space complexity classes the situation is quite different, quantum computing has not brought an asymptotical decrease in the space resources needed.

5.2.1 Basic quantum versus classical complexity classes

Definitions of the complexity classes to consider for quantum complexity theory parallel that for randomized complexity classes. The main new parameter is the way measurements are made and, consequently, probabilities are calculated, as well as what types of amplitudes are allowed.

Types of QTM. In order to define the time complexity classes, one-tape multitrack QTM are considered. To define the space complexity classes, off-line multitape QTM are considered with one-way, read-only, input tape, a working tape, and one-way, write-only, output tape. Such a model is needed to investigate sublinear complexity classes. In both cases only such QTM are considered all amplitudes of which are rational. The case of more general amplitudes is discussed separately.

Measurements and probabilities. It has been shown by Bernstein (1997), (see also Exercise 5.1.7), that in order to study the time complexity classes, it is sufficient to consider only computations in which the measurement is done only after the machine comes into the halting state. On the other hand, to study the space complexity of multitape QTM (see Watrous, 1997a), a measurement is done each time a symbol is written on the output tape. More precisely, to study time complexity classes, it is sufficient to consider only such multitrack QTM $\mathcal{M}$ where the last track alphabet is $\{\lambda, 0, 1\}$ and if $\mathcal{M}$ runs with input string x on the first track and with all other track empty, then the last track is observed each time $\mathcal{M}$ gets into the halting state, which can be determined by a measurement that has no effect on the computation, and the probability p with which we see 1 on the starting cell of the last track is the overall probability x is accepted by $\mathcal{M}$.[1]

Acceptance. Table 5.1 summarizes the definitions of the main complexity classes. Each

	EQ	RQ	BQ	NQ	PrQ
$x \in L$	1	$\geq \frac{1}{2} + \varepsilon$	$\geq \frac{1}{2} + \varepsilon$	> 0	$> \frac{1}{2}$
$x \notin L$	0	0	$\leq \frac{1}{2} - \varepsilon$	0	$\leq \frac{1}{2}$

Table 5.1: Definitions of the main complexity classes. The occurrence of ε in a column of the table means that there is an $\varepsilon > 0$ such that the corresponding condition for acceptance and rejection holds for all x in the alphabet of L.

of the classes is defined as the class of languages L such that for $x \in L$ and for $x \notin L$ certain conditions for the acceptance probabilities are satisfied. E stands here for "error-free computation"; R stands for "one-sided bounded-error computation" (Monte Carlo); B stands for "bounded-error computation"; N for "nondeterministic computation" and Pr for unbounded-error computation. (In the Appendix, in Section 9.3.2, this class is denoted **PP** in the case of the acceptance in polynomial time.)

Resources. If $t : \mathbf{N} \to \mathbf{N}$ is a (time-constructible[2]) function and **CQ** a complexity class, then $\mathit{CQTime}(t(n))$ is the class of languages that belong to **CQ** and can be accepted by a CQ-Turing machine[3] in time $t(n)$. **CQP** stands for acceptance in polynomial time.

Similarly, if $s : \mathbf{N} \to \mathbf{N}$ is a (space-constructible) function and **CQ** a complexity class, then $\mathit{CQSpace}(s(n))$ is the class of languages that belong to **CQ** and can be accepted by a CQ-Turing machine within $s(n)$-space.

Quantum time complexity classes

Quantum versions of the classical time complexity classes **P** and **BPP** are classes **EQP** and **BQP**. The class **BQP** can also be defined as the class of languages $L \subseteq \Sigma^*$ such that a QTM can decide, with probability at least $\frac{2}{3}$ for each $x \in \Sigma^*$, whether $x \in L$.

In quantum computing the class **BQP** is regarded as the class of languages (problems) efficiently decidable (on quantum computers).

[1] One can define observation/acceptance also for space-bounded computations the same way as for time bounded computation. In this case the resulting space-bounded classes correspond to the "halting" classes, i.e., classes for which QTM halts absolutely

[2] A function $f : \mathbf{N} \to \mathbf{N}$ is $t(n)$-time-constructible and $s(n)$-space-constructible if the function $f' : \{1\}^* \to \{0,1\}^*$, defined by $f'(1^{(n)}) = bin^{-1}(f(n))$—binary representation of $f(n)$—is computable by a $t(n)$-time-bounded and $s(n)$-space-bounded 2-tape TM. f is called time-constructible (space-constructible) if f is f-time-constructible (f-space-constructible).

[3] A CQ-Turing machine is a QTM such that the language it accepts is accepted in a way satisfying the definition of the class CQ.

The following basic relations between the main classical and quantum time complexity classes hold:

$$\mathbf{P} \subseteq \mathbf{EQP} \subseteq \mathbf{BQP} \qquad \mathbf{BPP} \subseteq \mathbf{BQP} \subseteq \mathbf{PSPACE}.$$

However, in order to prove these inclusions we need to show more precisely how the acceptance of languages by QTM is defined and how the probability of acceptance is determined.

We consider only such multitrack QTM $\mathcal{M}$ where the last track alphabet is $\{\lambda, 0, 1\}$ and if $\mathcal{M}$ runs with input string x on the first track and with all other tracks empty, then the last track is observed when $\mathcal{M}$ gets into the halting state (which can be determined by a measurement that has no effect on the computation) and the probability p with which we see 1 on the starting cell of the last track is the overall probability x is accepted by $\mathcal{M}$. (Here the polynomial time restriction translates as follows: there exist a polynomial p, such that the halting state is observed after at most $p(|x|)$ steps.)

Relations between P, EQP, BPP, BQP and PSPACE

Basic inclusions between the main classical and quantum complexity classes are easy to show (see Bernstein and Vazirani (1997) for the first detailed proofs adapted here).

Theorem 5.2.1 $\mathbf{P} \subseteq \mathbf{EQP} \subseteq \mathbf{BQP}$

Proof. If $L \in \mathbf{P}$, then there is a DTM $\mathcal{M}_L$ such that $\mathcal{M}_L$ produces 1 on an input x if $x \in L$ and produces 0 otherwise. By synchronization theorem, there is a stationary, normal-form, two-track QTM which reads x from the first track and produces in polynomial time as the output, on the second track, the symbol 1 if $x \in L$ and 0 otherwise. Hence $L \in \mathbf{EQP}$. The inclusion $\mathbf{EQP} \subseteq \mathbf{BQP}$ is trivially true. ∎

Theorem 5.2.2 $\mathbf{BPP} \subseteq \mathbf{BQP}$

Proof. If $L \in \mathbf{BPP}$, then there exists a polynomial $p(n)$ and a PTM $\mathcal{M}$ with output 0 or 1 which has the following properties. For an input x, $\mathcal{M}$ computes for different $y \in \{0,1\}^{p(n)}$, as the content of a tape with random bits, a multiset S_x of $2^{p(n)}$ bits where the proportion of 1s in S_x is at least $\frac{2}{3}$ if $x \in L$ and at most $\frac{1}{3}$ otherwise. $\mathcal{M}$ will be used in the QTM $\mathcal{M}'$ accepting L as follows. $\mathcal{M}'$ first generates an equally weighted superposition of all basis states $|x, y\rangle$, $y \in \{0,1\}^{p(n)}$ and then $\mathcal{M}$ is applied to each basis state. This can be done as follows.

By composing a stationary normal form RTM which produces $0^{(p(n))}$ on the second track for an input x on the first track, with a stationary, normal-form QTM from Example 4.2.28, which applies Hadamard transformation to the contents of the second track, we get a stationary QTM $\mathcal{M}_0$ which produces on the input x the superposition

$$\frac{1}{\sqrt{2^{p(n)}}} \sum_{y \in \{0,1\}^{p(n)}} |x, y\rangle.$$

By composition of $\mathcal{M}_0$ with a stationary, normal form version of $\mathcal{M}$, designed according to synchronization theorem, we get a polynomial time QTM which on the input x produces the final superposition

$$\frac{1}{\sqrt{2^{p(n)}}} \sum_{y \in \{0,1\}^{p(n)}} |x, y, \mathcal{M}(x, y)\rangle,$$

where $\mathcal{M}(x, y)$ is the output of $\mathcal{M}$ on inputs x and y. By the assumption about $\mathcal{M}$ the proportion of 1s in $\{\mathcal{M}(x, y) \,|\, y \in \{0,1\}^{p(n)}\}$ is at least $\frac{2}{3}$ if $x \in L$ and at most $\frac{1}{3}$ otherwise. L is therefore in $\mathbf{BQP}$. ∎

Theorem 5.2.3 BQP $\subseteq$ PSPACE

Proof. Let $L \in$ **BQP** and $\mathcal{M} = \langle \Sigma, Q, q_0, \delta \rangle$ be a QTM that accepts L in polynomial time with the probability at least $\frac{2}{3}$. By Theorem 4.2.23, any QTM $\mathcal{M}'$ which is $(\varepsilon/(24|Q|\,|\Sigma|p(n)))$-close to $\mathcal{M}$ will simulate $\mathcal{M}$ for $p(n)$ steps with accuracy ε. Therefore, if $\mathcal{M}$ is simulated with accuracy $\frac{1}{12}$ the probability of success is still at least $\frac{2}{3} - \frac{1}{12} = \frac{7}{12}$. It is therefore sufficient to use such QTM $\mathcal{M}'$, amplitudes of which are given only to $\lg(288|\Sigma|\,|Q|p(n))$ bits.

It is now easy to see that $\mathcal{M}'$ can be simulated in polynomial space by a DTM. Indeed, in order to compute the amplitude of a configuration of $\mathcal{M}'$ after $p(n)$ computational steps, it is sufficient to make a depth-first search of the configuration tree to the depth $p(n)$. Since each amplitude has at most $\lg(288|\Sigma|\,|Q|p(n))$ bits, the overall amplitude of each configuration can be computed in polynomial space. In order to determine the overall probability of acceptance, it is sufficient to compute the sum of the squared magnitudes of those configurations at time $p(n)$ that are in the halting state and have blanks in all cells of the distance at most $p(n)$ from the starting cell. This can be done again within polynomial space. ∎

The above result has been improved (see Bernstein and Vazirani, 1997, credited to L. Valiant), and by Adleman et al. (1997), as follows.

Theorem 5.2.4 BQP $\subseteq$ PP $\subseteq$ $\mathbf{P}^{\#\mathbf{P}}$ $\subseteq$ PSPACE

Open problem 5.2.5 *Determine which of the inclusions* **P** $\subseteq$ **EQP** $\subseteq$ **BQP** *and* **BPP** $\subseteq$ **BQP** $\subseteq$ **PP** $\subseteq$ $\mathbf{P}^{\#\mathbf{P}}$ $\subseteq$ **PSPACE** *are proper.*

It follows from the above inclusions that one cannot expect a proof that QTM are more time efficient than the classical TM, in the sense that some strict inclusions of the classical and the corresponding quantum polynomial time complexity classes would be shown, unless there is some breakthrough in the classical complexity theory.

Impacts of amplitudes on quantum time complexity classes

The type of amplitudes allowed in QTM can have an essential impact on particular complexity classes, as shown by Adleman et al. (1997). If A is a set of amplitudes and C a complexity class, then let C_A denote the variant of **C** obtained by considering only such QTM all transition amplitudes of which are from A.

If $A = \mathbf{C}$, the set of complex numbers, then the class $\mathbf{BQP}_{\mathbf{C}}$ has uncountable cardinality[4] and $\mathbf{BQP}_{\mathbf{Q}} \subsetneq \mathbf{BQP}_{\mathbf{C}}$, where **Q** is the set of rational numbers. On the other hand $\mathbf{EQP}_{\mathbf{C}} = \mathbf{EQP}_{\mathbf{Q}}$. Moreover, if **A** is the set of algebraic numbers[5], then $\mathbf{EQP}_{\mathbf{A}} \subseteq \mathbf{PP}$.

As follows from the results of Section 5.1, when considering the class **BQP** it is sufficient to use QTM with a single quantum operation, a rotation by the angle $2\pi\sum_{i=1}^{\infty} 2^{-2^i}$. In this spirit let $\mathbf{EQP}_\theta$ denote the sets in **EQP** accepted by QTM with a single quantum operation—rotation by the angle θ. It has been shown for θ such that $\cos\theta$ is transcendental that $\mathbf{EQP}_\theta = \mathbf{P}$.

Open problem 5.2.6 *Is it true that* $\mathbf{EQP}_{\mathbf{A}} = \mathbf{BQP}_{\mathbf{A}}$*?*

[4] In addition, the class **BQP** contains non-recursive sets of all Turing degrees.

[5] Algebraic numbers are roots of polynomials with integer coefficients.

Position of the class NQP

The class **NQP**, a quantum analogue of **NP**,[6] was introduced by Adleman et al. (1998). A series of results (see Adleman et al. 1998, Fortnow and Rogers, 1998, and Fenner et al. 1998), about the power of the classes **NQP**, culminated by the result that

$$\mathbf{NQP} = \mathbf{co\text{-}C_{=}P}, \tag{5.3}$$

where $\mathbf{C_{=}P}$ is the class of decision problems that determine whether the number of accepting computable paths (of a NTM) equals that of rejecting paths.[7] This is the first result showing that a quantum complexity class can be characterized by a counting complexity class. In addition, Yamakami and Yao (1998) have shown that $\mathbf{NQP}_K = \mathbf{co\text{-}C_{=}P}$ for any K such that $\mathbf{Q} \subseteq K \subseteq \mathbf{C}$. Since the class $\mathbf{BQP_C}$ has uncountable cardinality this implies $\mathbf{BQP_C} \neq \mathbf{NQP_C}$.[8]

There is an analogy between **NP** and **NQP** also with respect to a "complete problem" characterization, see Fenner et al. (1998). Namely, a language L is in **NP** (**NQP**) if and only if there is a PTM (QTM) $\mathcal{M}$ and a polynomial p such that

$$x \in L \Leftrightarrow Pr[\mathcal{M} \text{ accepts } x \text{ in } p(|x|) \text{ steps}] \neq 0.$$

The above characterization of the class **NQP** has also brought a deeper understanding how hard are some basic problems of quantum computing. Fenner et al. (1998) have shown that the following problems are hard for polynomial time hierarchy: to determine whether a quantum computation has non-zero probability of acceptance and to determine whether a basis state will appear with no-zero amplitude in a superposition.

Quantum space complexity classes

Concerning space efficiency there is no essential difference between the classical and quantum computation.

The first result along these lines, for reversible DTM, was obtained by Bennett (1989), who showed that RTM require at most quadratic increase of space (at the cost of only slight increase in time) comparing to DTM. This result has been improved by Lange et al. (1997). They proved that RTM can simulate DTM with a constant factor increase of space, at the expense of the exponential increase in time.

The space efficiency of QTM, in comparison to PTM, was investigated by Watrous (1997). The main result is that unbounded-error space $\mathcal{O}(s(n))$-bounded QTM and PTM are equivalent in power. Since PTM can be simulated by a DTM with only a quadratic increase of space, this implies that any space $s(n)$-bounded QTM can be simulated by a DTM in space $\mathcal{O}(s^2(n))$.

Formally, the main result is stated as follows.

Theorem 5.2.7 *If a language L is accepted by an $s(n)$ space-bounded QTM with unbounded error, then L can be accepted by an $\mathcal{O}(s(n))$ space-bounded PTM with unbounded error. (In short $PrQSpace(s(n)) \subseteq PrSpace(s(n))$.)*

[6]Formally, in a sense, **NQP** can be seen as quantum analogue of the class **NP** even if such a definition does not fully correspond to the basic idea behind the class **NP**: guess and verify. A different view of **NQP**, represented as a complexity class called quantum **NP** was presented by Kitaev at AQIP'99. He defines this class in terms of quantum witnesses and verifiers—see Section 9.3 for the classical version of this approach.

[7]This class, introduced by Wagner (1986), contains, for example, the graph nonisomorphism problem, which is not known to be in **NP**. Equality 5.3 therefore implies that there is a QTM that can accept in polynomial time with nonzero probability a description of two graphs if and only if they are nonisomorphic.

[8]A space-bound versionb of equality 5.3 was shown by Watrous (1997).

Sketch of the proof. The following two technical results, for the first one see Allender and Ogihara (1996), the second to Watrous (1997a), are the base of the proof.

1. If $w_{A,B}$ denotes a binary encoding of integer matrices A, B, then $L = \{w_{A,B} \mid det(A) > det(B)\} \in PrSpace(\lg n)$.
2. Let $\mathcal{M}$ be an $s(n)$-space bounded multitape QTM. Then for each input x there are integer matrices A, B of degree $2^{\mathcal{O}(s(n))}$, elements of which are integers of length $2^{\mathcal{O}(s(n))}$, such that the following properties hold:
   - $det(A) > det(B)$ if and only if $\mathcal{M}$ accepts x with probability $> \frac{1}{2}$.
   - There exists a DTM $\mathcal{M}$ which on input x and an integer $k \in 2^{\mathcal{O}(s(n))}$, initially written on its working tape, computes the kth bit of $w_{A,B}$ in space $\mathcal{O}(s(n))$.

Let $\mathcal{M}$ be a QTM running in space $s(n)$, and let A and B be matrices the existence of which follows from Claim 2. Since both matrices are of degree $2^{\mathcal{O}(s(n))}$ and all their elements have $2^{\mathcal{O}(s(n))}$ size, the encoding $w_{A,B}$ can be assumed to have asymptotically at most the same length.

According to Claim 1, there is a lg-space bounded PTM $\mathcal{M}_1$ accepting the string $w_{A,B}$ such that $det(A) > det(B)$ with probability $> \frac{1}{2}$. On this base we can design a PTM $\mathcal{M}_2$ which works as follows. On an input x, $\mathcal{M}_2$ simulates $\mathcal{M}_1$ and keeps recording the position of the head of $\mathcal{M}_1$ on $w_{A,B}$. (For that $\mathcal{M}_2$ needs $\mathcal{O}(s(n))$ space.) On the base of this position, using the machine $\mathcal{M}_0$ the existence of which is assumed in Claim 2, second item, $\mathcal{M}_2$ computes in $\mathcal{O}(\lg 2^{\mathcal{O}(s(n))}) = \mathcal{O}(s(n))$ space the bit of $w_{A,B}$, which $\mathcal{M}_1$ needs to inspect. Since $det(A) > det(B)$, if $\mathcal{M}$ accepts x with probability $> \frac{1}{2}$, $\mathcal{M}_2$ also accepts x with probability $> \frac{1}{2}$. ∎

Open problem 5.2.8 *Are QTM with algebraic amplitudes equivalent to those with rational amplitudes with respect to space efficiency?*

As it follows from Theorem 5.2.1 and results of Watrous (1998), bounded-error probabilistic computations can be simulated by QTM either in time-efficient way or in space-efficient way. An open problem is whether they can be simulated by QTM in both time and space efficient way. The existence of such simulations for a class of important graph problems was shown by Watrous (1998).

Remark 5.2.9 Quantum version QNC of the classical parallel time complexity class NC is also of interest (see, for example, Moore and Nilsson, 1998a). Due to the enormous problems decoherence causes, it is of special interest to find out what can be computed in quantum parallel polylogarithmic time. (However, from a complexity-theoretic point of view (e.g. with a uniformity condition analogous to the classical case) a fully satisfactory definition of **QNC** seems to be still a challenge.)

5.2.2 Relativized quantum complexity

The most significant demonstration of the computational power of quantum evolution was in Chapter 3 on the black-box promise problems. The idea of the black-box computations is on the level of QTM captured, similarly as in the classical case, by the concept of **oracle QTM**.

Oracle QTM

In classical computing an oracle can be seen as a special subroutine each call of which costs only one time unit (see page 384 in the Appendix, Section 9.3, for classical oracle TM). In the context of QTM, subroutine calls have to satisfy a special requirement which has no classical parallel. It is necessary that subroutine calls do not leave around any garbage, only their outcomes, because computational paths with the same result, but different garbage behind do not interfere.

The simplest case of an oracle QTM is that with a Boolean oracle $f : \{0,1\}^* \rightarrow \{0,1\}$. An oracle QTM $\mathcal{M}$ with the oracle f has a special query (oracle) track, with the alphabet $\{\lambda, 0, 1\}$, on which the machine writes its question to the oracle in the form of a string xb with $x \in \{0,1\}^*, b \in \{0,1\}$ and it has two distinguished query states: a pre-query state $q?$ and a post-query state $q!$. If $\mathcal{M}$ enters the state $q?$ and the oracle track contains at that moment a string xb, where $x \in \{0,1\}^*$, $b \in \{0,1\}$, then $\mathcal{M}$ enters, in one step, the state $q!$ and the content of the oracle tape is changed to $x \cdot (b \oplus f(x))$, where $\cdot$ is here a symbol for concatenation. In other words, the XOR operation is performed with b and $f(x)$ as arguments.

In a more general case the oracle is represented by a unitary operator U and the oracle call changes the state $|x\rangle$ into $|Ux\rangle$.

As in the classical case, one can define such complexity classes as $\mathbf{C}^A$, where $\mathbf{C}$ is a quantum complexity class and A is an oracle and also classes $\mathbf{C}^{\mathcal{A}}$, where $\mathcal{A}$ is a set of oracles.

Separation of quantum complexity classes by oracles

As in the classical setting, various separation results have been obtained between relativized quantum complexity classes. The basic separation results are due to Berthiaume and Brassard (1992, 1992a, 1992b).

Theorem 5.2.10 *There are oracles A and B such that*

$$\mathbf{P}^A \subsetneq \mathbf{EQP}^A, \quad \mathbf{EQP}^B \not\subseteq \mathbf{NP}^B \cup \mathbf{co\text{-}NP}^B.$$

Proof. We prove in detail the existence of an oracle A with the above property. The existence of oracle B can be shown in a similar way and we only show the main new trick.

The construction of A is done recursively. In doing that we use two integer-to-integer functions, $p(n)$ and $s(n)$, where

$$p(1) = 1, \quad p(n) = 2^{p(n-1)} \text{ if } n > 1$$

and $s(n)$ is any recursive function which maps $\mathbf{N}$ into $\mathbf{N}$ and takes each value infinitely many times.

Let $\{\mathcal{M}_i\}_{i=1}^{\infty}$ be a fixed enumeration of oracle DTM with input alphabet $\Sigma = \{0,1\}$. The oracle A is defined by $A = \bigcup_{i=1}^{\infty} A_i$, where $A_1 = \emptyset$ and, for $n > 1$, $A_{n+1} = A_n \cup R_n$, with R_n is chosen as follows.

Let the machine $\mathcal{M}_{s(n)}$ be simulated on the input $1^{p(n)}$ for $2^{p(n)-1}$ steps with A_n as the oracle. If $\mathcal{M}_{s(n)}$ does not stop within $2^{p(n)-1}$ steps, or rejects the input within $2^{p(n)-1}$ steps, then $R_n = \emptyset$. Otherwise, let Q_n be the set of queries asked during the computation of $\mathcal{M}_{s(n)}$ on the input $1^{p(n)}$. Clearly, $|Q_n| \leq 2^{p(n)-1}$. In this case R_n is chosen as any subset of $\Sigma^{p(n)}$ of size $2^{p(n)-1}$ such that $R_n \cap Q_n = \emptyset$. Such a set must exist because at most half of the binary strings of length $p(n)$ can be query strings in the computation of $\mathcal{M}_{s(n)}$ on $1^{p(n)}$ for $2^{p(n)-1}$ steps.

Let now
$$S_A = \{1^n \mid A \cap \Sigma^n = \emptyset\}.$$

We show first that there is an oracle QTM with the oracle A which accepts S_A in polynomial time.

Observe that $1^n \in S_A$ if and only if there are no words of length n in A. Let $f : \{0,1\}^* \to \{0,1\}$ be a recursive function such that $f(x) = 1$ if and only if $x \in A$. For any $n \in \mathbf{N}$ let f_n be the restriction of f to inputs of length n. From the definition of A it follows that f_n is either constant or balanced for each n. Hence $1^n \in S_A$ if and only if f_n is not balancedbalanced function. For a given n this can be decided by a quantum computer in one step using the Deutsch–Jozsa algorithm. It therefore requires maximally linear time (to write down 1^n) on a quantum computer to decide whether $1^n \in S_A$.

Suppose now that S_A can be accepted by an oracle DTM M_i with oracle A in polynomial time. Let n be such that $s(n) = i$. We claim that $\mathcal{M}_i^A$ accepts the input $1^{p(n)}$ and uses at least $2^{p(n)-1}$ steps to do so.

Suppose first that $\mathcal{M}_i^A$ rejects the input $1^{p(n)}$. Then $1^{p(n)} \notin S_A$ because $\mathcal{M}_i^A$ should recognize S_A. Therefore, $A \cap \Sigma^{p(n)} \neq \emptyset$. However, this means, by the definition of oracle A, that $\mathcal{M}_i^A$ accepts $1^{p(n)}$ within $2^{p(n)-1}$ steps using oracle A_n. By the definition of A_n, the set A_n does not contain any oracle queries asked during such a computation. In addition, none of these queries can be longer than $2^{p(n)-1}$—there would be no time to write down such a query. Therefore, the sets A and A_n can differ in strings smaller than $2^{p(n)}$ only in strings from R_n that are in A, but not in A_n. Therefore the same accepting path in $\mathcal{M}_i^{A_n}$ for input $1^{p(n)}$ exists if A_n is replaced by A. Consequently, M_i^A accepts S_A—a contradiction. Hence S_A cannot be accepted by an oracle DTM with the oracle A.

Suppose now that $\mathcal{M}_i^A$ accepts $1^{p(n)}$ with less than $2^{p(n)-1}$ steps. Then $1^{p(n)} \in S_A$, since $\mathcal{M}_i^A$ is supposed to accept S_A. Hence $A \cap \Sigma^{p(n)} = \emptyset$ and therefore A and A_n are identical when restricted to strings of length $2^{p(n)}$ or less. Consequently, it does not matter whether we use as oracle A or A_n for the first $2^{p(n)-1}$ steps of $\mathcal{M}_i$. Hence $\mathcal{M}_i^{A_n}$ also accepts $1^{p(n)}$, and therefore $A \cap \Sigma^{p(n)} \neq \emptyset$ by the construction of A—a contradiction.

In a similar way we can show the existence of an oracle B and a set $Y_B \in \{1\}^*$ such that $Y_B \subseteq \mathbf{EQP}^B$, $Y_B \notin \mathbf{NP}^B$ and the set $Y_B \cup \{0^n \mid 1^n \notin Y_B\}$ is neither in $\mathbf{NP}^B$ nor in $\mathbf{co-NP}^B$, but it is in $\mathbf{EQP}^B$. ∎

In addition, a variety of relativized results concerning **BQP** have been obtained by Fortnow and Rogers (1998) that seem even to suggest an hypothesis that **BQP** actually contains no interesting complexity class outside **BPP**. For example, there is a relativized setting where $\mathbf{P} = \mathbf{BQP}$ and the polynomial time hierarchy (see, for example, Gruska (1997)) is finite. And there is a relativized setting in which **BQP** does not have complete sets.

Simon (1994) showed the existence of an oracle relative to which **BQP** cannot be simulated by a PTM in $2^{n/2}$ steps.

Another key question is whether $\mathbf{NP} \subseteq \mathbf{BQP}$. Bennett et al.(1997) showed that relative to a random oracle, with probability 1, the class **NP** cannot be solved on a QTM in time $o(2^{n/2})$. This bound is tight due to Grover's result in Section 3.3.[9]

[9] These results do not rule out the possibility that $\mathbf{NP} \subseteq \mathbf{BQP}$. (It is not even clear whether $\mathbf{BQP} \subseteq \mathbf{BPP}^{\mathbf{NP}}$; i.e. whether nondeterminism+randomness is sufficient to simulate QTM.) These results only imply that there is no "black-box approach" to solving **NP**-complete problems utilizing some uniquely quantum mechanical features of QTM (see Bennett et al. 1997).

As already mentioned, oracles of QTM have to fullfil special conditions in order not to have undesirable effects. An important task is therefore to determine how powerful oracles can be used. Bennett et al. (1997) showed that each BQP-machine[10] $\mathcal{M}$ can be modified into another equivalent BQP-machine whose final superposition consists almost entirely of a tape configuration containing just the input and a single bit answer. On this base they have shown that $\mathbf{BQP}^{\mathbf{BQP}} = \mathbf{BQP}$. In addition, Fortnow and Rogers (1998) have shown that $\mathbf{PP}^{\mathbf{BQP}} = \mathbf{PP}$.

In the classical complexity one considers, in addition to various classes $\mathbf{P}$, $\mathbf{NP}$,... of decision problems, also their F-versions—classes of functions computable with similar resources. The same can be done also in quantum computing. For example, Aharonov et al. (1998) have introduced the class **FQP** of functions computable by uniform quantum circuits with polynomial size and depth. For this class they have shown that $\mathbf{FQP}^{\mathbf{FQP}} = \mathbf{FQP}$.

One-way functions and the power of quantum computing

Another important concept of classical complexity theory and also of classical cryptography is that of one-way functions.

A polynomial time computable function $f : \Sigma^* \rightarrow \Sigma^*$, for an alphabet Σ, is one-way if it is one-to-one, honest and not invertible in polynomial time. (Honesty means that there is a polynomial p such that $p(|f(x)|) > |x|$ (that is honest functions do not map long input strings into short ones).)

It is well known that one-way functions exists if and only if $\mathbf{P} \neq \mathbf{UP}$, where $\mathbf{UP}$ is the class of languages L that can be accepted by NTM which have exactly one accepting path for each $x \in L$.

The following result (due to Fortnow and Rogers, 1998), shows that the existence of one-way functions does not necessarily imply that quantum computing is more powerful than the classical one.

Theorem 5.2.11 *There is an oracle A relative to which one-way functions exist and* $\mathbf{P}^{\mathbf{A}} = \mathbf{BQP}^{\mathbf{A}}$,

Remark 5.2.12 A variety of results concerning the power of quantum computation have been presented in Chapters 3 to 5. In spite of all these insights we have to realize the fact that we are still far away from understanding well the computational power of quantum systems.

Remark 5.2.13 The study of quantum computational complexity can be of importance beyond quantum computing. Indeed, let us assume that we can prove that quantum systems can be, from the computational point of view, exponentially more powerful. In such a case, perhaps, using complexity oriented tools we could get a better insight concerning a borderline between classical and quantum physics.

5.3 Quantum Communication Complexity

There are two essentially different ways of using quantum phenomena to make communications more efficient.

[10]A QTM $\mathcal{M}$ is said to be a BQP-machine, if it halts in polynomial time and if it accepts (rejects) an input x, then it accepts (rejects) it with probability at least $\frac{2}{3}$.

In the so-called **entanglement model** only classical bits are communicated but communication is facilitated by an a priori distribution of entangled qubits among the communicating parties. It has already been demonstrated in Section 2.2 that this can bring asymptotic improvement compared to the classical communication complexity. [11]

In the so-called **qubit communication model**, discussed in this section, communicating parties exchange qubits. This model is quite a straightforward generalization of the model of communication in classical communication complexity.

Classical communication complexity theory (see Hromkovič, Juraj (1997), Kushilevitz and Nisan (1997), Gruska (1997)), has already been much developed. Its importance stems from the experience that it is to a large extent the complexity of the communication that is behind the complexity of parallel and distributed computing. Lower-bound results on communication complexity are often used to derive lower-bound results for computation complexity. On the other hand, the development of the quantum communication complexity theory is only just beginning and seems to be essentially more difficult.

It is not immediately clear whether qubits can reduce communication costs because one of the fundamental results of the quantum information theory, due to Holevo (1973), says that by sending n qubits one cannot convey faithfully more than n bits of information.

5.3.1 Classical and quantum communication protocols and complexity

The main concepts and definitions of quantum communication complexity theory parallel those for the classical communication complexity theory.

Classical communication protocols and complexity

The basic (two-parties) **communication problem** is to determine how many bits parties A and B, usually called Alice and Bob, have to exchange in order to compute $f(x,y)$, provided Alice knows x, Bob knows y, and both of them know the function f. It is assumed that in any communication both parties exchange bits and in doing that they follow a communication algorithm, usually called **protocol**. $C_\pi(f)$ denotes the minimal number of bits they need to communicate in order to compute f provided input data are divided between Alice and Bob according to a given partition π.

Example 5.3.1 *Let $x, y \in \{0,1\}^n$, $\pi = \{x, y\}$. If $f(x,y)$ is the parity of the string xy, then clearly $C_\pi(f) = 1$. However, if $f(x,y) = 1$ if and only if $x = y$, then it can be shown that $C_\pi(f) = n$.*

In the above setting we have considered a fixed partition of inputs. Another model of the classical communication complexity is to take least number of communication bits needed to solve the problem with respect to all **balanced** (equal size) **partitions** of inputs between Alice and Bob—notation $C_b(f)$.

Example 5.3.2 *(Addition of binary numbers.) Assume that parties A and B are to compute the sum of two n-bit numbers $x = a_n \dots a_1$, $y = b_n \dots b_1$, where n is even, and each of them knows exactly half of the input bits. Assume also that B is to compute $\frac{n}{2}$ of the least*

[11] Sometimes also a modification of the above entanglement communication model is considered in which qubits are used for communication. As shown in Section 6.4.4, with n entangled pairs of particles one can send n qubits by sending $2n$ bits. Entangled model therefore needs for communication at most twice as many bits as qubits, if enough of entangled pairs is available.

significant bits of the sum and A the rest. How many bits do they need to exchange? The answer largely depends much on how the input bits are divided between the two parties. Let us consider two possible cases.

1. If B knows $a_{\frac{n}{2}} \dots a_1, b_{\frac{n}{2}} \dots b_1$ *and A the rest of the input bits, then it is clearly enough that B sends to A the single bit, namely* 0*, if*

$$bin(a_{\frac{n}{2}} \dots a_1) + bin(b_{\frac{n}{2}} \dots b_1) < 2^{n/2}$$

and 1 *otherwise. A can then compute the remaining bits of the sum.*

2. However, if A knows $a_n \dots a_1$ *and B knows* $b_n \dots b_1$*, then it seems to be intuitively clear that B needs to get bits* $a_{\frac{n}{2}} \dots a_1$ *and A needs to get at least bits* $b_n \dots b_{\frac{n}{2}}$ *and an additional bit carrying information whether the sum of* $n/2$ *least significant parts of both numbers is or is not larger than* $2^{n/2}$*.*

Another communication model is to consider the minimum of communication bits needed for all **almost balanced partitions** of inputs—each party gets at least one-third and at most two thirds of the input bits.

In the case of **randomized communications** both parties are allowed to use their private random bits to decide in each communication round which bits they communicate. In such a case, the outcome of a communication is a random distribution on outputs and it is required that the outcome is correct with the probability $1 - \varepsilon$, for some $\frac{1}{2} > \varepsilon \geq 0$.

There are several ways complexity of communication is defined in the case of randomized communications and a fixed distribution of inputs. We talk about **zero-error** or **Las Vegas** communication complexity of a communication problem P, notation $C_0(P)$, if the result is always correct. **Bounded-error communication complexity** $C(P)$ is in the case all results are correct with the probability at least $\frac{2}{3}$.

Example 5.3.3 *Let* $f, g : \{0,1\}^n \rightarrow \{0,1\}$ *and let Alice get as the input* f *and Bob* g*. The following communication problems belong to the basic ones.*

- $EQ(f,g) = \bigwedge_{x \in \{0,1\}^n} (f(x) = g(x))$,
- $IP(f,g) = \bigoplus_{x \in \{0,1\}^n} (f(x) \wedge g(x))$,
- $DISJ(f,g) = \bigvee_{x \in \{0,1\}^n} (f(x) \wedge g(x))$,

where these functionals represent equality, inner product and disjointness[12] *of two functions, respectively.*

The following results hold; see references in Gruska (1997) and Buhrman et al. (1998).

$$C_0(\mathrm{EQ}) = 2^n, C(\mathrm{EQ}) = \mathcal{O}(n), C(\mathrm{IP}) = \Omega(2^n), C(\mathrm{DISJ}) = \Omega(2^n).$$

Quantum communication protocols and complexity

Communication protocols are defined in the quantum case in a similar way to the classical case, the only difference being that complexity of communication is considered by the number

[12] As defined above $\mathrm{DISJ}(f,g)$ stands for the complement of the set of disjointness problems. However, for the model considered here the communication complexity of DISJ problem and its complement are equal (Buhrman et al. 1998).

of qubits exchanged. Similarly with randomized classical communication protocols, the result of quantum communications are random distributions on outputs.

A special feature of quantum communication is that due to the fact that parties cannot make copies of their qubits they have sometimes sent them forth and back.

Several quantum communication complexity measures are considered as analogous to the classical ones. For a communication problem P, let $QC_\varepsilon(P)$ denote the minimal number of qubits that need to be exchanged if the outcomes are to be correct with probability $1-\varepsilon$. $QC_0(P)$ then stands for Las Vegas, error-free, communication complexity and $QC(P) = QC_{1/3}(P)$ for bounded-error quantum communication complexity.

5.3.2 Quantum communication versus computation complexity

A simple technique, presented below, due to Buhrman et al. (1998), allows the transfer of quantum algorithms for black-box computations to quantum communication protocols. This technique can be used in two ways.

1. To derive upper bounds for quantum communication complexity from the upper bounds for quantum computation complexity.

2. To derive lower bounds for quantum computational complexity from the lower bounds for quantum communication complexity .

Let $\mathcal{F}_n$ denote the set of Boolean functions $f : \{0,1\}^n \to \{0,1\}$.

Theorem 5.3.4 *Let $F : \mathcal{F}_n \to \{0,1\}$ and $L : \{0,1\} \times \{0,1\} \to \{0,1\}$. L induces a mapping $\mathcal{L} : \mathcal{F}_n \times \mathcal{F}_n \to \mathcal{F}_n$ such that $\mathcal{L}(g,h)(x) = L(g(x),h(x))$ for all $x \in \{0,1\}^n$. If there is a quantum algorithm $\mathcal{A}$ to compute $F(f)$ with t calls of f, then there is a $t(2n+4)$-qubit quantum communication protocol $\mathcal{P}$ for the following problem: Alice gets g, Bob gets h and the aim is for Alice to determine $F(\mathcal{L}(g,h))$ by communication with Bob. In addition, the probability that communication according to $\mathcal{P}$ produces the correct result is the same as for the algorithm $\mathcal{A}$.*

Proof. Let $\mathcal{A}$ be a quantum algorithm that computes $F(f)$ with t calls of f. In the communication protocol Alice makes use of the algorithm $\mathcal{A}$ with $f = \mathcal{L}(g,h)$. She needs to communicate with Bob only when an $\mathcal{L}(g,h)$-call is to be made because she does not know h. To get the result of an $\mathcal{L}(g,h)$-call Alice initializes the following communication protocol with Bob, where $x \in \{0,1\}^n$, $y \in \{0,1\}$, is the current argument for which $\mathcal{L}(g,h)$ is to be computed.

1. Alice performs the unitary operation $|x,y,0\rangle \to |x,y,g(x)\rangle$ and sends $n+2$ qubits to Bob;

2. Bob performs the unitary operation $|x,y,g(x)\rangle \to |x, L(g(x),h(x)) \oplus y, g(x)\rangle$ and sends $n+2$-qubits to Alice;

3. Alice uncomputes $g(x)$ to produce $|x, L(g(x),h(x)) \oplus y, 0\rangle$.

Since there are t calls for which such a communication is needed, the total amount of exchanged qubits is $t(2n+4)$. $\square$

Exercise 5.3.5 *(a) Does the protocol in the proof of Theorem 5.3.4 make use of entanglement? If yes, then where? If not, could we do the whole protocol classically? (b) Does Bob need to send Alice $n+2$ qubits? Is it not sufficient for him to send back only one qubit with $L(g(x), h(x)) \oplus y$?.*

Example 5.3.6 *The upper bound $QC(DISJ) = \mathcal{O}(n\sqrt{2^n})$ follows from Theorem 5.3.4, by taking L to be binary AND function and F to be 2^n-ary OR function, because for computation of 2^n-ary OR-function we have the upper bound $\sqrt{2^n}$, see Section 3.3.*

Example 5.3.7 *An exponential gap between the exact classical and quantum communication complexity has been shown by Buhrman et al. (1998) for the following problem.*

Let $f, g : \{0,1\}^n \to \{0,1\}$ and let $\Delta(f,g)$ be the Hamming distance between f and g which equals the Hamming distance of 2^n-bit strings $f(0)f(1)\ldots f(2^n-1)$ and $g(0)g(1)...g(2^n-1)$. Let EQ' be the partial function defined by

$$EQ'(f,g) = \begin{cases} 1, & \text{if } \Delta(f,g) = 0, \\ 0, & \text{if } \Delta(f,g) = 2^{n-1}; \end{cases}$$

and undefined for other arguments. (In the case of partial functions we require that communication yields the correct outcome only for arguments at which the partial function is defined.)

The upper bound $QC_0(EQ') = \mathcal{O}(n)$ is the consequence of Theorem 5.3.4. Indeed, take L to be XOR function and F to be the 2^n-ary OR function restricted to balanced or zero functions. The upper bound now follows from the analysis of the Deutsch–Jozsa algorithm in Section 3.1. It has been shown, by Buhrman et al. (1998), that $C_0(EQ') = \Omega(2^n)$.

Example 5.3.7 shows a communication task for which the number of bits needed to communicate in order to compute a given function with zero-error is exponentially larger then the number of qubits that need to be communicated. However, for this task there is a classical randomized communication protocol that achieves the same result with small error and requires to communicate the same number of bits as the number of qubits needed for the best quantum communication protocol.

The first fully exponential gap between classical bounded-error randomized communication and quantum communication has been shown by Ambainis et al. (1998a) for the following sampling task. Alice has a subset $A \subseteq \{1, 2, \ldots, n\} = S$ of cardinality k and Bob task is to pick up another subset $B \subseteq S$ of cardinality k disjoint with A (if possible), The result was obtained as a byproduct of a method to deal with the following important communication primitive.

Definition 5.3.8 *(Sampling) Let $f : X \times Y \to \{0,1\}$ and let D be any probability distribution on $X \times Y$. A communication protocol $\mathcal{P}$ is said to sample f according to $\mathcal{D}$ with error $\varepsilon > 0$, if the distribution the protocol induces on $\{(x,y,z)\}$ is ε-close, in the total variation distance, to the distribution $(\mathcal{D}, f(\mathcal{D}))$ obtained by picking first (x,y) according to $\mathcal{D}$ and then computing $z = f(x,y)$.*

For the disjointness problem discussed above nad $k = \Theta(\sqrt{n})$ they give a quantum protocol in which Alice sends $\mathcal{O}(\lg n \lg \frac{1}{\varepsilon})$ qubits to Bob and they allow him to sample from a distribution ε-close to the desired uniform distribution on subsets of S disjoint to A. In addition, they show that each classical randomized protocol needs $\Omega(\sqrt{n})$ bits to be exchanged between Alice and Bob.

5.4 Computational Power of quantum non-linear mechanics

It is one of the basic axioms of quantum mechanics that quantum evolution is fully described by the linear Schrödinger equation. For many classical phenomena linear models are just a first approximation of the reality, and when a deeper insight has to be obtained nonlinear models have to be considered. It has therefore been natural to expect that the same could apply to quantum phenomena. However, so far all quantum experiments confirm linearity. In spite of that some believe that small nonlinearities in quantum evolution have not yet been completely ruled out.

There are various models of quantum nonlinear mechanics and it is of interest to explore how much computational power their validity would bring. Some of the nonlinear models have turned out to be "too weird"—within them effects are possible that are commonly considered as "unphysical". For example, faster than light effects or even an arbitrarily fast influence between non-interacting systems. However, there are also nonlinear models that are claimed to be free of such effects.

Abrams and Lloyd (1998) described two ways how to use nonlinearity of quantum mechanics to solve **NP**-complete problems in polynomial time. We present one of them and we show that there exists a polynomial time algorithm, which uses some nonlinear quantum operations, the existence of which is fully within a model of nonlinear quantum mechanics, to decide, given a function $f : \{0,1\}^n \rightarrow \{0,1\}$ whether there is an x such that $f(x) = 1$.

The algorithm makes use of the following nonlinear operator on two qubits .

$$|00\rangle + |11\rangle \rightarrow |01\rangle + |11\rangle, \qquad |01\rangle + |10\rangle \rightarrow |01\rangle + |11\rangle, \qquad |00\rangle + |10\rangle \rightarrow |00\rangle + |10\rangle \tag{5.4}$$

In order to simplify presentation we assume that there is at most one i such that $f(i) = 1$.

Algorithm 5.4.1 *Given* $f : \{0,\dots,2^n - 1\} \rightarrow \{0,1\}$

1. In the superposition
$$|\psi\rangle = \frac{1}{\sqrt{2^n}} \sum_{i=0}^{2^n-1} |i, f(i)\rangle$$
group all basis states $|i, f(i)\rangle$ into 2^{n-1} pairs depending on the values of the qubits number 2 to n. This way we create 2^{n-1} subspaces of dimension 4 for the first and $(n+1)$st qubit. For each subgroup the first and $(n+1)$st qubits will be in one of the following three states
$$|00\rangle + |11\rangle, \qquad |01\rangle + |10\rangle, \qquad |00\rangle + |10\rangle \tag{5.5}$$
(Observe that at most one of the 2^{n-1} superpositions will not be in the third state.) Apply now to these two qubits the nonlinear transformation (5.4).

2. Apply the previous construction $n-1$ times in such a way that at the ith repetition the ith and $(n+1)$th qubit are chosen to form the 4-dimensional space. These two qubits will always be, for any combination of values of other qubits, in one of the states (5.5).

   The key point now is that at each such iteration, and at each application of the operator (5.4), the number of components in the superposition which have the value 1 in the last qubit doubles. Therefore, after n iterations the last qubit is no longer entangled with the others and it is in the same state for all components of the superposition.

3. Measure the last qubit.

Nonlinear transformation (5.4) used in the algorithm seems to be artificial, chosen just to get the desirable result. However, Abrams and Lloyd (1998) demonstrated that virtually any deterministic nonlinear quantum theory will include such a nonlinear operator. In addition, they have shown that virtually any deterministic nonlinear operation can be recast into this form and the operation (5.4) can be constructed from an arbitrary unitary operation and a simple nonlinear operation $|0\rangle \to |0\rangle$ and $|1\rangle \to |0\rangle$.

Abrams and Lloyd (1998), developed their algorithms within Weinberg's (1989) model of nonlinear dynamics and they used operators that do not preserve scalar product of states. In such a case Weinberg's model can exhibit unphysical effects. By Czachor (1998) slight modifications of both algorithms work also in the model that is known to be free of unphysical influences.

Using quite a different model of quantum computing, Černý (1993) has shown how to solve one **NP**-complete problem, namely the traveling salesman problem, in quantum polynomial time, but using an exponentially large amount of energy.

Chapter 6

CRYPTOGRAPHY

INTRODUCTION

Secure communication is one of the areas of key importance for modern society in which quantum information transmission and processing seems to be able to bring significant contributions. For example, quantum cryptography may be the main defence against quantum code breaking in the future.

An important new feature of quantum cryptography is that security of quantum key generation and quantum cryptographic protocols is based on a more reliable fact, on the laws of nature as revealed by quantum mechanics, than in the case of classical cryptography, whose security is based on unproven assumptions concerning the computational hardness of some algorithmic problems.

It is difficult to overemphasize the importance of quantum cryptography for an understanding and utilization of quantum information processing. Quantum cryptography was the first area in which quantum laws were directly exploited to bring an essential advantage in information processing.

Closely related are quantum teleportation and quantum superdense coding—special ways of the transmission of quantum or classical information using one of the most puzzling phenomena of the quantum world—non-locality features of the quantum entanglement.

LEARNING OBJECTIVES

The aim of the chapter is to learn:

1. several methods of secret key generation by two parties;
2. a method of multiparty secret key generation;
3. the unconditional security of quantum key generation;
4. the basic quantum cryptographic protocols;
5. the problems related to the security of quantum cryptographic protocols;
6. the main principles, circuits and some applications of quantum teleportation;
7. the quantum superdense coding procedure.

"Amazing, Holmes."
"Elementary, my dear Watson, elementary."

Security of a significant portion of the public-key cryptography and cryptographic protocols in use today depends directly or indirectly on the essential difficulty of either factoring of integers, or computing discrete logarithms. Shor's results therefore indicate that quantum mechanics has the potential to ruin a large part of the contemporary cryptography.[1] That is also a reason why Shor's results were taken as very significant by all those whose well-being can be much influenced by significantly new tools for breaking codes. This is so far also the main reason why quantum computing research finds support in institutions much depending on security of communication.

However, quantum mechanics allows also to design cryptographic systems of a completely new type and thereby also new cryptography—significantly more secure than so far. One can say that quantum cryptography brought a new dimension to security of communications. Its main new feature are mechanisms for detecting eavesdropping. The very basic idea is simple. If classical information is encoded in a special quantum way, for example, with nonorthogonal quantum states, and transmitted, then any attempt of an eavesdropper to access transmitted information spoils, with high probability, information irreversibly in a way which can be detected by the legal receiver. Quantum cryptography provides a technologically feasible approach to harnessing quantum physics for communication and security purposes. In doing that quantum cryptography makes a positive use of fundamental limitations of quantum physics—for example of Heisenberg's uncertainty principle.

While security of classical cryptosystems is based on unproven assumptions of computational complexity (and can be jeopardized by progress in technology, computing theory and algorithmic design), security of quantum cryptographic systems is based on the laws of nature—or at least on basic assumptions of quantum mechanics (such as its completeness).

The idea of quantum cryptography goes back to S. Wiesner around 1970, but it was practically only in 1983 when his ideas received an attention, mainly due to Ch. Bennett. Subsequent development was very fast and already in 1994 transfer of quantum information along an ordinary optic fibre of several kilometers length had been performed.

Interestingly enough, the main reason why quantum cryptography went from science fiction to science in such a short term was not technological progress but the discovery of how to make reasonable use of quantum mechanics for cryptographic purposes.

[1] Secret-key cryptography, such as DES, has not yet been shown to be efficiently breakable by quantum computing. Whether this can be done is an interesting but hardly practical problem because DES is expected to be obsolete before any quantum computer is built.

6.1 Prologue

Quantum cryptography is like classical cryptography, a continuous fight between good and bad. The basic setting is that Alice tries to send a quantum system to Bob and an adversarial eavesdropper, Eve, tries to learn, or to change, as much as possible without being detected.

An eavesdropper has this time an especially hard task. Quantum states cannot be copied and cannot be measured without causing, in general, a disturbance.

The key problem can be formulated as follows (see Fuchs and Peres, 1996). Alice prepares a quantum system in a specific way unknown to Eve and sends it to Bob. The question is how much information can Eve extract of that quantum system and what the cost of that information is, in terms of the disturbance of the quantum system. Two concepts are therefore crucial here: *information* and *disturbance*.

Let us consider two extreme cases within the scheme that Alice sends a state $|\psi\rangle$ to Bob. The first is that Eve has no information about how $|\psi\rangle$ was prepared. The only thing Eve can then do is to choose some basis $\{|e_i\rangle\}$ of orthonormal states and to use the corresponding projection measurement on $|\psi\rangle$. In such a case $|\psi\rangle$ collapses into one of the states $|e_i\rangle$ and by that the only information Eve has learned is that $|\psi\rangle$ is not orthogonal to $|e_i\rangle$. At the same time $|\psi\rangle$ can get much destroyed.

The second extreme case is that Eve knows that $|\psi\rangle$ is one of the states of the basis $\{|e_i\rangle\}$. In such a case Eve gets, by measuring $|\psi\rangle$ with respect to the basis $\{|e_i\rangle\}$ full information about $|\psi\rangle$ because $|\psi\rangle$ collapses into itself. No disturbance to $|\psi\rangle$ occurs.

The most interesting, important and hard case for quantum cryptography is the third case, where Eve knows that $|\psi\rangle$ is one of the states $|\psi_1\rangle, \ldots, |\psi_n\rangle$ that are mutually nonorthogonal and p_i is the probability that $|\psi_i\rangle$ is the state Alice sends. In this case the question is how much information Eve can get by a measurement and how much disturbance she causes by that.

How to measure information gain of Eve? Shannon entropy $\sum_{i=1}^{n} p_i \lg p_i$ is a measure of her ignorance about the system before the transmission takes place. She can try to decrease this entropy by some measurement. What she gets is called *mutual information*.

In the case of eavesdropping Bob does not get a pure state but a mixed state specified by a density matrix ρ_i for the case Alice sends $|\psi_i\rangle$. The **disturbance** detectable by Bob is given by $D = 1 - \langle\psi_i|\rho_i|\psi_i\rangle$.

An important case is that $n = 2$, $p_1 = p_2$ and $|\psi_1\rangle$ and $|\psi_2\rangle$ are states of H_2 such that a measurement of $|\psi_b\rangle$ with respect to $|\psi_{\bar{b}}\rangle$ produces both states with the same probability. In such a case Eve has a 50% chance to make a correct guess for her measurement. In the case of the correct guess she gets $|\psi\rangle$ for 100%, otherwise for 50%. In the average she gets 75% of information. Can she do better? By doing some other measurements? As we shall see in this chapter, she can get up to 85% of information but not more, no matter what she does. This may sound pretty good, but not enough for cryptography, as we shall see.

Of course, when Alice sends a sequence of states Eve has also other options for eavesdropping than to try to measure immediately, one-by-one each sent state. For example, she can postpone measurements to the end of the transmissions.

In such a general case the problem of Eve's information gain versus disturbance is one of the central one for security of quantum cryptography protocols.

How to send quantum states in general and qubits in particular is another general problem. Transmission of polarized photons is so far one of the basic tools in quantum cryptographic protocols for sending qubit states.[2] The main property to be used is that if a

[2]For more about polarization of photons see Section 9.1.2 and 9.2.6.

θ-polarized photon is subjected to a polarization measurement along axis ϕ, it behaves like a ϕ-polarized photon with probability $\cos^2(\theta-\phi)$ and like a $(\phi+90)^\circ$-polarized photon with probability $\sin^2(\theta-\phi)$. Moreover, such a measurement can be performed by using a ϕ-oriented piece of birefringed material, such as calcite, to split the incoming light beam into two beams (polarized at ϕ and $\phi+90^\circ$).

6.2 Quantum Key Generation

Quantum secret key generation discussed next is so far the most important application of quantum mechanics to cryptography. A secret key can be generated by several parties. The most important case is that of two parties.

6.2.1 Basic ideas of two parties quantum key generation

One of the main problems of cryptography is the so-called **key distribution problem**. How can two parties, Alice and Bob, come into the possession of a secret key in such a way that the third party, an eavesdropper Eve, cannot acquire useful information about the key.

Indeed, the main idea of the classical **secret key cryptography** is to use cryptosystems S in which both **encrypting algorithms** e_S and **decrypting algorithms** d_S are publicly known and if two parties, Alice and Bob, want to be engaged in a secure communication, then they have to share a **secret random key** k. With the key k the **encryption** of a **plaintext** w and the **decryption** of a **cryptotext** c are performed by the computations

$$c = e_S(k, w) \quad \text{and} \quad w = d_S(k, c).$$

This way such well-known and practically important cryptosystems as DES and ONE-TIME PAD work. The second cryptosystem is **perfectly secure** under the assumption that each key has the same length as the plaintext w and each time the cryptosystem is used with a new randomly chosen key. In the case of binary plaintexts and keys the encryption and decryption algorithms of ONE-TIME PAD cryptosystem are very simple

$$c = k \oplus w \quad \text{and} \quad w = k \oplus c,$$

where $\oplus$ stands for a component-wise exclusive-or operation. (Observe that using ONE-TIME PAD cryptosystem we need one secret bit to transfer securely one bit—quite a price.)

Since the search for limitations concerning security is also one of the main tasks of the quantum computing research, let us look a bit more closely into the problem in which sense ONE-TIME PAD cryptosystem is perfectly secure. Let $\mathbf{w}$ and $\mathbf{c}$ be random variables for plaintexts and cryptotexts. ONE-TIME PAD cryptosystem is perfectly secure in the sense that $Pr(\mathbf{w}|\mathbf{c}) = Pr(\mathbf{w})$, or, equivalently $Pr(\mathbf{c}|\mathbf{w}) = Pr(\mathbf{c})$.

Exercise 6.2.1 *Show why it is necessary for the security of ONE-TIME PAD cryptosystem to require that each key is used only once.*

Classical cryptography can therefore be so much secure, how secure key distribution methods are available.

The main practical importance of **public key cryptography** is therefore for key distribution. It provides **computationally secure** key distribution under some unproven assumptions of computational complexity. For example, on the assumption that no polynomial time algorithm exists for integer factorization.

In the classical setting no unconditionally secure key distribution is possible. On the other hand, as discussed in this and the next section, unconditionally secure key generation is possible, with respect to our current knowledge of quantum physics, using communications through quantum channels.

In the classical setting we speak about the key distribution because the basic scheme is that one party (Alice) sends (distributes) the key to the other party (Bob) in a secure way. In quantum setting we speak about quantum key generation (QKG) because in such a process both parties play a more symmetrical role, and only in the course of their communication do they generate (or agree on), a secret key.[3]

Quantum key generation can be done in such a way that

- no eavesdropper can obtain, with non-negligible probability, the transmitted key,
- any attempt of an eavesdropper to obtain a transmitted key can be detected (with as large probability as needed),
- both parties can be as sure as needed that they share a secret key.

Any interception by an eavesdropper is actually a set of measurements on the communication channel. Classical key distribution can in principle be passively monitored (no matter how difficult this may be technologically), without the legitimate users being aware that any eavesdropping has taken place. On the other hand, quantum cryptography provides means to detect any attempt of an eavesdropper to tap the channel.

Protocols

The basic idea behind key generation is simple. If Alice and Bob want to share a binary key of length n, then each of them first generates independently a private random binary sequence of length $m \gg n$. In order to extract from this random sequence a common key of length n, Alice prepares a sequence of m tokens, one type for bit 1 and another type for bit 0 and sends through a sequence of such tokens her random sequence to Bob. Bob reports to Alice the order, but not the value, of those of Alice's bits that are the same as his. From this sequence of bits they select n bits, say the first n.

The idea is simple but not secure in the classical setting because Eve can tamper transmissions and read Alice's tokens and Bob's report.

In the quantum setting the above idea can be implemented in such a way that its security is based on the following basic principle of quantum mechanics.

1. Non-orthogonal states cannot be copied.[4]
2. Any measurement of states can change them, with high probability, irreversibly and can create a significant and detectable rate of errors.

[3]In the literature, for historical reasons, the term "quantum key distribution" (QKD) is mostly used even though the term "quantum key generation" is more logical.

[4]Encodings with a nonorthogonal set of states, for example $\{|0\rangle, |1\rangle, |0'\rangle, |1'\rangle\}$ were first used by Wiesner (1983) to suggest money that could not be counterfeit without detection. The idea is simple, The bank embeds qubits, randomly chosen from a nonorthogonal set into the bank notes and puts notes into circulation, keeping the record which quantum states were embedded into notes. When a note returns to the bank its qubit is meausred, according to the record. If there was no attempt to counterfeit the note, the readout corresponds to the record and, in addition, the embedded state is not disturbed. (The main problem with this idea, not solved yet, is how to store qubits for a long time.)

The quantum version of the basic idea for key generation goes as follows. Alice sends Bob a sequence of photons randomly polarized. Bob measures them using a randomly chosen basis. This necessarily disturbs some of the photons. Any eavesdropping by Eve introduces additional disturbance. After all transmission of photons Alice and Bob determine, in a public communication, the amount (probability) of eavesdropping. If it is not too much, they select from the sent and received photons a shared secret key.

Three QKG protocols are discussed in the following. All of them are both historically and practically of importance.

1. Protocol BB84, or 4-state protocol, with encodings and decodings based on the existence of two non-commuting observables (Bennett and Brassard, 1984). BB84 has been the first fully successful attempt to exploit quantum laws to obtain a fundamental advantage in information processing.

2. Protocol E91 with encodings based on quantum entanglement (Ekert, 1991).

3. Protocols B92, or 2-state protocol, with encodings based on two non-orthogonal states (Bennett, 1992).

6.2.2 Security issues of QKG protocols

The security of QKG protocols is interesting, important and a very complex problem, both practically and theoretically. The ultimate aim is to achieve unconditional security, i.e., to develop QKG protocols that withstand all attacks quantum mechanics allows, according to our current knowledge of it. It is far from easy to determine how secure are particular QKG protocols. Several security claims in the past were found later to be wrong.

Problems concerning security of quantum cryptographic protocols brought up new fundamental questions concerning quantum information processing. For example, how much information can be measured from a quantum system for a given amount of expected disturbance? All such questions are of fundamental importance and not only for quantum cryptography.

Security problem

The aim of a QKG protocol is to allow two parties, Alice and Bob, that share no information initially, to share a secret key (a binary string) at the end.

There are two potential obstacles to overcome. First, the communication channel between Alice and Bob can be noisy and faulty (some photons can get lost). Second, communication during the key extraction phase has to be assumed to be performed "before the eyes" of Eve, who can do her best to achieve that Alice and Bob do not meet their aim. (Eve can try to learn the key Alice and Bob generate, or at least to get some information about it. She can also try to achieve that at the end of the generation protocol Alice and Bob actually do not share the same key.)

We shall not consider here the case that Eve can just disrupt communication between Alice and Bob. We shall consider only the more difficult case that Eve tries to meet her aims without being detected. We shall also not consider the case that Eve tries to alter public communication between Alice and Bob or to pretend to be one of those parties. Alice and Bob can use some authentication protocols for communication, to avoid such an interference.

We shall also consider the worst case, the usual one for cryptography, that Eve knows which protocol is used. All she does not know is the private random keys of the parties.

What are the means Eve can use to achieve her goals? First observe she cannot be only a passive eavesdropper because the key is extracted only from bits Bob receives. She cannot "tap" quantum transmissions. A single photon cannot be split and no quantum system can clone nonorthogonal quantum states. Therefore it would seem that the only thing Eve can do is to measure transmitted states (according to one of the observables Bob uses[5]) and then forward to Bob the states she gets as results of her measurement. In such a case, as already discussed, she has a 25% chance to make an error. (As a consequence, if Alice sends n bits and Eve measures all, then there is only a $(\frac{3}{4})^n$ chance that there will be no error introduced by her. For the case $n = 100$ there is therefore probability only $3 \cdot 10^{-13}$ that no error will be introduced by Eve.) However, this is not the whole story. There are various attacks/measurements she can make.

Several types of attack have been identified so far. The two extreme types are the following ones.

1. **Intercept–reset attacks**. Eve tries to learn as much as possible from particular transmissions of Alice, qubit by qubit (photon by photon) using von Neumann measurements.

2. **Coherent** or **joint attacks**. Instead of measuring the particles while they are in transit from Alice to Bob, one–by–one, Eve regards all the transmitted particles as a single entity. She then couples this entity with a simple auxiliary system (ancilla), prepared in a special state, and creates the compound system. Afterwards, she sends the particles to Bob and keeps the ancilla. After the end of the public interactions between Alice and Bob (for error detection, error correction and privacy amplification), Eve extracts from her ancilla some information about the key. Such attack are directed against the final key. They represent the most general type of attacks that is possible. (However, no particular attack of this type has been suggested so far.)

Error rate

In practice eavesdropping is not the only source of errors in transmission. Imperfections of source, channels and detectors may also produce errors, usually up to a few per cent. The number of such errors, as a fraction of the total number of detected bits, is called the **quantum bit error rate**, and it is one of the parameters that characterizes how well a transmission system works.

When a noisy channel is used to transmit quantum states the problem is to detect eavesdropping. One way out for Alice and Bob is first to calculate the likely error rate caused by a noisy channel, and then to consider the real error rate to be suspicious if it is higher than estimated. Of course this is not a very secure method. Usually the best is to assume the worst case, namely that all errors are due to an eavesdropper.

Test for eavesdropping and errors

What kind of tools do Alice and Bob have to detect an eavesdropping or/and errors? They can, for example, measure the key generation rate and the error rate in a portion of extracted keys and they can make some tests of randomness of the extracted key (for example 0/1 bias). However, in order to make a mathematical proof of unconditional security of some QKG protocols a much more general view of testing methods they can use is needed. This will be discussed later in this chapter.

[5]It is assumed that Eve knows the protocol and therefore observables Bob can use.

Secure key generation

The basic method is straightforward.[6] From the set of extracted bits Bob chooses, randomly, a subsequence, sufficiently large, and makes it public. Alice compares these bits with her and if the error rate is too large, they conclude that an eavesdropping took place and they repeat the key generation protocol. Otherwise, the remaining part of the extracted bits form the key they generate. Details of the method depend on the degree of security Alice and Bob want, or need, to achieve.

Error correction

One way to deal with the problems of noisy channels and faulty detectors is for Alice to first encodes the sequence she wants to transmit using an error correction code.

Privacy amplification

Privacy amplification is a tool developed by Bennett, Brassard and Robert (1988) to select a short and very secret binary string s from a longer but less secret string s'.

The main idea is simple. If $|s| = n$, then one picks up n random subsets $S_1, \ldots, S_n$ of bits of s' and let s_i, the ith bit of S, be the parity of S_i. One way to do it is to take a random binary matrix of size $|s| \times |s'|$ and to perform multiplication Ms'^T, where s'^T is the binary column vector corresponding to s'.

The point is that even in the case where an eavesdropper knows quite a few bits of s', she will have almost no information about s.

More exactly, if Eve knows parity bits of k subsets of s', then if a random subset of bits of s' is chosen, then the probability that Eve has any information about its parity bit is less than $2^{-(n-k-1)}/\ln 2$.

Of particular importance is the case that a linear error correcting code is used to encode at first the transmitted sequence and then a syndrome of it is distributed by public channel. In such a case if Eve knows already t bits of s', and if no more than $n-t-r-1$ bits are given to Eve as the syndrome of s', where r is a security parameter, then the expected amount of information Eve knows on the parity of a random subset of bits of s' is less than $\frac{2^{-r}}{\ln 2}$.

6.2.3 Quantum key generation protocols BB84 and B92

Protocols BB84 and B92 will be described in terms of Hilbert space constructs and illustrated using transmissions of single photons randomly polarized. Protocols BB84 and B92, for generation of a key of length n, have several phases:

Preparation phase

BB84 protocol. Alice generates independently two private random binary sequences of length $m \gg n$ bits and Bob generates one such private sequence of random bits.

B92 protocol. Both Alice and Bob generate their two private random binary sequences of length m.

Quantum transmission phase

[6] We assume again that Alice and Bob need an n bit key but Alice starts the protocol by transmitting $m \gg n$ photons.

Assumptions: Alice is assumed to have four transmitters of photons in one of the following four polarizations—$0, 45, 90$ and 135 (or -45) degrees—in the case of BB84 protocol (see Figure 6.1a); and in one of two polarizations—90 and 135 degree—in the case of B92 protocol (see Figure 6.1b).[7]

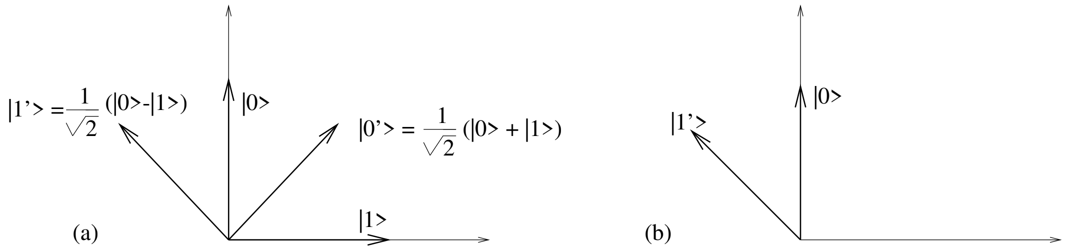


Figure 6.1: Polarizations of photons for BB84 and B92 protocols

In accordance with the laws of quantum physics Bob has a detector that can be set up to distinguish between rectilinear polarizations (0 and 90 degrees) or can be quickly reset to distinguish between diagonal polarizations (45 and 135 degrees). However, again in accordance with the laws of quantum physics, there is no detector that could distinguish between unorthogonal polarizations. In a more formal setting, Bob can use either the standard observable $\mathcal{B} = \{|0\rangle, |1\rangle\}$ or the dual observable $\mathcal{D} = \{|0'\rangle, |1'\rangle\}$, discussed on page 44, to measure the incoming photon.

Transmissions

BB84 protocol: To send a bit 0 (1) of her first random sequence through a quantum channel[8], Alice chooses, on the base of her second random sequence, one of the encodings $|0\rangle$ or $|0'\rangle$ ($|1\rangle$ or $|1'\rangle$), i.e., the standard or dual basis, and sends the photon of the corresponding polarization.

It is assumed here and in the following that photons are sent one by one in regular intervals. As a consequence Bob knows when some photon does not get through and the order index of all received bits.

Bob chooses, each time on the base of his private random sequence, one of the observables $\mathcal{B}$ or $\mathcal{D}$ to measure the photon he is to receive. Bob records the results of his measurements and keeps them secret. Observe that there are three situations Bob can encounter: photon is not received, Bob uses the correct basis (with respect to Alice's choice) for his measurement, and Bob uses the incorrect basis for measurement.

Figure 6.2 shows the possible results of the measurements and their probabilities. When Bob guessed correctly the polarization chosen by Alice he obtained for 100% the same bit as Alice had sent. However, when Bob failed to guess polarization, and did not used the corresponding observable, he obtained the correct result only with probability $\frac{1}{2}$. An example of an encoding–decoding process is in the Figure 6.3.

[7]Expressed in a more general form, Alice uses for encoding states from the set $\{|O\rangle, |1\rangle, |0'\rangle, |1'\rangle\}$a in the case of the BB84 protocol and staes from the set $\{|0\rangle, |1'\ rangle\}$ in the case of the B92 protocol.

[8]Quantum channel is a transmission medium that isolates the quantum state from interactions with the environment.

Alice's encodings	Bob's observables	Alice's state relative to Bob	the result and its probability	correctness
$0 \rightarrow \|0\rangle$	$0 \rightarrow \mathcal{B}$	$\|0\rangle$	0 (prob. 1)	correct
	$1 \rightarrow \mathcal{D}$	$\frac{1}{\sqrt{2}}(\|0'\rangle + \|1'\rangle)$	0/1 (prob. $\frac{1}{2}$)	random
$0 \rightarrow \|0'\rangle$	$0 \rightarrow \mathcal{B}$	$\frac{1}{\sqrt{2}}(\|0\rangle + \|1\rangle)$	0/1 (prob. $\frac{1}{2}$)	random
	$1 \rightarrow D$	$\|0'\rangle$	0 (prob. 1)	correct
$1 \rightarrow \|1\rangle$	$0 \rightarrow \mathcal{B}$	$\|1\rangle$	1 (prob. 1)	correct
	$1 \rightarrow \mathcal{D}$	$\frac{1}{\sqrt{2}}(\|0'\rangle - \|1'\rangle)$	0/1 (prob. $\frac{1}{2}$)	random
$1 \rightarrow \|1'\rangle$	$0 \rightarrow \mathcal{B}$	$\frac{1}{\sqrt{2}}(\|0\rangle - \|1\rangle)$	0/1 (prob. $\frac{1}{2}$)	random
	$1 \rightarrow D$	$\|1'\rangle$	1 (prob. 1)	correct

Figure 6.2: Quantum cryptography with BB84 protocol

1	0	0	0	1	1	0	0	0	1	1	Alice's random sequence
$\|1\rangle$	$\|0'\rangle$	$\|0\rangle$	$\|0'\rangle$	$\|1\rangle$	$\|1'\rangle$	$\|0'\rangle$	$\|0\rangle$	$\|0\rangle$	$\|1\rangle$	$\|1'\rangle$	Alice's polarizations
0	1	1	1	0	0	1	0	0	1	0	Bob's random sequence
$\mathcal{B}$	$\mathcal{D}$	$\mathcal{D}$	$\mathcal{D}$	$\mathcal{B}$	$\mathcal{B}$	$\mathcal{D}$	$\mathcal{B}$	$\mathcal{B}$	$\mathcal{D}$	$\mathcal{B}$	Bob's observable
1	0	R	0	1	R	0	0	0	R	R	outcomes

Figure 6.3: Quantum transmissions in the BB84 protocol—R stands for the case that the result of the measurement is random

B92 protocol. Alice uses encodings $0 \rightarrow |0\rangle$ and $1 \rightarrow |1'\rangle$ and sends each bit by a photon in one of the two nonorthogonal states.

Bob chooses, on the base of his random sequence, observable $\mathcal{D}$ for 0 and $\mathcal{B}$ for 1, and checks whether the photon he has received was polarized as $|0'\rangle$ or $|1\rangle$.[9] He records the results of his measurements and keeps them secret.

Alice's bit	Alice's encoding	Bob's bit	Bob's test for	Test's result and probability	Correctness
0	$\|0\rangle$	0	$\|0'\rangle$	Yes/No (prob. $\frac{1}{2}$)	random
		1	$\|1\rangle$	No (prob. 1)	correct
1	$\|1'\rangle$	0	$\|0'\rangle$	No (prob. 1)	correct
		1	$\|1\rangle$	Yes/No (prob. $\frac{1}{2}$)	random

Figure 6.4: Encodings/decodings with B92 protocol

Table 6.4 shows the possible results of the measurements and their probabilities. Where the corresponding bits of their random sequences are different, the test fails with probability 1. Otherwise, it fails with probability $\frac{1}{2}$. An example of an encoding/decoding procedure is in Figure 6.5.

Exercise 6.2.2 *Could we modify B92 protocol in such a way that Bob makes test not for $O'\rangle$ and $1\rangle$ but for $|0\rangle$ and $|1'\rangle$ or for $|0\rangle$ and $|1\rangle$?*

[9] Yes, this way, this is not an error. It has a perfect sense.

1	0	0	0	1	1	0	0	0	1	1	Alice's random seq.
$\lvert 1'\rangle$	$\lvert 0\rangle$	$\lvert 0\rangle$	$\lvert 0\rangle$	$\lvert 1'\rangle$	$\lvert 1'\rangle$	$\lvert 0\rangle$	$\lvert 0\rangle$	$\lvert 0\rangle$	$\lvert 1'\rangle$	$\lvert 1'\rangle$	Alice's polarizations
0	0	1	1	0	1	1	0	0	1	0	Bob's random sequence
$\mathcal{D}$	$\mathcal{D}$	$\mathcal{B}$	$\mathcal{B}$	$\mathcal{D}$	$\mathcal{B}$	$\mathcal{B}$	$\mathcal{D}$	$\mathcal{D}$	$\mathcal{B}$	$\mathcal{D}$	Bob's observables
No	R	No	No	No	R	No	R	R	R	No	outcomes of test

Figure 6.5: Quantum transmissions within B92 protocol

Basic key extraction phase

BB84 protocol: Bob makes public the sequence of observables he used to measure the photons he received—but not the results of the measurements —and Alice tells Bob, through a classical channel, in which cases he has chosen the same basis for observable as she did for encoding. The corresponding bits then form the basic key both parties agree on.

B92 protocol: Those bits for which Bob's tests don't pass he takes as the key being extracted and reports their positions to Alice through a public channel.

B92 protocol is simpler because only two polarizations of photons are used and that is why this protocol is sometimes said to be the "minimal protocol" for QKG.

The basic description of the BB84 and B92 protocols is by that finished. In the following we describe a more involved test for errors and a more involved protocol for key extraction in the case of the BB84 protocol. This will be used later when discussing unconditional security of the BB84 protocol. The basic ideas presented below (see Mayers and Yao, 1998), are of a general importance for making quantum key generation protocols more robust to the noise of various types.

Test for errors

Denote by $A_{\mathcal{B}}$ and $B_{\mathcal{B}}$ vectors of the bases Alice and Bob have chosen, respectively, and by A_b (B_b) the sequences of bits Alice sent (Bob got as the outcomes of his measurements).

The test is done with respect to the following two parameters Alice and Bob had to agree on before the transmission:

δ —tolerated error rate,

β —key extraction parameter.

1. Bob makes public the vector $B_{\mathcal{B}}$ (but not B_b). Alice lets Bob know the set $T = \{i \mid A_{\mathcal{B}}[i] = B_{\mathcal{B}}[i]\}$ of those cases where Bob chose the correct basis (and therefore he should get the same bit Alice sent).

2. Bob chooses randomly a set R of $\frac{n}{2}$ indices $i \in [1, 2, \ldots, n]$ and makes public the set $\{(i, B_b[i]) \mid i \in T \cap R\}$. Alice verifies whether the number of positions $i \in T \cap R$ such that the number of $A_b[i] \neq B_b[i]$ is smaller than $\delta\frac{|T|}{2}$. If not, they stop the key generation process because of the suspiciously large rate of errors. Otherwise, they continue to find out whether there are still enough bits to use for key extraction.

3. Alice and Bob verify whether $|T \cap R| \geq \bar{n} = (\frac{1}{4} - \beta)n$. If this is not the case, the protocol is stopped. The parameter β is needed to make sure, using Chernoff's bound[10] that $|T - R| \geq \bar{n}$ with probability larger than $(1 - e^{-2\beta^2 n})$.

4. If $|T - R| \geq \bar{n}$, a set E of size $\bar{n}$ is randomly chosen from the set $T - R$.

Final key extraction

The set E is used to extract the key.[11] To do that two procedures are used: one for error correction, the second for privacy amplification.[12] The procedure for error correction is needed to correct errors among bits in the strings $A_b[E] = \{A_b[i] \mid i \in E\}$ and $B_b[E] = \{B_b[i] \mid i \in E\}$. The privacy-amplifying procedure is used to reduce the chance that Eve could make use of the few bits she could learn.

When Alice uses a linear code to encode her randomly chosen sequence of bits, she computes and sends to Bob the syndrome $s = F \cdot A_b[E]$, where F is an $r \times \bar{n}$ parity check matrix—see Section 7.4.1—of the chosen linear code. Bob uses the syndrome s to correct errors in $B_b[E]$ to get exactly $A_b[E]$. As the outcome, Alice and Bob share a common key $A_b[E]$.

As the last step, a privacy-amplifying procedure is used. For example, a random binary matrix K of dimension $m \times \bar{n}$ is used and $K \cdot A_b[E]$ is computed to yield the final common key of the length m.

Efficiency improvements

Protocols BB84 and B92 were presented above in their most basic form. Several generalizations, modifications and improvements have already been developed (see Brassard, 1994, and Brassard and Crépeau, 1996, for older references). For example, B92 protocol can be based on any two nonorthogonal states $\cos\theta|0\rangle + \sin\theta|1\rangle$ and $\cos\theta|0\rangle - \sin\theta|1\rangle$. Bruß (1998) explored the security and efficiency of a generalized BB84 protocol in which three bases have been used: classical, dual and circular.

A modification of the BB84 protocol that can almost double its efficiency was developed by Ardehali et al. (1998). The basic idea behind such an increase of efficiency is very simple and will now be presented.

In the original BB84 protocol, as presented above, both Alice and Bob choose their bases with equal probabilities. As a consequence in about 50% of the cases Bob uses a different polarization than Alice and therefore about 50% of polarized photons are discarded.

Two modifications were suggested in the protocol developed by Ardehali et al. (1998).

To select the basic key Alice chooses her polarizations with probabilities p and $1 - p$ and Bob with probabilities p' and $1 - p'$, $0 < p, p' < 1$. If $p \neq \frac{1}{2} \neq p'$ this creates a larger probability that the same basis is used for encoding and for measurement. This way efficiency can be in the limit doubled, to approach 100%.

[10] **Chernoff's bound**. Let $X_1, \ldots, X_n$ be independent Bernoulli variables and $S = \sum_{i=1}^n X_i$. If $Pr(x_i = 1) = p_i$ for $1 \leq i \leq n$, then for all $0 < \varepsilon \leq 1$ it holds: (a) $Pr(S - pn \leq -\varepsilon n) \leq e^{\frac{-n\varepsilon^2}{2n}}$; (b) $Pr(S - pn \leq -\varepsilon n) \geq e^{\frac{-n\varepsilon^2}{3n}}$.

[11] Observe that key is extracted only from those bits Bob received. This fact makes such QKG protocols robust against losses during transitions.

[12] Error correction is used mainly to protect transmissions against errors due to imperfections of the quantum channel and Bob's detectors. In such a case Alice does not transmit to Bob a randomly chosen sequence of bits but Alice first encodes her random sequence of bits using a linear error correcting code both Alice and Bob agree on beforehand.

To make the protocol secure a refined error analysis is used. Instead of creating one sequence of agreed-upon bits and computing a single error rate, two sequences are created: one if both of them use the rectilinear polarization; the other for the case where they use the diagonal polarization. The error rate is considered small if it is small for both of these sequences.

6.2.4 Multiparty key generation

There are various applications at which several parties need to possess a secret key. One approach to such a multiparty key generation, that works even in the case of one unreliable party, due to Hillery, Bužek and Berthiaume (1998), will now be presented.

At first the story: Alice, Bob and Charles want to generate a common secret key and Alice knows that one of the other two is not to be trusted, but she does not know which one. In addition, she knows that if the other two act together then the honest one can make the second one behave properly.

Assume that Alice, Bob and Charles share three particles P_a, P_b and P_c in the state $|\psi\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$ and each of them measures her (his) particle with respect to the randomly chosen basis (either the standard basis $\{|0\rangle, |1\rangle\}$, or the dual basis $\{|0'\rangle, |1'\rangle\}$ or the circular basis $\{|0''\rangle, |1''\rangle\}$). Since $|\psi\rangle$ can be expressed in any of the following four ways:

$$\begin{aligned}
|\psi\rangle &= \frac{1}{\sqrt{8}}((|0'\rangle|0'\rangle + |1'\rangle|1'\rangle)(|0\rangle + |1\rangle) + (|0'\rangle|1'\rangle + |1'\rangle|0'\rangle)(|0\rangle - |1\rangle)) \\
&= \frac{1}{\sqrt{8}}((|0'\rangle|0''\rangle + |1'\rangle|1''\rangle)(|0\rangle - i|1\rangle) + (|0'\rangle|1''\rangle + |1'\rangle|0''\rangle)(|0\rangle + i|1\rangle)) \\
&= \frac{1}{\sqrt{8}}((|0''\rangle|0'\rangle + |1''\rangle|1'\rangle)(|0\rangle - i|1\rangle) + (|0''\rangle|1'\rangle + |1''\rangle|0'\rangle)(|0\rangle + i|1\rangle)) \\
&= \frac{1}{\sqrt{8}}((|0''\rangle|0''\rangle + |1''\rangle|1''\rangle)(|0\rangle - |1\rangle) + (|0''\rangle|1''\rangle + |1''\rangle|0''\rangle)(|0\rangle + |1\rangle)),
\end{aligned}$$

once Alice and Bob make their measurements of the first two qubits, the state of Charles particle is uniquely determined. If Charles measures his particle, then from the outcome he can deduce whether Alice and Bob used for their measurements the same basis or not. In addition, a knowledge of the bases they used allows him to determine whether their outcomes were the same or not.

Observe also that if Alice and Bob use the same (different) basis for their measurements and Charles uses the dual basis (circular basis), then he gets his outcome with probability 1 (probability $\frac{1}{2}$, i.e. randomly).

Let us now assume that all three of them make public which of the bases they used for their measurements (but not the results of their measurements). If Charles's basis is not the one that provides the correct measurement, as happens 50% of times if he chooses his basis randomly, then they discard the measurements outcomes and repeat the process. Otherwise, if Charles's choice of measurement is a proper one, then from his outcome he knows whether the results of Alice's and Bob's measurements were the same or not.

Moreover, by exchanging the outcomes of their measurements, Bob and Charles can determine the outcome of Alice's measurement and then that result can be taken as specifying the bit all three of them agree on.

However, we know that either Bob or Charles can be dishonest. In order to deal with this problem the announcement of the choices of the bases should be done as follows: First both Bob and Charles send their bases for measurement to Alice and then she sends all of them to both Bob and Charles. This ensures that neither Bob nor Charles, the one who is dishonest, postpones his choice of the bases until he learns which bases were chosen by Alice and the other one.

Exercise 6.2.3 *(a) Verify the first of the following decompositions and complete the second one.*

$$\begin{aligned}|000\rangle + |111\rangle &= \frac{1}{2}(|0\rangle + |1\rangle)(|00\rangle + |11\rangle) + (|0\rangle - |1\rangle)(|00\rangle - |11\rangle) \\ &= \frac{1}{2}(|0\rangle + i|1\rangle)(?...?)\end{aligned}$$

(b) Show that

$$\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = \frac{1}{2}[e^{-i\pi/4}(|0'\rangle|0''\rangle + |1'\rangle|1''\rangle) + e^{i\pi/4}(|0'\rangle|1'\rangle + |1'\rangle|0''\rangle)].$$

(c) Show that in the following cheating scheme for the multiparty key generation, the probability of error is 25%. Bob succeeds to get Charles's particle. Alice measures her particle using either the dual or the circular basis. Bob measures his two particles either in $(|00\rangle \pm |11\rangle)/\sqrt{2}$ basis or in the $(|00\rangle \pm i|11\rangle)/\sqrt{2}$ basis and after that he sends one particle to Charles. Both Bob and Charles measure now their particles with respect to either the dual or the circular basis.

6.2.5 Entanglement-based QKG protocols

In 1991 A. Ekert discovered a new type of QKG protocols, security of which is not based on Heisenberg's uncertainty principle, as in the case of BB84 and B92 protocols, but on the completeness of quantum mechanics. Completeness (see Section 9.1.8), here means that quantum mechanics provides maximum possible information about any quantum system and an eavesdropping can then be seen as "an introduction of some elements of physical reality (see Section 9.1.7), to the measurement" (see Ekert, 1991). To each specific entanglement-based QKG protocol a (generalized) Bell inequality is formed and by a statistical test of this inequality, on the basis of probability statistics on rejected data, it is possible to determine whether eavesdropping took place. (Eve plays here the role of a hidden variable and disturbs the quantum nature of the correlations between probabilities of occurrence of various outcomes for various measurements.)

In the following we first describe a general scheme of entanglement-based QKG protocols, then two examples, and finally we demonstrate how entanglement-based protocols can be reduced to protocols which do not use entangled pairs of particles and security of which is not based on the Bell theorem.

Protocols

1. **Preparation phase.** Alice chooses three vectors $\alpha_1, \alpha_2, \alpha_3$ in H_2 and Bob vectors β_1, β_2 and β_3, also from H_2, such that they have at least one vector in common.

2. **Key generation phase.** A source produces a sequence of maximally entangled states

(say maximally entangled pairs of photons), for example in the state

$$\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle),$$

and sends one particle of each pair to Alice, second to Bob.

Both Alice and Bob measure their particles with respect to one of the randomly chosen vectors of their sets of vectors.

Both Alice and Bob make public sequences of vectors they used for measurements. From the outcomes of measurements obtained using the same vector for measurements they establish the common key.

3. **Test for eavesdropping.** How secure is the above protocol? Eve has no chance to get some information about the key from the particles while they are in transit because there is no information encoded there. She has two possibilities:

   (a) To measure one or both particles on their way from the source to Alice and Bob and by that to disturb the protocol and to ensure that Alice and Bob do not share at the end a common key.

   (b) To substitute her own, carefully prepared, particles for those generated by the ideal source.

Let us discuss at first the case of eavesdropping. If vectors α_1, α_2 and α_3 and β_1, β_2 and β_3 are properly chosen, then one can design the so-called generalized Bell inequality which can be used to detect eavesdropping. Bell inequality uses as its parameters values $Pr(C, i, b)$–the probability that the party C (Alice or Bob) gets by measuring with respect to the ith vector the value $b \in \{0, 1\}$.

Bell inequality is formed in such a way that it is violated if there is no eavesdropping and satisfied if there is an eavesdropping.

Examples

Example 6.2.4 (Ekert, 1991) *The source emits spin-*$\frac{1}{2}$ *particles in the state* $\frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$*. Alice performs her measurement with respect to the angles* $0°$, $45°$, $90°$ *and Bob with respect to angles* $45°$, $90°$ *and* $135°$.

Denote $E(i, j, b_1, b_2)$ *the probability that if Alice measures with respect to vector* α_i*, Bob with respect to* β_j*, then Alice (Bob) gets as the outcome* b_1 (b_2)*. Let us also denote*

$$E(i, j) = E(i, j, 1, 1) + E(i, j, 0, 0) - E(i, j, 1, 0) - E(i, j, 0, 1).$$

In such a case quantum mechanics requires that for

$$E(1, 1) - E(1, 3) + E(3, 1) + E(3, 3) = S$$

it holds $S = -2\sqrt{2}$ *but if an eavesdropping introduces an element of physical reality, then it holds*

$$-2 \leq S \leq 2.$$

Example 6.2.5 (Phoenix and Townsend, 1995) *The source emits pairs of photons linearly polarized in the state* $\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$, *where* $|0\rangle$ $(|1\rangle)$ *corresponds to the vertical (horizontal) polarization. Both Alice and Bob perform measurement with respect to the bases corresponding to polarizations* $\{0°, 90°\}$, $\{30°, 120°\}$ *and* $\{60°, 150°\}$.

Denote by $D(i, j)$ *the difference between probabilities of obtaining a different outcome and the probability to obtain the same outcome if Alice uses* i*th basis and Bob uses* j*th basis for measurement and let*

$$B = 1 + D(2, 3) - |D(1, 2) - D(1, 3)|.$$

From quantum mechanics one derives $B = -\frac{1}{2}$ *and Bell inequality is* $B \geq 0$.

Entanglement protocols seem to be very different from BB84 and B92 protocols. However, this is not really so. Bennett, Brassard and Merman (1992) showed that a simplified version of such a protocol, in which both Alice and Bob randomly choose for their measurement 0° or 90°, is actually equivalent to the BB84 protocol. In addition, as shown by Barnett and Phoenix (1993) and Phoenix and Townsend (1995), and as will be illustrated below, one can have with only one particle a protocol equivalent to an entanglement-based protocol which has the same level of security.

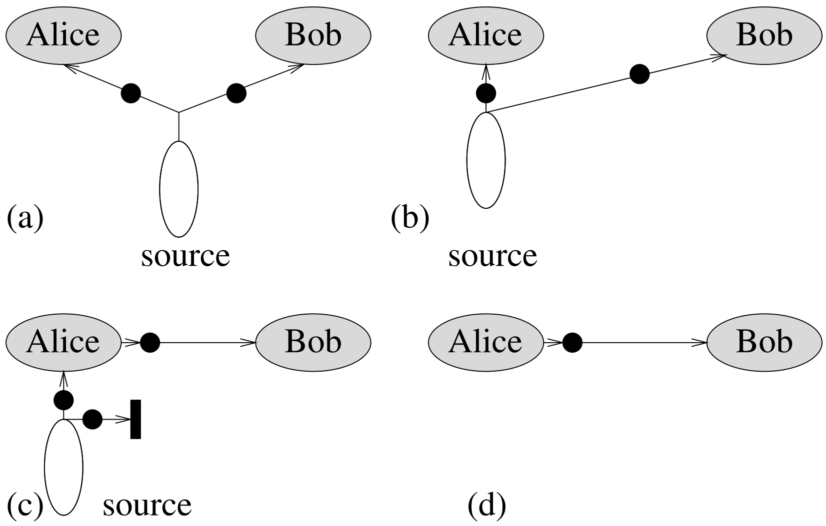


Figure 6.6: Entanglement-based QKG protocols without entanglement

The basic setting discussed above is illustrated in Figure 6.6a. A far-away source of the maximally entangled particles sends one photon of each pair to Alice and the second to Bob, and they perform their measurements. The same protocol can be used and the same claim about security and Bell inequality holds, if the source of photons is in Alice's environment (see Figure 6.6b), or if the second particle is not sent to Bob, (see Figure 6.6c), but after Alice's measurement she makes a copy of the to-be-Bob particle in the corresponding state (Alice knows this state after her measurement) and sends it to Bob who performs a measurement on it. Again the same holds about Bell inequality and its validity as in the original case. Now it is evident that Alice actually does not need at all entangled particles. She can randomly choose one of the six possible states Bob's original particle could be in[13] and she can send the particle in such a state to Bob—see Figure 6.6d.

Similarly as in the case of BB84 and B92 protocols, the question arises whether one can have entanglement-based protocol secure even in the case of noise and eavesdropping. Entanglement purification technique introduced in Section 8.3.2 is a way to deal with these problems.

6.2.6 Unconditional security of QKG*

Once the first QKG protocol was designed, the problem of unconditional security of QKG protocols was approached by a variety of authors and with respect to a variety of assumptions concerning the tools the key generating parties and also the eavesdroppers can use. Step-by-step proofs covered a larger and larger range of situations. This process culminated with the so-called unconditional security proof due to Mayers (1998) and Mayers with Yao (1998). We discuss here the main ideas, methods and results of their approach.

Concerning the environment, it is assumed that the initial coding into quantum channel is perfect, i.e., Alice has a perfect source of photons. The case of a less perfect source will be discussed later. Concerning the detector, it is only assumed that whether or not the received system (photon) is detected is independent of the observable used to measure the system.

Security criteria

Two security criteria are considered: the **privacy criterion** and the **security against tampering criterion**. Privacy means that the eavesdropper cannot learn the key, no matter what she does. More exactly, that the eavesdropper is able to obtain only negligible information (less than one bit), about the final key. Security against tampering means that the eavesdropper cannot make Alice and Bob believe that they share a secret key if they do not.

Usually, Alice and Bob perform a test and decide accordingly whether or not they share a secret key. If the result of the test is negative the key generation process is repeated. A very general view of the test will be used in the following —anything Alice and Bob consider to decide whether or not they share a secret key. Denote by $\mathcal{P}$ such a test.

In the formal definitions of the security criteria we use the concept of a quantity Q_n which is **exponentially small**. By that is meant that there are $c, \varepsilon > 0$, such that $Q_n \leq c2^{-\varepsilon n}$ for almost all n. (n will be used in this context as a security parameter. For example, the number of qubits to be transmitted.)

Definition 6.2.6 *A key generation protocol is secure against tampering if the joint probability that each test P defined above is fulfilled and the keys of Alice and Bob are different is exponentially small.*

The basic idea of the privacy criterion defined below is that whatever Eve's attack is, information i she can gather is such that either the test P is not passed or i has negligible value.

To be more formal, let $\mathbf{k}$ be the random variable values of which are potential keys and $\mathbf{i}$ the random variable each value of which is an information i Eve can obtain either from the public communication or through her measurements of the transmitted qubits. For a

[13] As the consequence of the possibility for Alice to choose one of the three measurement bases.

particular key k, let $Pr(k)$ be the probability Alice and Bob generate the key k and let $Pr(i|k)$ be the probability Eve gathers information i if the key generated has the value k. The formal definition is based on the concept of the "noninformative information" of Eve.

Definition 6.2.7 *Let $0 \leq \theta < 1$. The information i of Eve is θ-noninformative if there exists a real number p such that for every key k*

$$|Pr(i|k) - p| \leq \theta Pr(k). \tag{6.1}$$

In addition, $N_\theta(i)$ will denote the event that the information i of Eve is θ-noninformative.

The basic idea behind the inequality 6.1 is that the difference $Pr(i|k) - p$ should be small. However, $Pr(i|k)$ itself is small and therefore it would not be sufficient to ask only that $|Pr(i|k) - p|$ is small. It is needed that this difference is small even with respect to small $Pr(k)$.

Definition 6.2.8 *A key generation protocol is secure with respect to privacy, if there are two exponentially small positive real numbers γ and θ such that $Pr(P \cap \overline{N_\theta(i)}) \leq \gamma$.*

Informally, Definition 6.2.8 says that except with the probability θ, Eve's activity cannot be such that the test P pass and Eve obtains "informative" information.

A variety of other security criteria have been investigated. Those presented above seem to be strong enough. They take into consideration all known types of attacks including coherent attacks. The development of these criteria has been a significant step that allows to study the idea of "unconditional security" of QKG protocols.

Security requirements

It has been shown, as discussed below, that QKG protocol BB84 is secure and its security holds:

1. against all attacks of an eavesdropper allowed by quantum mechanics—this is the so-called **unconditional security**;
2. against imperfections of the channel and of the detectors.

Main result

In order to illustrate some proof methods we show first security against tampering of a special version of BB84 protocol with the following two properties.

1. The set T has size exactly $\frac{n}{2}$. (This can be achieved if Bob first stores all received photons and only after Alice announces $A_{\mathcal{B}}$, Bob chooses $B_{\mathcal{B}}$ such that the set T has the size $\frac{n}{2}$).
2. When the size of $T - R$ is not large enough, the protocol does not stop and the error correction is done on all bits with indices in $T - R$.

Lemma 6.2.9 *For every $\varepsilon > 0$ and $\delta > 0$, it holds*

$$Pr(P \wedge \bar{\mathcal{E}}) \leq \mu(\varepsilon),$$

where $\mathcal{E}$ is the event that the Hamming distance on bits with indices in $T - R$ between A_b and B_b is smaller than $(\delta + \varepsilon)\frac{n}{4}$ and

$$\mu(\varepsilon) = e^{\frac{-n\varepsilon^2}{32(\delta+\frac{\varepsilon}{2})}}.$$

Proof is based on the fact that the set R is random and remains secret until Alice's quantum transmissions are over. Eve is therefore not able to distinguish between the sets $R \cap T$ and $R - T$ when an error is created. We can therefore consider as fixed the sequences $A_{\mathcal{B}}$, $B_{\mathcal{B}}$, A_b and B_b, and therefore also T. Let $e = A_b \oplus B_b$.

The rest of the proof is based on the fact that the weight of $e(T)$ (see Section 7.4.1), namely $w(e(T))$, is either larger or smaller than $(\delta + \frac{\varepsilon}{2})\frac{n}{2}$. In the first case, the probability that P holds is small. In the second case, the probability that $\bar{\mathcal{E}}$ holds is small. In both cases, the probability that $P \wedge \bar{\mathcal{E}}$ holds is small, as required.

Case 1. $w(e(T)) \geq (\delta + \frac{\varepsilon}{2})\frac{n}{2}$. The test P was defined in such a way that P holds only if $w(e(T \cap R)) \leq \frac{\delta n}{4}$. R is constructed in such a way that every $i \in T$, and also every i with $A_b[i] \neq B_b[i]$, belongs to $T \cap R$ with probability $\frac{1}{2}$. Therefore, each of the $(\delta + \frac{\varepsilon}{2})\frac{n}{2}$ errors is in $T \cap R$ with probability $\frac{1}{2}$. According to Chernoff's bound (see footnote, page 226), the number of errors in $T \cap R$ is smaller than

$$\left[\frac{1}{2} - \frac{\varepsilon}{4(\delta + \frac{\varepsilon}{2})}\right](\delta + \frac{\varepsilon}{2})\frac{n}{2} = \frac{\delta n}{4}$$

with the probability less than $\mu(\varepsilon)$.

Case 2. $w(e(T)) < (\delta + \frac{\varepsilon}{2})\frac{n}{2}$. In this case it follows from Chernoff's bound that the number of errors in $T - R$ is greater than

$$\left[\frac{1}{2} + \frac{\varepsilon}{4(\delta + \frac{\varepsilon}{2})}\right](\delta + \frac{\varepsilon}{2})\frac{n}{2} = (\delta + \varepsilon)\frac{n}{4}.$$

$\Box$

Lemma 6.2.9 sounds very technical. Less formally it says that only with the exponentially small probability it may happen that the test P succeeds on $R \cap T$ and yet the number of errors in $T - R$ is greater than $(\delta + \varepsilon)\frac{n}{4}$. This further implies that if an adequate error-correcting technique is used, then Eve cannot succeed to make Alice and Bob believe that they share the key if they do not.

The main result about security of BB84 protocol, proven by Mayers and Yao (1998a) has the following form.

Theorem 6.2.10 *Let $\varepsilon, \varepsilon' > 0$ and*

$$\gamma = \left[\frac{e^{-4\beta^2 n}}{1 - e^{-2\beta^2 n}} + 2e^{\frac{(4\beta-1)n\varepsilon^2}{32(\delta+\varepsilon/2)}}\right]^{\frac{1}{2}}.$$

Security against tampering*: The probability that the test P passes and there is more than $(\delta + \varepsilon)\bar{n}$ errors in E is smaller than γ^2.*

Security with respect to privacy*: If*

$$m + r \leq [1 - H(2(\delta + \varepsilon)) - \varepsilon'] \cdot \bar{n},$$

where $H(x)$ is Shannon entropy of x and the matrix K (used for privacy amplification), is random, then with probability at least $1 - 2^{-\varepsilon'\bar{n}}$ we get K such that

$$I(\mathbf{i}, \mathbf{k}') \leq \frac{2\eta}{\ln 2} + m\gamma,$$

where $\mathbf{k}' = \mathbf{k} \oplus \mathbf{w}$ and w is a random string chosen by Alice and announced to Bob, $I(\mathbf{i}, \mathbf{k}')$ is mutual information and $\eta = 2^m(2\sqrt{\gamma} + \gamma)$.

Remark 6.2.11 Theorem 6.2.10 is both a significant achievement and a questionable result. Technically, it is correct. The problem is only whether such a complicated result fully justifies its interpretation as "the ultimate proof of unconditional security of the BB84 protocol". Because of its very complex claim and dependence on several parameters, it is unlikely that the result will be fully accepted soon by the whole quantum computing community as a final step concerning the unlimited security of the BB84 protocol. At the same it is unclear whether a significantly simpler, more elegant and equally powerful proof is possible.

Security of the BB84 protocol and the quality of the photon source

Unconditional security of the BB84 protocol was obtained only under the assumption that there is a perfect source of photons. It is known that imperfect sources may seem to behave quite normally and, at the same time, seriously compromise security of the BB84 protocol. The security problem for an imperfect source is a difficult problem to deal with. An interesting and promising step in this direction was done by Mayers and Yao (1998a). They proposed a concrete design for a new concept of a **self-checking source**. It is required, from the manufacturer of a photon source, to provide certain tests designed in such a way that if they pass, then the source is guaranteed to make the BB84 protocol secure.

6.2.7 Experimental quantum cryptography

Experimental QKG has been based so far mainly on the BB84 and B92 protocols. However, Ekert's idea (1991) to use non-locality has also received attention, as a principally different approach to QKG, especially after the development of quantum purification techniques. They allow one to exchange imperfect quantum entanglement through a noisy channel and then to distil from it almost perfect entanglement (see Deutsch et al. 1996, Bennett et al. 1996b). A survey of other approaches to QKG can be found in Brassard and Crépeau (1996).

The first experimental quantum bit transmission by photons, in a tube over a distance of 32 cm, was done in 1989 (see Bennett and Brassard, 1989, and Bennett et al. 1992). A series of longer distance experimental transmissions of qubits by photons followed soon. Marand and Townsend (1995) reported transmission over 30 km of commercial optical fibre in the same room. Muller, Zbinden and Gisin (1995) reported transmissions using optical cable below Lake Geneva over the distance of 22.7 km, with bit error 3.4%. The experiments with outdoor transmissions culminated so far in the transmissions over an outdoor optical path of almost 1 km (Buttler et al. 1998), with error rate 1.5% in night-time. (The key problem is the quality of the transmission and detection of single photons against a high background through a turbulent medium; because of the presence of background photons and difficulties of directing photons into a remote detector, this problem was considered of immense difficulty.) On the base of these experiences QKG between a ground station and a satellite in a low-earth orbit seems to be feasible (Buttler et al. 1998).

Photons seem to be the best medium to carry qubits. They are relatively easy to produce and photons of certain wavelength can be sufficiently reliably transmitted using existing optical fibres. They are also quite well detected. Photons of wavelength 1.3μ m can travel 10 km in a fibre before half of them get absorbed. This should be enough to perform QKG in local networks. (Unfortunately amplifiers cannot be used. This follows from "No-cloning theorem".) However, one could use "quantum repeaters" (see Briegel, 1998).

In the experimental quantum state transmissions two main methods of encoding of quantum states by photons are through polarization of photons (using photons of shorter wavelength) and through photon's phase. (The last set-up was used by BT (British Telecom) (see Phoenix and Townsend, 1995) and in Los Alamos National Laboratory (see Hughes, 1995).)

A practically very useful technique to realize QKG protocols was introduced by Bennett (1992). The technique uses a Mach–Zehnder interferometer. Alice and Bob each controls one of the phase modulators on one of the arms of the interferometer. Encoding and measurements are done by setting the corresponding phase modulator. We describe very briefly the basic ideas only for the BB92 and BB84 protocols. For implementation details see Hughes et al. (1995).

An interferometric implementation of the QKG scheme for B92 protocol is shown in Figure 6.7. Alice has a single photon source that she can use to send photons into a Mach–Zehnder interferometer, in which she controls the phase ϕ_A along one of the optical paths. Bob has a single photon detector at one of the interferometer's outputs and controls the phase ϕ_B along the other optical path. (In Figure 6.7 we indicate the phases corresponding to Alice's and Bob's random bits.) The probability that a photon sent by Alice is detected by Bob is

$$P = \cos^2(\frac{\phi_A - \phi_B}{2})$$

and it depends on both paths. If Alice and Bob use the phase angles $(\phi_A, \phi_B) = (0, \frac{3\pi}{2})$ for their 0-bits, and $(\phi_A, \phi_B) = (\frac{\pi}{2}, \pi)$ for their 1-bits, we get the situation the B92 protocol requires.

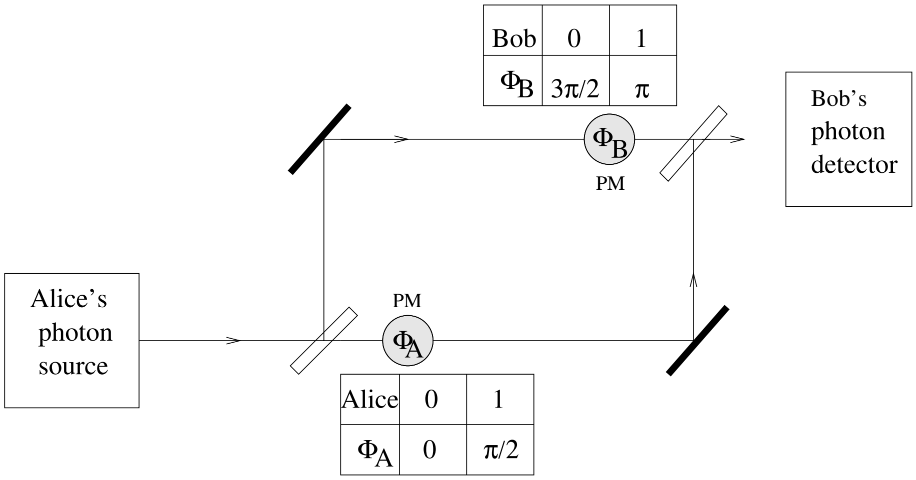


Figure 6.7: B92 protocol implementation

In the case of the BB84 protocol, both Alice and Bob use identical Mach–Zehnder interferometers with one path longer than the other (see Figure 6.8), and a phase modulator (PM) on the shorter path. In order to send a random bit Alice randomly adds, through her PM, a phase shift of 0, $\frac{\pi}{2}$, π or $\frac{3\pi}{2}$, to her photon. Bob can add only phase shifts 0 or $\frac{\pi}{2}$. By adding 0 ($\frac{\pi}{2}$) shift, Bob can detect whether Alice's phase has a phase shift 0 or π ($\frac{\pi}{2}$ or $\frac{3\pi}{2}$). If shifts 0 and $\frac{\pi}{2}$ are interpreted as 1 and other two as 0, we get a basic setting for the BB84 protocol.

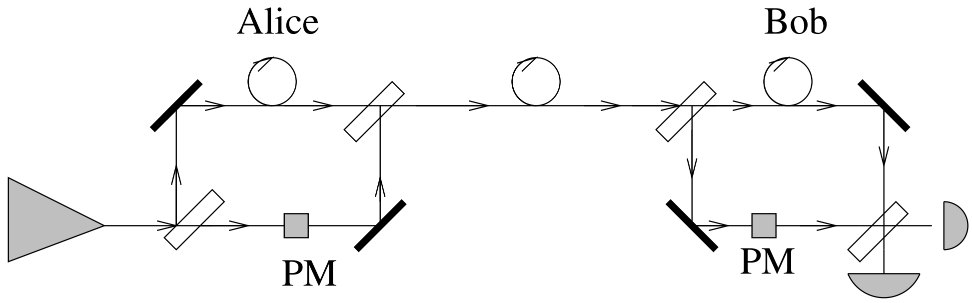


Figure 6.8: Mach–Zehnder interferometer implementation of BB84 protocol, adapted from Tittel et al. (1998a)

Both theoretical and experimental work proceeds also to develop multiuser quantum cryptographical networks. Biham, Huttner and Mor's (1996) approach makes a use of non-locality and assumes a quantum memory is kept in a transition centre to which users "bring their particles". An approach of Phoenix et al. (1995) uses optical networks and has been experimentally tested, see also Phoenix and Townsend (1995).

From experimental to practical quantum cryptography

Progress in two areas seems to be of key importance for quantum cryptography to go from the experimental stage to practical developments: reliability and transmission rate. While the progress in the first area is already significant, transmission rate is still an issue. The highest transmission rate of 1.2 MHz so far exhibited the system tested by BT.

Moreover, in order to succeed for quantum cryptography as a fully competitive method of secure data transmission, it seems that quantum cryptography must be capable of implementation on multiuser optical quantum networks.

6.3 Quantum Cryptographic Protocols

Cryptographic protocols are specifications how two parties, say Alice and Bob, should prepare themselves for a communication and how they should behave during a communication in order to achieve their goal and be protected against an adversary (Eve). It is assumed that parties involved in a protocol know and follow it fully. Two parties participating in a protocol may want to share their secrets in order to compute some value, generate jointly random numbers, convince each other of their identity, simultaneously sign a contract or participate in a secret voting. Cryptographic protocols that accomplish such tasks have radically changed our views of what mutually distrustful parties can accomplish over networks.

A variety of interesting and important classical cryptographic protocols have been designed (see Schneier, 1996, Gruska, 1997). Three of them are of a special importance because they are often the crucial ingredients of more sophisticated protocols.

In the **coin-flipping protocols** , Alice and Bob can flip coin over a distance ("by phone") in such a way that neither of them can determine the outcome of the flip but both can agree on the outcome in spite of the fact that they do not trust each other.

In the **bit commitment protocols** (BCP), Alice can choose a bit and **get committed** to it in the following sense. Bob has no way of knowing what Alice has chosen, and Alice has no way of changing her commitment once she has made it; say after Bob announces his guess as to what Alice has chosen.

Design of a BCP consists of two procedures *commit*(x) and *open*(x) that have the following properties.

If Alice has a bit b in mind to which she wants to commit herself, then she can, through the procedure *commit*(b), provide Bob with a piece of evidence that she has the bit b in her mind to which she commits herself. *commit*(b) should be such that Bob has no way to use it to get b. At a later time, if needed, Alice can use the protocol *open*(b) to reveal x and to prove to Bob that the information she sent, as the consequence of the procedure *commit*(b), really corresponds to that bit.

Alice can choose the probability distribution for b before her commitment process starts. The commitment obtained after the procedure *commit*(b) is realized is said to be **binding** if Alice cannot change the distribution of the probability for b, and it is called **concealing** if Bob cannot obtain any information about b without Alice's help. The commitment is **secure** if it is both binding and concealing. The commitment is **unconditionally secure** if it is secure against any cheater, Alice or Bob, with unlimited computational power.

An example of a "pre-computer era" BCP is that Alice writes her commitment on a piece of paper, locks it in a box, sends the key to Bob and, in the opening phase, she sends also the box to Bob. A "complexity era" solution is that Alice chooses a one-way function f and an integer x, say even for the commitment "1" and odd, otherwise, and then sends to Bob $f(x)$ and f.

There are two types of the oblivious transfer protocols (OTP). In the 1**-out-of-**2 **oblivious transfer protocols** (due to Even et al. 1983), Alice transmits two messages, m_1, m_2, to Bob who can choose whether to receive m_1 or m_2, but cannot learn both, but Alice has no idea which message Bob chosed to receive.

In the **standard oblivious transfer protocols**, in short oblivious transfer protocols, due to Rabin (1981), Alice can send a message to Bob in such a way that Bob receives the message with probability $\frac{1}{2}$ and a garbage (no information) with probability $\frac{1}{2}$. Moreover, at the end Bob knows whether he got a message or garbage, but Alice has no idea which of them Bob has received.

It has been shown by Crépeau (1987) that both types of OTP are equivalent in the sense that either of them can be implemented from a primitive that implements the other.

From the practical point of view, there is one essential difference between QKG protocols and protocols for coin-flipping, bit commitment and oblivious transfer. The first ones are of practical interest only for longer distance transmissions; the other three have perfect sense also for short distances.

In the classical setting using an oblivious transfer protocol one can design a bit commitment protocol and with it a coin-flipping protocol. The reversed reductions are believed to be impossible. OTP are therefore very significant protocols in spite of the fact that at the first glance they look almost like a bizarre idea.

Exercise 6.3.1 *Show how to design a bit commitment protocol once an oblivious transfer protocol is given.*

None of the classical BCP and OTP is absolutely secure. Their computational security is always based on some unproven assumptions of the computability theory.

The history of cryptographic protocols started with the following Blum's coin-flipping protocol (1981):

Protocol 6.3.2 (Coin-flipping by telephone)

1. *Alice chooses two large primes p, q, sends Bob $n = pq$, and keeps p, q secret.*
2. *Bob chooses a random number $y \in \{1, \ldots, \lfloor \frac{n}{2} \rfloor\}$ and sends Alice $x = y^2 \bmod n$.*
3. *Alice computes four square roots $(x_1, n - x_1)$ and $(x_2, n - x_2)$ of x. (Alice can compute them because she knows*[14] *p and q.) Let $x_1' = \min\{x_1, n - x_1\}$, $x_2' = \min\{x_2, n - x_2\}$. Since $y \in \{1, \ldots, \lfloor \frac{n}{2} \rfloor)\}$, either $y = x_1'$ or $y = x_2'$. Alice then guesses whether $y = x_1'$ or $y = x_2'$ and tells Bob her choice (for example, by reporting the position and the value of the leftmost bit in which x_1' and x_2' differ).*
4. *Bob tells Alice whether her guess was correct (head) or not correct (tail).*

Later, if necessary, Alice can reveal p and q, and Bob can reveal y.

Blum's coin-flipping protocol has actually been also the first non-trivial example of the bit commitment protocol. In the field of the classical cryptographic protocols BCP play an important role. For example, with such a protocol one can construct zero-knowledge proofs for a variety of statements.[15]

6.3.1 Quantum coin-flipping and bit commitment protocols

Also the history of quantum cryptographic protocols started with the following coin-flipping protocol by Bennett and Brassard (1984).

Protocol 6.3.3 (Quantum coin-flipping protocol) *—see Figure 6.9*

1. Alice randomly chooses a sequence of bits (for example 1000 should be enough) and a polarization (rectilinear or diagonal—standard or dual). Finally, Alice sends the resulting sequence of the polarized photons to Bob.
2. Bob chooses, for each received photon, randomly, an observable, $\mathcal{B}$ or $\mathcal{D}$, and measures the incoming photon. He records the result into two tables—one for the observable $\mathcal{B}$ and the second for the observable $\mathcal{D}$. Since some photons can get lost during the transmissions, there can be holes in both tables. At the end of all transmissions, Bob makes a guess whether Alice chose rectilinear or diagonal polarization and announces his guess to Alice. He is to win if the guess is correct and to lose otherwise.
3. Alice tells Bob whether he won or lost by telling him the polarization she chose. She can certify her claim by sending Bob the random sequence of bits she chose at Step 1.

[14] See, for example, Gruska (1997).
[15] Actually for the whole class **IP** (see, for example, Gruska, 1997).

Alice's random bits	1	0	1	1	0	1	1	0	0	1	1
Alice polarization choice						rectilinear					
photons sent	$\|1\rangle$	$\|0\rangle$	$\|1\rangle$	$\|1\rangle$	$\|0\rangle$	$\|1\rangle$	$\|1\rangle$	$\|0\rangle$	$\|0\rangle$	$\|1\rangle$	$\|1\rangle$
Bob's observable	$\mathcal{B}$	$\mathcal{D}$	$\mathcal{D}$	$\mathcal{B}$	$\mathcal{B}$	$\mathcal{B}$	$\mathcal{B}$	$\mathcal{D}$	$\mathcal{B}$	$\mathcal{B}$	$\mathcal{D}$
Bob's table for $\mathcal{B}$	1			1			1		0	1	
Bob's table for $\mathcal{D}$		1	1					0			0
Bob's guess of Alice's pol.						rectilinear					
Alice's message						you WON					
Alice's original bits	1	0	1	1	0	1	1	0	0	1	1
comparison with $\mathcal{B}$	Y			Y			Y		Y	Y	
comparison with $\mathcal{D}$		N	Y					Y			N

Figure 6.9: Illustration of a quantum coin-flipping protocol

4. Bob verifies Alice's claim by comparing his records in the table for the basis she claims to choose. There should be a perfect agreement with the entries in that table and no correlation with the other table.

Can Alice or Bob cheat? Bob is not able to cheat. Indeed, he would be able to cheat only in the case he would be able to determine, with probability larger than $\frac{1}{2}$, on the base of the sequence of photons he received, which polarization Alice has chosen. However, it can be shown (see, for example, Glaubner, 1988), that any measuring device capable to distinguish polarization from the stream of incoming photons, could be used, together with the EPR phenomenon, to transmit information faster than light.

Alice has two possibilities to cheat: either at Step 1 or at Step 3. Let us first discuss the second case. The only way she could cheat is by sending a sequence of bits that would match entries of Bob's table for others from the two possible polarizations (in our example for the diagonal one). However, she has no way to find out the results of the measurements Bob made using his observables, because they were just randomly chosen. She can only guess, but the probability of making correct guesses gets exponentially small with length of the transmitted sequence and, consequently, the probability that her cheating could be discovered approaches fast 1.

Alice also cannot cheat at Step 1 by sending a mixture of diagonally and rectilinearly polarized photons. In such a case, almost certainly, she would not be able in Step 3 to ensure that the sequence of bits she sent agrees with any of Bob's tables. However, there is a special way Alice can cheat in Step 1—by making a clever use of the entangled photons.

In Step 1, instead of sending a sequence of photons polarized in one way or another, she produces pairs of polarization-entangled photons, each pair in the state $\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$, sends to Bob one member of each pair, and stores the other one. After Bob announces his guess in Step 2, she measures her photons in the opposite basis as was Bob's guess. By that she receives a sequence of bits perfectly correlated with Bob's table corresponding to the basis he did not choose as his guess in Step 2 and completely uncorrelated with the other table. She then announces the sequence in Step 3.

The cheating Alice can do this way is theoretically perfectly acceptable. However, practical problems with storing entangled photons for a longer period are large. Moreover, every error in this process could result, with high probability, in an error Bob could discover and that would make him, at least, suspicious about the whole process.

Quantum bit commitment protocols

The coin-flipping protocol presented above has actually been the first quantum bit commitment protocol published. An important further step in developing QBCP was protocol BCJL, due to Brassard, Crépeau, Jozsa and Langlois (1993), an improvement to the protocol due to Brassard and Crépeau (1991), which played an important role in our understanding of the problems of security of the cryptographic protocols.

In order to describe the BCJL protocol we need to describe its subprotocols *commit* and *open*. The protocol *commit* uses, as a parameter, an $\varepsilon > 0$—an upper bound on the error rate of the quantum channel. ε is the probability that an encoded bit b is decoded as $\bar{b}$.

Protocol 6.3.4 (*commit*(x))

1. Bob chooses a Boolean matrix G as a generator matrix (see Section 7.4.1), of a binary linear (n, k, d)-code C such that $\frac{d}{n} > 10\varepsilon$ and $\frac{k}{n} = 0.52$ and announces it to Alice.
2. Alice chooses:
   (a) a random string r of length n and announces it to Bob;
   (b) a random k-bit vector s, such that $r \cdot c = x$, where $c = sG$;
   (c) a random sequence b of length n of the polarizations, $\mathcal{B}$ or $\mathcal{D}$, and sends to Bob a sequence of n photons with the polarization of the ith photon $P_{b_i}(c_i)$, where $P_0(0) = 0°$, $P_0(1) = 90°$ and $P_1(0) = 45°$, $P_1(1) = 135°$.
3. Bob chooses a random string b' of n bits and measures the ith photon according to the basis $M(b'_i)$, where $M(0) = \mathcal{B}$ and $M(1) = \mathcal{D}$. Let c' be the n-bit vector where c'_i is the result of the measurement of the ith photon.

Alice keeps the bit x and vectors c and b secret, until the opening takes place, and Bob keeps vectors b' and c' secret.

To open the commitment x, Alice initiates the following protocol:

Protocol 6.3.5 (*open*(c, b, x, c', b'))

1. Alice sends c, b and x to Bob.
2. Bob verifies that c is a codeword of C and computes $B = \sum_{\{i \mid b'_i = b_i\}} \frac{c_i \oplus c'_i}{n/2}$, in order to verify that the error rate is under the limit of those pairs of outgoing and measured bits that were polarized/measured by the same basis.
3. if $B < 1.4\varepsilon$ and $x = r \cdot c$, then Bob accepts, otherwise Bob rejects.

There are two features of this protocol that need an explanation. The use of a linear code C and bounds 10ε and 0.52 on its parameters.

Both C and bounds were chosen in order to be able to prove that the protocol is secure also in the case of a noisy channel. They have been chosen well to show that Bob can obtain only exponentially small amount of Shannon information about b. However, not only the original proof that Alice cannot cheat had a flaw but, as shown in Section 6.3.4, this flaw cannot be corrected. The way Alice can cheat is in principle similar to the one described in the protocol for coin-flipping.

In the BCJL protocol, Alice first chooses a random string r and a codeword c such that $r \cdot c = x$. She sends r to Bob through a classical channel and c by a quantum channel in a similar way as in the case of QKG protocols. As already mentioned, see page 217, this way Bob can obtain correctly roughly 75% bits. We show in Section 6.3.3 that cheating Bob could obtain as much as 85% of bits, but this is the best possible outcome for Bob.

The binary code C was chosen in such a way that there are exponentially many codewords around the vector c'—the result of Bob's measurements. To show that the bound $\frac{d}{n} = 0.52$ was used. This bound can also be used to show that if G is chosen randomly, then it defines, with large probability, an (n, k, d)-linear code.

6.3.2 Quantum oblivious transfer protocols

The so-called "quantum multiplexing" (due to Wiesner, 1983), can be seen as the first protocol close to a quantum oblivious transfer protocol (QOTP).

It is easy to design a simple QOTP which is perfectly secure provided transmissions and detectors are perfect and no party cheats.

Protocol 6.3.6 (Ideal one-photon standard QOTP)

1. Alice chooses a bit b and sends it to Bob through one photon encoded using a randomly chosen basis—standard or dual.

2. Bob measures the photon with respect to a randomly chosen basis—standard or dual.

3. Alice lets Bob know the basis she chosed. □

At the end Bob has a 50% chance to know b for sure and he knows whether he knows b for sure. Alice has no information whether Bob knows the bit for sure.

There are two problems with this protocol.

1. An imperfect Alice's source, or a noisy channel, or a faulty Bob's detector could much affect the probability of success of Bob's measurement.

2. Bob could cheat by making his measurement in Breidbart basis (see page 244, for details). This way he could learn b with large probability $\cos^2 \frac{\pi}{8} \approx 0.85$.

The first nontrivial QOTP was due to Crépeau and Kilian (1988). A more robust version of this protocol (due to Bennett et al. 1991), will now be presented. At first in an idealized form, where polarized photons are used to transmit bits. A more practical version of the protocol will be briefly discussed next.

Let b_0 and b_1 be Alice choices of bits and c be Bob guess.

Protocol 6.3.7 (BBCS—1-out-of-2 QOTP)

1. **Parameters–agreeing phase.** Alice and Bob engage in a public communication concerning the quality of the source, the channel and the detectors and into a process of determining and verifying their parameters. During this process four parameters are determined or chosen.

   (a) ε—the expected error rate

   (b) α—the fraction of photons Bob is expected to detect successfully;

   (c) n—the security parameter specifying the number of photons to be transmitted;
   (d) C—a binary linear error-correcting code capable of correcting, with large probability, n-bit words transmitted with the expected error rate ε.

2. **Transmission phase.** Alice chooses a random binary string of length $\frac{2n}{\alpha}$ and sends each of the bits through a polarized photon using randomly either the standard/rectilinear (0° or 90° degrees) or the dual/diagonal polarization (45° or 135°).

3. **Measurement phase.** Bob measures each incoming photon by a randomly chosen basis—either standard or dual—and records both the bases chosen and the results of the measurements into tables.

   In this step Bob is expected to receive $2n$ photons. In the case he gets more, he ignores those over $2n$. In the case he gets less, he completes the table to $2n$ by chosing the remaining entries randomly. At the end Bob reports to Alice the arrival times of all $2n$ photons—so she knows which photons got through—but neither the bases nor the results of his measurements.

4. **Bases-revealing phase.** Alice tells Bob, through a public channel, the bases she used to encode her random sequence of bits.

5. **Design of good and bad sequences phase.** Bob partitions his $2n$ bits into two sequences, each of length n. Into the "good" sequence he puts as much as possible of bits he obtained when he used the correct basis for measurements. Into the "bad" sequence he puts as much as possible of bits he obtained when using the incorrect basis for measurements.

   Bob then tells Alice "addresses" (or indices) of the particular bits of both sequences, but not which one is good and which bad.

   At this point Bob shares with Alice a binary word of his good sequence (with respect to an expected error rate not greater than ε). Concerning the bad sequence, Bob shares almost nothing with Alice.

   Of course, the above process is not ideal. There can be some errors introduced by the fact that Bob did not use exactly n times the correct basis for his measurement. However, the number of errors introduced this way should be negligible, with respect to the expected noise in the transmissions.

6. **Error-correction phase.** Using the error-correcting code C, Alice computes syndromes of her words corresponding to the good and bad sequence of Bob, and she sends syndromes to Bob. (Alice knows which bits of her correspond to the good and bad sequences of Bob, even she does not know the values of these bits.) Using these syndromes Bob performs error correction on his good and bad sequences.

7. **Privacy-amplification phase.** Alice chooses randomly two subsets of bits, one from her "good" and one from her "bad" sequence and computes their parities. She let Bob know the "addresses" of bits she chose, but not their values. This way Bob can compute the parity of the corresponding subset of his good sequence, but he will have no idea about the parity Alice obtained for the subset of bits corresponding to his bad sequence. Alice knows both, but she has no idea which one Bob knows. Let x_1 and x_2 be parity bits Alice knows and $\bar{c}$ the one Bob knows.

8. **Oblivious transfer phase**

(a) Bob tells Alice whether or not $c = \bar{c}$.

(b) If $c = \bar{c}$, then Alice sends Bob bits $x_0 \oplus b_0$ and $x_1 \oplus b_1$, in this order. If $c \neq \bar{c}$, Alice sends Bob bits $x_0 \oplus b_1$ and $x_1 \oplus b_0$, again in the given order.

(c) Bob computes b_c out of two bits he got from Alice.

Remark 6.3.8 The paper by Bennett et al. (1991) discussed also the ways how to derive and verify parameters mentioned in Step 1 of the above protocol. In addition, it is there assumed that instead of single photons dim light pulses are used for transmission, to have a more realistic setting.

The security of the BBCS protocol will be discussed in Section 6.3.3. The main result shown by Bennett et al. (1991) is:

Theorem 6.3.9 *Let Δ be data Bob obtained by the protocol. At least one of $H(b_0|\Delta, b_1)$ or $H(b_1|\Delta, b_0)$ is exponentially close (in n) to 1; in no case does Alice learn something (where $H(b_0|\Delta, b_1)$ is the conditional Shannon entropy).*

For another improvement of the ideas presented by Crépeau and Kilian (1991) see to Crépeau (1994),

General cryptographic protocols

A two-party cryptographic protocol, called also two-party oblivious circuit evaluation protocol, is a communication algorithm that allows to two parties, say Alice and Bob, to compute $f(x, y)$, for a publicly known function f, in the case Alice knows, but does not want to reveal, x, and Bob knows y, but keeps it secret, in such a way that Alice learns nothing about y and Bob learns nothing about x, except for what can they deduce from their private inputs and the public value of $f(x, y)$.

It has been shown by Kilian (1988) that in the classical setting 1-out-of-2 oblivious transfer protocol can be used to implement any two-way cryptographic protocol securely. In the quantum setting it is not clear whether there is unconditionally secure QOTP and, consequently, it is not clear whether the standard reduction of a general quantum cryptographic protocol to OTP works, concerning the security, in the quantum setting.

Exercise 6.3.10 *Design a quantum protocol* identify(x, y) *for Alice and Bob to decide whether strings x and y are the same provided Alice knows only x and she does not want to reveal it and Bob knows only y and does not want to reveal it. Assume they communicate through: (a) noiseless channel; (b) noisy channel.*

6.3.3 Security of the quantum protocols

The quantum cryptographic protocols presented in the previous section are simple. On the other hand, problems concerning their security (unconditional and provable) are complex and the attempts to solve them have led to new fundamental questions in both quantum mechanics and computing.

The difficulty of the problems lies in the richness, not fully understood yet, of **cheating strategies** quantum mechanics offer. (A party is said to be cheating if it deviates from the protocol while appearing to follow it.)

Security of the BCJL protocol

The results presented below (due to Brassard et al. 1993), refer to the BCJL protocol, but both the results and their proofs are of a broader interest for the verification of the security of cryptographic protocols.

The first result shows that Bob can obtain maximum information about the vector c sent to him through polarization 0° and 45°, by the measurement with respect to the so-called Breidbart basis

$$\mathcal{B}_0 = \{\theta_0, \theta_1\},$$

where

$$\theta_0 = \cos\frac{\pi}{8}|0\rangle + \sin\frac{\pi}{8}|1\rangle; \quad \theta_1 = -\sin\frac{\pi}{8}|0\rangle + \cos\frac{\pi}{8}|1\rangle$$

Lemma 6.3.11 *The measurement with the basis $\mathcal{B}_0$ provides Bob with the maximum possible information about x.*

Proof. Density matrices ρ_0 and ρ_1 describing the mixed states representing bits 0 and 1, respectively, have the form

$$\rho_0 = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|0'\rangle\langle 0'| = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix} \quad \rho_1 = \frac{1}{2}(|1\rangle\langle 1| + |1'\rangle\langle 1'|) = \begin{pmatrix} \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} \end{pmatrix}.$$

Let us now denote by *commit*' the protocol which is like the protocol *commit*, but it has instead of the step 2.c the following step:

> Alice chooses a binary sequence b of length n randomly in such a way that 0 (1) is chosen with probability $\cos^2\frac{\pi}{8}$ ($\sin^2\frac{\pi}{8}$), and sends Bob a sequence of n photons with polarizations $\mathcal{B}_0(c_i \oplus b_i)$.

It is easy to verify that for the protocol *commit*' the density matrices ρ_0' and ρ_1' that describe the quantum mixtures representing the states 0 and 1 are identical to matrices ρ_0 and ρ_1. Namely,

$$\rho_0' = \cos^2\frac{\pi}{8}|\theta_0\rangle\langle\theta_0| + \sin^2\frac{\pi}{8}\langle\theta_1\rangle\langle\theta_1| = \rho_0.$$

and, similarly, $\rho_1' = \rho_1$.

Moreover, if we denote by ρ_c (ρ_c') the density matrix associated with the mixture of pure states used by the procedure *commit* (*commit*') to send c, then

$$\rho_c = \bigoplus_{i=1}^{n} \rho_{c_i} = \bigoplus_{i=1}^{n} \rho_{c_i}' = \rho_c'.$$

Since $|\{c \in C \,|\, c \cdot s = 0\}| = |\{c \in C \,|\, c \cdot s = 1\}| = 2^{k-1}$, the density matrices ρ_0, ρ_1, ρ_0' and ρ_1', describing the quantum mixture of all states sent to Bob to commit to 0 (to 1), have the form

$$\rho_0 = \sum_{\{c\in C\,|\,c\cdot r=0\}} \frac{\rho_c}{2^{k-1}} = \sum_{\{c\in C\,|\,c\cdot r=0\}} \frac{\rho_c'}{2^{k-1}} = \rho_0', \quad \rho_1 = \sum_{\{c\in C\,|\,c\cdot r=1\}} \frac{\rho_c}{2^{k-1}} = \sum_{\{c\in C\,|\,c\cdot r=1\}} \frac{\rho_c'}{2^{k-1}} = \rho_1'.$$

Since mixed states represented by the same density matrices cannot be distinguished by any quantum measurement, the above result implies that Bob is able to get about c and

x the same information in both protocols *commit* and *commit*'. The point now is that the measurement performed in the protocol *commit*' maximizes Bob's information about c (and therefore about x), because in the measurement performed in that protocol Bob gets all information available! Hence the optimal measurement for Bob in protocol *commit* is the same. This implies that no coherent measurement on all photons could provide more information for Bob. □

As the next step we show that even if Bob performs the optimal measurement, he can get only a very little information about x, and therefore he cannot cheat. As the first step we show that the code c' received by Bob must be quite away from the vector c sent by Alice.

Lemma 6.3.12 *Even if Bob performs the optimal measurement, there exists an* $0 < \alpha < 1$ *such that the probability that* $hd(c, c') < \gamma n$*, where* $\gamma = H^{-1}(\frac{1}{2}) = 0.1100279$*, is at most* α^n*.*

Proof. Let us assume the most ideal situation for Bob—a noiseless channel. If Bob performs the optimal measurement, then $Pr(c_i = c'_i) = \cos^2 \frac{\pi}{8}$ and $Pr(c_i \neq c'_i) = \sin^2 \frac{\pi}{8}$. Hence the Hamming distance $hd(c, c')$ is expected to be $\sin^2 \frac{\pi}{8} n \approx 0.14644n$. In order to estimate the probability that the number of errors will be less than γn, we use Bernstein law of large numbers[16] as follows: $hd(c, c') = \sum_{i=1}^{n} x_i$, where $x_i = c_i \oplus c'_i$ and $Pr(x_i = 1) = \sin^2 \frac{\pi}{8}$. Hence the probability that $hd(c, c') < \gamma n$ can be estimated as follows

$$Pr(\sum_{i=1}^{n} \frac{x_i}{n} \leq \gamma) \leq Pr(|\sum_{i=1}^{n} \frac{x_i}{n} - \sigma^2| \geq \sigma^2 - \gamma) \leq 2e^{-n(\sigma^2-\gamma)^2} \approx 2e^{-0.001326},$$

and $2e^{-0.001326} < \alpha^n$, for an $\alpha < 1$ and an n sufficiently large. □

The main result about Bob's chance to learn the commitment bit has now the form:

Theorem 6.3.13 *Even if Bob knows the Hamming distance* $hd(c, c') = d$ *he would have asymptotically small information about* x *if* $d > \gamma n$*.*

Sketch of the proof. The number of codewords of length n at Hamming distance d from c' is $\binom{n}{d}$. Using the assumption $d > \gamma n$ and a standard/clever approximation of $\binom{n}{d}$ one can derive that the average number of codewords at distance d from c' is greater than $\frac{2^{k-\frac{n}{2}+\alpha n}}{\sqrt{n}}$ except with probability $2^{-\alpha n}$ for any $\alpha > 0$.

The codeword c is one of the $\frac{2^{k-\frac{n}{2}+\alpha n}}{\sqrt{n}}$, at least, many equally likely codewords at distance d from c'. The following lemma (due to Bennett et al., 1998),will be used to determine the number of bits Bob can learn.

Lemma 6.3.14 *If* E *is the set of equally probable candidates for* c *and a random subset of bits of* c *is chosen, then the expected amount of Shannon information available to Bob about the parity of this subset is less than* $\frac{2}{|E| \ln 2}$ *bits.*

It follows from the above lemma that the number of bits of information Bob can learn about Alice's commitment, after seeing c', is less than $\frac{2\sqrt{n}}{2^{k-n/2-\alpha n} \ln 2}$. This number is exponentially small if $k > \frac{n}{2} + \alpha n$. Since $\frac{k}{n} = 0.52$ we have for $\alpha = 0.1$, $\frac{2\sqrt{n}}{2^{k-n/2-\alpha n} \ln 2} \leq \frac{2^{-0.1n}\sqrt{n}}{\ln 2}$.

[16] Bernstein law of large numbers: Let $x_1, x_2, \ldots, x_n$ be independent Bernoulli variables. If $Pr(x_i = 1) = p$ for $1 \leq i \leq n$, then for all $0 < \delta \leq p(1-p)$ we have $Pr(|\sum_{i=1}^{n} \frac{x_i}{n} - p| \geq \delta) \leq 2e^{-n\delta^2}$.

Security of the quantum oblivious transfer protocols

The first protocol (due to Crépeau and Kilian, 1988), was considered as secure provided neither party could store photons for a longer time and only projection measurements were used by Eve. Mayers (1998) made the final contribution to the numerous attempts to show that there is an unconditionally secure quantum oblivious transfer protocol provided there is an unconditionally secure QBCP. At that time this was considered as very encouraging results for quantum cryptography because it was believed that unconditionally secure QBCP do exist.

The result mentioned above, and discussed in the next section, namely that unconditionally secure QBCP is impossible, implies that one cannot have an unconditionally secure QOTP the security of which is based on security of a QBCP. However, this result does not rule out the possibility that there is unconditionally secure QOTP.

The BBCS protocol was shown secure (see Bennett et al. 1991) even against cheating by Bob with unlimited computing power under the assumption that Bob measures each photon (or pulse), before the next one arrives, using a projection measurement, or else he loses the opportunity to measure it at all.

Let us now discuss security of the BBCS protocol against the so-called photon (pulse) storing attacks (due to Bennett et al. 1991). The basic idea is that Bob does not measure the incoming photons in Step 3, he only stores them and waits with the measurement until Alice makes clear in Step 4 which bases she used. In this way it seems that Bob could present Alice with two good sequences and therefore he could get both bits b_0 and b_1.

From the practical point of view this attack is very unlikely to succeed. First of all it is technically hard to store photons (pulses) for a longer period. Secondly, even if such a storage were to be available this would not be sufficient. The problem is that Bob needs to tell Alice in Step 3 which photons (pulses) arrived successfully and were measured. However, no technique is available or foreseeable to determine whether a measurement will succeed without actually doing the measurement.

In addition, it is possible to change the BBCS protocol in such a way that it is fully secure against any photons (pulses) storage attack, providing there exists (unconditionally) secure QBCP. The basic idea goes as follows.

Alice sends to Bob not $\frac{2n}{\alpha}$, but at least $\frac{3n}{\alpha}$ of photons to achieve that $3n$ of them arrive successfully. Then, before Step 4, Bob is required to use a QBCP to commit himself both to bases he used in his measurements and to the outcomes of the measurements. Immediately after that Alice would choose randomly n of the reported successful measurements and ask Bob to unveil his commitments. This would allow Alice to check whether Bob's commitments are correct (subject to the error rate ε) when his commmited bases are correct and uncorrelated otherwise. In addition, this way Alice could be sure that Bob's measurements took place before Step 4 and that he used bases as he was required to do.

6.3.4 Security limitations of the quantum cryptographic protocols

Security of QBCP has been one of the most intriguing and important problems in the area of the security of quantum cryptography.

The first QBCP (see page 238), was shown insecure already by its authors. The second important one, the BCJL protocol, see page 240, has been claimed to be unconditionally secure by its authors and it was believed for a few years that their claim is true. A flaw in the proof was discovered in 1995 by Mayers and, independently, by Lo and Chau (1996). In addition, Lo and Chau were the first to argue that unconditionally secure QBCP may not exist. This suspicion was then shown to be valid by Mayers (1998).[17]

The very basic idea behind breaking BCJL is that Alice can use ancilla to create a compound quantum state that allows her to cheat as follows: She sends a part of the state to Bob and keeps the rest. By measuring her part appropriately, without touching Bob's part, she can modify her state in such a way that she can cheat concerning her commitment.

General form of quantum bit commitment protocols

The main idea behind the Mayers–Chau–Lo proof that unconditionally secure QBCP do not exist, as presented by Chau and Lo (1998) and Mayers (1998), assumes that each QBCP has the following form:

1. **Commitment phase.**
   (a) Alice and Bob put particles in their hands to some prescribed initial states.
   (b) Alice and Bob repeat several times the following steps:
      i. Depending on her commitment b Alice applies a unitary transformation U_b on her particles and sends *some* of her particles to Bob.
      ii. After receiving particles from Alice, Bob applies a unitary transformation to particles in his hands and then sends *some* of his particles to Alice.
2. **Opening phase.**
   (a) To open her commitment Alice sends all her particles to Bob.
   (b) After receiving particles from Alice Bob performs some measurements on particles in his hands to verify Alice's honesty.

In the terms of Hilbert space concepts the above general scheme of QBCP has the following transcription.

Let H_A and H_B be Hilbert spaces of Alice and Bob, and let H_C correspond to their communication channel and the environment. They execute their QBCP in $H = H_A \otimes H_B \otimes H_C$.

As the first step Alice prepares a state $|0_A\rangle$ or $|1_A\rangle$ in $H_A \otimes H_C$, according to her commitment, and Bob prepares a state $|a\rangle$ in $H_B \otimes H_C$. The overall initial state is then $|b_A\rangle \otimes |a\rangle$.

In Step 1.b, in each communication round, each party $D \in \{A, B\}$ performs a unitary transformation on $H_D \otimes H_C$ (and therefore also on H).

The key insight is now that since each product of unitary transformations is again a unitary transformation the whole communication process can be characterized by a single unitary transformation U applied to $|b_A\rangle \otimes |a\rangle$. Since both Alice and Bob know the protocol they also know U. Bob can therefore readily verify Alice's commitment after she sends him all her particles.

Remark 6.3.15 As is often the case, once the impossibility proof was made public attempts started to show that it does not cover all cases, all possible QBCP. Two ideas were explored: to introduce also classical communications, and to make Alice use classical BCP to commit

[17] For a detailed treatment of the history of the attempts to deal with security of QBCP see Brassard et al. (1998b).

herself, during the protocol, to some values, steps or measurements. All these attempts failed. For some discussion of such ideas see, for example, Brassard et al. (1997). In addition, Brassard et al. (1998b), have shown that even unconditionally secure classical BCP does not help.

Exercise 6.3.16 *Show that both QBCP on page 238 and 240 are special cases of the above general scheme of QBCP.*

Cheating

We show now that either Bob or Alice can cheat. More precisely, it will be shown that if Bob cannot learn Alice's commitment with high probability, then Alice can change her commitment at the beginning of the opening phase, without Bob noticing it, and therefore she can cheat—provided she has a quantum computer to perform unitary transformations.

Without loss of generality we can consider H_C as a part of H_A or H_B, depending on who is just making a transformation on H and therefore let $H = H_A \otimes H_B$.

The key tool to do cheating is the Schmidt decomposition theorem—see page 374. According to this theorem, the total state of H at the end of the commitment phase can be seen as having the following form in the case of the commitment to 0:

$$|0_{final}\rangle = \sum_i \sqrt{\alpha_i}|e_i, \phi_i\rangle \tag{6.2}$$

and in the case of the commitment to 1:

$$|1_{final}\rangle = \sum_i \sqrt{\beta_i}|e'_i, \phi'_i\rangle, \tag{6.3}$$

where $\{e_i\}_i$ and $\{e'_i\}_i$ are orthonormal bases of H_A.

The quantum state of Bob's particles, without the extra information coming from Alice, is described by the density matrix $\rho_B = Tr_{H_A}(|b_{final}\rangle\langle b_{final}|)$, if the commitment bit was b. By (6.2) and (6.3) we have

$$\rho_0 = \sum_i \alpha_i|\phi_i\rangle\langle\phi_i|, \qquad \rho_1 = \sum_i \beta_i|\phi'_i\rangle\langle\phi'_i|.$$

If ρ_0 and ρ_1 are very different, then Bob can learn, with high probability, Alice's commitment and therefore Bob can cheat. If ρ_0 and ρ_1 are not too different there are still two cases to consider.

Ideal case: $\rho_0 = \rho_1$. In such a case[18] for all i

$$\alpha_i = \beta_i \quad \text{and} \quad |\phi_i\rangle = |\phi'_i\rangle$$

and therefore

$$|1_{final}\rangle = \sum_i \sqrt{\alpha_i}|e'_i, \phi_i\rangle.$$

[18]It is assumed here that all eigenvalues are non-degenerate. The case of degenerate eigenvalues can be considered in a similar way.

Alice can therefore cheat by mapping $|0_{final}\rangle$ into $|1_{final}\rangle$ by applying on $|0_{final}\rangle$ a unitary transformation (on H_A only!) that maps, for all i, $|e_i\rangle$ into $|e'_i\rangle$.

This means that at the beginning of the commitment phase Alice can proceed by the protocol as she would make commitment 0 and at the end of the commitment phase, or better at the beginning of the opening phase, she can, without getting caught by Bob, change her commitment if she wishes to do so.

Non-ideal case: the difference between matrices ρ_0 and ρ_1 is small, with respect to the fidelity $F(\rho_0,\rho_1)$, defined as follows.

$$F(\rho_0,\rho_1) = \max\{|\langle\theta_0|\theta_1\rangle \mid \theta_i \text{ is a purification of } \rho_i, i = 0,1\}.$$

This measure of fidelity has the following property: to any purification θ_1 of ρ_1 there exists a purification θ_0 of ρ_0 such that $F(\rho_0,\rho_1) = |\langle\theta_0|\theta_1\rangle|$. Clearly, $0 \leq F(\rho_0,\rho_1) \leq 1$.

Let us now assume that there is a small $\delta > 0$ such that $F(\rho_0,\rho_1) = 1-\delta$. In such a case there is a state θ_1 which is a purification of ρ_0 and

$$|\langle\theta_1|1_{final}\rangle| = F(\rho_0,\rho_1) = 1-\delta \tag{6.4}$$

It is now clear that Alice's strategy for cheating in this non-ideal case can resemble that in the ideal case. Namely, Alice chooses 0 as her "preliminary commitment" at the beginning of the commitment phase and performs the commitment phase according to the protocol. If, at the beginning of the opening phase, Alice decides to cheat, she makes public that her commitment was 1 and she applies a local unitary transformation to change $|0_{final}\rangle$ to $|1_{final}\rangle$ such that $\langle 0_A|1_{final}\rangle = 1-\delta$. Since the difference between $|0_A\rangle$ and $|1_{final}\rangle$ is very small, Bob is not able to distinguish it and Alice can cheat with large probability.

6.3.5 Insecurity of quantum one-sided two-party computation protocols

Let us consider again the problem that Alice knows x, Bob knows y and they both want to compute $f(x,y)$ for a function f they both know—see Section 5.3.1 and page 243. However, this time the condition is that at the end of the protocol Alice learns nothing (about y and $f(x,y)$), Bob learns $f(x,y)$, but nothing about x (except what can be learned from x and $f(x,y)$ alone)—the so-called discrete function evaluation problem.

Lo (1998) has shown that in any quantum protocol to solve this problem Bob can cheat in the sense that he can learn $f(x,y)$ for all y. Cheating can be done as follows: Bob starts with some y_1 and uses the protocol to learn $f(x,y_1)$. Afterwards Bob changes y_1 to y_2 and learns $f(x,y_2)$ and so on. Such a cheating works for the following reason. When learning $f(x,y_1)$ by a measurement Bob will not change the measured state because it has to be the eigenstate of $f(x,y_1)$—this follows from the requirement that Bob learns $f(x,y)$ for 100%. Bob can therefore apply a unitary transformation to change y_1 to y_2 because Alice is supposed to learn nothing about y_1 when cooperating on computation of $f(x,y_1)$. In other words, the state of all particles of Alice and Bob when computing $f(x,y_1)$ and $f(x,y_2)$ is related by a unitary transformation that involves only Bob's particles.

6.4 Quantum Teleportation and Superdense Coding

Quantum teleportation[19], due to Bennett et al. (1993), allows transmission of quantum information to a distant place in spite of the impossibility of measuring or broadcasting information to be transmitted. It is an interesting demonstration of quantum entanglement as a communication resource.

6.4.1 Basic principles

We start with an informal presentation of the basic ideas behind quantum teleportation. Technical details, as well as the teleportation circuit, follow.

Assume that Alice and Bob share a pair of entangled particles A and B and let Alice be given a particle C in an unknown state $|\psi\rangle$. What should she do if she wishes Bob has $|\psi\rangle$ instead of her? She cannot measure the particle to learn the information which she could then transmit to Bob over a classical channel because such a measurement would almost certainly spoil the information irreversibly.[20] However, she can make the particle C in the unknown state $|\psi\rangle$ interact, in the proper way, with her part A of the entangled pair and then she can measure a shared property of both C and A. As the result, the state of Bob's particle B of the entangled pair instantaneously becomes a replica of $|\psi\rangle$—up to a rotation. At the same time Alice's particles C and A lose their information but she obtains, as the result of her measurement, two purely random bits of classical information that tell her which operation (rotation) Bob should perform on his particle B to make the particle get into the state $|\psi\rangle$. Alice then has to communicate these two bits to Bob over a classical channel, because Bob's particle B remains indistinguishable from a purely random qubit until he performs the required rotation, to get the particle into the state $|\psi\rangle$. Alice can therefore divide information encoded in $|\psi\rangle$ into two parts, classical and nonclassical, and send them to Bob through two different channels, a classical one and an EPR channel. One can also say that by performing a proper measurement on C and A a part of quantum information in C is immediately transmitted to B and Alice gets the remaining part of information on $|\psi\rangle$ in the classical form, and this can then be used by Bob to make his particle B get to the unknown state $|\psi\rangle$.

At the end neither Alice nor Bob knows the state $|\psi\rangle$ but both of them know that at the end of the teleportation not Alice but Bob has a particle in the state $|\psi\rangle$. This way the unknown quantum state $|\psi\rangle$ can be disassembled into, and later reconstructed from two classical states and an entangled, purely nonclassical, state. Observe also, that this way an intact and unknown quantum state can be "teleported" from one place to another by a sender who does not need to know—for teleportation itself—either the state to be teleported or the location of the intended receiver.

Finally, observe that the above process cannot be used to transmit information faster

[19]The term "teleportation" comes originally from science fiction and refers to making a person or an object disappear while an exact replica appears somewhere else. An underlying implicit assumption is that the teleported object does not traverse directly to its destination. Only necessary information for its assembly is first extracted, then transferred and finally used to assemble the original. This idea has been discarded by scientists as deferring physical laws, for example, Heisenberg's uncertainty principle, that does not allow measurement of all needed information about the object to be teleported. "Quantum teleportation", on the other hand, defers no physical law and therefore it is a term that sounds perhaps stronger than its real meaning is. Of course, this does not mean that it is not a very attractive concept (see Bennett et al, 1993).

[20]She could do that were she to have a whole set of particles all in the state $|\psi\rangle$. In such a case Alice could perform measurements on all these particles and determine $|\psi\rangle$ pretty well and then send this information to Bob who could prepare his source of qubits to produce $|\psi\rangle$.

than light but it can be argued that part of the information that was present in the particle C is transmitted instantaneously (except two random bits that needed to be transported at the speed of light at most).

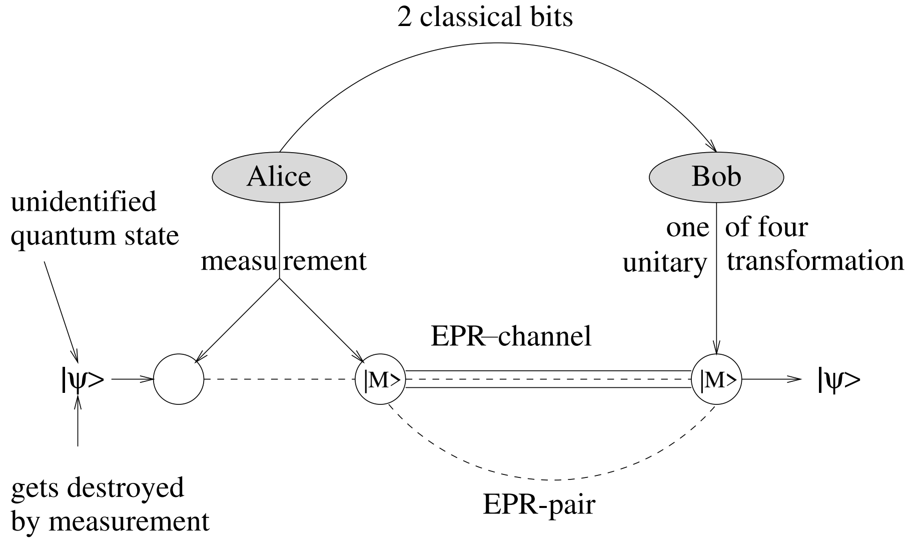


Figure 6.10: Quantum teleportation

Mathematical details are as follows: Assume that Alice and Bob share the EPR pair $|EPR\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. Let $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ be the unknown quantum state of the particle owned by Alice. She first couples her particle A with C to create the state

$$|\phi\rangle = |\psi\rangle|EPR\rangle = \frac{1}{\sqrt{2}}(\alpha|000\rangle + \alpha|011\rangle + \beta|100\rangle + \beta|111\rangle).$$

The key point for teleportation is now that $|\phi\rangle$ can be expressed in a special way using the Bell basis $\{|\Phi^{\pm}\rangle, |\Psi^{\pm}\rangle\}$.

Indeed, since

$$|00\rangle = \frac{1}{\sqrt{2}}(|\Phi^+\rangle + |\Phi^-\rangle), \qquad |01\rangle = \frac{1}{\sqrt{2}}(|\Psi^+\rangle + |\Psi^-\rangle),$$
$$|10\rangle = \frac{1}{\sqrt{2}}(|\Psi^+\rangle - |\Psi^-\rangle), \qquad |11\rangle = \frac{1}{\sqrt{2}}(|\Phi^+\rangle - |\Phi^-\rangle),$$

we get

$$\begin{aligned}
\alpha|000\rangle &= \alpha|00\rangle|0\rangle = \frac{1}{\sqrt{2}}(\alpha|\Phi^+\rangle|0\rangle + \alpha|\Phi^-\rangle|0\rangle),\\
\alpha|011\rangle &= \alpha|01\rangle|1\rangle = \frac{1}{\sqrt{2}}(\alpha|\Psi^+\rangle|1\rangle + \alpha|\Psi^-\rangle|1\rangle),\\
\beta|100\rangle &= \beta|10\rangle|0\rangle = \frac{1}{\sqrt{2}}(\beta|\Psi^+\rangle|0\rangle - \beta|\Psi^-\rangle|0\rangle),\\
\beta|111\rangle &= \beta|11\rangle|1\rangle = \frac{1}{\sqrt{2}}(\beta|\Phi^+\rangle|1\rangle - \beta|\Phi^-\rangle|1\rangle)
\end{aligned}$$

and therefore

$$\begin{aligned} |\phi\rangle &= \Phi^+\frac{1}{\sqrt{2}}(\alpha|0\rangle + \beta|1\rangle) + \Psi^+\frac{1}{\sqrt{2}}(\beta|1\rangle + \alpha|1\rangle) \\ &\quad +\Phi^-\frac{1}{\sqrt{2}}(\alpha|0\rangle - \beta|1\rangle) + \Psi^-\frac{1}{\sqrt{2}}(-\beta|0\rangle + \alpha|1\rangle). \end{aligned}$$

If Alice now makes a measurement of the first two qubits of $|\phi\rangle$, with respect to the Bell basis, then she will get one of the four possible outcomes: 00+, 01+, 00− and 01−, and therefore two classical bits of information, and $|\phi\rangle$ gets reduced to one of the states

$$\frac{1}{\sqrt{2}}(\alpha|0\rangle + \beta|1\rangle),\quad \frac{1}{\sqrt{2}}(\beta|0\rangle + \alpha|1\rangle),\quad \frac{1}{\sqrt{2}}(\alpha|0\rangle - \beta|1\rangle),\quad \frac{1}{\sqrt{2}}(\beta|0\rangle + \alpha|1\rangle), \tag{6.5}$$

and to the same state Bob's particle gets into. In order to tell Bob into which of the above four states $|\phi\rangle$ got reduced, she needs to send him two classical bits of information. Bob needs them to know which of the following four unitary transformations

$$U_{00} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\quad U_{10} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},\quad U_{01} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\quad U_{11} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

to apply to his particle in order to transform it to the original unknown state $|\phi\rangle = \alpha|0\rangle + \beta|1\rangle$.

The sending of two classical bits plays the key role in quantum teleportation. Indeed, it can be shown that if it was sufficient to send less than two classical bits of information in the above teleportation scheme, this could be used by Bob to send messages faster than light—see Bennett et al. (1993)—i.e. to send the so-called *superluminal messages.*

6.4.2 Teleportation circuit

Despite the fact that the idea of teleportation sounds more like science fiction, it is expected to be implemented sooner than some nontrivial quantum processors. The hardest problem to solve for teleportation seems to be, for most technologies, the development of an efficient long-term storage for quantum information, especially for entangled particles.

Quantum circuit to perform teleportation can be made very simple (see Brassard, 1998), and his circuit will now be presented.

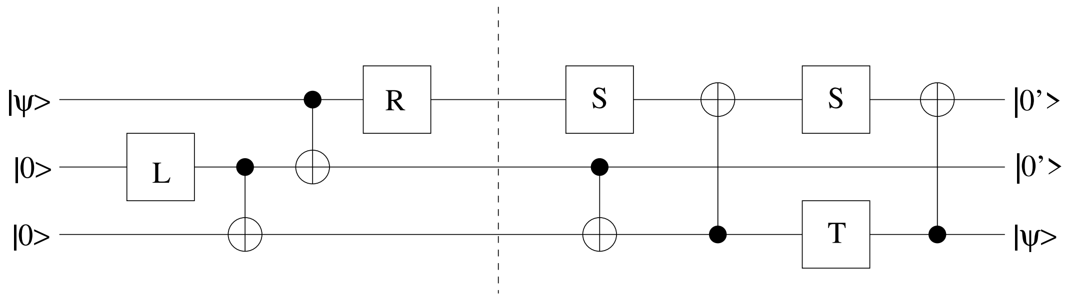


Figure 6.11: Brassard's teleportation circuit

Consider the circuit in Figure 6.11 where L, R, S, and T are the gates implementing the following unitary transformations

$$R = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \quad L = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \quad S = \begin{pmatrix} i & 0 \\ 0 & 1 \end{pmatrix} \quad T = \begin{pmatrix} -1 & 0 \\ 0 & -i \end{pmatrix}$$

Let $|\psi\rangle$ be a qubit state. If the state $|\psi 00\rangle$ is processed by the circuit from Figure 6.11, i.e. $|\psi\rangle$ is put on the topmost input and $|0\rangle$ on the other two, then the output will be $|0'0'\psi\rangle$.

Exercise 6.4.1 *Design unitary matrices corresponding to circuits: (a) in Figure 6.12a; (b) in Figure 6.12b; (c) in Figure 6.11 before the dashed line; (d) in Figure 6.11 after the dashed line; (e) for the whole circuit in Figure 6.11.*

Exercise 6.4.2 *Determine the intermediate states of the computation of the circuit from Figure 6.11 on input $|\psi 00\rangle$ after all gates.*

In the state of the circuit at the dashed line all three qubits are entangled. A measurement of the two upmost qubits provides two random classical bits, say u, v. Surprisingly enough (verify it), if these two bits are "returned" into the circuit, i.e. if the computation of the circuit to the right of the dashed line starts with input $|uvz\rangle$, where z is the state of the third qubit of the dashed line, then the output of the circuit will be $|uv\psi\rangle$.

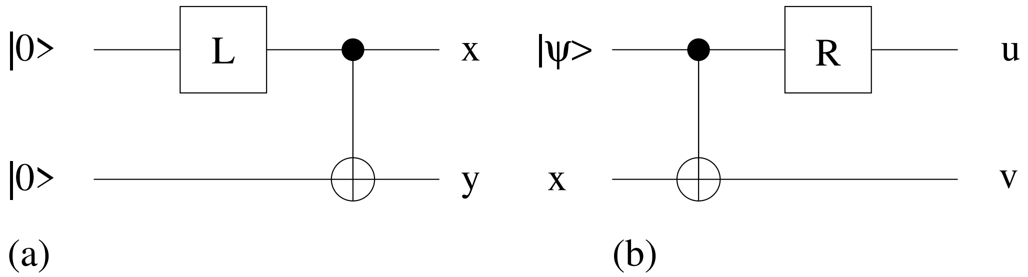


Figure 6.12: A teleportation device

Out of the teleportation circuit shown in Figure 6.11 we can make, by a "cut along the dashed line", two circuits, one for Alice, one for Bob, to use for teleportation (see Figure 6.13).

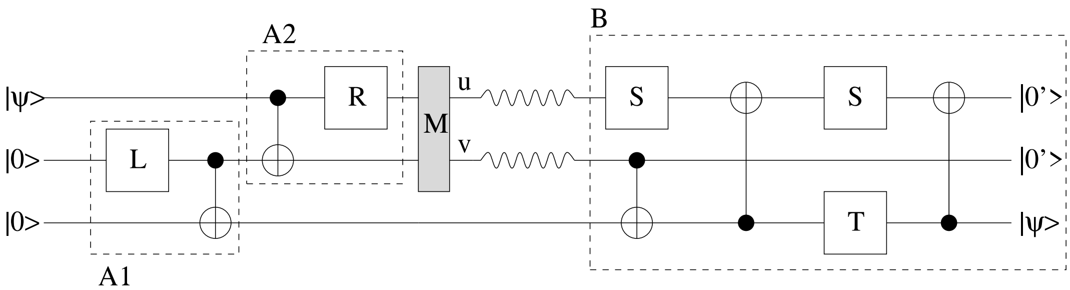


Figure 6.13: Two teleportation circuits of Alice and Bob

Alice's circuit consists of two subcircuits. Using the first one, A1 (see also Figure 6.12a), with the initial states $|0\rangle$, Alice can create a pair of particles Q2 and Q3 in the entangled state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. Alice keeps Q2 and sends Q3 to Bob.

At some time later, let Alice want to teleport to Bob the unknown state $|\psi\rangle$ of her new particle Q1.[21] She can then use the second subcircuit A2 (see Figure 6.12b), to entangle $|\psi\rangle$ with particles Q2 and Q3. (Observe that at the output of the second subcircuit all three qubits are entangled.)

Alice now makes a measurement of Q1 and Q2, with respect to the Bell basis. As the result she gets two classical bits of information, the states of Q1 and Q2 will collapse and Q3 will get into one of the states shown in (6.5). The two bits she gets Alice sends, using a classical channel, to Bob.

If Bob gets two bits he can add them as inputs to his teleportation circuit with Q3 as the third input. Bob's circuit is to choose, on the basis of two inputs, the proper rotation to apply to Q3 and to perform the required rotation to have his qubit Q3 in the state $|\psi\rangle$.

Applications? Let us first explore a natural question whether teleportation could be useful for quantum information processing. Yes, because teleportation can provide another way to transmit information inside quantum computers and information systems, which can be of interest especially if information has to be kept secret and should never be transmitted over an insecure channel. Actually, it is in the area of quantum computers and communication systems where the first applications of quantum teleportation are expected.

In addition, if Bob already possesses the state $|\psi\rangle$, then teleportation can be used by Bob to determine $|\psi\rangle$ more completely by making measurements on both copies of $|\psi\rangle$. Moreover, teleportation is possible without Alice knowing the exact position of Bob. It is sufficient to broadcast classical bits to all the possible locations Bob could be in (or to send him an email).

Methods of quantum teleportation have been improved to work with arbitrarily high fidelity even if the quantum channel is imperfect and the quantum noise is too strong to use some quantum error-correction techniques (see Section 8.3.2 and Bennett et al. 1996a).

Partial implementations (without the last stage—Bob's transformations), of quantum teleportation over macroscopic distance have already been reported by Bouwmeester et al. (1997), for the distance of 1m, and Boschi et al. (1998), using optical systems and photons. A complete implementation of quantum teleportation over inner-atomic distance using liquid state NMR technology was reported by Nielsen, Knill and Laflamme (1998).

Remark 6.4.3 In spite of the remarkable power of quantum entanglement for quantum teleportation, and also as a substitute for communication, see Section 7.4.1, the power of entanglement to facilitate direct communication between two parties is quite restricted. For example, let two entangled particles be possed by Alice and Bob. If Alice receives an unknown bit of information there is no operation she can perform on her particle in such a way that Bob could then get the bit by performing an appropriate operation on his particle.

6.4.3 Quantum secret sharing

The idea of teleportation can be used to solve communication problems in several ways, see especially Chapter 8. One of them will now be discussed. It provides quantum realization of the classical secret sharing ideas.

[21]More exactly Alice does not need to know the state $|\psi\rangle$ in order to teleport it to Bob. Without loss of generality we can therefore assume the state is unknown to her.

Splitting of information

There is a simple method, due to Hillery, Bužek and Berthiaume (1998), how Alice can teleport a (secret) qubit $|\phi\rangle = \alpha|0\rangle + \beta|1\rangle$ to Bob and Charles in such a way that they have to cooperate in order to have $|\phi\rangle$.

The basic idea is that Alice couples a given particle P in the state $|\phi\rangle$ with the state $|\psi\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$ of three particles P_a, P_b and P_c she shares with Bob and Charles and then performs a measurement on the state of particles P and P_a, with respect to the Bell basis $\{\Phi^\pm, \Psi^\pm\}$. Since

$$\begin{aligned}|\phi\rangle|\psi\rangle &= \frac{1}{2}(|\Phi^+\rangle(\alpha|00\rangle + \beta|11\rangle) + |\Phi^-\rangle(\alpha|00\rangle - \beta|11\rangle) \\ &\quad +|\Psi^+\rangle(\beta|00\rangle + \alpha|11\rangle) + |\Psi^-\rangle(-\beta|00\rangle + \alpha|11\rangle)),\end{aligned}$$

the outcome of the measurement is that particles P_b and P_c get into one of the states

$$\frac{1}{\sqrt{2}}(\alpha|00\rangle + \beta|11\rangle), \frac{1}{\sqrt{2}}(\alpha|00\rangle - \beta|11\rangle), \frac{1}{\sqrt{2}}(\beta|00\rangle + \alpha|11\rangle), \frac{1}{\sqrt{2}}(-\beta|00\rangle + \alpha|11\rangle)$$

and Alice gets two bits to tell her about which of these four cases happened. However, neither Bob nor Charles has information about which of these four states their particles are in.

Bob now performs a measurement of his particle with respect to the dual basis. He gets out of it one bit of information and Charles's particle P_c gets into one of 8 possible states, which is uniquely determined by bits both Alice and Bob got as the results of their measurements, and which can be transformed into the state $|\phi\rangle$ using one or two applications of Pauli matrices.

Exercise 6.4.4 *(a) Determine the density matrix of Charles's particle after Alice's measurements; (b) determine* 8 *possible states into which Charles's particle can get after Alice's and Bob's measurements; (c) determine transformations Charles has to perform in order to have his qubit in state* $|\phi\rangle$ *in dependence on bits learned by Alice and Bob in their measurements.*

Exercise 6.4.5 *Show how to generalize the idea of splitting information between two parties to the case of (a)* 3 *parties; (b)* n *parties.*

Quantum secret sharing among n parties

A more general problem is to design methods how one can "partition" or "share" a quantum "secret" (state) among n parties in such a way that, for a fixed k, any k of parties can, by cooperation, reconstruct the secret, but no combination of $k-1$ of parties is able to do that.

Exercise 6.4.6 *Show (Cleve et al. 1999), that using the mapping*

$$\alpha|0\rangle+\beta|1\rangle+\gamma|2\rangle \rightarrow \alpha(|000\rangle+|111\rangle+|222\rangle)+\beta(|012\rangle+|120\rangle+|201\rangle)+\gamma(|021\rangle+|102\rangle+|210\rangle)$$

one can distribute a "secret qutrit" to three qutrits in such a way that if each qutrit is owned by a different party, then any two of them can reconstruct the secret state, but no single party can do that alone.

The secret sharing problem, a quantum analogue of the classical one, is in full generality solved by Cleve et al. (1999). They showed that such a secret sharing does exist if $n < 2k$. The last restriction is due to "No-cloning theorem".

6.4.4 Superdense coding

Teleportation uses two classical bits to transmit one qubit. Superdense coding, Bennett and Wiesner (1992), is a dual to teleportation. It uses a single qubit, i.e. a single photon, to transmit two classical bits. This can be done as follows—see Figure 6.14.

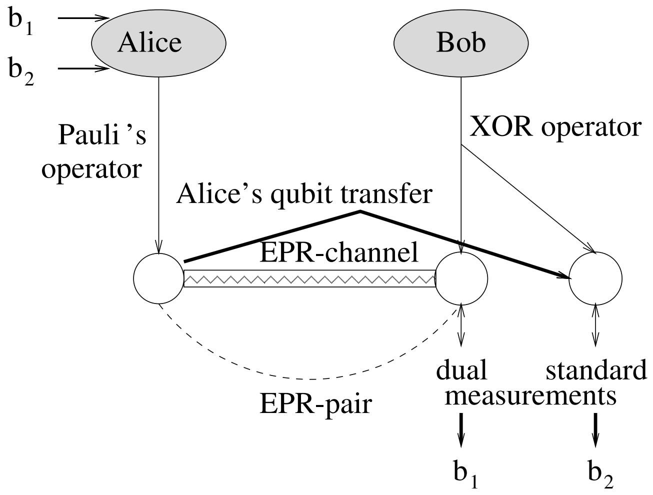


Figure 6.14: Superdense coding

Assume that Alice and Bob share two particles in the EPR state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, which forms the EPR-channel. If Alice receives two classical bits, b_1, b_2, she performs on her particle one of the Pauli rotations as shown in the second column of Figure 6.15. The resulting state is shown in the third column and this state she then sends to Bob. He performs first on both states XOR operation and this way he disentangles the state with the result shown in column 5. Finally, Bob performs the measurement of Alice's qubit in the dual bases and of his qubit in the standard basis to get two bits (see column 6), Alice has sent him in one qubit. In a simplified form the resulting system is depicted in Figure 6.16.

Alice's bits	Pauli's rotations	Alice's particle: new state	$\rightarrow$	Bob's XOR transformation	Bob's bases $\mathcal{D}, \mathcal{B}$	Bob's bits
00	I	$\frac{1}{\sqrt{2}}(\|00\rangle + \|11\rangle)$		$\frac{1}{\sqrt{2}}(\|0\rangle + \|1\rangle)\|0\rangle$	00	00
01	σ_x	$\frac{1}{\sqrt{2}}(\|10\rangle + \|01\rangle)$		$\frac{1}{\sqrt{2}}(\|0\rangle + \|1\rangle)\|1\rangle$	01	01
10	σ'_y	$\frac{1}{\sqrt{2}}(-\|10\rangle + \|01\rangle)$		$\frac{1}{\sqrt{2}}(\|0\rangle - \|1\rangle)\|1\rangle$	11	10
11	σ_z	$\frac{1}{\sqrt{2}}(\|00\rangle - \|11\rangle)$		$\frac{1}{\sqrt{2}}(\|0\rangle - \|1\rangle)\|0\rangle$	10	11

Figure 6.15: Superdense coding steps

Remark 6.4.7 1. Quantum superdense coding transmission was first put into practice in

Innsbruck by Zeilinger's group with polarization-entangled photons (see Matte et al. 1996).

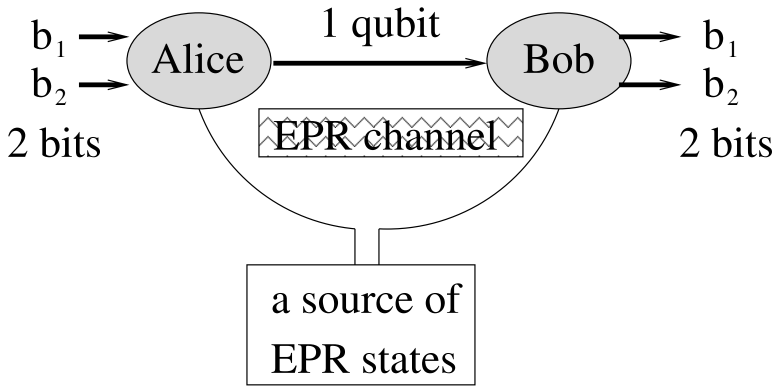


Figure 6.16: Superdense coding scheme

2. In the superdense coding presented above, it is essential that Alice and Bob use a maximally entangled state. Indeed, Barenco and Ekert (1995) and Hausladen et al. (1996) have shown that the amount of information communicated by the superdense coding decreases from its maximum, 2 bits per qubit, with the decrease in the amount of entanglement, and it becomes 1 bit when the entanglement is zero. When the initial state of the entangled pair of qubits is mixed, capability to do superdense coding in terms of various measures of entanglement has been investigated by Bose et al. (1998).

Remark 6.4.8 Experimental progress in the quantum key generation, creation of entangled pairs over a long distance and in quantum teleportation has been such that the vision of small quantum networks does not have to be far away. This puts the problems of quantum multiparty communications and quantum distributed computing into a promising research agenda.

Chapter 7

PROCESSORS

INTRODUCTION

Theoretical investigations concerning quantum algorithms, automata, complexity, information theory and in cryptography are of great interest and importance. However, progress in the experimental efforts to design quantum information-processing systems is crucial for seeing properly the overall perspectives of the future designs of real and powerful quantum computers, and for isolating and solving the problems that need to be dealt with if powerful quantum computers are ever to be built.

It has been realized, from the very early days of research in quantum computing, at least by some, that powerful evolution of isolated quantum systems is hard to utilize in real quantum processors, because of their interaction with the environment that can destroy very large but fragile quantum superpositions; and because of the natural imperfections of (inherently analogue) quantum devices. In addition, quantum error correction was considered impossible.

Fortunately, several developments brought the vision of quantum computers closer to reality. Quantum computation stabilization methods and quantum error correction codes have turned out to be possible and efficient. Techniques for fault-tolerant quantum computing have been developed. Finally, some promising technologies to design quantum gates, circuits, and processors have been identified and are being experimentally tested.

LEARNING OBJECTIVES

The aim of the chapter is to learn:

1. the early proposals for quantum computers;
2. the impacts of imprecision, dissipation and decoherence on quantum computing;
3. the methods of designing and using quantum error-correcting codes;
4. the basics of quantum fault-tolerant techniques;
5. the basics of the main current technologies used, or considered, to develop experimental quantum processors.

> In theory there is no difference between theory and practice, but in practice, there is.

Theoretical developments concerning quantum algorithms, automata, complexity, information and cryptography have been accompanied by an intensive theoretical as well as experimental effort to identify and overcome problems that need to be dealt with in order to design successfully (experimental) quantum processors.

A real quantum computer is just a physical system whose evolution can be interpreted as performing some specific quantum computation. In order to design such systems, frontiers of current technology have to be explored and insights have to be developed into the essence of such crucial problems as imprecisions during quantum computations and decoherence, which occurs when quantum information is sent, in time or space, through a noisy quantum channel. These problems are either specific for quantum computing or have in this case a very different nature than in the classical computing. Finally, methods have to be developed to deal with these problems. Some of such methods can be seen as quantum generalizations of classical ones, but also methods had to be developed that have no classical analogue.

Progress has been achieved in studying the main obstacle to practical quantum computing — decoherence — and formidable successes have been achieved in developing quantum computing and storage-stabilizing techniques: quantum error-correcting codes, entanglement purification (or distillation) techniques, and so on. In addition, fault-tolerant techniques have already been developed for quantum computing. It is clear today that arbitrarily long quantum computations can be performed reliably, in principle, provided that the average probability of error per quantum gate is less than a certain threshold. Therefore, internal imprecisions and external decoherence do not have to be any longer considered as an obstacle to quantum computation we would not be able to cope with. Theoretical results indicate that it might even be possible to build inherently fault-tolerant quantum hardware. All that makes the vision of real quantum computers much closer than expected a few years ago.

Significant progress has been achieved in the experimental development and testing of several technologies that seem to have a potential use in the design of small experimental quantum processors.

The main theoretical developments behind the design of experimental quantum processors as well as the main principles of technologies being currently explored for this purpose are discussed in this chapter.

7.1 Early Quantum Computers Ideas

The first four historically important attempts to describe how a quantum computer could function are briefly summarized in this section. They also illustrate a shift from the first half-classical/half-quantum Turing machine model of Benioff, through Feynman's universal quantum simulator—a quantum version of Boolean circuits—to a fully quantum Turing machine model of Deutsch.

7.1.1 Benioff's quantum computer

Benioff (1982, 1982a) made the first attempts to develop "quantumized" Turing machines. In his models the tape of a QTM consists of a sequence of qubits (spin states), each one being in one of the basis states $|0\rangle$ or $|1\rangle$ after each elementary computation step. This way a binary input and intermediate results were encoded. The "head" of a TM had the form of a device that could change the values of qubits. Finite control, with the program to be performed, was implicitly contained in the Hamiltonian of the carefully designed Schrödinger equation.

During each computational step Benioff's computer could be in a superposition of quantum states, but at the end of each step the tape had to be in a basis state representing a sequence of classical states and no characteristically quantum property of the model (superposition of basis states, parallelism, entanglement) could be detected. His model was developed within the framework of quantum kinematics and dynamics, but a classical Turing machine could simulate it effectively and easily.

There have been several variants of Benioff's model. A major problem with the early models was that in order to design a Hamiltonian to mimic a particular Benioff's computer, not only the program but also the complete set of computation paths had to be given. In other words, "one would need to know the answer to the problem one had to compute in order to design a Benioff's computer to solve it".

The main implementation problem of Benioff's computer was how to realize an interaction between the head and qubits that could be far apart.

7.1.2 Feynman's quantum computer

Feynman (1982, 1986) took at first glance a very simplified approach to quantum computing. His approach was based on the understanding, due to the results of Fredkin and Toffoli (1982), that there are universal reversible Boolean gates. His main contribution was a general method to design the Hamiltonian for a quantum circuit in which each gate performs a unitary operation.

At first it seemed (see Deutsch, 1985) that Feynman's approach is a restrictive one. "Programming" of his simulator actually means designing a new quantum circuit and then setting its inputs. However, due to Yao's (1993) result, it is now clear that from the computational power point of view the quantum circuits model has the same power as quantum Turing machines.

Let us assume that we have a quantum circuit consisting of a serial connection of k quantum gates, each performing a unitary operation, say $U_1, \ldots, U_k$. Feynman dealt with such a case in its full generality and showed the form a Hamiltonian H in the Schrödinger equation

$$U(t) = e^{-\frac{i}{\hbar}Ht}$$

should have for such a circuit. Feynman thereby found a systematic, though not very efficient, way to transform a quantum circuit description of a quantum computer to the dynamical Schrödinger equation that simulates computation steps of the circuit.

Feynman used extra k qubits, for so-called "program counter sites" as well as the "creation" operators c_i $(i = 1, \ldots, k)$ and the "annihilation operators" a_i $(i = 0, \ldots, k-1)$. Each creation operator c_i "sets the ith counter qubit to 1", the annihilation operator to 0. The overall Hamiltonian has then the form

$$H = \sum_{i=0}^{k-1} (c_{i+1} \cdot a_i \cdot U_{i+1} + (c_{i+1} \cdot a_i \cdot U_{i+1})^*)$$

where "$\cdot$" denotes the product of matrices. The first terms take care of a sequential execution of all gates. (One needs to add the conjugate terms because the resulting Hamiltonian has to be Hermitian.) In total, Feynman's approach needs a register with $m + k - 1$ qubits to deal with k counters and m input qubits.

More exactly,

$$c = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \qquad a = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

and c_i (a_i) is just c (a) applied to the ith counter qubit. (Observe that c maps $|0\rangle \to |1\rangle$ and $|1\rangle$ to a "null state"; similarly a maps $|1\rangle \to |0\rangle$ and $|0\rangle$ to a "null state".)

Computation on such a circuit begins by putting the input bits into the input register and the pointer to occupy the site 0. One then checks, at site k, that site k is empty, or that the site has a pointer. Once the cursor is found, it is removed so that it cannot return down the program line. At that moment the register contains all outputs that just need to be measured. Termination is not taken care of by such a quantum computer itself. It has to be decided from the outside when a measurement is to be performed.

In Feynman's model all of the quantum uncertainty of the computation is concentrated in the time needed for computation to be completed, and not at all in the correctness of the outcomes. Namely, if a computation is done, and a certain bit indicates it, the result obtained is always correct.

7.1.3 Peres' quantum computer

Peres (1985) improved some of Benioff's and Feynman's ideas on how a quantum computer should be designed. For example, he improved Feynman's model in such a way that the result of the calculation shows up at the precise time and place. In spite of the fact that his model, similarly to Benioff's, is effectively classical—inherently quantum phenomena (as superposition) occur only in the dynamics and not in the logic of the computer—Peres has already brought up the idea of having a computer "continuous quantum logic"—with states being a superposition of basis states.

Like all previous attempts to describe a quantum computer, Peres concentrates on Hamiltonians: how to write a Hamiltonian H for a quantum system such that the time evolution $e^{-\frac{iHt}{\hbar}}$ represents the execution of a computation. The main new contribution is an analysis of possible errors (in the Hamiltonian and in measurements) and especially an idea of how to use quantum error-correcting codes and how to incorporate into Hamiltonians quantum error correction procedures.

Peres has considered the case that one qubit is encoded by three ($|0\rangle \to |000\rangle$, $|1\rangle \to |111\rangle$) and the potential error states have the form

$$\alpha_0|000\rangle + \beta_0|100\rangle + \gamma_0|010\rangle + \delta_0|001\rangle$$

or

$$\alpha_1|111\rangle + \beta_1|011\rangle + \gamma_1|101\rangle + \delta_1|110\rangle,$$

for suitable amplitudes $\alpha_i, \beta_i, \gamma_i, \delta_i$, $i = 0, 1$.

Peres also considered a way of using Stern–Gerlach magnets for error detection and error correction. In addition, he realized that errors can be corrected by unitary operations. Finally, he has considered ways error correction can be incorporated in the Hamiltonian, so that the probability of error can be made arbitrarily small.

In spite of the fact that Peres' error-correcting code is not good enough, see page 280, because it does not use entanglement to protect quantum information, it was the first attempt to consider quantum error-correcting codes.

7.1.4 Deutsch's quantum computer

Deutsch's basic philosophical standpoint was that underlying the Church–Turing principle there is an implicit physical assertion. On this base he developed the so-called physical version of the Church–Turing principle:

> Every finitely realizable[1] physical system can be perfectly simulated by a universal model computing machine operating by finite means.

Deutsch (1985) presented a general, fundamentally new, and fully quantum model of quantum computation. The tape ($\mathbf{t}$) of Deutsch's Turing machine $\mathcal{U}$ consists of an infinite sequence of qubits and its finite control consists of a finite sequence of qubits ($\mathbf{m}$). In addition there is an observable $\mathbf{x}$, which has any integer from $\mathbf{Z}$ as its potential value—a pointer to the currently scanned tape cell. Deutsch deals with the problem of infinitely long tape by assuming that tape is not rigid and there is a mechanism that can move the tape according to signals transmitted at finite speed between adjacent segments. The state of the quantum computer $\mathcal{U}$ is therefore a unit vector in the space spanned by basis vectors $|\mathbf{x}, \mathbf{t}, \mathbf{m}\rangle$.

The dynamics of $\mathcal{U}$ is given by a constant unitary operator U and for the evolution of the state $|\psi(t)\rangle$ it holds

$$|\psi(t)\rangle = U^t|\psi(0)\rangle \quad \text{where}|\psi(0)\rangle = \sum_n \lambda_n|0, 0, \mathbf{t}\rangle,$$

and only finitely many λ_i are non-zero if an infinite number of elements in $\mathbf{t}$ are non-zero. U has to satisfy a special condition in order to perform operations "by finite means".

Deutsch was fully aware of such features of quantum computing as quantum parallelism and entanglement. To explain these features he used Everett's many-world interpretation, Section 9.1.7, because, as he explains, "the intuitive explanation of these properties places an intolerable strain on all interpretations of quantum theory other than Everett's".

[1] By "finitely realizable physical system" any physical object is meant upon which experimentation is possible.

Deutsch also described the universal quantum computer capable of simulating every finitely realisable physical system, therefore also any other quantum computer, with arbitrarily high precision. To design the universal quantum computer Deutsch made use of the fact that if α is any irrational multiple of π, then the four transformations

$$\begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}, \quad \begin{pmatrix} \cos\alpha & i\sin\alpha \\ i\sin\alpha & \cos\alpha \end{pmatrix}, \quad \begin{pmatrix} e^{i\alpha} & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & e^{i\alpha} \end{pmatrix}.$$

form a group dense in the group of all unitary transformations in H_2.

7.2 Impacts of Imperfections

There are two main sources of imperfections that may influence computations on real quantum processors and which one may ignore when ideal quantum computation is considered: internal and external.

Internal imperfections have to be seen more as a rule than an exception: they are imperfections in the setting of the initial state (input data), imperfections of the quantum evolution itself (of the gate operations), and, finally, imperfections of the final measurement. Internal imperfections can be seen as "hardware imperfections".

As we shall see, these imperfections do not influence quantum computing dramatically. Actually, with respect to the imperfections of this type quantum processors are much more robust than classical ones, in which impacts of internal imperfections grow exponentially.

On the other hand, the external imperfections, due to decay (dissipation), and, especially, decoherence, are much more serious and they keep threatening the very basis of existence of the successful and powerful quantum computers.

7.2.1 Internal imperfections

At first we explore impacts of errors that one can expect to occur, with high probability, in the preparation and execution of quantum computations on real quantum computers.

There is a variety of reasons why such imperfections are practically inevitable, directly opposite to the situation in classical computing. The main reason is that in all three main stages: setting of the initial state; quantum evolution; and final measurements; one has to set up complicated, sensitive and in the essence analogue devices about which there is practically no likelihood of absolute precision. Fortunately, these imprecisions are not fatal for quantum computing.

Imprecisions in the initial state

The initial state $|\psi_0\rangle$ for the quantum evolution $|\psi(t)\rangle = U^t|\psi_0\rangle$, may be one of the basis states, or a quantum superposition of such states. To set up such a state various sensitive devices (i.e. polarizers, magnets, devices for timing of laser pulses in the ion trap computers) have to be used. Their setting and performance may be a source of errors. The question is now how much such errors propagate and accumulate during a quantum evolution—how much is quantum evolution stable with respect to the initial uncertainties. In other words, how much can we trust the results of quantum evolution in the case of errors in the initial state?

Fortunately, due to the linearity of quantum evolution, such errors do not grow with time. To illustrate this case let us consider the situation that instead of the initial state

$|\psi_0\rangle = |0\rangle$ we have the state $|\psi_{err}\rangle = \sqrt{1-\varepsilon^2}|0\rangle + \varepsilon|1\rangle$. After t computation steps we have

$$U^t(\sqrt{1-\varepsilon^2}|0\rangle + \varepsilon|1\rangle) = \sqrt{1-\varepsilon^2}U^t(|0\rangle) + U^t(|1\rangle) = \sqrt{1-\varepsilon^2}|\psi(t)\rangle + \varepsilon U^t(|1\rangle)$$

For small ε the first term is almost the state $|\psi(t)\rangle$ and the second one can be negligible. (On the other hand, many quantum algorithms are so sensitive to small changes in the amplitudes of superpositions that the above result should not be overestimated.)

Operational imperfections

Another source of imperfections, that results in the errors in the Hamiltonian of the system and consequently in the quantum evolution, is an unavoidable inaccuracy of quantum computer components. This is basically due to the fact that quantum computer components are mainly analogue type devices. As a consequence, the state of a quantum superposition depends on several continuous parameters. For example, gates very often used are those performing a rotation by an angle θ. If such a gate is applied there is naturally some inaccuracy in θ. Errors of this type are caused by unitary transformations, actually by over-rotations or under-rotations.

There are several ways to estimate impacts of operational imperfections.

In Section 4.2.3 we have seen, by analyzing imprecisions of Turing machine computations, Theorem 4.2.21, that imprecisions during computations only add and do not grow exponentially. In addition, it was shown that $\mathcal{O}(\lg t)$ bits of precision in transition amplitudes are sufficient to support t steps of a QTM with required precision.

Another way to approach the problem is to consider the case that instead of correct Hamiltonian H there is slightly different Hamiltonian $H' = H + H_{err}$. An analysis by Williams and Clearwater (1997) shows that errors grow at most quadratically in time.

Imperfections in the measurement

Finally, let us discuss the last case, namely that we do not know exactly when the result of a quantum computation is ready (as at the Grover's search algorithms), and we have to perform n measurements at the equally spaced time intervals.

An analysis by Williams and Clearwater (1997) implies that more measurements will not make computation less secure. This is, however, only a partial result. A deeper and more general analysis of imperfections due to measurements is still to be done.

7.2.2 Decoherence

Two essential properties of quantum information processing that efficient quantum algorithms essentially exploit are the existence of quantum superpositions and entanglement—non-local correlations between different parts of physical systems. It is an elementary but fundamentally important fact that both of them are ultrasensitive to interactions with the environment. The enormous fragility of quantum states used to process information is the main problem for any design of quantum processors.

In order to perform a successful quantum computation one has to maintain a coherent unitary evolution until the completion of the computation. However, technologically it is not possible to ensure that a quantum register is completely isolated from the environment.

There are several reasons why it is practically not possible to avoid interactions of a real quantum computer with the environment. Quite sophisticated technology has to be used to create quantum registers, and this infrastructure technology cannot be fully isolated from the register. Secondly, there is an unavoidable coupling of quantum systems with the thermal environment. (Even such phenomena as cosmic rays may have their impacts.)

There are several ways interactions of quantum systems with their environment materialize. For example, the **decay** (**dissipation**) is a process by which a quantum system dissipates energy into the environment. For example, if an excited (high energy) state is to represent $|1\rangle$ in a qubit and the lower energy state represents $|0\rangle$, then the system (qubit) can, spontaneously, make a transit from $|1\rangle$ to $|0\rangle$, emitting a photon in the process. A qubit flip (from $|1\rangle$ to $|0\rangle$ only), is then the net result of such a spontaneous transition.

The **decoherence** is a general term for process of coupling of a quantum system (processor) with its environment, if it is not perfectly isolated. As a consequence, quantum state of the system is modified due to the interactions of the system with the environment. Such interactions mean that the quantum dynamics of the environment is also relevant to the operations of the quantum computer and its states become entangled with the states of the environment. The effect of such an entanglement with the environment can be seen as if the environment applied a measurement to the system. This has destructive impacts on superpositions and interference. Decoherence tends to destroy irreversibly information in a superposition of states in a quantum computer in a way we cannot control. It spoils both constructive and destructive interferences that are essential for quantum computing and long computations seem to be impossible. As the result, quantum information initially encoded in a register becomes encoded instead in the correlations between the quantum computer and its environment and we can no longer access the information by observing only the computing device. Decoherence can therefore be seen as a physical process, in which quantum systems lose, due to their interaction with environment, some of their key quantum properties of importance for being able to have quantum information processes more efficient than classical ones.

Especially delicate are entangled states. Making a measurement on an entangled state will usually cause a collapse of it to a less entangled state. Small interactions with the environment provide a sort of continuous measurement of the system. As a quantum system gets larger this is harder and harder to ignore. System starts to decohere more and more and starts to look and behave more and more as a classical system. Decoherence is why quantum world looks classical at the human level (Gottesman, 1997).

Decoherence is the most fundamental obstacle so far preventing the design of real quantum computers. For example (see Barenco, 1996), the effects of decoherence in time on the qubit state $|\phi\rangle = \alpha|0\rangle + \beta|1\rangle$, represented by the density matrix

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

can be discussed by the time-dependent density matrix

$$\rho = \begin{pmatrix} |\alpha|^2 & e^{-\frac{t}{\tau}}\alpha\beta^* \\ e^{-\frac{t}{\tau}}\alpha^*\beta & |\beta|^2 \end{pmatrix},$$

in which τ is so-called decoherence time. Matrix ρ_τ converges to the diagonal matrix

$$\begin{pmatrix} |\alpha|^2 & 0 \\ 0 & |\beta|^2 \end{pmatrix}.$$

Example 7.2.1 *(Barenco, 1996) Decoherence can affect the probability distribution of the possible outcomes of computations. In order to demonstrate that let us consider the situation* *that the Hadamard matrix H is applied to the state $|0\rangle$ of a qubit twice. We get*

$$|0\rangle \xrightarrow{H} \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \xrightarrow{H} |0\rangle. \tag{7.1}$$

If (7.1) is reformulated in terms of the corresponding density matrices we see this evolution as follows:

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \xrightarrow{H} \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \xrightarrow{H} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{7.2}$$

If there is no decoherence, the measurement of the final state yields the outcome $|0\rangle$ with probability 1. Let us now assume that decoherence occurs between the two applications of the operator H and annihilates completely off-diagonal elements of the second matrix in (7.2).[2] *Instead of (7.2) we get the evolution*

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \xrightarrow{H} \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \xrightarrow{decoherence} \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \xrightarrow{H} \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

At the measurement we get both $|0\rangle$ and $|1\rangle$ with probability $\frac{1}{2}$.

It is clear that decoherence is a major problem. It is less clear how really big a problem it is and whether we can successfully deal with it.

The first estimations of decoherence impacts were very pessimistic. The basic message was that decoherence is so large that the probability of getting a correct result decreases exponentially. Indeed, the error rate due to decoherence is time-dependent and approximately modeled by the function

$$1 - e^{-\frac{t}{\tau_{dec}}},$$

where τ_{dec} is the so-called decoherence time. Therefore, by Barenco (1996), if τ_{dec} is the typical decoherence time for a single qubit, the probability of getting the correct result for a quantum computation with the input of size n is

$$P \approx P_0 e^{-is(n)t(n)/\tau_{dec}},$$

where P_0 is the probability of the result with no errors, $s(n)$ is the total number of qubits necessary to perform the computation and $t(n)$ is the time needed to perform the computation. This implies, for example, that the decoherence problem cannot be efficiently dealt with by simply increasing the number of runs.

It has slowly turned out that a more detailed analysis of the decoherence problem can bring a different and less pessimistic view of decoherence, in some situations at least. For example, the decoherence problem is usually considered under the assumption that each of the qubits interacts with a different environment. However, if a different assumption is taken, for example that all qubits interact with the same environment, then it can be shown (Duan and Guo, 1996), that for some entangled initial states no decoherence occurs at all.

Table 7.1, due to DiVincenzo (1995), displays estimates for gate switching time t_s, decoherence time τ_{dec}, as well as the number of steps that can be performed without losing coherence, for several technologies of potential interest.

Experiments indicate (see DiVincenzo and Terhal, 1998), that using a single trapped beryllium ion, decoherence time is about 1 ms and with NMR technology around 1 s.

[2] In general, decoherence is often a process that eliminates off-diagonal elements of the density matrix of mixed states.

Quantum system	t_s	τ_{dec}	comput. steps
Mössbauer nucleus	10^{-19}	10^{-10}	10^9
GaAs electrons	10^{-13}	10^{-10}	10^3
Au electrons	10^{-14}	10^{-8}	10^6
Trapped indium ions	10^{-14}	10^{-1}	10^{13}
Optical microcavity	10^{-14}	10^{-5}	10^9
Electron spin	10^{-7}	10^{-3}	10^4
Electron quantum dot	10^{-6}	10^{-3}	10^3
Nuclear spin	10^{-3}	10^4	10^7

Table 7.1: Switching time t_s, decoherence τ_{dec}, both in seconds, and the number of computation steps performed before decoherence impacts occur

How to fight decoherence? Firstly, one should try to use technologies with low decoherence time. It is believed that there is still much to discover along these lines.

Secondly, it is well known that under certain circumstances decoherence is smaller. For example, some systems need more time to decohere at very low temperatures.

Ingenuity of experimental physicists is expected to bring significant improvements concerning decoherence. However, the main way to fight decoherence seems to be indirect one —a "software approach"—to try to avoid damage caused by decoherence by undoing its bad effects using, similarly as in the classical case, quantum error correcting codes, and other methods presented in Section 7.4.

The third main source of environmental impacts is in some sense inverse to the first one. They are impacts of external forces such as cosmic rays, or residual gas molecules, that can hit a qubit and change its state. These impacts are much out of our control.

7.3 Quantum Computation and Memory Stabilization

Efficient methods for the stabilization of quantum computations and memory are of crucial importance for computing and communication.

Redundancy is the main idea behind all methods of stabilization. In classical computing redundancy is utilized through redundant computations combined with the majority voting stabilization or by redundant coding using error detection and correction codes. Unfortunately, none of these methods can be used directly in the quantum setting.

The facts that quantum bits cannot be measured, without risking the collapse of their states, and cannot be copied, bring a completely new situation, compared with the classical case, into the problem of quantum error detection and correction. Even where we know what the state of a quantum bit should be we cannot test whether this is the case.

There is also no direct way to use the majority voting method. Indeed, even it is the case the states $|\phi_1\rangle, \ldots, |\phi_n\rangle$ are all equal, any measurement of them, with respect to an observable for which they are not eigenvectors, can each time give a different result. A majority voting method in quantum setting cannot therefore be used either to determine the majority state or to reset computations to the correct state. Because of that the idea of quantum error correcting codes looked hopeless. However, by utilizing in a novel way quantum entanglement, the field of quantum error correcting codes developed excellently.

In this section we present a sort of quantum analogue of the majority voting method to make use of the redundancy for quantum computation stabilization in a way that has no parallel in classical computing.

Before presenting the method itself there is still one point to discuss concerning stabilization: efficiency. Only such computation/memory stabilization methods are of interest that are efficient enough. Computations which require an exponentially increasing precision, or exponential amount of time, space or energy or of some other physical resource are considered as unfeasible. The same is true for polynomial resource-bounded computations, provided they do not have polynomial resource-ounded stabilization methods. From a quantum computing/memory stabilization technique it is therefore required that it makes polynomial resource-bounded computations out of any polynomial time-bounded computations.

Another fact of importance in this connection is that in some cases, for example for decoherence, error probability grows exponentially with the size of input. It is therefore of vital importance to determine whether we can have quantum stabilization that would make polynomial algorithms stay polynomial even under such unfavourable circumstances.

The majority voting method, brought to classical computing by von Neumann, is very successful in classical computing for a very simple reason: it is extremely good. If r computations are performed and each has probability $1-\theta$ of being successful, then by Chernoff bound the probability that the method fails is less than $e^{-\theta^2 r/6}$ and therefore the probability decreases exponentially with the number r of redundant computations.

Such a performance we can hardly expect from a quantum computation stabilization method. Fortunately, as will be shown, even much more modest stabilization methods may still be useful. Indeed, let us have a polynomial time algorithm that performs each step of the computation correctly with probability $1-\varepsilon$. After t steps the probability of successful computation is $(1-\varepsilon)^t \approx e^{-\varepsilon t}$. Suppose now that we have a computation stabilisation method available which, using redundancy r, reduces the error in each step modestly, by the factor $\frac{1}{r}$. After t steps the probability of success is $e^{-\frac{\varepsilon t}{r}} \approx 1-\delta$ if $r = \frac{\varepsilon t}{-\lg(1-\delta)}$, for any $\delta > 0$. In addition, r is polynomial in t and therefore also in the size of input.

We can therefore see that even with a moderately successful stabilization method, as the one presented in this section, which is due to Barenco et al. (1997), we can stabilize exponentially growing errors using only polynomial time computations.

7.3.1 The symmetric space

The very basic idea behind stabilization of computation by symmetrization is simple. In the case of r redundant computations in a Hilbert space H the resulting composed state will be, in the ideal error-free case, $\bigotimes_{i=1}^{r} |\psi\rangle$, i.e. a symmetric state that surely belongs to the smallest subspace of $H^{(r)} = \bigotimes_{i=1}^{r} H$ containing all such states. In the case of any erroneous computation, the state after r computations can be $\bigotimes_{i=1}^{r} |\psi_i\rangle$, with all $|\psi_i\rangle$ different, that is not symmetric. However, after adding additional "software redundancy" and making a projection to the smallest subspace containing all symmetric states[3] we can expect to get rid of erroneous computations.

Definition 7.3.1 *If H is an n-dimensional Hilbert space, then $\mathcal{SYM}_H^{(r)}$ is the smallest subspace of $H^{(r)} = \bigotimes_{i=1}^{r} H$ containing all states of the form $\bigotimes_{i=1}^{r} |\psi\rangle$, where $|\psi\rangle \in H$.*

[3]Projection always has to be done to a subspace, and states of the form $\bigotimes_{i=1}^{r} |\psi\rangle$, $|\psi\rangle \in H$ do not form by themselves a subspace of $H^{(r)}$.

The above definition does not bring a sufficient insight into the space $\mathcal{SYM}_H^{(r)}$—one does not see immediately that states $|\psi_1\rangle|\psi_2\rangle + |\psi_2\rangle|\psi_1\rangle$ are in $\mathcal{SYM}_H^{(2)}$ if $|\psi_1\rangle, |\psi_2\rangle \in H$. The following theorem offers an alternative and equivalent definition.

Theorem 7.3.2 *$\mathcal{SYM}_H^{(r)}$ is the subspace of all states in $H^{(r)}$ which are symmetric in the sense that they are unchanged under the interchange of sites for any pair of positions in the tensor product of their basic states.*

It is important that the subspace $\mathcal{SYM}_H^{(r)}$ is small. Its dimension is $\mathcal{O}(r^{n-1})$.

7.3.2 Stabilization by projection into the symmetric subspace

The basic idea of the stabilization method discussed below is that error-free components of states from $H^{(r)}$ are always in $\mathcal{SYM}_H^{(r)}$ and therefore by a projection into this subspace the error-free components remain to be included and the components with errors will be removed. This way not all errors will be removed because $\mathcal{SYM}_H^{(r)}$ contains also states not of the form $|\psi\rangle|\psi\rangle \ldots |\psi\rangle$. Nevertheless the error probability will be decreased by the factor of r.

The projection algorithm presented below is efficient: the number of operations performed is polynomial with respect to r and $\lg n$.

In the algorithm we assume that the set $\{\sigma_i\}_{i=1}^{r!}$ contains all permutations of r elements. Let $|\phi_1\rangle, \ldots, |\phi_r\rangle \in H^{(r)}$.

Algorithm 7.3.3 (Projection into $\mathcal{SYM}_H^{(r)}$)

1. *Apply to the first $\lceil \lg r! \rceil$ qubits of the register $|0^{(\lceil \lg r! \rceil)}, \phi_1, \ldots, \phi_r\rangle$. the transformation*

$$U : |0^{(\lceil \lg r! \rceil)}\rangle = \frac{1}{\sqrt{r!}} \sum_{i=0}^{r!-1} |i\rangle$$

2. *Apply, to the state $|\psi\rangle = \frac{1}{\sqrt{r!}} \sum_{i=0}^{r!-1} |i, \phi_1, \ldots, \phi_r\rangle$, the unitary transformation which transforms $|i\rangle|\phi_1\rangle \ldots |\phi_r\rangle$ into $|i, \phi_{\sigma_i(1)}, \ldots, \phi_{\sigma_i(r)}\rangle$' to get*

$$|\psi_1\rangle = \frac{1}{\sqrt{r!}} \sum_{i=0}^{r!-1} |i, \psi_{\sigma_i(1)}, \ldots, \phi_{\sigma_i(r)}\rangle.$$

3. *Apply U^{-1} to the first $\lceil \lg r! \rceil$ qubits of $|\psi_1\rangle$. The resulting state is then*

$$|\psi_2\rangle = \frac{1}{\sqrt{r!}} \sum_{i=0}^{r!-1} |i, \xi_i\rangle,$$

where $|\xi_i\rangle \in H^{(r)}$. Since U transforms $|0^{(\lceil \lg r! \rceil)}\rangle$ into an equal-amplitude superposition of all $|i\rangle$, U^{-1} transforms each $|i\rangle$ back to $|0^{(\lceil \lg r! \rceil)}\rangle$ with equal-amplitude. The coefficient at $|0^{(\lceil \lg r! \rceil)}\rangle$ will therefore be an equal amplitude superposition of all permutations of $|\phi_1\rangle, \ldots, |\phi_r\rangle$, i.e., the required symmetrized state.

4. *Measure the first $\lceil \lg r! \rceil$ qubits with respect to the standard basis. If the result is $0^{(\lceil \log r! \rceil)}$, then the state $|\psi_0\rangle = |\phi_1\rangle \ldots |\phi_r\rangle$ has been successfully projected into $\mathcal{SYM}_n^{(r)}$; otherwise the symmetrization failed.*

Due to the linearity of all processes the algorithm can be applied to the general state of $H^{(r)}$ and not only to a basis state as illustrated above.

Exercise 7.3.4 *Show that for the algorithm 7.3.3 the total number of operations is* $\mathcal{O}(r^2 \lg n + (r \lg r)^2)$ *and therefore the algorithm can be considered as efficient — in spite of the fact that we need to consider a creation of superpositions with* $r!$ *members.*

Exercise 7.3.5 *Design a symmetrization network of the size* $\mathcal{O}(r^2)$ *for* r *redundant computations.*

It has been shown by Barenco et al. (1997), by a detailed analysis, that both for unitary errors and for errors due to decoherence, the stabilization by symmetrization of r-redundant computations reduces the error in each step of stabilization by the factor $\frac{1}{r}$.

7.4 Quantum Error-Correcting Codes

No really significant progress in classical computing and communication would have been possible without error-correcting codes. They help to deal successfully and efficiently with the noise and disruptions one encounters in classical computing and communication.

Practically till 1995 it has been believed that error-correcting codes cannot be used in quantum computing, and this had been one of the main arguments of those with serious doubts concerning the overall feasibility of quantum computing. The reasoning against quantum error-correcting codes went as follows.

All states of the form

$$\alpha|0\rangle + \beta|1\rangle \quad \text{with} \quad |\alpha|^2 + |\beta|^2 = 1$$

are legal states. There are uncountably many such states. Since quantum evolution is basically an analogue and therefore a continuous process, and since all the above states are legal, there is no way to distinguish a state $|\phi\rangle$ from the one obtained from $|\phi\rangle$ by adding some noise. Moreover, the number of possible errors seemed to be infinite and it seemed that an error destroys a state in such a way that the original state cannot be recovered. In addition, each attempt to make a restriction to a discrete set of states seemed to bring an essential restriction to quantum computing.

However, it was shown first by Shor (1995) and Steane (1996), and soon by many others, that good quantum error-correcting codes exist and they can protect qubits against general types of error (which may be caused by imperfections or interactions with the environment). Quantum error-correcting codes have to be based on different principles than classical error-correcting codes, but they do exist. Since then the progress in the development of quantum error-correcting codes has been remarkable—this has been one of the most successfully developing areas of quantum computing and it is to a large extent due to these results, and results discussed in the next section, on fault-tolerant quantum computing, that the vision of real quantum computing is much closer.

The very basic idea of quantum computation with error-correcting codes goes as follows. The evolution of the quantum computer is restricted to a subspace of the Hilbert space carefully chosen in such a way that if quantum states are encoded using states of the chosen subspace, then all departures from this subspace, due to errors, lead to mutually orthogonal subspaces. After a quantum state is entangled with the environment and an "error" occurs, one can determine, by a measurement, but without destroying the erroneous state, into which of the error subspaces the erroneous state has fell, and the error can be undone using a unitary transformation.

However, it is far from trivial how to implement such an idea and how to utilize redundancy for that. (It is well known that redundancy is not very useful in analogue computing.) The ingenious idea of Shor and Steane was to use quantum entanglement for the design of quantum error-correcting codes (QECC). The discovery of QECC caused much excitement because it converted large-scale quantum computation from an impossibility to a possibility.

There are three new types of problem concerning quantum error-correcting codes, compared to the classical situation:

1. The variety of possible errors seems to be much broader.

2. The assumption that encoding and decoding are error-free is much less realistic.

3. Since decoherence decreases the probability of success of quantum computation exponentially, quantum error-correction codes need to have some "exponential efficiency". They have to be able "to fight exponential decoherence in polynomial time".

A desirable error correction process can be seen as having the following form: the sender, Alice, encodes a to-be-sent quantum state into a new quantum state which is then sent through a noisy channel on which an error/noise operator acts and changes the transmited state. Encodding has to be such that even if the error operator changes the state being trasmitted, it cannot entangle it with the environment and, as a consequence, the receiver Bob, who can act on the state he receives, but not on the environment, is able first to detect which error operator was applied and then he can undo its effect and to receive the original state.

As discussed in more detail in Chapter 8, in order Bob is able to recover the state Alice has sent, no information about her state should leak into the environment. Quantum codes have to be therefore such that they hide information from the environment. The idea is to use encodings of such types that encoded quantum information of k qubits is spread out over n qubits in a non-local way through an entangled state in such a way that environment which can access only a small number of qubits can gain no information about the overall state being transmited and this way transmited quantum information is protected.

Quantum error-correcting theory is a crucial part of quantum information theory. A variety of quantum error-correction methods have their analogues in classical error-correcting codes and rely heavily on their properties. That is why we start with a short summary of the very basic concepts and methods of the classical error-correcting codes codes. For more see Hill (1986), Hoffman et al. (1991), van Lint (1995) and MacWilliams and Sloane (1977).

7.4.1 Classical error-detecting and -correcting codes

The aim of the error-detecting and -correcting codes is to help to transmit safely messages through noisy channels using words of a q-ary alphabet $\Sigma_q = \{0, 1, \ldots, q-1\}$.

The **Hamming distance** of two words u and v, notation $hd(u, v)$, is the number of symbols in which u and v differ. The **(Hamming) weight**, $hw(w)$, of a codeword w is the number of symbols in w different from 0.

A **code** C is a subset of Σ_q^n for some n; its elements are called **codewords**. For error detection and correction the minimal distance $d(C)$ of a code C is of importance.

$$d(C) = \min\{hd(u,v) \mid u, v \in C, u \neq v\}.$$

The distance $d(C_1, C_2)$ of two codes C_1 and C_2 is defined by

$$d(C_1, C_2) = \min\{hd(w_1, w_2) \mid w_1 \in C_1, w_2 \in C_2\}.$$

This allows us to formulate one of the most basic results of the error-detecting and -correcting codes.

Theorem 7.4.1 *(1) A code C can detect up to s errors in any codeword if and only if $d(C) \geq s+1$; (ii) A code C can correct up to t errors if and only if $d(C) \geq 2t+1$.*

Definition 7.4.2 *An (n, M, d)*-**code** *is a code of M words of length n and minimal distance d. $A_q(n,d)$ denotes the largest M such that there exists a q-nary (n, M, d)-code.*

Exercise 7.4.3 *Show that $A_q(n,1) = q^n$, $A_q(n,n) = q$.*

The equivalence of codes is defined as follows:

Definition 7.4.4 *Two q-nary codes are called* **equivalent** *if one can be obtained from the other by a combination of the following operations:*

- *permutation of the positions of the codewords;*
- *permutation of the symbols appearing in a fixed position in the codewords.*

One of the aims of the coding theory is to find perfect codes. In order to define them let, for $u \in \Sigma_q^n$ and $r \geq 0$, $S(u,r) = \{v \in \Sigma_q^n \mid hd(u,v) \leq r\}$ be the sphere of the radius r.

Exercise 7.4.5 *Show that a sphere of radius $0 \leq r \leq n$ in Σ_q^n contains exactly*

$$\binom{n}{0} + \binom{n}{1}(q-1) + \binom{n}{2}(q-1)^2 + \ldots + \binom{n}{r}(q-1)^r$$

words.

Exercise 7.4.6 *Show that a q-nary $(n, M, 2t+1)$-code satisfies the following* **space packing** *or* **Hamming inequality (bound)***;*

$$M\{\binom{n}{0} + \binom{n}{1}(q-1) + \ldots + \binom{n}{t}(q-1)^t\} \leq q^n.$$

Definition 7.4.7 *An $(n, M, 2t+1)$-code C is called* **perfect**[4] *if the equality holds in 7.4.6.*

[4]it is already pretty well known which codes are perfect.

Exercise 7.4.8 *Explain why is perfect a perfect code.*

Extension of a code is an important technique to design new codes. A code C is extended by adding one bit to each codeword. For example, this bit, the parity check bit, is 0 if the weight of the code is even and 1 otherwise.

Several classes of codes are of special basic importance. The simplest of them is the $(n, 2, \lceil (n-1)/2 \rceil)$ **repetition code** $\{0^{(n)}, 1^{(n)}\}$. **Linear codes**, which we deal with next, are the most often used family of codes.

Linear codes

In order to formulate elegantly basic concepts and results of linear codes it is useful to consider words $u_1 u_2 \dots u_n$ of Σ_q^n as vectors $(u_1, u_2, \dots, u_n)$ of length n with elements from $\mathbf{Z}_q$. The set of all such vectors is denoted $V(n, q)$. Component-wise addition of two vectors and scalar/vector multiplications in $V(n, q)$ are done in $\mathbf{Z}_q$.

Definition 7.4.9 *A code C over $V(n, q)$ is called* **linear** *if C is a subspace of $V(n, q)$,qprime.*

Exercise 7.4.10 *Show that a subset $C \subseteq V(n, q)$ is a linear code if and only if: (1) $u + v \in C$ for all u, v in C; (2) $au \in C$ for all $u \in C$, $a \in \mathbf{Z}_q$.*

If the dimension $dim(C)$ of a linear code C in $V(n, q)$, as that of the subspace C, is k then C is said to be an $[n, k]$-code. In addition if C is of distance d, then it is said to be $[n, k, d]$**-code**. In other words, an $[n, k, d]$-code is a code by which n bits can store k bits of information in such a way that correction of up to $\lfloor (d-1)/2 \rfloor$ errors is always possible.

The **rate** of an $[n, k, d]$-linear code C is $\frac{k}{n}$. This is the ratio of the information content of a codeword to the information content of an arbitrary string of length n.

If C is a linear code, then $C^\perp = \{w \mid u \cdot w = 0 \text{ if } u \in C\}$ is called the **dual code** to C. A code C is **self-dual** if $C^\perp = C$.

Exercise 7.4.11 *Show that all codewords of a binary self-dual code have an even number of ones.*

Exercise 7.4.12 *Show that if C is an $[n, k]$ code over $\mathbf{Z}_q$, then $C^\perp$ is a $[n, n-k]$-code over $\mathbf{Z}_q$.*

Exercise 7.4.13 *Show that two $k \times n$ matrices generate equivalent linear $[n, k]$-codes over $\mathbf{Z}_q$ if one matrix can be obtained from the other by a sequence of operations of the following type: (1) permutation of rows; (2) multiplication of a row by a non-zero scalar; (3) addition of a scalar multiple of one row to another; (4) permutation of columns; (5) multiplication of any column by a non-zero scalar.*

A matrix G whose rows are all vectors of a basis of a linear code C (as a subspace) is said to be a **generator** of C. A generator matrix H of the dual code $C^\perp$ is called the **parity-check** matrix of C.

If G is a generator matrix of an $[n,k]$-code C, then $C = \bigcup_{v \in \{0,1\}^k} vG$. The name "parity-check matrix" is derived from the fact that a parity matrix H of a code C can be used to test whether a given word w is in C. Indeed, $w \in C$ if and only if $Hw^T = \mathbf{0}$.

The following theorem, easy to show, provides a simple way to construct a parity-check matrix of a linear code with a given generator matrix and vice versa.

Theorem 7.4.14 *If G is the generator matrix of an $[n,k]$-code C written in the form $G = [I_k|A]$, where I_k is the $k \times k$ unit matrix, then a parity-check matrix for C is $H = [-A^T|I_{n-k}]$.*

It follows from the definition of the parity-check matrix H of a code C that for each $w \in C$

$$wH^T = \mathbf{0^T}, \quad \mathbf{0} = Hw^T$$

and if $w \notin C$, $w = w_1 + w_e$, $w_1 \in C$, then $w_1H^T = \mathbf{0}^T, \quad w_1^TH = 0$. This means that the row space of H is orthogonal to C. In addition, $GH^T = \mathbf{0}^T, \quad \mathbf{0} = HG^T$.

Exercise 7.4.15 *Denote by $w(u)$ the* **(Hamming) weight** *of a word u. For a code C let $w(C) = \min\{w(u) \,|\, u \in C - \{0\}\}$. Show that $d(C) = w(C)$ for any linear code C.*

Encoding with linear codes. If C is an $[n,k]$-code over $\mathbf{Z}_q$ with a generator matrix G, then C contains q^n codewords and therefore it can be used to communicate up to q^k distinct messages. Let us identify messages with words in $V(n,q)$. Encoding of a message u is done by the matrix multiplication uG.

Syndrome decoding with linear codes is also easy, but several new concepts are needed to formulate an efficient algorithm.

Definition 7.4.16 *If C is an $[n,k]$-code over $\mathbf{Z}_q$ and a is any vector in $V(n,q)$, then the set $a + C = \{a + x \,|\, x \in C\}$ is called the* **coset** *of C. A vector of a coset with the minimum weight is its* **leader** *(which does not have to be unique).*

Exercise 7.4.17 *Suppose C is an $[n,k]$-code over $\mathbf{Z}_q$. Show: (a) every vector of $V(n,q)$ is in some coset of C; (b) any coset contains exactly q^k vectors; (c) two cosets of C are either identical or disjoint.*

Suppose H is a parity-check matrix of an $[n,k]$-code C. For any $y \in V(n,q)$ the row vector $S(y) = yH^T$ is called the **syndrome** of y (with respect to C).

As discussed above, if $w = w_1 + w_e$ with $w_1 \in C$, then $S(w) = S(w_e)$ and therefore the syndrome only depends on the word (vector) w_e. That means that a syndrome specifies an error without revealing anything about the codeword w_1 itself. This is an important property of the syndromes of linear codes that plays the key role in several quantum error correcting codes.

Exercise 7.4.18 *Show that two vectors are in the same coset if and only if they have the same syndrome.*

Decoding is now easy once we have constructed the so-called **standard array** for an $[n,k]$-code C. It is a $q^{n-k} \times q^k$ array of all vectors in $V(n,q)$. The first row contains all codewords of C starting with the word $0^{(n)}$. The first column contains leaders of all cosets. All other entries in the array are sums of elements in the first row of the corresponding column and in the first column of the corresponding row. In addition, one column is added with ith element being the syndromes for the cosets of the ith row.

Algorithm 7.4.19 (Syndrome decoding for linear codes) *Given a word y to decode do the following;*
1. *compute $S(y) = yH^T$;*
2. *Decode y as $y - l_y$, where l_y is the coset leader in the coset with the syndrome $S(y)$.*

In order to make decoding, or error correction, one needs to flip the erroneous bits, i.e., to apply NOT operation to them.

Exercise 7.4.20 *(Singleton bound) Show that if C is an (n, k, d) linear code, then $d \leq n - k + 1$.*

The Hamming bound, page 273, and Exercise 7.4.20 provide upper bounds on the size of linear (n, k, d)-codes with the given distance. The **Gilbert–Varshamov bound**

$$\sum_{j=0}^{d-2} \binom{n-1}{j} < 2^{n-k}$$

provides an achievable lower bound—for binary codes.

Hamming codes

The Hamming codes are interesting and important examples of the single error-correcting linear codes with easy encoding and decoding.

Definition 7.4.21 *Let $r \in \mathbf{N}^+$ and H be an $r \times (2^r - 1)$ matrix whose columns are distinct non-zero vectors of $V(r, 2)$. The code having H as its parity-check matrix is called a* **binary Hamming code** *and denoted by $Ham(r, 2)$.*

Example 7.4.22 *$Ham(3, 2)$ is the Hamming code with the parity-check matrices*

$$H = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix} \quad or \quad H' = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

By Theorem 7.4.14 a generator matrix for $Ham(3, 2)$ is

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

The main theoretical results on Hamming codes are summarized in the following theorem:

Theorem 7.4.23 *The Hamming code $Ham(r, 2)$ has the following properties: (1) it is a $[2^r - 1, 2^r - 1 - r]$-code; (2) it has minimum distance 3; (3) is a perfect code.*

Hence $Ham(3,2)$-code is an $[7,4,3]$-code. The dual to a Hamming code is called **simplex code** or **maximal-length feedback shift register code**.

Exercise 7.4.24 *Show that dual code to the Hamming* $[7,4,3]$*-code consists of Hamming code codewords of even weight.*

Dual of the extended Hamming code is also an important code: **first-order Reed–Muller code**. Since each Hamming code is linear, encoding with it is easy as described above. To describe decoding observe first that all coset leaders are exactly all vectors of weight ≤ 1. The syndrome of each such coset leader $(0,\ldots,1,\ldots,0)$ with 1 in the jth position is just the transpose of the jth column of H. Therefore, if the columns of H are arranged in order of the increasing binary numbers they represent, as in the example above, we have the following decoding method:

Algorithm 7.4.25 (Decoding algorithm for Hamming code)
Input*: a vector* y*.*
1. *Calculate the syndrome* $S(y) = yH^T$*.*

2.*If* $S(y) = 0$ **then** y *is assumed to be the codeword itself;*
else $S(y)$ *gives the binary representation of the error position.*

Example 7.4.26 *If* $y = 1110011$*, then* $S(y) = 001$ *and therefore* y *is decoded as* 0110011.

Cyclic codes

Definition 7.4.27 C *is a* **cyclic code** *if it is a linear code and if* $a_{n-1}a_{n-2}\ldots a_0 \in C$ *implies that* $a_0a_{n-1}\ldots a_1 \in C$*.*

Cyclic codes have an interesting algebraic structure. To see it let us identify a polynomial $a_0 + a_1x + a_2x^2 + \ldots + a_{n-1}x^{n-1}$ with each codeword $a_{n-1}a_{n-2}\ldots a_0$.

Moreover, denote by R_n^p the set of all polynomials of one variable over $\mathbf{Z}_p$, with p a prime, taken modulo the polynomial $x^n - 1$. For a polynomial $f(x) \in R_n^p$ denote $\langle f(x)\rangle = \{r(x)f(x) \,|\, r(x) \in R_n^p\}$. It holds:

Theorem 7.4.28 *(1) For any* $f(x) \in R_n^p$*, the set* $\langle f(x)\rangle$ *is a cyclic code (generated by* $f(x)$*). (2) If* C *is a non-zero cyclic code in* R_n^p*, then there is a polynomial* $g(x)$ *such that* $C = \langle g(x)\rangle$ *and* g *is a factor of* $x^n - 1$*.*

Two special types of cyclic codes are of importance.

Definition 7.4.29 *A cyclic code of codewords of length* n *over* $\mathbf{Z}_q$*,* $q = p^r$ *is called* **BCH code** *of distance* d *if its generator* $g(x)$ *is the least common multiple of the minimal polynomials of* ω^l*,* $\omega^{l+1},\ldots,\omega^{l+d-2}$ *for some* l*, where* ω *is the primitive* n*th root of unity.*[5] *If* $n = q^m - 1$ *for some* m*, then the BCH code is called primitive. A* **Reed–Solomon code** *is a primitive BCH code with* $n = q - 1$*. Reed-Solomon codes are self-dual.*

[5] A polynomial p is said to be minimal for a complex number $x \in \mathbf{Z}_q$ if $p(x) = 0$ and p is irreducible over $\mathbf{Z}_q$.

Classical codes as quantum states

A binary code $C = \{c_1, \dots, c_m\}$ of codewords of length n specifies a subspace H_C of Hilbert space H_{2^n} spanned by the vectors $\{|c_1\rangle, \dots, |c_m\rangle\}$ and can be represented by the quantum state

$$|\phi_C\rangle = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} |c_i\rangle.$$

If $|\phi_C\rangle$ is represented in the dual basis, then this representation can be seen as defining a new code—C_{dual}. Let C_{dual} have m' codewords and distance d' and let C have distance d. Clearly $m \leq A(n,d)$, $m' \leq A(n,d')$. According to the inequality 2.5, page 73, we have

$$A(n,d)A(n,d') \geq 2^n.$$

If C is a $[n,k]$-linear code, then C_{dual} has a neat form, which is easy to show:

Theorem 7.4.30 *If C is a linear code, then $C_{dual} = C^{\perp}$.*

According to Exercise 7.4.12, C_{dual} is a $[n, n-k]$-code with 2^{n-k} codewords. Linear codes are therefore codes for which the equality holds in the inequality (2.5).

Classical error-correction techniques cannot be directly applied to quantum information processing for two main reasons: (1) it is not possible, in general, to copy or measure qubits without causing undesirable effects; (2) it is not sufficient to correct 0/1 values of qubits—also amplitudes need to be preserved and this is a completely new feature with which quantum error-correcting techniques have to be able to deal.

7.4.2 Framework for quantum error-correcting codes

In order to be able to deal with the problem of correcting the effects of noise (errors) at the communication through noisy channels we need a formal description of the noise or error process (see Steane, 1998, and Knill and Laflamme, 1997).

Error model

Inaccuracies, noise and decoherence can be described in terms of the most general quantum operators—superoperators—or, equivalently, in terms of the unitary operators on the system and its environment.

There is a large variety of possible quantum errors. [6] However, to consider QECC successfully it is quite sufficient to make several simple, but (quasi-)realistic assumptions concerning the character, frequency and types of errors. [7]

We shall assume that errors, due to decoherence and inaccuracies, on different qubits or on the same qubit in different times are random and statistically uncorrelated. Namely, that they are **locally independent** (errors in different qubits or gates are not correlated) and **sequentially independent** (subsequent errors on the same qubit or in the same gate are not correlated). [8] (In other words, it is assumed that there are no interactions between

[6] The term "error" is used here in a special way. As pointed out by Peres (1996) "A computer is a physical system, subject to the laws of nature. No errors occur in the application of these laws. What we call an error is a mismatch between what the computer does and what we wanted it to do."

[7] Without assumptions on how error occurs it is not possible to prove nontrivial results on error correction.

[8] If additional information is available about the error process, more efficient quantum error-correcting procedures can be developed to deal with errors of such processes. See, for example, Plenio, Vedral and Knight (1996), for the case that the error process is a spontaneous emission.

environments of different qubits and also between environments of the same qubit at different time steps.) No knowledge about their nature will be assumed. [9] (As a consequence, an error operator on n qubits can be written at each time step as a tensor product of errors on particular qubits.) If these conditions are satisfied, then it is believed that errors are correctable provided the error rate is below 10^{-5} per qubit and clock cycle (DiVincenzo and Terhal, 1998). As a consequence, the above error model implies that a correlation between errors on different qubits can exists only in the case of qubits interacting through a quantum gate.

Error types

A general interaction between a qubit $\alpha|0\rangle + \beta|1\rangle$ and its environment leads to the evolution of the form:

$$|e\rangle(\alpha|0\rangle + \beta|1\rangle) \rightarrow \alpha(|e_{00}\rangle|0\rangle + |e_{01}\rangle|1\rangle) + \beta(|e_{11}\rangle|1\rangle + |e_{10}\rangle|0\rangle), \tag{7.3}$$

where $|e\rangle$, $|e_{00}\rangle$, $|e_{01}\rangle$, $|e_{10}\rangle$ and $|e_{11}\rangle$ are states of the environment.

The right-hand side of (7.3) can now be written in the form

$$(|e_{0+}\rangle I + |e_{0-}\rangle\sigma_z + |e_{1+}\rangle\sigma_x - |e_{1-}\rangle i\sigma_y)(\alpha|0\rangle + \beta|1\rangle), \tag{7.4}$$

where

$$|e_{0+}\rangle = \frac{1}{2}(|e_{00}\rangle + |e_{10}\rangle) \qquad |e_{0-}\rangle = \frac{1}{2}(|e_{00}\rangle - |e_{10}\rangle), \tag{7.5}$$

$$|e_{1+}\rangle = \frac{1}{2}(|e_{01}\rangle + |e_{11}\rangle) \qquad |e_{1-}\rangle = \frac{1}{2}(|e_{01}\rangle - |e_{11}\rangle), \tag{7.6}$$

and $\sigma_x, \sigma_y, \sigma_z$ are Pauli matrices.

The key ideas behind quantum error-correction codes can be seen looking carefully on the state (7.4): (1) any error can be seen as being composed of four basic errors and therefore if we are able to correct any of these four types of error, we can correct any error; (2) error model resembles more a discrete one than a continuous one; (3) the resulting state of the environment is independent of the state on which an error process acts and depends only on the type of error operator being applied. This also suggest the following error-detection and -recovery process: (1) To compute which type of error has occurred (error "syndrome" is computed); (2) to undo errors.

The impact of Pauli matrices on a qubit $|\phi\rangle = \alpha|0\rangle + \beta|1\rangle$ is shown in Figure 7.1.

$\lvert\phi\rangle$	$\sigma_x\lvert\phi\rangle$	$\sigma_y\lvert\phi\rangle$	$\sigma_z\lvert\phi\rangle$	$i\sigma_x\sigma_y\lvert\phi\rangle$
$\alpha\lvert 0\rangle + \beta\lvert 1\rangle$	$\alpha\lvert 1\rangle + \beta\lvert 0\rangle$	$i\alpha\lvert 1\rangle - i\beta\lvert 0\rangle$	$\alpha\lvert 0\rangle - \beta\lvert 1\rangle$	$\alpha\lvert 1\rangle - \beta\lvert 0\rangle$

Figure 7.1: Impacts of quantum errors

σ_x therefore stands for the **bit (flip) error** (or "amplitude error"), σ_z for the **sign (flip) error** (or "phase (shift)" error). σ_y and $\sigma_x\sigma_z$ for a **bit-sign (flip) error** (or "bit-phase" error)—a combination of the bit error and the phase error.

[9] Sometimes the so-called **no leakage** assumption is made: a physical system which implements a qubit has access only to the two-dimensional Hilbert space defined by the qubit. A photon with two basis states represented by the horizontal and vertical polarizations is an example of a system which, without modifications, does not satisfy this assumption (photons have a tendency to be scattered or absorbed and in this way lost for the computation).

Observe that $\sigma_y = i\sigma_x\sigma_z$. That is why sometimes a slightly different error model is used, with three types of errors represented by matrices $X = \sigma_x$, $Z = \sigma_z$ and $Y = \sigma_x\sigma_z$.

In the case of an n-qubit register the general type of error is therefore represented by the matrix

$$M = \bigotimes_{i=1}^{n} M_i,$$

where $M_i \in \{X, Y, Z, I\}$ is an error matrix for the ith qubit.

Example 7.4.31 *Let us explore perhaps the simplest idea for a quantum error-correcting code (Peres, 1985, Aharonov, 1998), namely encoding of the basis states*

$$|0\rangle \rightarrow |000\rangle \qquad |1\rangle \rightarrow |111\rangle,$$

which results in the encoding of the general one-qubit state $\alpha|0\rangle + \beta|1\rangle \rightarrow \alpha|000\rangle + \beta|111\rangle$. Unfortunately, such an encoding is not good enough because it does not protect the quantum state even against one error. Indeed, let us assume that a noise operator operates on the first qubit and the environment $|e\rangle$ in such a simple way that it does not changes the first qubit, but it changes the environment depending on the value of the qubit:

$$|e\rangle(\alpha|000\rangle + \beta|111\rangle) \rightarrow \alpha|e_0\rangle|000\rangle + \beta|e_1\rangle|111\rangle.$$

The resulting state is entangled with the environment and Bob cannot disentangled it by a local action and cannot recover the original state.

Basic ideas of quantum error correction

Suppose we want to encode uniquely k qubits into $n > k$ qubits. This means we want to map, using a unitary transformation, any state $|\phi\rangle$ over k qubits into a state over n qubits. This can be done if we introduce $n-k$ new qubits, initially in the state $|0^{(n-k)}\rangle$, for example, and then apply a unitary encoding transformation E as follows.

$$E(|\phi\rangle|0^{(n-k)}\rangle) \rightarrow |\phi_E\rangle,$$

where $|\phi_E\rangle$ is said to be the **quantum code** of $|\phi\rangle$ defined by E. The encodings, or of the basis states of k qubits are called **codewords** and they form an orthonormal basis of a 2^k-dimensional subspace of H_{2^n}.

If an error occurs in a codeword $|\phi_E\rangle$, then $|\phi_E\rangle$ is altered by some linear transformation, superoperator, $\mathcal{E}$ and

$$|\phi_E\rangle \stackrel{\mathcal{E}}{\rightarrow} |\mathcal{E}\phi_E\rangle.$$

($\mathcal{E}$ is not required to be unitary; there is a need to correct also non-unitary errors.)

An error-correction process (ECP) can now be modeled by unitary transformations that first entangle the erroneous state $|\mathcal{E}\phi_E\rangle$ with an **ancilla** (an auxiliary state of auxiliary qubits), and then transform the resulting entangled state into a tensor product of the original state $|\phi_E\rangle$ and a new state $|A_\mathcal{E}\rangle$ of the ancilla:

$$|\mathcal{E}\phi_E\rangle|A\rangle \stackrel{ECP}{\rightarrow} |\phi_E\rangle|A_\mathcal{E}\rangle.$$

Since the state $|\phi_E\rangle|A_\mathcal{E}\rangle$ is not entangled we can measure $|A_\mathcal{E}\rangle$ without disturbing $|\phi_E\rangle$ and this way we can determine a transformation which has to be applied to $\mathcal{E}|\phi_E\rangle$ to get $|\phi_E\rangle$.

Let us now look into the error-creation and -correction process in more detail for the important case where erroneous states have the form

$$\sum_{s=1}^{l} M_s|\phi_E\rangle \quad \text{or} \quad \sum_{s=1}^{l} |\psi^s_{env}\rangle M_s|\phi_E\rangle, \tag{7.7}$$

where each M_s is a tensor product of n error matrices from the set $\{X, Y, Z, I\}$ (and it is called an an "**error operator**" or an "**error**" and $|\psi^s_{env}\rangle$ are states of the environment. (As discussed more in Section 7.4.5, such error operators generate a group which will be denoted by $\mathcal{G}_n$.)

The basic task is to determine, without disturbing $|\phi_E\rangle$ in an irreversible way, an operation that has to be performed in order to get $|\phi_E\rangle$ out of $\mathcal{E}|\phi_E\rangle$. The basic idea is to compute, as in the case of linear codes, syndromes of errors without disturbing $|\phi_E\rangle$. In order to do that additional qubits of ancilla are introduced in a special initial state, for example in the state $|0^{(n-k)}\rangle$. In order to compute syndromes, a syndrom-extraction operator S is applied with the effect

$$S(M_s|\phi_E\rangle|0^{(n-k)}\rangle) = M_s|\phi_E\rangle|s\rangle$$

for all M_s from a set of "correctable errors" (which depends on the encoding). Here s is just a number indicating with which error operator we are dealing. (States $|s\rangle$ have to be orthogonal!) Where the noisy state (7.7) contains only correctable errors, after the syndrome extraction operator is applied we get as the resulting state

$$\sum_{s=1}^{l} |\psi^s_{env}\rangle(M_s|\phi_E\rangle|s\rangle). \tag{7.8}$$

Since the states $|s\rangle$ are orthogonal we can measure the ancilla qubits in the basis $\{|s\rangle\}$ to get:

$$|\psi^{s_0}_{env}\rangle(M_{s_0}|\phi_e\rangle|s_0\rangle)$$

for a single, randomly chosen, s_0. This is excellent: instead of a complicated erroneous state we have now only one error operator M_{s_0} and by applying $M^{-1}_{s_0}$ we get as the result the state $|\psi^{s_0}_{env}\rangle|\phi_E\rangle|s_0\rangle$. Therefore, the state $|\phi_E\rangle$ has been reconstructed—it is no longer entangled.

As shown above, the quantum error-correcting processes processes are, surprising, discrete processes, not continuous ones. The key discretization step is projection measurement.

Remark 7.4.32 Actually, it is not necessary to measure ancilla qubits to get the syndrome. Indeed, after the syndrome extraction one can apply a unitary operator C such that $C(|x\rangle|s\rangle) = M_s^{-1}|x\rangle|s\rangle$ to the sum in (7.8). The final state then would be $|\phi_E\rangle\sum_{s=1}^{l}|\psi^s_{env}\rangle|s\rangle$—entanglement between the state and the environment is transferred into the entanglement between the environment and the ancilla.

.

A quantum t error-correcting code code is a (unitary) mapping of k qubits into a subspace of a quantum space of $n > k$ qubits such that errors in any of the t qubits can be corrected; i.e. the original quantum state can be perfectly recovered from the remaining $n - t$ qubits.

The simplest case to consider is $k = 1$, even it is quite clear that "more efficient" QECC are expected to exist for $k > 1$.

Sufficient and necessary conditions for quantum error-correcting codes

Let $\mathcal{S}_\mathcal{E}$ be a set of errors and we are interested to find necessary and sufficient conditions for an error-correcting code to be able to correct all errors from $\mathcal{S}_\mathcal{E}$.

In order to be able to correct any two errors E_a and E_b from $\mathcal{S}_\mathcal{E}$ we must be able to distinguish the case E_a is acting on a basis codeword $|\psi_i\rangle$ from the case E_b is acting on another basic codeword $|\psi_j\rangle$, $i \neq j$. This can be done for sure if and only if the states $E_a|\psi_i\rangle$ and $E_b|\psi_j\rangle$ are orthogonal. Thus, it has to hold

$$\langle\psi_i|E_a^* E_b|\psi_j\rangle = 0 \tag{7.9}$$

if $i \neq j$ and E_a, E_b are any errors from $\mathcal{S}_\mathcal{E}$. In other words, errors on different basis codewords have to result in orthogonal states.

How about different errors on the same basis codeword? We could require the same condition to hold. Namely, that (7.9) holds also for $i = j$ and all E_a, E_b from $\mathcal{S}_\mathcal{E}$. However, the condition (7.9) is too strong.

What is needed for an error-correcting code is that when we make a measurement to find out about an error, we must learn nothing about the actual state of the coding space on which the error was made. Would we learn something we could damage the superposition of the basis states and we would not be able to correct an arbitrary codeword. How we learn information about a codeword? By measuring $\langle\psi_i|E_a^* E_b|\psi_i\rangle$. This value has therefore to be the same for all basis codewords. In other word, for any correctable errors (i.e. from $\mathcal{S}_\mathcal{E}$) E_a and E_b and any $i \neq j$ it has to hold:

$$\langle\psi_i|E_a^* E_b|\psi_i\rangle = \langle\psi_j|E_a^* E_b|\psi_j\rangle. \tag{7.10}$$

Conditions (7.9) and (7.10) are therefore necessary for a code to be able to correct a given set $\mathcal{S}_\mathcal{E}$ of errors. Actually, they are also sufficient conditions as shown by Knill and Laflamme (1997) and Bennett et al. (1996a). In this case the proof is more subtle and we sketch here only its basic ideas.

Conditions (7.9) and (7.10) can be combined into a single condition

$$\langle\psi_i|E_a^* E_b|\psi_j\rangle = c_{a,b}\delta_{ij}, \tag{7.11}$$

where c_{ab} is a constant independent on i and j. Let M be a matrix such that $M[a,b] = c_{ab}$. The matrix is Hermitian and therefore it can be diagonalized. If this is done and errors $\{E_a\}$ are properly transformed to a get a new basis $\{F_a\}$ for the space of possible errors, we get that

$$\langle\psi_i|F_a^* F_b|\psi_j\rangle = \delta_{ab}\delta_{ij} \tag{7.12}$$

or

$$\langle\psi_i|F_a^* F_b|\psi_j\rangle = 0, \tag{7.13}$$

depending on a and b. (The new basis does not have to contain only errors that are tensor products of errors from the set $\{X, Y, Z, I\}$). Errors of the type (7.13) annihilate any codeword and therefore the probability they occur is 0. The other errors always produce orthogonal states and therefore we can detect them by a measurement and then they can be corrected. Hence the code satisfies the condition (7.11) if and only if it can correct all correctable errors from $\mathcal{S}_\mathcal{E}$.

If for all errors E_a and E_b and any basis states $|\psi_i\rangle, |\psi_j\rangle$,

$$\langle\psi_i|E_a^* E_b|\psi_j\rangle = 0 \tag{7.14}$$

then the code is called **orthogonal** or **non-degenerate**, otherwise it is called **degenerate**.

Remark 7.4.33 Observe that whether a code is degenerate or not depends on the set of errors that is expected to get corrected. For example, a degenerate two error-correcting code can be nondegenerate when considering it as a one error-correcting code.

Exercise 7.4.34 *How should condition (7.11) be changed would we want from a code only the capability to detect an error or a set of errors?*

Bounds on quantum error-correcting codes

As for linear codes, one can derive a quantum version of the Hamming bound on parameters n, k, t for quantum non-degenerate codes mapping k qubits into n and correcting up to t errors (see Steane, 1996, and Ekert and Macchiavello, 1996).

There are 2^k basis states of k qubits. Since there are 3 possible errors (X or Y or Z) the number of possibilities for having i errors on the codeword of n qubits is $3^i\binom{n}{i}$ and for all basis states and for i from 0 to t there are $2^k\sum_{i=0}^{t} 3^i\binom{n}{i}$ possible error states. If the code is non-degenerate all error states obtained from the original codeword have to be orthogonal. Since the overall dimension of the space of n qubits is 2^n we have the basic bound:

$$2^k \sum_{i=0}^{t} 3^i \binom{n}{i} \leq 2^n. \tag{7.15}$$

and for large n and t

$$\frac{k}{n} \leq 1 - \frac{t}{n}\lg_2 3 - H(\frac{t}{n}),$$

where H is Shannon entropy, and therefore $\frac{k}{n}$ is close to 1 if $\frac{t}{n} \approx 0.1892$.

For the case $k = 1$ and $t = 1$ we have, as a special case, the bound

$$2(3n+1) \leq 2^n \tag{7.16}$$

and 5 is the minimal n satisfying (7.16). Three questions now immediately come up:

1. Is the bound 7.15 satisfied also by non-degenerate (non-orthogonal) codes?
2. Are there infinitely many codes that can reach the bound (i.e., codes for which equality holds in (7.15))?
3. Can we get bound (7.15) into a more transparent form?

The first problem is still open. Concerning the second problem, Gottesman (1996) has constructed infinitely many codes reaching the bound. Concerning the third problem, attention has focused on the task to determine the best possible rate $\frac{k}{n}$.

Calderbank and Shor (1995) have shown quantum error-correcting codes with asymptotic rate $\frac{k}{n} = 1 - H(\frac{2t}{n})$, and have derived, on the basis of the results of Bennett et al. (1996a), concerning the bounds on the capacity of quantum channels, the upper bound

$$\min\left[1 - H_2\left(\frac{2t}{3n}\right), H_2\left(\frac{1}{2} + \sqrt{(1-\frac{t}{n})\frac{t}{n}}\right)\right],$$

for $\frac{t}{n} < \frac{1}{2}$ and 0 for $\frac{t}{n} \geq \frac{1}{2}$.

Bit versus sign errors

There is a simple relation between bit errors, represented by the matrix $X = \sigma_x$ and the phase error, represented by the matrix $Z = \sigma_z$. Namely,

$$Z = HXH \qquad \text{and} \qquad X = HZH \tag{7.17}$$

where H is the Hadamard matrix, an application of which transforms the states expressed in the standard basis to the dual basis and vice versa. In other words a sign error in the standard basis is the bit error in the dual basis and vice versa.

Together with the fact that the classical error-correcting codes are used to correct bit errors, the identities (7.17) suggest a simple approach to quantum error correction for the case that only bit and sign errors are expected.

A classical error-correcting code C_1 is used to correct bit errors. In order to correct sign errors the erroneous state is first transformed to the dual base, sign errors are then corrected, if by the basis transformation we get codewords with support from some error-correcting code C_2. The resulting state is then transformed back to the standard basis. This will be described later.

Notation

$\{n, k, d_1, d_2\}$ is used to denote quantum codes C that with n qubits encode k qubits and $\lceil (d_1 - 1)/2 \rceil$ ($\lceil (d_2 - 1)/2 \rceil$) bit (sign) errors can be corrected by the code. Notation $\{n, k, d\}$, or $[[n, k, d]]$, is used for $\{n, k, d, d\}$ codes. Finally, notation $\{n, k, d\}^+$ denotes codes codewords of which are superpositions all basis states of which have as amplitude 1.

Remark 7.4.35 Interestingly enough, quantum error-correcting codes are able to fight decoherence caused by entanglement of quantum systems with the environment, using again entanglement as the main tool. As pointed out by Preskill (1998), "we can fight entanglement with entanglement".

7.4.3 Case studies

Design and use of quantum error-correcting codes brings a variety of new problems and methods. This will now be illustrated using some of the first good quantum error-correcting codes—see table in Figure 7.2.

Code	$\|0_E\rangle$	$\|1_E\rangle$
Shor's 9 qb code	$(X)(X)(X)$ $X = \|000\rangle + \|111\rangle$	$(Y)(Y)(Y)$ $Y = \|000\rangle - \|111\rangle$
Steane's 7 qb code	$\|0000000\rangle + \|1010101\rangle + \|0110011\rangle$ $+\|1100110\rangle + \|0001111\rangle + \|1011010\rangle$ $\|0111100\rangle + \|1101001\rangle$	$\|1111111\rangle + \|0101010\rangle + \|1001100\rangle$ $+\|0011001\rangle + \|1110000\rangle + \|0100101\rangle$ $+\|1000011\rangle + \|0010110\rangle$
LMPZ's 5 qb code	$+\|00000\rangle + \|11100\rangle - \|10011\rangle - \|01111\rangle$ $+\|11010\rangle + \|00110\rangle + \|01001\rangle + \|10101\rangle$	$-\|00011\rangle + \|11111\rangle - \|10000\rangle + \|01100\rangle$ $+\|11001\rangle - \|00101\rangle - \|01010\rangle + \|10110\rangle$
Barenco's 3 qb code	$\|000\rangle + \|011\rangle + \|101\rangle + \|110\rangle$	$\|111\rangle + \|100\rangle + \|010\rangle + \|001\rangle$

Figure 7.2: Examples of 1-qubit quantum error-correcting codes; all superpositions are equally weighted, but amplitudes are omitted in the table

Quantum error-correcting code design

The first problem is how to design a good QECC. Several general methods of designing quantum error-correcting codes are presented in the Sections 7.4.4 and 7.4.5. Here we discuss only how codes presented in Figure 7.2 were (or it could be) discovered. Shor's 9-qubit code was historically first (see Shor, 1995), Steane's 7-qubit code next (Steane, 1996) and LMPZ's code (see Laflamme, Miquel, Paz and Zurek (1996)[10] is the best possible code, from the point of view of the minimal number of qubits and also codewords of the support, see inequality (7.15), which can correct all three types of errors when one qubit is encoded. A simple 3-qubit code shown in Figure 7.2 is good for those case where only sign errors occur.

> **Exercise 7.4.36** *(Shor, 1995) (a) Show that if p is the probability of one qubit error, then the probability that at least 2 qubits out of 9 are erroneous is approximately $36p^2$; (b) Show that if k qubits are encoded using Shor's 9-qubit code, then the probability that $9k$-qubits can be decoded is $(1-36p^2)^k$; (c) determine probabilities as in the cases (a) and (b) for LMPZ's code.*

Shor's code was obtained from the repetition code $\{|000\rangle, |111\rangle\}$ by replacing $|0\rangle$ with $\frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$ and $|1\rangle$ with $\frac{1}{\sqrt{2}}(|000\rangle - |111\rangle)$, two states of the Bell's basis in H_8:

$$\frac{1}{\sqrt{2}}(|000\rangle \pm |111\rangle), \frac{1}{\sqrt{2}}(|001\rangle \pm |100\rangle), \frac{1}{\sqrt{2}}(|010\rangle \pm |101\rangle), \frac{1}{\sqrt{2}}(|010\rangle \pm |101\rangle)$$

The starting classical code for Steane's code was the simplex $(7, 8, 4)$-code

$$C \;=\; \{0000000, 1010101, 0110011, 1100110, 0001111, 1011010, 0111100, 1101001\}.$$

Using the standard-to-dual basis transformation we get $C_{dual} = C^{\perp} = Ham(3, 2)$—page 276. The resulting Steane's code is now obtained as

$$|0_E\rangle = \sum_{\substack{c \in Ham(3,2) \\ c \text{ even}}} |c\rangle, \qquad |1_E\rangle = \sum_{\substack{c \in Ham(3,2) \\ c \text{ odd}}} |c\rangle,$$

The point is that the code C has distance 4 and $C^{\perp}$ has still large enough distance, 3, to correct one bit error. Observe also that in both codes, $|0_E\rangle$ and $|1_E\rangle$, the last three bits of any codeword uniquely determine the first four. (They are actually the parity bits.)

LMPZ's code was obtained experimentally, by analysis of the orthogonality conditions, as discussed in the next subsection, and by a computer search (Laflamme et al. 1996).

Finally, Barenco's 3-qubit code can be obtained from the repetition code $\{000, 111\}$, using the Hadamard rotation on all three qubits providing:

$$\begin{aligned}
|0\rangle \rightarrow |000\rangle \quad &\rightarrow \quad \frac{1}{\sqrt{8}}(|000\rangle + |001\rangle + |010\rangle + |011\rangle + |100\rangle + |101\rangle + |110\rangle + |111\rangle) \\
&= \quad \frac{1}{\sqrt{2}}(|0_E\rangle + |1_E\rangle), \\
|1\rangle \rightarrow |111\rangle \quad &\rightarrow \quad \frac{1}{\sqrt{8}}(|000\rangle - |001\rangle - |010\rangle + |011\rangle - |100\rangle + |101\rangle + |110\rangle - |111\rangle)
\end{aligned}$$

[10] LMPZ code is equivalent up to a change of basis of individual qubits to the code independently discovered by Bennett et al. (1996a).

$$= \frac{1}{\sqrt{2}}(|0_E\rangle - |1_E\rangle).$$

Exercise 7.4.37 *Determine which bit, sign and bit–sign errors map the following codes into mutually orthogonal states: (a) Shor's* 9*-qubit code; (b) Steane's* 7*-qubit code; (c) LMPZ's* 5*-qubit code; (d) Barenco's* 3*-qubits code.*

Encoding circuits—encoders

To use a quantum code with mappings $|0\rangle \rightarrow |0_E\rangle$, $|1\rangle \rightarrow |1_E\rangle$, a quantum circuit is needed to transform an arbitrary quantum state $\alpha|0\rangle + \beta|1\rangle$ into the state $\alpha|0_E\rangle + \beta|1_E\rangle$. Encoding circuits for Steane's and LMPZ's codes are shown in Figures 7.3a,b.

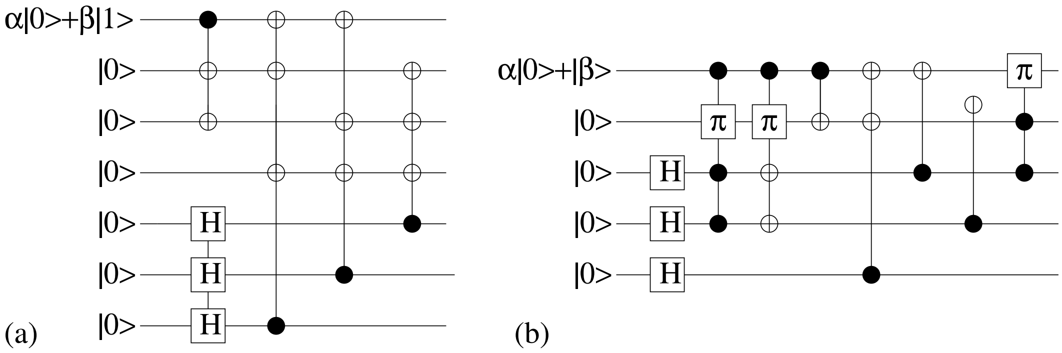


Figure 7.3: Encoding circuits for the Steane's and LMPZ's codes; π-gate realizes π-rotation

The circuit to produce Steane's code is simple (Preskill, 1998) and makes use of the fact that codewords of the support of $|1_E\rangle$ can be obtained from the codewords of the support for $|0_E\rangle$ by flipping the bits. The first two XORs produce the state $\alpha|0000000\rangle+\beta|1110000\rangle$. The three Hadamard gates produce equally weighted superposition of all eight possible values for the last three bits. These three bits uniquely determine the first four bits and this is taken care by the remaining multiple XOR gates.

Encoding circuit for LMPZ's code is more tricky (Laflamme et al. 1996). The first three bits of each codeword uniquely determine the last two in each of the codewords. This is then easy to implement using multiple XOR gates. The problem is with the signs, and they are taken care by π-gates that represent rotation by π (i.e. multiplication by $e^{i\pi}$).

Exercise 7.4.38 *Design encoding circuits for: (a) Shor's* 9 *qubit code; (b) Barenco's* 3*-qubit code.*

Error syndrome computation

Efficient syndrome computation is the key problem in using quantum error-correcting codes.

In the case of Shor's code, Shor (1995), it is instructive to see the impact of the decoherence process

$$|e_0\rangle|0\rangle \to |a_0\rangle|0\rangle + |a_1\rangle|1\rangle, \qquad |e_1\rangle|1\rangle \to |b_0\rangle|0\rangle + |b_1\rangle|1\rangle,$$

where $|a_0\rangle$. $|a_1\rangle$, $|b_0\rangle$, $b_1\rangle$ are states of the environment, on the first qubit of the states $\frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$ and $\frac{1}{\sqrt{2}}(|000\rangle - |111\rangle)$. The resulting states are shown in Figure 7.4a,b.

$$\begin{array}{ll} \frac{1}{\sqrt{8}}(|a_0\rangle + |b_1\rangle)(|000\rangle + |111\rangle) & \frac{1}{\sqrt{8}}(|a_0\rangle + |b_1\rangle)(|000\rangle - |111\rangle) \\ +\frac{1}{\sqrt{8}}(|a_0\rangle - |b_1\rangle)(|000\rangle - |111\rangle) & +\frac{1}{\sqrt{8}}(|a_0\rangle - |b_1\rangle)(|000\rangle + |111\rangle) \\ +\frac{1}{\sqrt{8}}(|a_1\rangle + |b_0\rangle)(|100\rangle + |011\rangle) & +\frac{1}{\sqrt{8}}(|a_1\rangle + |b_0\rangle)(|100\rangle - |011\rangle) \\ +\frac{1}{\sqrt{8}}(|a_1\rangle - |b_0\rangle)(|100\rangle - |011\rangle) & +\frac{1}{\sqrt{8}}(|a_1\rangle - |b_0\rangle)(|100\rangle + |011\rangle) \\ \text{(a)} & \text{(b)} \end{array}$$

Figure 7.4: Decoherence impact on Shor's code

Observe an important fact that the states of the environment $(|a_0\rangle + |b_1\rangle, \ldots)$, are the same for both error states arising from encodings of $|0_E\rangle$ and $|1_E\rangle$.

It is now clear that using XOR gates with several ancillary qubits and comparing three triplets one can easily determine whether an error occurred and in which qubit. By a measurement with respect to the Bell basis the type of error can be determined. (By such a measurement the overall state collapses but what remains is good enough to restore the initial state.)

In the case of Steane's code one can make use of the way syndrome is computed for Hamming codes, page 277. Namely, the matrix H, page 276, shows that the error syndrome for the bit error is uniquely determined by three bits: $(b_4 \oplus b_5 \oplus b_6 \oplus b_7, b_2 \oplus b_3 \oplus b_6 \oplus b_7, b_1 \oplus b_3 \oplus b_5 \oplus b_7)$, where $(b_1, \ldots, b_7)$ denote the seven bits of the codewords.

To correct also sign errors the code has to be transformed first from the standard to the dual basis. The key point is now that the resulting classical code of codewords is $Ham(3, 2)$ or $(7, 4, 3)$-code. Now

$$|0'_E\rangle = \frac{1}{\sqrt{2}}(|0_E\rangle + |1_E\rangle), \qquad |1'_E\rangle = \frac{1}{\sqrt{2}}(|0_E\rangle - |1_E\rangle),$$

and the same technique as above can be used to detect sign errors (and consequently also bit–sign errors). Once this is done the code is transformed to the standard basis.

Exercise 7.4.39 *Design error syndrome computation network for Steane's code.*

Syndromes for LMPZ's code can be computed with the same circuit as for code generation; it is only necessary to run this circuit backward. A relation between syndromes and errors is shown in Figure 7.5a (Laflamme et al. 1996).

Finally, in the case of the Barenco's 3-qubit code let us denote by $|0^j_E\rangle$ and $|1^j_E\rangle$, $1 \le j \le 3$ states obtained from $|0_E\rangle$ and $|1_E\rangle$ by a sign error on the jth qubit. Error symbols can be computed using two projection operators $\mathcal{P}_1$ on the subspace $\{|0_E\rangle, |1_E\rangle, |0^1_E\rangle, |1^1_E\rangle\}$ and $\mathcal{P}_2$ on the subspace $\{|0_E\rangle, |1_E\rangle, |0^2_E\rangle, |1^2_E\rangle\}$. The syndrome table is shown in Figure 7.5b.

error type	syndrome s_1, s_2, s_3, s_4	resulting state
no	0000	$\alpha\|0\rangle + \beta\|1\rangle$
BS3	1011	$-\alpha\|1\rangle + \beta\|0\rangle$
BS5	1111	$-\alpha\|0\rangle + \beta\|1\rangle$
B2	1000	
S3	0101	$\alpha\|0\rangle - \beta 11\rangle$
S5	0011	
BS2	1010	
B5	1100	
S1	0001	$-\alpha\|0\rangle - \beta\|1\rangle$
S2	0010	
S4	0100	
B1	0110	
B3	1110	
B4	1101	$-\alpha\|1\rangle + \beta\|0\rangle$
BS1	1110	
BS4	1001	

(a)

$\mathcal{P}_1$	$\mathcal{P}_2$	error
true	**true**	no
true	**false**	S1
false	**true**	S2
false	**false**	S3

(b)

Figure 7.5: Syndrome tables for the LMPZ's code. (B (S) stands for bit (sign) error and the number specifies the qubit, and 3-qubit code.

Correcting circuits

Once error symbols are computed, the error correction is done by applying the appropriate unitary transformation on the erroneous state. Let us discuss it in some detail only for Shor's 9-qubit code.

At the measurement during the syndrome computation, with respect to Bell basis, the erroneous state collapses into one of the states

$$\begin{array}{l}(|a_0\rangle + |b_1\rangle)((\alpha+\beta)|000\rangle + (\alpha-\beta)|111\rangle),\\ (|a_0\rangle - |b_1\rangle)((\alpha+\beta)|000\rangle - (\alpha-\beta)|111\rangle),\\ (|a_1\rangle + |b_0\rangle)((\alpha+\beta)|000\rangle + (\alpha-\beta)|011\rangle),\\ (|a_1\rangle - |b_0\rangle)((\alpha+\beta)|000\rangle - (\alpha-\beta)|011\rangle),\end{array} \tag{7.18}$$

In order to make the correction to each of such resulting states a unitary transformation has to be applied that produces the original state

$$\alpha(|000\rangle + |111\rangle) + \beta(|000\rangle - |111\rangle).$$

Exercise 7.4.40 *Determine unitary transformations to get from the states in (7.18) to the original state* $\alpha(|000\rangle + |111\rangle) + \beta(|000\rangle - |111\rangle)$

Technology tailored quantum error correction

As first illustrated by van Enk et al. (1997), specific technologies can allow to use more efficient error-correction methods than those illustrated above. They prepared a quantum optic scheme to transfer perfectly qubits from one ion trap to another. This techniques allows to use two qubits to encode one and uses both quantum teleportation and sending of classical bits to achieve transmissions. (This example indicates that potentials of quantum mechanics for fighting quantum mechanics (decoherence) are still far from being explored.)

Experiments with quantum error correction

The first experimental implementation of QECC, a simple 3 qubit one phase error-correcting code, using NMR technology was reported by Cory et al. (1998).

7.4.4 Basic methods to design quantum error-correcting codes

There are two basic approaches to the design of quantum error-correcting codes: an adaptation of some classical error-correcting codes and a design of completely new, inherently quantum, methods.

Base changes

The first method we are going to discuss is very simple but surprisingly useful and an important component of some more general methods discussed later.

Each quantum code mapping k qubits into n qubits specifies a subspace of dimension 2^k in H_{2^n}. A particular code is usually given by providing a particular basis of the code subspace. By changing the basis a (potentially infinite) number of different code representations can be obtained.

Of special importance is Hadamard rotation, which can switch a bit-correcting code to a sign-correcting code and vice versa.. Sometimes a variety of useful code modifications can be obtained by applying Hadamard rotation to a few carefully chosen qubits.

Example 7.4.41 (DiVincenzo and Shor, 1996)) *Let us start with the codewords*

$$\begin{aligned}
|0_E\rangle &= \frac{1}{4}(|00000\rangle \\
&+ |11000\rangle + |01100\rangle + |00110\rangle + |00011\rangle + |10001\rangle \\
&- |10100\rangle - |01010\rangle - |00101\rangle - |10010\rangle - |01001\rangle \\
&- |11110\rangle - |01111\rangle - |10111\rangle - |11011\rangle - |11101\rangle)
\end{aligned}$$

and $|1_E\rangle$ obtained from $|0_E\rangle$ by flipping all bits in the basis states. (Observe that basis states of both superpositions are classical cyclic codes.) By changing the basis for all qubits using the Hadamard rotation we get the codewords

$$|0'_E\rangle = \frac{1}{\sqrt{2}}(|0_E\rangle + |1_E\rangle) \qquad |1'_E\rangle = \frac{1}{\sqrt{2}}(|0_E\rangle - |1_E\rangle), \tag{7.19}$$

and if the Hadamard rotation is again applied to the first two qubits of the codewords $|0_E\rangle, |1_E\rangle$, then exactly LMPZ's code from Figure 7.2 is obtained.

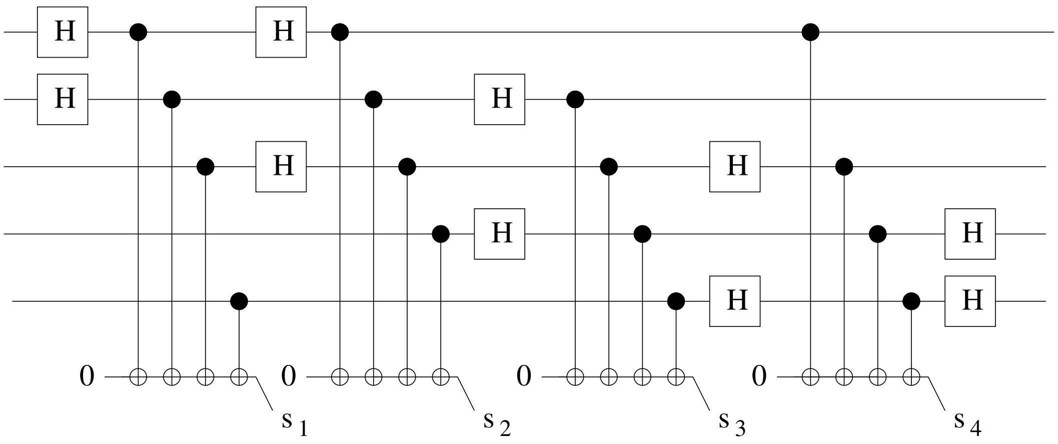


Figure 7.6: A circuit to compute syndromes for the code (7.19)

Exercise 7.4.42 *Verify, in detail, the above claim how one can get LMPZ's code.*

Changing a quantum code by changing the basis with Hadamard rotation is also an important tool to design error syndrome computing circuits. For example, Figure 7.6 shows such a syndrome computation circuit (due to DiVincenzo and Shor, 1996), for the code with $|0'_E\rangle$ and $|1'_E\rangle$ given above as codewords. A systematic structure of this network indicates that once the first two qubit rotations have been found useful, one could deduce the rest of the circuit by a careful inspection of the codes obtained each time additional qubits bases were changed (and then transformed back).

Exercise 7.4.43 *Design an error-correcting table showing how syndromes s_1, s_2, s_3, s_4, provided by the network in Figure 7.6, determine errors for the code (7.19).*

CSS-codes—converting classical codes into quantum codes

Another simple and powerful method of designing one quantum error-correcting code out of two, or out of one (dual), classical error-correcting linear codes, was developed by Calderbank and Shor (1995) and Steane (1996a).

Let C_1 be an $[n,k]$ linear binary code with the generator matrix G. Denote by H_{C_1} the subspace of H_{2^n} spanned by the set of vectors $\{|c\rangle \mid c \in C_1\}$. In this subspace we construct a new basis as follows. To each $w \in \{0,1\}^n$ we associate a quantum state:

$$|c_w\rangle = \frac{1}{\sqrt{2^k}} \sum_{v\in\{0,1\}^k} (-1)^{vG\cdot w}|vG\rangle \tag{7.20}$$

$$= \frac{1}{\sqrt{2^{n-k}}} \sum_{u\in C^\perp} (-1)^{u\cdot w}|v+w\rangle. \tag{7.21}$$

The last equality, which expresses $|c_w\rangle$ in the dual basis follows from the properties of the superpositions $|c_w\rangle$ summarized in the following exercises.

Exercise 7.4.44 *Show the following properties of the superpositions* $|c_w\rangle$. *(a)* $\sum_{v\in\{0,1\}^k}(-1)^{vG\cdot w} = 0$ *unless* $vG \cdot w = 0$ *for all* $v \in \{0,1\}^k$; *(b) if* $w_1 + w_2 \in C_1^{\perp}$, *then* $|c_{w_1}\rangle = |c_{w_2}\rangle$; *(c) if* $w_1 + w_2 \notin C_1^{\perp}$, *then* $c_{w_1} \cdot c_{w_2} = 0$; *(d)* $\{|c_w\rangle \,|\, w \in \{0,1\}^k\}$ *contains exactly* 2^k *different states and all are mutually orthogonal (and therefore form a basis for* H_{C_1}*); (e) cosets of* $C_1^{\perp}$ *are natural elements to index states* $\{|c_w\rangle \,|\, w \in \{0,1\}^k\}$.

If $C_2 \subseteq C_1$ is another code, then we define a quantum code $Q_{C_1,C_2} = \{|c_w\rangle \,|\, w \in C_2^{\perp}\}$. Clearly, $dim(Q_{C_1,C_2}) = dim(C_1) - dim(C_2)$. If C is a self-dual (n,k,d) code, then $Q_C = Q_{C,C^{\perp}} = \{|c_w\rangle \,|\, w \in C^{\perp}\}$ and $dim(Q_C) = k - (n-k) = 2k - n$.

Exercise 7.4.45 *Design* Q_{C_1,C_2} *code for the case* $C_1 = Ham(3,2), C_2 = C_1^{\perp}$.

Exercise 7.4.46 *Show that if* C *is a self-dual code then in the dual bases* $|c_w\rangle$ *has the form*

$$|d_w\rangle = \frac{1}{\sqrt{2^{n-k}}} \sum_{c\in C^{\perp}} |w + c\rangle. \tag{7.22}$$

The following theorem (see Calderbank and Shor, 1996), provides a justification for an introduction of the codes Q_{C_1,C_2}.

Theorem 7.4.47 (Quantum Correction Theorem, 1996)) *If* C_1, C_2 *are linear* $[n,k,d]$*-codes and* $C_2^{\perp} \subset C_1$, *then* $Q_{C_1,C_2^{\perp}}$ *is a* t *error-correcting quantum code for* $t = \lceil (d-1)/2 \rceil$.

It is easy to see that Q_{C_1,C_2} codes can be used to decode bit errors in one basis and sign errors in the dual basis, and that these two error correction steps do not interfere—Q_{C_1,C_2} can be therefore used also to correct bit–sign errors.

Using this approach $\{13,5,3\}^+$-code, $\{14,6,3\}^+$-code, $\{17,7,3\}^+$-code and $\{20,9,3\}^+$-code were designed by Steane (1997c). He has also found, using a systematic change of the signs of the basis states of codewords, the following codes: $\{8,3,3\}$-code, $\{10,4,3\}$-code and $\{11,5,3\}$-code.

An interesting/important example of CSS codes are quantum Reed-Muller Codes, see Section 7.4.1 for classical versions, due to Steane (1996b) and Zhang and Fuss (1997).

Steane (1998a) has also developed a method how to convert certain CSS codes into quantum codes with better parameters.

Analysis of orthogonality conditions and search for signs of the basis states

This method, introduced by Laflamme et al. (1996), is simple in principle, but elaborate, and was used to get the first "perfect" encoding of one qubit by 5 qubits and 8 basis states for each of the codewords —see Table 7.2.

Exercise 7.4.48 *(Laflamme et al. 1996) Assume that an encoding has the form*

$$|0\rangle \to \sum_{i=0}^{2^n-1} \alpha_i |i\rangle, \qquad |1\rangle \to \sum_{i=0}^{2^n-1} \beta_i |i\rangle$$

and that all states obtained by one qubit error are orthogonal. Show that in such a case the following equalities have to hold:

$$\sum_{i_k=0,i_l=0} |\alpha_i|^2 = \sum_{i_k=0,i_l=1} |\alpha_i|^2 = \sum_{i_k=1,i_l=0} |\alpha_i|^2 = \sum_{i_k=1,i_l=1} |\alpha_i|^2. \tag{7.23}$$

for all $0 \le k \ne l < n$ (where j_s denotes sth bit of j expressed in binary).

Encodings of $|0_E\rangle$ and $|1_E\rangle$ should be such that all states obtained from them by one qubit error should be orthogonal. This gives rise to a bunch of conditions the amplitudes have to satisfy, such as one in (7.23). These conditions then have to be analysed to determine (to guess) the support for the code. As reported in Laflamme et al. (1996), this way they found the support for the code shown in Figure 7.7.

If one then takes a natural assumption of simplicity, namely that all amplitudes of the basis states are either +1 or −1, then a computer search can find out quantum codes with a given support and amplitudes ±1. Laflamme et al. (1996) discovered that all such minimal encodings of one qubit have to have two −1 as amplitudes in one codeword and four in other. An example of a modified version of their code (see Williams and Clearwater, 1998), has the form shown in Table 7.7:

Code	$\vert 0_E\rangle$	$\vert 1_E\rangle$
Modified LMPZ's code	$+\vert 00000\rangle - \vert 11100\rangle + \vert 10011\rangle - \vert 01111\rangle$ $+\vert 11010\rangle + \vert 00110\rangle + \vert 01001\rangle + \vert 10101\rangle$	$-\vert 00011\rangle + \vert 11111\rangle - \vert 10000\rangle + \vert 01100\rangle$ $+\vert 11001\rangle - \vert 00101\rangle - \vert 01010\rangle + \vert 10110\rangle$

Figure 7.7: Modified version of LMPZ code

Exercise 7.4.49 *Determine whether each code with the same string-support as LMPZ code has to have four amplitudes −1 at basis states in one of codewords and two at the other.*

7.4.5 Stabilizer codes

A general class of quantum error-correcting codes was introduced by Gottesman (1996) and Calderbank and Shor (1997) and included all QECC known till that time. See also Gottesman (1997) for a detailed treatment of the subject.

Basic concepts—code stabilizer

The basic elements to work with at this approach are error operators

$$E_1 \otimes E_2 \otimes \ldots \otimes E_n, \tag{7.24}$$

where each $E_i \in \{I, X, Y, Z\}$ and $X = \sigma_x$, $Z = \sigma_z$ and $Y = XZ = i\sigma_y$, where σ_x, σ_y and σ_z are Pauli matrices. An error operator (7.24) is said to have **weight** t if it has t component matrices different from I.

Denote by X_i, Y_i and Z_i operator/matrices X, Y and Z acting on the ith qubit and let $\mathcal{G}_n$ be the group generated by $3n$ of such matrices.

Properties of $\mathcal{G}_n$: $X_i^2 = Z_i^2 = I = -Y_i^2$ and $Y_i = -X_iZ_i = Z_iX_i$, for $1 \leq i \leq n$. $\mathcal{G}_n$ has order $2 \cdot 4^n = 2^{2n+1}$ (there are four possible elements for each i and the sign). In addition, each element of $\mathcal{G}_n$ squares to ± 1 and if $A, B \in \mathcal{G}_n$, then A and B either commute or anticommute, i.e., either $[A, B] = 0$ or $\{A, M\} = 0$, see page 371.

Let C be a quantum error-correcting code of H_{2^n}. C spans a subspace of H_{2^n}. The group $\mathcal{G}_n$ can be seen as acting on states of C. A **stabilizer** $\mathcal{S}_C$ of C is the set

$$\mathcal{S}_C = \{M \in \mathcal{G}_n \mid M|\phi\rangle = |\phi\rangle \quad \text{if} \quad |\phi\rangle \in C\}.$$

The following property is of crucial importance for the "stabilizer codes" to be defined later:

If $M \in \mathcal{G}_n$ and $S \in S_C$ are such that $\{M, S\} = 0$, then for any $|\phi\rangle$, $|\psi\rangle \in C$,

$$\langle\phi|M|\psi\rangle = \langle\phi|MS|\psi\rangle = -\langle\phi|SM|\psi\rangle = -\langle\phi|M|\psi\rangle$$

and therefore $\langle\phi|M|\psi\rangle = 0$. The code herefore satisfies the condition 7.11 whenever $M_a{}^*M_b$ anticommute with some element of S.

M_1	Z	Z	I	I	I	I	I	I	I
M_2	Z	I	Z	I	I	I	I	I	I
M_3	I	I	I	Z	Z	I	I	I	I
M_4	I	I	I	Z	I	Z	I	I	I
M_5	I	I	I	I	I	I	Z	Z	I
M_6	I	I	I	I	I	I	Z	I	Z
M_7	X	X	X	X	X	X	I	I	I
M_8	X	X	X	I	I	I	X	X	X

(a)

M_1	X	X	X	X	I	I	I
M_2	X	X	I	I	X	X	I
M_3	X	I	X	I	X	I	X
M_4	Z	Z	Z	Z	X	X	X
M_5	Z	Z	I	I	Z	Z	I
M_6	Z	I	Z	I	Z	I	Z

(b)

M_1	X	Z	Z	X	I
M_2	I	X	Z	Z	X
M_3	X	I	X	Z	Z
M_4	Z	X	I	X	Z

(c)

M_1	X	X	X	X	X	X	X	X
M_2	Z	Z	Z	Z	Z	Z	Z	Z
M_3	I	X	I	X	Y	Z	Y	Z
M_4	I	X	Z	Y	I	X	Z	Y
M_5	I	Y	X	Z	X	Z	I	Y

(d)

Figure 7.8: Stabilizers

Example 7.4.50 *Figures 7.8a,b,c (Gottesman, 1997), shows generators of the stabilizers for Shor's* 9*-qubit code, Steane's* 7*-qubit code and LMPZ's* 5*-qubit code. Figure 7.8d show generators of the stabilizer for an* $[[8, 3, 3]]$*-code due to Gottesman (1996). Let us discuss design and use of the stabilizer for Steane's code. Error vectors in Figure 7.8b can be discovered in a straightforward way from how one detects a single bit or sign error for this code. Indeed, to detect a bit error in a state* $|\psi\rangle$ *on one of the first three qubits it is sufficient* *to compare the first qubit with second and then the first qubit with the third in $|\psi\rangle$. One way of doing that is to measure $|\psi\rangle$ with respect to M_1 and M_2 as observables. A similar role play error vectors M_3 to M_6. The last two error vectors can be used to detect sign errors.*

Exercise 7.4.51 *Show that if $\{M_1, \ldots, M_s\}$ is a stabilizer for a code C, then so is $\{M_1, \ldots, M_{i-1}, M_iM_j, M_{i+1}, \ldots, M_s\}$, for any $1 \leq i, j \leq s$.*

Exercise 7.4.52 *Show that if a code encodes k qubits using n qubits, then its stabilizer has dimension 2^{n-k}.*

Implications? Let E and F be error operators, both of weight t or less. If the operator F^*E anticommutes with any operator in $\mathcal{S}$, then vectors $E|\phi\rangle$ and $F|\psi\rangle$ are orthogonal for any $|\phi\rangle, |\psi\rangle \in C$. However, this is exactly the requirement a non-degenerate quantum error-correcting code should satisfy. In other words, to get a non-degenerate code we just need to find a code C and the corresponding stabilizer $\mathcal{S}_C$ such that any non-identity error operator of $\mathcal{G}_n$ of weigth at most $2t$ anticommutes with some element from $\mathcal{S}_C$.

The above result suggests two methodologies to design non-degenerate codes: To choose C and to look for $\mathcal{S}_C$ or vice versa.

In order to facilitate the second methodology let us look for some properties $\mathcal{S}_C$ should have to be a stabilizer.

$\mathcal{S}_C$ is clearly a subgroup of $\mathcal{G}_n$. In addition, for any $|\psi\rangle \in C$, $S \in \mathcal{S}_C$, $S^2|\psi\rangle = S|\psi\rangle = |\psi\rangle$ and therefore $S^2 = I$. Moreover, for any $|\phi\rangle \in C$ and any $M, N \in \mathcal{S}_C$, $MN|\psi\rangle = NM|\psi\rangle = |\psi\rangle$ and therefore $[M, N]|\psi\rangle = 0$. We have therefore that either $[M, N] = 0$ or $\{M, N\} = 0$. In the case $\{M, N\} = 0$ we would have $[M, N] = 2MN$, but this contradicts the property $[M, N]|\psi\rangle = 0$ for any $|\psi\rangle \in C$. Hence $[M, N] = 0$. $\mathcal{S}_C$ has therefore to be an Abelian group. $S^2 = I$ has to hold for any of its elements S. It can be shown that these conditions are sufficient for the existence of a non trivial code C for which $\mathcal{S}_C$ is a stabilizer (provided S_c is not too big).

Let C be a code of codewords of length n and $\mathcal{S}_C$ be its stabilizer. The **centralizer** of $\mathcal{S}_C$, $\mathcal{C}(\mathcal{S}_C)$ is the set of elements of $\mathcal{G}_n$ that commute with all elements of $\mathcal{S}_C$, Clearly $\mathcal{S}_C \subseteq \mathcal{C}(\mathcal{S}_C)$. Define $\mathcal{F}_\mathcal{S} = \mathcal{S}_C \cup (\mathcal{G}_n - \mathcal{C}(\mathcal{S}_C))$. The code C corrects any set $\mathcal{E}_0$ of errors such that for any $M_a, M_b \in \mathcal{S}_\mathcal{E}$, $M_a^* M_b \in \mathcal{S}_C \cup (\mathcal{G}_n - \mathcal{C}(\mathcal{S}_C))$.

For the stabilizer codes there are straightforward ways to design networks for encoding—see Gottesman (1997)—and also for syndrom computation, see DiVincenzo and Shor (1995).

Gottesman (1996) has developed a methodology for choosing $\mathcal{S}$ so that error operators of length at most $2t$ will anticommute with some member of $\mathcal{S}$. On this base he designed an infinite class of codes saturating the quantum Hamming bound. One of them is $\{8, 3, 3\}$-code whose stabilizer is shown in Figure 7.8d.

The concept of the stabilizer has much simplified the design of quantum error correcting codes—see Gottesman (1996), Steane (1996) and Calderbank et al. (1996, 1997).

Special notation

In order to describe stabilizers also a different notation is used which is especially handy in the case of the quantum error-correcting codes derived from the classical codes. An error operator $M_1 \otimes \ldots \otimes M_n$ is written in the form $X(x_1, \ldots, x_n)Z(z_1, \ldots, z_n)$, where $x_i = 1$ if $M_i \in \{X, Y\}$, and 0 otherwise, and $z_i = 1$ if $M_i \in \{Y, Z\}$, and 0 otherwise. Moreover, in such a case also the notation $(X_x|Z_z)$ is used, where x_i and z_i are defined as above.

A set of error operators forming generators of a stabilizer can then be written in the matrix form

$$(X|Z), \tag{7.25}$$

For example, stabilizers for the LMPZ code and for the {8,3,3 }code mentioned above are in this form described in Figure 7.9a,b.

If X and Z are $(n-k)\times n$ matrices for some n and k then $(X|Z)$ forms a stabilizer if $XZ^T + ZX^T = 0$ (Calderbank et al. 1997).

Stabilizers for CSS-codes

Let us consider a CSS code created out of two classical linear codes C_1 and C_2 and let P_{C_1} and P_{C_2} be their parity-check matrices.

A quantum code to correct just bit errors can be designed using the set of generators S_{C_1} obtained from vectors of P_{C_1} by replacing each 1 with Z. To that we add the set of generators obtained from the parity-check matrix P_{C_2} for C_2 with X replacing 1's this time. These generators can identify the sign errors. Together they identify also bit/sign errors. In the case $C_2^\perp \subseteq C_1$, $C_1^\perp \subseteq C_2$ these two generators can be combined into a single set of generators for the code.

$$\begin{pmatrix} 1&0&0&1&0&|&0&1&1&0&0\\ 0&1&0&0&1&|&0&0&1&1&0\\ 1&0&1&0&0&|&0&0&0&1&1\\ 0&1&0&1&0&|&1&0&0&0&1 \end{pmatrix} \qquad \begin{pmatrix} 1&1&1&1&1&1&1&1&|&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&|&1&1&1&1&1&1&1&1\\ 0&0&0&0&1&1&1&1&|&0&0&1&1&0&0&1&1\\ 0&0&1&1&0&0&1&1&|&0&1&0&1&0&1&0&1\\ 0&1&0&1&0&1&0&1&|&0&0&1&1&1&1&0&0 \end{pmatrix}$$

(a) (b)

Figure 7.9: Generators of two CSS-codes

Beyond stabilizer codes

The first example of a non-stabilizer code that is better than any other stabilizer code of the same type is due to Rains et al. (1997). It is a {5,6,2}code that is defined (as a subspace) by the following projection operators.

$$\begin{aligned} P \;=\; & \frac{1}{16}[3(I\otimes I\otimes I\otimes I\otimes I) + (I\otimes\sigma_z\otimes\sigma_y\otimes\sigma_y\otimes\sigma_z)_{cyc} + (I\otimes\sigma_x\otimes\sigma_z\otimes\sigma_z\otimes\sigma_x)_{cyc} \\ & -(I\otimes\sigma_y\otimes\sigma_x\otimes\sigma_x\otimes\sigma_y)_{cyc} + 2(\sigma_z\otimes\sigma_x\otimes\sigma_y\otimes\sigma_y\otimes\sigma_x)_{cyc} - 2\sigma_z\otimes\sigma_z\otimes\sigma_z\otimes\sigma_x\otimes\sigma_z] \end{aligned}$$

where the subscript "*cyc*" denotes that all five cyclic shifts of the corresponding error operator have to be taken into the sum. The code was discovered by a combination of careful reasoning and computations. The code is better than any other stabilizer code of the same type in the sense that for each stabilizer $(n,6,2)$-code we have $n\le 4$.

7.5 Fault-tolerant Quantum Computation

There are five main reasons why the discovery of quantum error-correcting codes is not by itself a sufficient guarantee that quantum error-correction will be practically successful on "noisy quantum computers".

1. Quantum error-correcting codes help to deal only with the problem of reliable storage and transmission of quantum information. They are by themselves insufficient to have fault-tolerant quantum information processing.

2. Each use of quantum error correction methods brings additional requirements on quantum memory, hardware and computation time. Indeed, to establish necessary redundancy, additional qubits, and thereby additional quantum memory, are needed. In addition encoding, error recovery and decoding operations require additional qubits, gates and circuits. That may slow down the overall computation.

3. Quantum error-correction processing networks are themselves non-trivial quantum networks and therefore, naturally, errors can occur during their performance.

What is therefore needed for quantum error-correction to be practically successful are such networks for error-correction processing that are sufficiently robust to succeed with high reliability even if some errors are made during error-correction processing operations themselves. In short, fault-tolerant quantum networks are what is practically needed. What is especially needed are fault-tolerant networks to compute directly with encoded qubits without any need to decode them.

It was first shown by Shor (1995) that for any polynomial size quantum circuit an equivalent and polynomial size fault-tolerant version can be constructed which can tolerate after t steps inaccuracies and decoherence averaging $\mathcal{O}(\frac{1}{\lg^c t})$ per step for a small constant c.

4. It is not sufficient that we can encode quantum information in such a way that it is stored or transferred reliably for some time. Of key importance for real quantum computing is that we can store and transmit quantum information reliably for a long time and through a long distance.

5. It is also not sufficient that we can process quantum information for some time in a fault-tolerant way. What is badly needed is to be able to do that for a sufficiently long time.

The so-called concatenated codes (see Section 7.5.3), represent a way to deal with the last two problems. Concatenated codes allow us to store and transmit a qubit with maximal error ε, provided gates with errors at most $c\varepsilon$ (where c is a constant not dependent on ε) and storage or channel elements with errors at most ε are given, independently of how long we must store a state or how far we need to transmit it.

Remark 7.5.1 A way to cope with the main drawback of concatenated codes—still high requirements on tolerable error probabilities for transmissions and local operations—has been suggested by Briegel et al. (1998). They suggest to use "quantum repeaters to form entanled pairs for an arbitrarily large distance. The main idea is to use a sequence of imperfectly entangled pairs of particles and a new (nested) purification protocol, to create a single distant pair of particles of high fidelity.

Methods presented in this section allow us to cope with inaccuracies and decoherence problem in three ways: (1) Fault-tolerant circuits are used to perform quantum gates information processing; (2) in between two quantum gates applications to qubits quantum error detection and recovery circuits are used to take care of their recovery from errors; (3) concatenated codes and quantum repeaters are used to perform quantum information processing and transmission over long time and distance.

7.5.1 Fault-tolerant quantum error correction

A network is usually called fault-tolerant if it can recover from some errors during its operations. Of course, there are no perfect networks. In addition, to any network it can happen that all gates malfunction. That is why we need to take a modest but still useful view of what is meant by saying that a network is considered as being fault-tolerant.

We shall be in the following happy to consider a network as fault-tolerant if the probability that the network malfunctions is at most ε^2, where ε is the probability that one gate or qubit error occurs.

Let us first analyze possible ways and circumstances in which error-correcting codes operations may fail, and what has to be changed to have fault-tolerant networks for error-correction operations (Preskill, 1998).

Error propagation

The first main problem we have to learn to deal with is **error propagation** due to entanglement. If a two-qubit gate is applied and one of qubits is erroneous, then the error can propagate to the second qubit.

Example 7.5.2 *As we could see in the previous section, XOR gates are of key importance for quantum error-correcting codes processing operations, especially for error syndrome computation. It is evident that if a error occurs in an XOR's gate control qubit, then the error propagates ("forwards") to its target qubit. Less obvious, but very important for quantum error-correction processing networks, is the fact that under certain circumstances an error can propagate through an XOR gate also "backwards", from the target qubit to the control qubit. This is easily seen from Figure 2.11, where a relation between an XOR gate and its reverse is shown at which control and target bits are interchanged. This network is a very typical component of quantum error correction networks. Indeed, Hadamard rotations applied to qubits change the basis from the standard to the dual (and vice versa) and interchange bit and sign errors. Therefore if a sign error occurs on the target qubit, it can become a bit error of the source qubit.*

Backward error propagation in XOR gates discussed in the above example implies that it is not good to use for error syndrome computation such circuits as in Figure 7.10a, in which one ancilla qubit is used "to xor" information from several data qubits. Indeed, a spontaneous phase error at an ancilla qubit, a far from an unexpected event, could propagate to a bit error in several code qubits. That is why a way of gathering information for syndrome computation depicted in Figure 7.10b is surely superior with respect to error propagation (even if it requires more qubits).

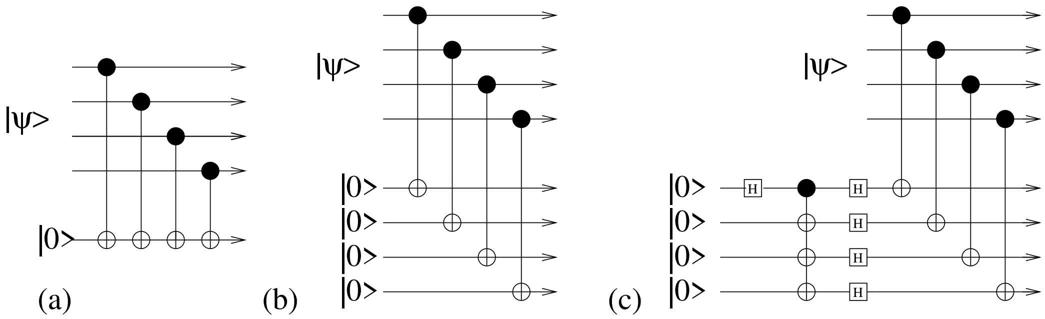


Figure 7.10: Three ways of gathering information from data to ancilla: bad, better and the best

Fault-tolerant ancillas

Unfortunately, even the way of gathering information shown in Figure 7.10b is not good enough for fault-tolerant quantum error correction. The point is that code qubits keep being entangled with the ancilla qubits and therefore a measurement of ancilla qubits can destroy the encoded state $|\psi\rangle$.

Example 7.5.3 *In the case of Steane's codendexcode!Steane's we have seen that information about its last* 4 *qubits is needed to get the first bit of the syndrome* $(x_4 \oplus x_5 \oplus x_6 \oplus x_7)$. *This information can be obtained using a four-qubit ancilla, with all qubits initially in the state* $|0\rangle$, *one qubit of ancilla for each bit of the syndrome. However, if the ancilla qubits yield, after the measurement, the result* 0101, *then in* $|\psi\rangle$ *the state* $|0_E\rangle$ *collapses to the state* $|1010101\rangle$, *and the state* $|1_E\rangle$ *collapses to the state* $|0100101\rangle$, *which is not acceptable because the state lost all protection against phase errors..*

There are three main operations to work with quantum error-correcting codes: **encoding** (code generation), **error recovery** (syndrome computation and code correction) and **decoding**. Networks to perform these operations, especially code generation and error recovery normally require an ancilla. It is therefore of crucial importance for the design of a fault-tolerant quantum network to find a way of copying information into the ancilla without distractive effects on the state being "copied".

One way out was found by Shor (1995) and it is illustrated in Figure 7.10c for the case of an ancilla with 4 qubits. Information is copied into the ancilla with the initial (Shor) state

$$|\phi\rangle = \frac{1}{\sqrt{8}}\sum_{i=0}^{7}|parity(i)\rangle$$

of the equal superposition of all even-parity 4-bit codewords. This state can be created in the way shown in Figure 7.10c. (The first Hadamard gate and the multiple XOR create the "cat state" $\frac{1}{\sqrt{2}}(|0000\rangle + |1111\rangle)$. The next four Hadamard gates then create $|\phi\rangle$.

Why is this way of copping information into the ancilla in the state $|\phi\rangle$ better? What is needed in the case of a syndrome computation is only information about the parity of four bits copied into the ancilla. If this parity is 0, the state $|\psi\rangle$ is not changed; if the parity is one, then the state of ancilla changes into the equal superposition of all odd-weighted 4-bit codewords. If we therefore measure the ancilla state, with respect to the standard basis in

H_{16}, we get a four-bit word, randomly, and its parity is the corresponding syndrome. The key point is that after XOR operations in this case the state of the ancilla is not entangled with encoded qubit and therefore a measurement of ancilla does not hurt it!

Verification of the ancilla state

Another potential problem and a potential source of errors is that the initial ancilla state may be erroneous, and a sign error that occurs during the preparation of the ancilla may propagate to the encoded qubit.

In the case of the 4-qubit ancilla, and its circuit in Figure 7.10c, a single qubit error during the second or third XOR gate (of the multiple XOR gate) can result in two bit errors in the cat state and consequently in two sign errors in the initial state of the ancilla. That will cause errors during the syndrome measurement.

A suggested way to test the cat state is to perform two XOR gates on two randomly chosen qubits of the cat state, and two new qubits originally in the state $|0\rangle$, and discard the resulting state if the measurement of these two new qubits does not agree. In such a case the ancilla has to be constructed again. Of course this means that one needs to use additional qubits.

For the case of 4-qubits, as in Figure 7.10c, there is a simpler way of verifying the cat state. It is based on the observation that for all the ways a single XOR gate can cause two bit errors in the cat state, the first and the fourth qubit will have different values. To verify the cat state it is therefore sufficient to use two additional XOR gates as shown in Figure 7.11.

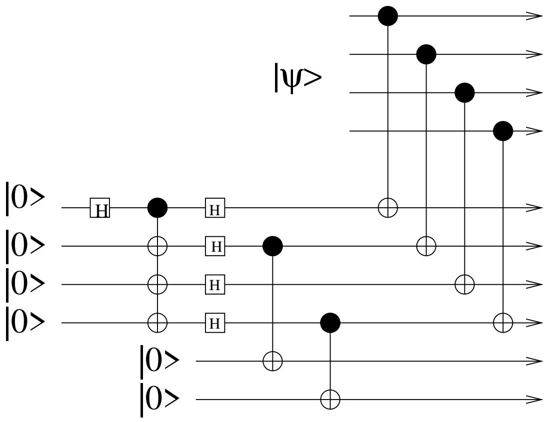


Figure 7.11: Ancilla state verification

Another possible case is that a single error at the cat state designing circuit could cause two sign errors in the cat state and therefore two bit-errors in the initial Shor's state. However, such errors do not damage the encoded qubits; they can only cause that the resulting syndrome measurement provides an incorrect result.

Testing of syndrome measurements

It has been shown that syndrome measurement circuits do exist, but this does not guarantee yet that syndrome measurement performed by such circuits will be correct. The main way to be sure with that is to repeat each syndrome measurement several times, until the same error syndrome is obtained k times in a row, where k is some reliability parameter.

Fault-tolerant syndrome computation for additive codes

A simple procedure for syndrome generation for stabilizer codes was developed by DiVincenzo and Shor (1996).

To each generator $X(x_1, \ldots, x_n)Z(z_1, \ldots, z_n)$ one stage in the syndrome computation circuit corresponds. Each stage consists of three phases: basis transformations on qubits, syndrome gathering and undoing of the previous basis. The basic idea is to transform the generator to the form

$$X(0, \ldots, 0)Z(z'_1, \ldots, z'_n).$$

In order to achieve it the basis of the ith qubit is done by the matrix I if $x_i = 0$, by Hadamard rotation if $x_i = 1$ and $z_i = 0$ and by the matrix $H' = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}$, if $x_i = z_i = 1$. After this change of the basis $z'_i = x_i \vee z_i$. The next part of the circuit for a given stage is to determine the parity of qubits with non-zero entries in Z. In order to do that fault-tolerantly one can use ancilla in the Shor state. Once this is done the stage ends by undoing the basis. This way the syndrome computing circuit for the LMPZ's code shown in in Figure 7.6 works.

Why does this work? The point is that each measurement of the quantum state, corresponding to one of the generators of the generator matrix G of the code, yields the eigenvalue of the quantum state with respect to that matrix. By that we can reduce the number of eigenspaces in which the quantum state may lie by a factor of 2. Therefore, if the measurement is done for every generator of G, the complete set of eigenvalues of the state with respect to G is known. By that one of the error eigenspaces is determined and that implies which unitary transformation is to be applied to undo the error.

Exercise 7.5.4 *Determine the number of gates needed to compute a syndrome in the fault-tolerant way for additive codes.*

7.5.2 Fault-tolerant quantum gates

A fault-tolerant realization of quantum error-correction operations is a step forward, with respect to the reliability of the storage and transmission of quantum information. It also offers one way to do reliably quantum computation: encoded qubits are first decoded, quantum gates are applied on decoded qubits and then qubits are again encoded. Clearly, this approach has a lot of drawbacks. What is really needed is the ability to apply quantum gates directly to encoded qubits and to work in the "protected" code subspace.. This means that what we really need in the first order are fault-tolerant gates.

There seem to be two ways to have fault-tolerant gates. The first way, discussed in more detail in the first part of this section, is to use "software methods": to design a set of quantum circuits that simulates, in a fault-tolerant way, a universal set of gates. The second way, so far only envisioned, is to use "hardware methods"—to use physically inherently fault-tolerant gates. This will be briefly discussed at the end of this section.

Fault-tolerant circuits for a universal set of gates

It will now be shown, following Shor (1996), that fault-tolerant quantum circuits exist for rotation gates, XOR gate and for the Toffoli gate for a special type of quantum codes, designed from the so-called punctured doubly even codes by the general method discussed on page 290.[11] The choice of these special codes allowed us to show in an easy way how fault-tolerant versions of the above set of gates can be implemented. The same can be shown for other codes, as discussed later.

Let us assume that C is a self-dual classical code of length n, dimension $k = \frac{n+1}{2}$ and distance d. Since the dimension of $C^{\perp}$ is $n-k$, we have, on page 290, that the number of different quantum codewords is $2^{2k-n} = 2$. Let us denote them $|0'_E\rangle$ and $|1'_E\rangle$. One of them consists of codewords with an even parity; the second with an odd parity.

Transferring these codes into into dual base yields:

$$|0_E\rangle = \frac{1}{\sqrt{2}}(|0'_E\rangle + |1'_E\rangle) \qquad |1_E\rangle = \frac{1}{\sqrt{2}}(|0'_E\rangle - |1'_E\rangle) \tag{7.26}$$

An important point is that transformation from one basis to another for $\{|0_E\rangle, |1_E\rangle\}$ codewords and $\{|0'_E\rangle, |1'_E\rangle\}$ codewords, can be done by applying Hadamard transformation qubitwise! This implies that the basis changing transformation is fault-tolerant. Indeed, when error-less gates are applied to individual qubits the number of erroneous qubits is preserved. Moreover an erroneous gate application will introduce only one error.

In a similar way it can be shown that $\frac{\pi}{2}$-rotations can be applied qubitwise and therefore in a fault-tolerant way.

Exercise 7.5.5 *Show that for punctured doubly even self-dual codes the following operations on logical qubits can be performed qubitwise: (a)* $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$*; (b)* $\begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$*; (c)* $\frac{\pi}{2}$*-rotation.*

XOR operation can also be performed qubitwise. Indeed, suppose we have two logical qubits $|d_a\rangle$ and $|d_b\rangle$, see page 291. Then

$$|d_a\rangle|d_b\rangle = \frac{1}{2^{n-k}}(\sum_{w \in C^{\perp}} |w+a\rangle)(\sum_{w' \in C^{\perp}} |w'+b\rangle).$$

By applying XOR from the ith qubit of $|d_a\rangle$ to the ith qubit of $|d_b\rangle$ we get the

$$\frac{1}{2^{n-k}}(\sum_{w \in C^{\perp}} |w+a\rangle)(\sum_{w' \in C^{\perp}} |w'+b+w+a\rangle) = |d_a\rangle|d_{a+b}\rangle,$$

because if $w \in C^{\perp}$, $w' \in C^{\perp}$, then also $w + w' \in C^{\perp}$.

Bitwise can also be done the phase changing operators $|d_a\rangle|d_b\rangle \rightarrow (-1)^{a \cdot b}|d_a\rangle|d_b\rangle$.

As the next step a fault-tolerant circuit for the Toffoli gate will be constructed. This will be a harder task.

Let us first assume that we have a fault-tolerant circuit that can perform transformation $T\colon |i_E, j_E\rangle \rightarrow |i_E, j_E, (i \wedge j)_E\rangle$. Using this circuit we can make another fault-tolerant circuit

[11] Doubly even codes are self-dual codes all codewords of which have a number of 1s divisible by 4. A punctured code is obtained from a self-dual code by deleting one bit-component.

which performs a similar transformation:

$$|i_E, j_E, k_E\rangle \rightarrow (i_E, j_E, (i \wedge j)_E, k_E\rangle.$$

Let us then attach $\mathrm{XOR}_{4,3}$ and H_4 to the previous circuit outputs. As the result a fault-tolerant circuit is obtained to perform the following transformation:

$$\begin{array}{lcllcl}
|0_E0_E0_E\rangle & \rightarrow & \frac{1}{\sqrt{2}}|0_E0_E0_E\rangle(|0_E\rangle+|1_E\rangle) & |0_E1_E0_E\rangle & \rightarrow & \frac{1}{\sqrt{2}}|0_E1_E0_E\rangle(|0_E\rangle+|1_E\rangle), \\
|1_E0_E0_E\rangle & \rightarrow & \frac{1}{\sqrt{2}}|1_E0_E0_E\rangle(|0_E\rangle+|1_E\rangle) & |1_E1_E0_E\rangle & \rightarrow & \frac{1}{\sqrt{2}}|1_E1_E1_E\rangle(|0_E\rangle+|1_E\rangle), \\
|0_E0_E1_E\rangle & \rightarrow & \frac{1}{\sqrt{2}}|0_E0_E1_E\rangle(|0_E\rangle-|1_E\rangle) & |0_E1_E1_E\rangle & \rightarrow & \frac{1}{\sqrt{2}}|0_E1_E1_E\rangle(|0_E\rangle-|1_E\rangle), \\
|1_E0_E1_E\rangle & \rightarrow & \frac{1}{\sqrt{2}}|1_E0_E1_E\rangle(|0_E\rangle-|1_E\rangle) & |1_E1_E1_E\rangle & \rightarrow & \frac{1}{\sqrt{2}}|1_E1_E0_E\rangle(|0_E\rangle-|1_E\rangle).
\end{array}$$

A measurement of the fourth qubit with the outcome 0 results in the state $(i_E, j_E, ((i \wedge j) \oplus k)_E)$, the one the Toffoli gate should produce. However, a measurement with outcome 1 produces the state $-|(i_E, j_E, ((i \wedge j) \oplus k)_E\rangle$. In such a case we need to change the sign, which can be done fault-tolerantly, as already mentioned.[12]

As the next step we show how to design a fault-tolerant circuit for the mapping $|i_E, j_E\rangle \rightarrow |i_E, j_E, (i \wedge j)_E\rangle$ provided we have a 3-qubit ancilla in the state

$$|A\rangle = \frac{1}{2}(|0_E0_E0_E\rangle + |0_E1_E0_E\rangle + |1_E0_E0_E\rangle + |1_E1_E1_E\rangle).$$

This can be done using the gates that have already been shown to have equivalent fault-tolerant circuits.

Let us first apply gates $\mathrm{XOR}_{3,1}$ and $\mathrm{XOR}_{4,2}$ to the state $|i_E, j_E\rangle|A\rangle$. This way we can realize the mapping

$$\begin{array}{lcl}
|0_E0_E\rangle|A\rangle & \rightarrow & \frac{1}{2}(|0_E0_E0_E0_E0_E\rangle + |0_E1_E0_E1_E0_E\rangle + |1_E0_E1_E0_E0_E\rangle + |1_E1_E1_E1_E1_E\rangle), \\
|0_E1_E\rangle|A\rangle & \rightarrow & \frac{1}{2}(|0_E1_E0_E0_E0_E\rangle + |0_E0_E0_E1_E0_E\rangle + |1_E1_E1_E0_E0_E\rangle + |1_E0_E1_E1_E1_E\rangle), \\
|1_E0_E\rangle|A\rangle & \rightarrow & \frac{1}{2}(|1_E0_E0_E0_E0_E\rangle + |1_E1_E0_E1_E0_E\rangle + |0_E0_E1_E0_E0_E\rangle + |0_E1_E1_E1_E1_E\rangle), \\
|1_E1_E\rangle|A\rangle & \rightarrow & \frac{1}{2}(|1_E1_E0_E0_E0_E\rangle + |1_E0_E0_E1_E0_E\rangle + |0_E1_E1_E0_E0_E\rangle + |0_E0_E1_E1_E1_E\rangle).
\end{array}$$

As the next step we measure the first two logical qubits. The outcome 0_E0_E results in the transformation $|i_E, j_E, i_E, j_E, (i \wedge j)_E\rangle$. The outcome 0_E1_E results in a mapping $|i_E, j_E, u_E, v_E, w_E\rangle$ such that

$$NOT_4XOR_{3,5}|i_E, j_E, u_E, v_E, w_E\rangle = |i_E, j_E, i_E, j_E, (i \wedge j)_E\rangle,$$

which is what is needed. In a similar way we can transform the results of other measurements outcomes, using XOR and NOT operations, to the desired mapping.

The last step is to show that the ancilla state $|A\rangle$ can be designed in a fault-tolerant way. Of course, this cannot be done using XOR and rotation gates only.

Let us consider two $3n$-qubit states

$$\begin{array}{lcl}
|A\rangle & = & \frac{1}{2}(|0_E0_E0_E\rangle + |0_E1_E0_E\rangle + |1_E0_E0_E\rangle + |1_E1_E1_E\rangle), \\
|B\rangle & = & \frac{1}{2}(|0_E0_E1_E\rangle + |0_E1_E1_E\rangle + |1_E0_E1_E\rangle + |1_E1_E0_E\rangle),
\end{array}$$

[12] One can make use of the trasformations $|i_Ej_Ek_E\rangle \rightarrow (-1)^{ij}(--1)^k|i_Ej_Ek_E\rangle$ which can be seen as a composition of operations $|i_Ej_E\rangle \rightarrow (-1)^{ij}|i_Ej_E\rangle$ and $|k_E\rangle \rightarrow (-1)^k|k_E\rangle$.

i.e., $|A\rangle = \mathrm{NOT}_3|B\rangle$. Moreover,

$$\frac{1}{\sqrt{2}}(|A\rangle + |B\rangle) = \frac{1}{\sqrt{8}}(|0_E\rangle + |1_E\rangle)(|0_E\rangle + |1_E\rangle)(|0_E\rangle + |1_E\rangle)$$

and the last state is easy to construct using gates, the fault-tolerance of which has already been demonstrated.

Let us first assume that we have fault-tolerant gates performing transformations

$$\begin{array}{lclclcl} |0^{(n)}\rangle|A\rangle & \rightarrow & |0^{(n)}\rangle|A\rangle, & & |1^{(n)}\rangle|A\rangle & \rightarrow & |1^{(n)}\rangle|A\rangle, \\ |0^{(n)}\rangle|B\rangle & \rightarrow & |0^{(n)}\rangle|B\rangle, & & |1^{(n)}\rangle|B\rangle & \rightarrow & -|1^{(n)}\rangle|B\rangle. \end{array}$$

Once this is done, $|A\rangle$ can be designed as follows. The state $\frac{1}{\sqrt{2}}(|0^{(n)}\rangle + |1^{(n)}\rangle)(|A\rangle + |B\rangle)$ is first designed and then the transformation given just above is applied to produce

$$\frac{1}{\sqrt{2}}(|0^{(n)}\rangle + |1^{(n)}\rangle)|A\rangle + \frac{1}{\sqrt{2}}(|0^{(n)}\rangle - |1^{(n)}\rangle)|B\rangle.$$

Finally, let us measure the first n qubits with respect to the basis

$$\{\frac{1}{\sqrt{2}}(|0^{(n)}\rangle + |1^{(n)}\rangle), \frac{1}{\sqrt{2}}(|0^{(0)}\rangle - |1^{(n)}\rangle)\}.$$

This measurement shows whether unmeasured qubits are in the state $|A\rangle$ or $|B\rangle$. In the second case we can obtain $|A\rangle$ from $|B\rangle$ by a NOT operation.

Remark 7.5.6 A general theory of fault-tolerant operations for stabilizer codes has been developed by Gottesman (1997). It is based on symmetries of the code stabilizer and it is shown that fault-tolerant universal computation is possible for any stabilizer code. Gottesman (1997) discussed in detail fault-tolerant computation based on LMPZ code.

Remark 7.5.7 Another general method, this time for a class of CSS-codes, to build a universal set of fault tolerant operations developed Steane (1998b). In addition, he has introduced new techniques to restrict accumulation of errors before and during recovery operations. On this basis he develops an optimistic evaluation of perspectives of quantum computing, from a certain point of view. Under the standard noise model of stochastic, uncorrelated errors, a quantum computer needs be only an order of magnitude larger than the logical machine contained within it in order to be reliable. For example, a scale-up by a factor 22, with error rate of order 10^{-5}, is sufficient to permit large quantum algorithms, such as factorization of thousand digit numbers.

Fault-tolerant gates

A different approach to fault tolerance of quantum gates was worked out by Kitaev (1997). He showed, theoretically, that there is a universal set of gates that are fault-tolerant by their physical nature and therefore they should be insensitive to local influences (and could be operated quite carelessly).

The possibility of a near-future physical realization of such gates is far from clear or, at least, it does not seem to be in sight. However, Kitaev's approach at least indicates that there may be essentially different ways to stabilization of quantum computations than those discussed above—to search for new physical principles that would lead to fault-tolerant quantum hardware.

Accuracy needed for fault-tolerant quantum computing

It is important to determine the upper bound on qubit and gate errors which still allows fault-tolerant quantum computing to be made sufficiently reliable.

If a code can correct e errors and the probability of one qubit or gate error is ε, then the probability of an error per gate working on encoded qubit is $\binom{m}{e+1}\varepsilon^{e+1}$, where m is the number of places an error can occur in an implementation of the gate, much smaller than ε. It would therefore seem that by choosing e large enough we can make the probability of an error per gate on encoded qubits as small as required. Unfortunately, this idea does not work because as e increases so does the complexity of the code and of all error processing-operations (especially syndrome computations).

For example, if the number of gates needed to perform the syndrome measurement increases as $\Theta(e^b)$, then the probability that $e+1$ errors accumulate before syndrome measurement is complete grows as $\Theta((e^b\binom{m}{e+1}\varepsilon)^{e+1})$.

In order to perform t steps of computation with a low probability of failure a quantum code is needed that can correct $\mathcal{O}(\lg t)$ errors. As analysed by Shor (1995), using the so-called self-dual Reed–Muller code, this can be achieved with codewords of length $\Theta(\lg^2 t)$. Using such codes, computation of syndromes requires the number of gates proportional to the number of 1s in the parity check matrix, which is $\Theta(\lg^3 t)$. If, to make computation reliable, syndrom computing is done $\Theta(\lg t)$ times, then the total number of gate operations needed is $\mathcal{O}(\lg^4 t)$. This implies that gates with error rate $\mathcal{O}(\frac{1}{\lg^4 t})$ are needed. (No much better estimation seems to be known.) This seems to imply that in any case there is a limit to how long a computation can be performed until an error is likely.

7.5.3 Concatenated coding

Fault-tolerant error-correcting codes and fault-tolerant gates are still not enough to ensure long reliable quantum computations. The problem is that these techniques require additional (often noisy) gates, additional computational overhead and, unless the basic error rate is very low, they may introduce more errors than they fix. In addition, the number of gates and time required may increase rapidly with the number of errors corrected. In total, error correction and fault-tolerant techniques are by themselves insufficient to cope with the decoherence and inaccuracies. What is required is a class of codes for which time needed to compute syndromes increases only slowly with error-correcting capabilities of the code and new methods and tools are needed that make long reliable quantum computations and transmissions possible.

Basic idea

The method of Knill and Laflamme (1996) shows one way how to deal with the problem. It allows one to transmit or store a qubit with an error at most ε, regardless of time and distance, provided the gates that are used work with an error at most $c\varepsilon$, for a constant c independent of ε, and storage or channel elements with an error ε at most, for some c and ε.

The basic idea is simple: to encode qubits recursively up to a certain level of recursion or hierarchy (see Figure 7.12), and to perform often recovery operations. The overhead of the method is polynomial in the time and storage and in the distance of transmission. The method could be implemented by having quantum repeaters used and spread at regular time or distance intervals of the quantum channel with sufficiently many parallel paths. The

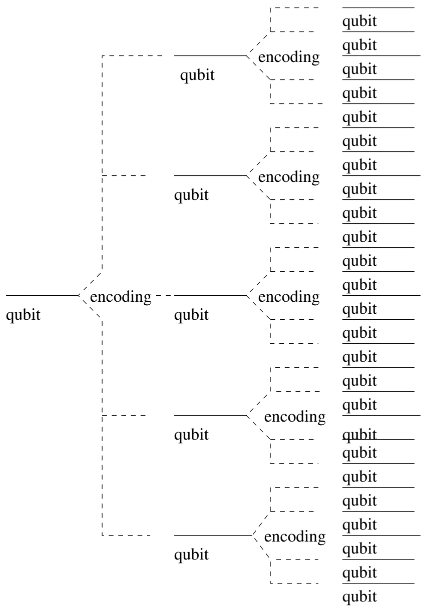


Figure 7.12: Concatenated codes

method can work in principle with any error-correcting code. Concerning the nature and frequency of errors it is only assumed that eerors occur randomly and independently and that "no leakage" assumption (see page 279) is satisfied.

For example, the data is first encoded using an $[[n, k, d]]$-code. Qubits of new codewords are then encoded using an $[[n_1, 1, d_1]]$-code. The resulting qubits are again encoded using some $[[n_1, 1, d_2]]$-code. This can continue up to a hth hierarchy level resulting in an $[[nn_1 \dots n_{h-1}, k, dd_1 \dots d_{h-1}]]$-code. In a special case all these encodings can be the same. The case $n = 5$, $k = 1$, $h = 2$ is illustrated in Figure 7.12. Computation of the error syndromes can be done in the time proportional to the sum of efforts to do so for all used codes provided we can do that in parallel for all blocks of each hierarchy level once starting with the last hierarchy level.

It is easy to see why several hierarchies of encodings can help. If a single encoding of one qubit into n is used, ε is a probability of an error per qubit, errors are uncorrelated and recovery is fault-tolerant, then the probability of error recovery failure is ε^2. However, if h level of encodings are used, then the probability of an error failure is ε^{2^h} (with the number of qubits n^h).

Recursive encoding has to be combined, in addition, with frequent recovery operations. The method can be simply described, for example (see Knill and Laflamme, 1996), in terms of the recursive concatenated coding procedure $\mathrm{CCP}_{h,r}$, where the parameter h specifies the depth of recursion and r the frequency of the error-recovery operations.

The basic level of the recursion, the procedure, $\mathrm{CCP}_{1,r}$, begins with one qubit, encodes it to n qubits using the code C, then applies the recovery operation $r-1$ times and, finally, decodes it back to a single qubit.

$\mathrm{CCP}_{h,r}$, $h > 1$, starts also with one qubit, encodes it into n qubits using the code C, applies $\mathrm{CCP}_{h-1,r}$ to each of n qubits of the code, applies $r-1$ times the recovery operation, then applies again $\mathrm{CCP}_{h-1,r}$ to each qubit and, finally, decodes the state into a single qubit.

Of course, two following recovery operations are to be performed only after some time or distance interval—after the so called recovery period.

Resources needed. If $\mathrm{CCP}_{h,r}$ is used, then the total number of (waiting or transmitting) periods is r^h and the total number of qubits needed is n^h. The number $p(h)$ of error-correcting operations (encodings, recoveries and decodings) is given by the recursion

$$p(1) = r + 1, \quad p(h+1) = r + 1 + rp(h)$$

with a solution $p(h) = \frac{r+1}{r-1}(r^h - 1)$.

The key problem is to determine the error threshold below which arbitrarily long computation is possible (Knill, Laflamme, Zurek (1996, 1997), Gottesman (1997)).

Example 7.5.8 *An analysis of the method for the case of LMPZ's code, made by Knill and Laflamme (1996), shows that a qubit can be stored for arbitrary amount of time or transmitted over arbitrary distance with error ε provided the following holds:*

$\varepsilon \geq \frac{1}{120}$**.** *The basic storage or channel elements have error at most $\frac{1}{120}$ and one or two qubit gates have error at most $\frac{1}{21600}$.*

$\varepsilon < \frac{1}{120}$**.** *The basic storage or channel elements have error at most ε and one or two qubit gates have error at most $\frac{\varepsilon(1-60\varepsilon^2)}{90}$.*

The concatenated coding method prefers to use such codes at which all gates used in the error-recovery operations can be performed **transversally**, i.e., a gate can be performed on a logical qubit by being performed on its particular qubits in a qubitwise manner. (For example Steane's 7-qubit code has this property.) If this is the case, then error-recovery operations can be performed simultaneously on all qubits at the highest level of hierarchy.

7.6 Experimental Quantum Processors -

It is still much too early to give a definite answer to the most basic question of quantum information processing from the practical point of view: will we soon (ever) have powerful quantum computers? At the moment the design of quantum gates and processors is in a very rudimentary stage—in its infancy. However, the experimental physics community have launched several projects, with various ingenious ideas, to build at least restricted working prototypes, and two very simple quantum algorithms, using 2 qubits, have already been "implemented in hardware" (see Chuang Chuang et al. 1998, 1998a, and Jones and Mosca, 1998), as well as a simple error-correcting method using 3 qubits (see Cory et al. 1998), and works on implementations of more qubit processors have been reported. The progress in the last years has been encouraging. However, there are still too many hard and basic problems to solve.

It has also been fully realized that the success of such academic projects may not be sufficient. Even if quantum computers could be built, scaling this up to computers large enough to be really significantly more useful than classical ones may cause difficulties hard to overcome.

On the experimental side the main current challenge actually is not to build a complete quantum computer right away, but rather to move from the experiments that exhibit quantum interference and entanglement to the experiments that can control these phenomena.

Currently it seems to be (quite) unfeasible to build a quantum computer sufficiently large and reliable to implement Shor's algorithm to factorize 200-digit integers—what would really be an achievement of great practical importance. However, even much more modest quantum processors would allow us to investigate problems of major scientific interest. It is expected that a 10-qubit processor would allow us to perform teleportation and a 100-qubit quantum processor could be of great importance for cryptography and would allow us to study problems of entanglement at very remote locations.

There are several good reasons why it is important:

1. to get better estimations whether and when and how powerful quantum computers could be available;

2. to have powerful quantum computers;

and therefore to invest a lot into quantum information processing research with the aim of answering these questions.

One important reason is that even a possibility of having a powerful quantum computer only in 20–40 years is of importance and danger already for current cryptography and information-processing security.

7.6.1 Main approaches

The first natural idea when the vision of designing quantum processors came up was, likely, to develop a natural successor of the current solid-state technology and to try to build quantum processors on the solid-state microchips. However, in the solid state devices the coupling with the environment seemed to be hopelessly strong. That is why the first experiments tried to use different technologies even if, as discussed later, a general feeling seems to converge to the understanding that the original idea has to get through if powerful quantum computers are ever built.

There have been a variety of early proposals and experimental attempts to design a quantum processor; for example Lloyd (1993). Some of them never got too far from the paper form; others brought at least some experimental experience. One interesting idea was that of heteropolymer based molecular computer developed in Stuttgart (see Teich, 1988) and at MIT (see Lloyd, 1994). For a detailed description see also Williams and Clearwater (1998).

Three the most interesting/promising current approaches seem to be to use quantum states of the laser-cooled ions in an magnetic trap (Cirac and Zoller, 1995, Monroe, 1995), cavity QED (Turchette et al. 1995), and, especially, nuclear magnetic resonance (Gershenfeld and Chuang, 1997, Cory et al. 1996).[13] These proposals are based on the-state-of-the-art experimental technologies developed already in quantum optics, atomic physics and magnetic resonance, and this makes them promising.

There are several sets of tasks or problems/difficulties that need to be overcome in order to be able to massage a technology to design a quantum processor.

1. Control. The available quantum states must be precisely identified. In addition, techniques have to be available for restricting states of the created quantum systems to the corresponding subspace of the Hilbert space. There must be a way to create quantum registers with qubits adequately isolated from interactions with the environment for the duration of computation.

[13] Of course a variety of other approaches are in different states of their "justifications" and "verifications".

2. Storage. Techniques have to be developed for storing quantum bits reliably for time needed to perform interesting computations.

3. Initial state. It has to be possible, in a reasonably simple way, to set the state vector into a given simple (initial) state; for example to the state $|0^{(n)}\rangle$. (This can sometimes be done by cooling the system into its ground state.)

4. Getting decoherence under a control. This requires to engineering quantum systems in which qubits effect each other, but not the environment. The coupling with the environment has to be sufficiently low in order not to destroy quantum interference and not to create undesirable entanglement.

5. Implementation of an adequate set of quantum gates. Actually, as already discussed in Section 2.3, it is sufficient to have good ways to realize all one-qubit gates and some control two-qubit gates. Quantum gates need to be implemented with sufficient precision. Rotation gates on single qubits are a more easy subproblem. To construct a controlllable "source of entanglement" a mechanism is needed by which two qubits can get entangled and this is usually a difficult task. (Practically, if one can implement with a technology the XOR gate, then "the door is open", at least in principle, to build any quantum circuit.)

6. Communications. The "wiring" or interconnection problem is far from trivial. On the contrary, in general it represents one of the most involved and uncertain features of quantum computing, that requires the development of new techniques not envisaged by former experimentalists. The key issue is to create controlled interactions of qubits.

   For an illustration, let us consider the network from Figure 1.14. When the first XOR gate is performed, there should be an interaction between qubits B and C and coupling between A and C should be zero. On the other hand, when the second XOR gate is performed, there should be a coupling between qubits a and c and zero coupling between a and b and b and c.

7. Measurement. The outcome of a quantum computation should be the result of a sequence of quantum measurements. It is essential that measurements can be made on specific subsystems (qubits) of the computational states; especially on each qubit separately.

7.6.2 Ion trap

The idea is due to Cirac and Zoller (1995), from the Innsbruck group, and was soon picked up by other groups: at NIST in Boulder, at the National Laboratory at Los Alamos, at IBM in Almaden and at the Max Planck Institute in Garching—at least. For a summarization of the efforts of the Boulder and Los Alamos groups and for their estimations of what seems to be achievable see Wineland et al. (1998, 1998a) and Hughes et al. (1998). The first demonstration of an ion trap XOR gate was due to Monroe et al. (1995). We discuss here mainly the Cirac–Zoller approach, in which a sequence of called ions is captured by a combination of static and electric fields in a so-called "Pauli trap" in a high vacuum. A single laser beam is split by beam splitters and acousto-optic modulators into many beam-pairs, one pair illuminating each ion. (The Boulder group used beryllium ions and the trap from radio-frequency waves.)

Problems to deal with:

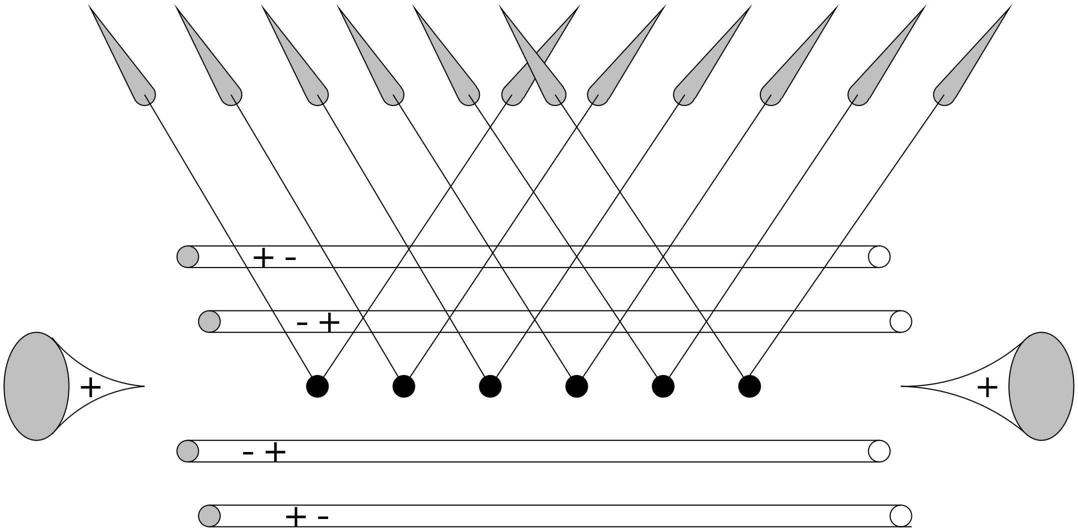


Figure 7.13: **Ion trap procesor**. Electrodes generate a time dependent electric field which generates an effective potential such that a string of ions is trapped—"stored in a linear trap". Each ion stores one qubit and it is addressed by a pair of laser beams. One-qubit operations are performed by shining precissely timed laser pulses on individual ions. The motional degree of freedom serves as a single "qubit bus" to transport quantum information among the ions. Two-qubit operations are performed by a sequence of laser pulses on the two particular ions. Initial state preparation is by optical pumping or laser cooling. Measurement is by electron shelving and resonance fluorerescence. This enables to measure the state of each ion with high signal to noise ratio (adapted from Steane, 1997, and Bennett, 1998).

Qubits. Each ion stores one qubit in its internal energy states. The quantum state of each ion is a superposition of the ground state (interpreted as $|0\rangle$) and the excited state (interpreted as $|1\rangle$).

Isolation. The ions are well isolated and spontaneous decay is the main source of decoherence. Other source of decoherence is the heat produced in the coupling between the charged ion strings and the noise voltage of electrodes.

Initial state. can be set up through optical pumping and laser cooling. This is a nontrivial technological problem because a very low temperature is needed.

One-qubit gates. Ions are sufficiently well separated and therefore they can be individually addressed by a pulsed laser. By shining precisely timed laser pulses and choosing the phase of the laser appropriately, any one-qubit unitary transformation can be performed.

Interaction of qubits. It is provided through the so-called joint vibrational mode. It is not easy to achieve and this is one of the major drawbacks of the ion-trap technology because this seems to prevent design of larger quantum registers. Two-qubit operations are performed by using a laser on one qubit to provide an impulse that ripples through a sequence of ions to the second qubit where another laser pulse stops the rippling and performs the operation. This way a single qubit "bus" is created to transport quantum information among the ions. Cirac and Zoller showed that the XOR gate can be implemented with ion-trap technology using altogether 5 pulses. In order to implement XOR-gate they encoded both qubits within a single beryllium ion.

Measurement., Readout is by electron shelving and resonance fluorescene. This allows the state of each ion to be measured with high reliability.

Speed. This seems to be a serious drawback of the ion-trap computers.

Performance conditions. They are very demanding. Extreme vacuum and extremely low temperature is needed.

The ion-trap approach does not seem to be able to deal easily with registers with a large number of qubits because of the problems with the interactions between ions. It has been proved difficult to progress beyond the one-qubit level. Estimates (guesses) for potentials of this technology much differ—from 10 or 12 till 47 qubits.

7.6.3 Cavity QED

An alternative technology to the ion-trap approach for the design of quantum processors is cavity QED (quantum electrodynamics) technology developed first at the Ecole Normale Supérieure in Paris (microwave cavity, by Horache's group, see Domokos et al. 1995) and then at Caltech (optical cavity, by Kimble's group, see Turchette et al., 1995).

The basic idea of the Caltech group approach is to trap several neutral atoms inside a high-fineness optical cavity. Quantum information is stored in the internal states of atoms or in the polarization of photons. In both cases it is reasonably easy to implement interactions of qubits via cavity QED.

To implement the XOR- gate (for a more detail description see, for example, Williams and Clearwater, 1998), both control and target qubits are implemented by beams of photons of different frequency and polarization (linearly polarized target qubit and circularly polarized control qubit), moving through an optical cavity (created by two mirrors). The target qubit is implemented by a beam of photons all of the same linear polarization. By lowering the intensity of the beam one can achieve that at any time there is at most one photon of the beam in the cavity. The control bit is implemented by a beam of photons all of either the right polarization or the left polarization. Again, by lowering the intensity of the beam one can achieve that at any time there is only one of those photons in the cavity.

To achieve an interaction of photons in the cavity, it is possible to make use of the fact that a linearly polarized photon can be seen as being in an equal superposition of the left and the right circularly polarized states. In order to help to make an interaction, a caesium atom drifts through the cavity. The distance between the mirrors of the cavity can be adjusted to a transition between two energy levels of the caesium atom and both photons.

When the target photon interacts with the caesium atom, its left-circular component is phase shifted by an amount that depends on the excited state of the caesium atom. (A linearly polarized target photon can be seen as being in an equal weight superposition of a left and right polarized state.) This in turn depends on which circular polarization control photon is. This way a conditional change of the state of the target qubit is achieved, depending on the state of the control photon. This rotation of the target qubit polarization is then measured.

7.6.4 Nuclear magnetic resonance (NMR)

A different approach is to adopt for quantum computing the nuclear magnetic resonance spectroscopy technology, that has already been used for years successfully in chemical analysis and for medical images.

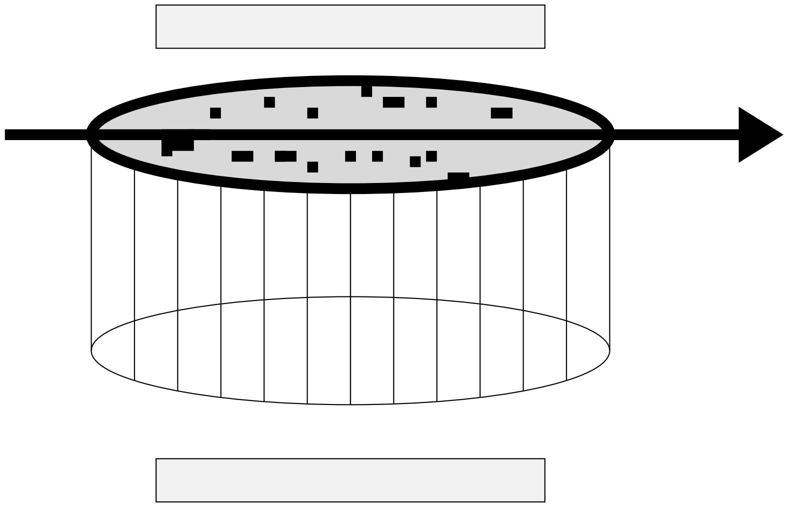


Figure 7.14: **NMR processor**. A liquid of $\approx 10^{20} - 10^{23}$ molecules ("subprocessors") is placed in a sensitive magnetometer which can generate oscillating magnetic fields and also detect the mean magnetic moment of the liquid. Each molecule is an independent processor. Qubits are stored and manipulated in the nuclear spin energy levels states. The spin energy levels of a given nucleus are influenced by neighbouring nuclei in the molecule. This allows XOR gate to be implemented. The initial state is one of thermal equilibrium. Only the average final state can be determined/ Experimental methods exist that can be used to extract out the final state—adopted from Steane (1997).

Research, development and even applications of the nuclear magnetic resonance technology already have long tradition and various experiments, in laboratories and hospitals, have already for many years routinely achieved spin-state manipulations and measurements equivalent in complexity to those required for quantum computing on a few qubits. In addition, this technology is known for relatively high decoherence time and capability to work at room temperature. It is therefore no wonder that with this technology it has proved relatively simple to design NMR quantum systems with 2 or 3 qubits, and it has been natural that the first two-qubits quantum algorithm implementations used NMR technology—as discussed in the example bellow.

The basic idea is that a quantum register is in this case a molecule containing a "backbone" of about ten atoms (with other atoms allowed so as to be able to create needed chemical bounds). Each qubit is realized in the spin orientation of an individual atomic nucleus—the direction of the nuclear magnetic dipole—in the atoms of the molecule. Each dipole can either reinforce or oppose an externally applied magnetic field. The first stage has lower energy than the second. The state can be changed by the absorption or emission of photons of right energy. The molecule is placed in a large magnetic field and the spin states of the nuclei are manipulated by applying oscillating magnetic fields in pulses of controlled duration.

The basic difficulty is that the spin state of the nuclei of a single molecule can neither be prepared nor measured. To overcome the problem not a single molecule but a cup of liquid containing about $10^{20} - 10^{23}$ identical molecules is used. The number of qubits of such an NMR computer is therefore equal to the number of backbone atoms per molecule. This implies that there is an enormous redundancy.

One approach is to encode a qubit in the average spin state of a large number of nuclei. The spin states are then manipulated by magnetic fields and the (average) spin state can be measured with the NMR spectroscopy technology.

Let us now deal in more detail with the problem of addressing and manipulating particular qubits. Using the so-called chemical shifts and spin–spin coupling one can identify the nuclear structure of the molecule of the sample. Single-qubit rotations are easy to implement. To them correspond rotations within a subspace corresponding to a single spin. Such rotations can be achieved using proper radio-frequency pulses. By using molecules with 2–3 magnetically active nuclei one can implement either XOR or Toffoli/Fredkin transformation. In both cases, first the current value of qubits is determined and then the corresponding sequence of radio-frequency pulses is applied.

The other major problem is how to extract results from such "massively parallel computers". The way out is to measure the average spin state. This seems to be a fundamental obstacle because quantum algorithms are probabilistic and averaging on an ensemble of molecules is not equivalent to the computation on a single device. The main new contribution of Gershenfeld et al. (1996) and Cory et al. (1998) was that they found a way how "effective pure state" could be prepared, manipulated and measured by performing suitable operations on the ensemble of molecules. One way to read out quantum information is to excite the spin system and to observe the resulting NMR spectrum. Different qubits correspond to different spins and give rise to signals at different resonance frequencies.

The NMR technology does not seem to scale up for several reasons: (1) an approach to distinguish qubits in a molecule by their chemical identities seems impossible for large molecules; (2) technology requires a too large redundancy; (3) it is hard to establish with sufficient precision an initial state. It is currently expected that with this technology one can hardly perform quantum computation with more than 12 qubits—unless radically new ideas appear.

The NMR technology was used to make the first three implementations of quantum algorithms: for the Deutsch problem discussed in Section 3.1 and for the Grover algorithm, see Section 3.3. In order to give a bit of insight into such an ventures, and a flavour of the expertise needed, let us provide a few basic details about the implementations of algorithms for Deutsch's problem—of slightly different algorithms from those presented in Section 3.1.

Example 7.6.1 *Jones and Mosca (1998) demonstrated that* ^{1}H *nuclei in partially deuterated cytosine can be used to implement a two-bit NMR quantum computer based on two coupled spins. They used a 50 mM solution of the pyrimidine base cytosine in* D_2O*. Rapid exchange of the two amine protons and the single amide proton with the deuterated solvent leaves two remaining protons forming an isolated two-spin system. All experiments used a home-made spectrometer with the* ^{1}H *operating frequency of 500 MHz. The first pair of Hadamard gates was replaced by so-called* 90°_y *pulses and the second pair by* 90°_{-y} *pulses. Also* U_f *transformation was implemented by a series of pulses. The final outcome was obtained by applying a* 90°_y *pulse and by observing the spectrum.*

Chuang et al. (1998) prepared the input state with a 200 mM, 0.5 ml sample of chloroform dissolved in d6-acetone, at room temperature and standard pressure. The $\mathcal{O}(10^{16})$

molecules in the bulk sample can be considered as independent quantum computers, all functioning simultaneously. Pulsed-radio frequence electromagnetic fields were applied to perform Hadamard rotations. The unitary mapping U_f was implemented using pulsed radio frequency and spin–spin interaction. The outcomes were read out by applying a special read-out pulses that transformed spin values into a voltage.

The first proposals for NMR implementations suffered from a signal-to-noise ratio that decayed exponentially with the number of qubits. An NMR implementation in which the signal-to-noise ratio depends only on features of NMR technology and not on the number of qubits was suggested by Schulman and Vazirani (1998). They gave also a new technology for preparing the initial state. In addition, they developed an abstract model of an NMR computer and proved several results.

Remark 7.6.2 For an introductory description of ion trapped and NMR technologies and implementations of NOT and XOR gates as well as basic gates needed to implement QFT (see Section 3.3), see Berman et al. (1998).

7.6.5 Other potential technologies

From other suggested technologies to build quantum gates and processors let us mention three approaches that are considered as promising. All of them have in common that they are solid-state technologies, and one could expect their scalability because the technology for making miniaturization of solid-state devices already exists. Decoherence problems are disadvantage.

1. Quantum-dot[14] ("artificial atoms") technologies. Quantum-dots are regions in a crystal that have artificial inhomogeneity and can be placed in a controlled manner. They can be used to trap single electrons and one can use the orientation of their spin to store information. For details see DiVincenzo (1995), Burkard, Loss and DiVincenzo (1998) and Barenco, Deutsch, Ekert and Jozsa (1995).
2. Super-conducting quantum interference devices (SQUIDs). In these systems the quantization of flux in a superconducting circuit is used to make a qubit.
3. A combination of the NMR technology with semiconductor physics (see Kane, 1998). The computer he suggest is made of the same components as the current semiconductor technology computers. The qubit is again a nucleus spin.

Moral: John von Neumann was one of the founders of modern quantum mechanics and also of modern computing; an extraordinary scientists who had great vision. We shall never know all his thoughts, but who can nowadays imagine von Neumann standing before ENIAC—weighting 60 tons and 30 m long—and envisioning a merger of quantum mechanics and computing? A good rule of thumb in searching for frontiers in computing is, as in life, never to underestimate the power of Nature, to keep learning from it as much as you can, to remember that an excellent theory is one of the most profitable products of mankind and that very fruitful results often come from combining two seemingly unrelated ideas into one.

---

[14] Quantum-dot technology is of importance also for classical computing in its attempts to develop single-electron transistors. It is already possible, even at room temperature, to transfer a single electron from a reservoir into a quantum dot in such a way that once inside, it blocks transfer of other electrons. The current through such a transistor depends on the number of electrons in the quantum dot. This allows us to "write" and "erase" information—see Berman et al. (1998).

Chapter 8

INFORMATION

INTRODUCTION

The development and the understanding of the basic concepts, methods and results of quantum information theory and of the faithful transmission of quantum information in time and space is the most fundamental problem of quantum information processing. In order to be able to understand and to utilize fully information processing available in nature the concepts of classical information theory need to be expanded to accumulate quantum information carriers. Three central problems concerning quantum information and its communication are dealt with, very briefly, in this chapter.

1. **Quantum information theory.** How to rebuild classical information theory on quantum foundations. How to define fundamental concepts of quantum information theory. How to identify and explore the inherently quantum elements of such a theory.
2. **Quantum transmissions theory.** How to use optimally quantum channels to send classical information and how to use optimally quantum and classical channels to transmit quantum information. How to define and determine the capacity of different quantum channels.
3. **Quantum entanglement theory.** How to quantify and manipulate entanglement and how to produce good entanglement.

LEARNING OBJECTIVES

The aim of the chapter is to learn:

1. the basic concepts of quantum information theory;
2. the basic concepts of quantum transmission;
3. the basic techniques of quantum data compression;
4. the basic techniques of communication through a noisy channel;
5. the basic modes and measures of entanglement;
6. the basic quantum entanglement concentration and purification techniques.

To do good and to cummunicate forget not

Hebrews ch 13, v 16

Quantum information theory and quantum information transmission problems are another area where quantum phenomena are of importance and bring new and surprising potentials, methods and results.

Quantum information theory exhibits on one side strong parallels and on the other side surprising and deep differences with respect to the classical information theory. The main outcomes of the classical information theory are the concepts of the source, channel, methods of encodings of information and results concerning capacities of quantum channels. Quantum information theory deals with quantum analogues of classical information theory concepts, methods and results. However, problems are this time much more complex. For example, there are two quantifiable kinds of information: classical and entanglement; there are at least four types of channel capacities and a variety of measures of entanglement.

In the classical communications we have only one communication primitive: transmission of a classical bit. In the quantum communication there are three communication primitives: transmiting a bit, transmitting a qubit and sharing an EPR-pair of qubits. Entanglement plays the central role in quantum information theory.

There are several fundamental problems to deal with in quantum information theory.

- How to use optimally quantum channels to send reliably classical information?
- How to use quantum and classical channels to transmit reliably quantum states?
- How classical and quantum information interact?
- How to develop qualitative and quantitative theory of entanglement?
- How to develop quantum error correction and fault-tolerant techniques?

With the first four problems we deal briefly in this chapter. Quantum error correction codes and networks as well as quantum computation fault-tolerant techniques were dealt with in Sections 7.4 and 7.5.

A special type of quantum/classical communication, on the basis of quantum teleportation and superdense coding, were dealt with in Section 6.4.

8.1 Quantum Entropy and Information

The basic problems of information theory are how to define such concepts as information and entropy formally, how to store information efficiently and faithfully and how to transmit efficiently information through a noiseless or noisy channel.

Classical information theory solved these basic problems elegantly and achieved enormous practical applicability. Modern communications, space exploration, and very high quality sound reproductions, for example, would be impossible without it.

8.1.1 Basic concepts of classical information theory

There are two basic approaches how to define the amount of information in an object. Shannon's approach defines the amount of information in an object as the number of bits that need to be transmitted in order to select the object from the previously agreed upon set of elements. Kolmogorov/Chaitin's approach considers the amount of information in an object as the number of bits needed to describe (to compress) the object. Theory of information developed using the second approach is called *algorithmic*. Even if some of the first attempts to apply the complexity point of view to physics (see Bennett, 1988a) used ideas close to Kolmogorov/Chaitin's approach so far Shannon's approach have been the basis of the concepts of quantum information theory.

If X is a random variable (a source), which takes a value x with probability $p(x)$, then the **(Shannon) entropy** of X is defined by

$$S(X) = -\sum_x p(x)\lg p(x)$$

and it is considered to be the **information content** of X or a measure of the lack of information about X. The maximum information which can be stored by an n-value variable is then $\lg n$.

It is natural to take as a unit of information the case that $S(X) = 1$, i.e., X can take two values, both with the same probability. A two-valued, or binary, variable can therefore contain one unit of information (called **bit**).

In the special case of binary variable, with two values, 1 with probability p and 0 with probability $1-p$, the **(binary) entropy function** $H(p)$ is defined by

$$H(p) = -p\lg p - (1-p)\lg(1-p).$$

The concepts of conditional probability $p(x|y)$ and mutual probability $p(x,y)$ influenced the definition of conditional and mutual entropy.

If X, Y are two random variables, then the **conditional entropy** $S(X|Y)$ is defined by

$$\begin{aligned} S(X|Y) &= -\sum_y p(y)\sum_x p(x|y)\lg p(x|y) \\ &= -\sum_y\sum_x p(x,y)\lg p(x|y), \end{aligned}$$

where $p(x,y) = p(y)p(x|y)$. $S(X|Y)$ can be seen as a measure of how much information, on average, would remain in X if we were to learn Y.

From the above relations one can easily deduce that

$$S(X|Y) = S(X,Y) - S(Y), \tag{8.1}$$

where $S(X,Y)$ is the information content of X and Y (i.e., the information one gains if, having a priori knowledge of neither, one learns the values of X and Y).

Another important concept is that of **mutual information** $I(X:Y)$ defined by

$$I(X:Y) = -\sum_x \sum_y p(x,y) \lg p(x:y),$$

where $p(x:y)$ is the mutual probability defined by $(p(x)p(y))/p(x,y)$. $I(X:Y)$ is a measure of how much X and Y contain information about each other. One can also say that $I(X:Y)$ is the amount of information about X that is obtained by determining the value of Y. It holds

$$\begin{aligned} I(X:Y) &= S(X) - S(X|Y) \\ &= S(X) + S(Y) - S(X,Y) \end{aligned}$$

and therefore $I(X:Y) = 0$ if and only if X and Y are independent.

Remark 8.1.1 The last identity can be used to describe the trade-off between entropy and information in measurements, if we write the above identity in the form

$$S(X) = S(X|Y) + I(X:Y).$$

Indeed, consider two random variables X, Y and a measurement of X and Y. $S(X|Y)$ can be seen as the entropy of X after having measured the values of X that become correlated by values of Y. $I(X:Y)$ is the information obtained about X via the measurement of Y. As it follows from the equality $S(X) = S(X|Y) + I(X:Y)$, $S(X|Y)$ and $I(X:Y)$ compensate each other such that $S(X)$ is unchanged. The second law of thermodynamic is not violated in a measurement—in spite of the decrease of $S(X|Y)$ (see Cerf and Adami, 1996).

Several simple inequalities are often useful.

1. Classical data processing inequality

$$X \to Y \to Z \Longrightarrow I(X:Z) \leq I(Y:Z)$$

says that if in a (Markov) process X develops to Y and later to Z, then Y cannot pass to Z more information about X than it has received.

2. If $X = X_1, \ldots, X_m$, $Z = Z_1, \ldots, Z_m$. then

$$S(X|Z) \leq \sum_{i=1}^{m} S(X_i|Z) \leq \sum_{i=1}^{m} S(X_i|Z_i).$$

8.1.2 Quantum entropy and information

The very basic concept of quantum information theory goes back to von Neumann (1927). Suppose that a quantum source (variable) X produces pure states $|\phi_1\rangle, \ldots, |\phi_m\rangle$ with probabilities $p_1, \ldots, p_m$. If ρ_X is the density matrix of the source, then[1]

$$QS(X) = QS(\rho_X) = -\mathrm{Tr}\rho_X \lg \rho_X$$

[1]In this section notation for quantum analogues of classical information and entropy concepts are derived from the classical ones using the prefix "Q". This Q is often ommitted in literature once the quantum case is clear from the context.

is called **von Neumann (quantum) entropy**[2] of ρ_X, or of X.[3] Quantum entropy of ρ expresses the degree of ignorance embodied in ρ_X.

Observe that in the case the states $|\phi_1\rangle, \dots, |\phi_m\rangle$ are mutually orthogonal, then ρ is a diagonal matrix. Let $p_1, \dots, p_m$ be elements of its diagonal. In such a case $QS(\rho) = S(p)$ for $p = (p_1, \dots, p_m)$, i.e. the quantum entropy is exactly the classical entropy. In general, however, the density matrix ρ has also off-diagonal elements and $QS(\rho_X) \leq S(p)$. Clearly $QS(\rho) = 0$ if and only if ρ is a pure state. and $QS(\rho) \geq 0$ in general. Quantum entropy can be arbitrarily large.

Basic properties of quantum entropy (see Wehrl, 1978): (1) **invariance:** $QS(\rho) = QS(U^*\rho U)$ for any unitary matrix U; (ii) **concavity:** if $\rho = \lambda_1\rho_1 + \lambda_2\rho_2$, $\lambda_1 + \lambda_2 = 1$, $\lambda_1 \geq, \lambda_2 \geq 0$, then $QS(\rho) \geq \lambda_1 QS(\rho_1) + \lambda_2 QS(\rho_2)$; (iii) **additivity:** $QS(\rho_1 \otimes \rho_2) = QS(\rho_1) + QS(\rho_2)$; (iv) **subadditivity**: if ρ is a density matrix of a bipartite system $A \otimes B$, $\rho_1 = TR_B\rho$, $\rho_2 = TR_A\rho$, then $QS(\rho) \leq QS(\rho_1) + QS(\rho_2) \leq QS(\rho_1 \otimes \rho_2)$; (v) **lower bound**: $S(\rho) \geq -\lg \lambda_1$, where λ_1 is the largest eigenvalue of ρ.

The relative quantum entropy $S(\rho_1|\rho_2)$ of two density matrices is defined as $S(\rho_1|\rho_2) = Tr(\rho_1(\lg \rho_1 - \lg \rho_2))$. For all density matrices ρ_1, ρ_2, $S(\rho_1|\rho_2) \geq 0$.

Though the concept of von Neumann quantum entropy has already been a standard one a for a long time a fully developed theory of quantum information does not yet exist. Only quite recent are attempts to provide quantum analogues of other basic concepts of information and entropy (see Cerf and Adami, 1996). Their approach will now be dealt with.

The **quantum conditional entropy** $QS(X|Y)$ is defined by a natural generalization of the classical case as

$$QS(X|Y) = -\mathrm{Tr}_{XY}[\rho_{XY} \lg \rho_{X|Y}],$$

where

$$\rho_{X|Y} = \lim_{n\to\infty}[\rho_{XY}^{\frac{1}{n}}(\mathbf{I}_X \otimes \rho_Y)^{-\frac{1}{n}}]^n \tag{8.2}$$

which should be an analogue to the conditional probability $p(x|y)$. In (8.2) $\mathbf{I}_A$ is the unit matrix in the Hilbert space for X and $\rho_Y = Tr[\rho_{XY}]$ denotes a "marginal" density matrix—an analogue to the "marginal" probability $p_y = \sum_x p(x,y)$.

The above definition of $QS(X|Y)$ allows to show that an analogue of the identity 8.1 holds also for quantum entropy.[4]

$$QS(X|Y) = QS(X,Y) - QS(Y).$$

In spite of the apparent similarity between quantum $QS(X|Y)$ and classical $S(X|Y)$ the fact that in the quantum case we deal with (density) matrices, rather than with numbers, as in the classical case, brings quite a different situation for quantum information theory and potential far exceeding the classical one.

[2] Actually, quantum entropy is defined as $QS = -k_B\, Tr\rho_X \lg \rho_X$, where k_B is Boltzman constant. By replacing k_B with 1 the entropy becomes dimensionless and has no direct physical meaning.

[3] Suppose that a classical random variable X has probability distribution $p(x)$. Let a quantum system be prepared in the state $|x\rangle$, dictated by the value of X, with the probability $p(x)$. In such a case it holds for the corresponding density matrix ρ that $QS(\rho)$ is an upper bound on the classical mutual information $I(X:Y)$ between X and the result Y of a measurement of an PVOM observable on the system (see Holevo, 1973, and Levitin, 1969).

[4] The quite complicated expression in (8.2), and the fact that a limit has to be used are due to the fact that the joint density matrix (ρ_{XY}) and the marginal matrices $\mathbf{I}_X \otimes \rho_Y$ do not commute in general. If they do commute, the whole expression gets much simpler, as discussed later.

The main point is that while $p(x|y)$ is a probability distribution in x (i.e. $0 \leq p(|y) \leq 1$), its quantum analogue $\rho(X|Y)$ is not a density matrix. It is Hermitian and positive but its eigenvalues can be larger than 1 and, consequently, the conditional entropy can be negative. This helps to explain the well-known fact that quantum entropy is non-monotonic, and it can be the case that $QS(X,Y) < QS(Y)$, i.e. the quantum entropy of the entire system can be smaller than the entropy of one of its subparts (what is not possible in the classical case). This happens, for example, in the case of quantum entanglement as shown in the following example.

Example 8.1.2 *(Cerf and Adami, 1996) Consider the Bell state* $|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle+|11\rangle)$ *of the Hilbert space* $H_{AB} = H_A \otimes H_B$*, where* $H_A = H_B = H_2$*. The density matrix* $\rho_{AB} = |\psi\rangle\langle\psi|$ *is shown in Figure 8.1a.* $\rho_A = \rho_B = Tr_{H_B}[\rho_{AB}] = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)$*, see Figure 8.1b.*

$$\rho_{AB} = \begin{pmatrix} 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 \end{pmatrix} \quad \rho_A = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \quad \rho_{A|B} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

(a) (b) (c)

Figure 8.1: Density matrices

Hence $QS(A) = QS(B) = 1$*. The density matrix* $\rho_{A|B} = \rho_{AB}(\mathbf{I}_A \otimes \rho_B)^{-1}$ *(because in this case the joint and marginal matrices* ρ_{AB} *and* $I_A \otimes \rho_B$ *commute) is shown in Figure 8.1c. Hence* $QS(AB) = QS(B) + QS(A|B) = 1 - 1 = 0$*, because* $QS(A|B) = -1$*.*

Quantum mutual information $QI(X:Y)$, as an analogue of the classical mutual information $I(X:Y)$ is defined on the base of "mutual" density matrix

$$\rho_{X:Y} = \lim_{n\to\infty}[(\rho_X \otimes \rho_Y)^{\frac{1}{n}}\rho_{XY}^{-\frac{1}{n}}]^n.$$

and the definition implies the standard relation

$$QI(X:Y) = QS(X) - QS(X|Y) = QS(X) + QS(Y) - QS(XY).$$

Observe that this definition is reduced to the classical one for the case ρ_{XY} is a diagonal matrix.

However, not all basic equalities and inequalities of the classical information theory transfer from the classical to the quantum case. For example, in the classical case we have

$$I(x:y) \leq \min[S(x), S(y)],$$

but in the quantum case the best upper bound possible is

$$QS(X:Y) \leq 2\min[QS(X), QS(Y)].$$

8.2 Quantum Channels and Data Compression

The question which is the most efficient way to code, to store and to transfer information is of large theoretical and especially practical interest. Coding is actually the central problem of the classical information theory.

8.2.1 Quantum sources, channels and transmissions

A classical (discrete, memoryless) channel is described by conditional probabilities p_{ij} of the output j for the input i. A quantum channel is described by a superoperator.

Let us now introduce quantum analogues of the classical concepts of the source and channel. A quantum **source** produces mixed states $\rho_1, \rho_2, \ldots$ with probabilities $p_1, p_2, \ldots$. (one also says it is an ensemble of mixed states each produced with certain probability.)

A **quantum noiseless channel** is any quantum system (sender) with the following properties:

1. it can be in any state of a finite dimensional Hilbert space;
2. it is entangled with some other (similar) quantum system (receiver);
3. it remains stable in this entangled state (while enroute the sender and the receiver).

In a similar way as a sequence of n bits can be used to transmit through a classical channel any of the 2^n messages from $\{0,1\}^n$, a sequence of n qubits can be used to transmit any state of H_{2^n}.

A **noisy quantum channel** is a quantum channel that interacts unitarily with an environment while enroute from the sender to the receiver. A noiseless channel is a special case of noisy channel. An equivalent way to define a noisy channel, as already mentioned, is that of an superoperator.

Quantum channels have a variety of implementations. A quantum channel can be a regular communication channel, for example an optic fibre, which gets on input a pure state $|\psi\rangle$, preserves coherence to some degree, and produces a mixed state $[\phi\rangle$ on the output. However, a quantum channel can be just "a flow of time" for a qubit "sitting in a quantum memory and interacting with its environment". It can also be the result of the operation of a noisy quantum gate, see Figure 8.2.

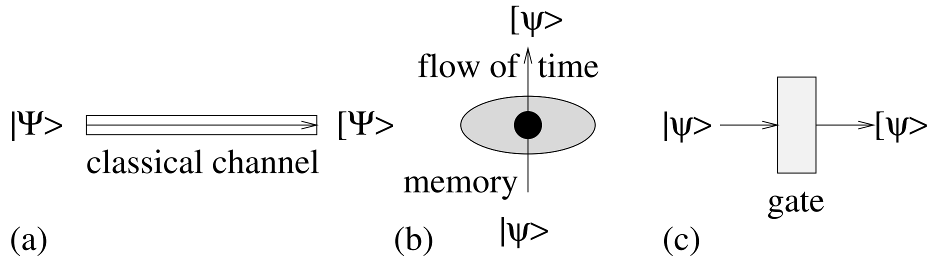


Figure 8.2: Quantum channels-80%

.

Transmissions through quantum channels

A quantum channel can be seen as a mean to transmit to the receiver states $\rho_i, i = 1, 2, \ldots$, produced by the source (or sent by the sender). Three cases are of special interest.

If the states ρ_i of the quantum source are orthogonal, then the source and the channel can be considered as classical. Indeed, in such a case complete information about states produced by the source can be extracted by a measurement of the sender, transmitted classically to the receiver who can then make arbitrarily many copies of the source states.

If the source states are pure and nonorthogonal, then no projection measurement can extract full information about a state. In addition, whenever a source state is transmitted through a quantum channel at most one faithful copy of the source state can be produced, and only if no faithful copy remains at the sender.

An interesting intermediate case is that the source states are nonorthogonal but commuting mixed states. Such source states ρ_i can be broadcast in the following sense. The two systems, A and B—that of sender and receiver—can be prepared in the joint state $\rho_i(AB)$ which is not a clone of the source states, i.e., $\rho_i(AB) \neq \rho_i(A) \otimes \rho_i(B)$, but can be obtained by tracing out each of the subsystems. Namely,

$$\rho_i = Tr_A \rho_i(AB) = Tr_B \rho_i(AB).$$

If density matrices of the source states do not commute, then the source can neither be cloned nor broadcasted (see Barnum et al. 1996), Bennett (1998a).

If a quantum channel is to transmit nonorthogonal states faithfully, it must operate on states that are sent through the channel "blindly" (in an "oblivious way")—without learning anything about them. This is due to the fact that quantum information can be neither read nor copied without disturbance.

Evolution of any quantum system can be seen as being done in the quantum channel of the time flow. This evolution remains deterministic till some information starts to leak into the environment. If this happens the channel gets noisy. The quantum state of the system gets randomized and entangled with the environment.

Transmission of an unknown quantum state requires the following resources.

1. A quantum resource (that cannot be cloned).
2. A directed communication resource (that cannot propagate faster than light).

There are two basic ways of transmission of quantum states. A direct unrestricted transmission of qubits at which the same particle provides both resources and functions. An indirect one, through teleportation, see Section 6.4, at which both functions are performed by an entangled pair of particles and by a communication through the classical channel.

Fidelity of transmissions

An important issue is how to measure quality of transmissions. (There are at least two reasons why a transmission will generally be imperfect: (i) data compression is needed because of the limited resources; (ii) noisy quantum channel corrupts the state being transmitted.) If source states are pure and a quantum channel produces on each pure input state $|\phi_i\rangle$, produced with probability p_i, an output (mixed state) W_i, then the quality of transmissions is measured by the fidelity (Jozsa and Schumacher, 1994)

$$F = \sum_i p_i \langle \phi_i | W_i | \phi_i \rangle. \tag{8.3}$$

F is the probability that a channel output state passes a test (conducted by someone who knows the inputs) for being the same as the input.

In the case the input states ρ_i are mixed the fidelity is defined as follows:

$$F = \sum_i p_i \left(Tr \left(\sqrt{\sqrt{\rho_i} W_I \sqrt{\rho_i}} \right) \right)^2. \tag{8.4}$$

As shown by Uhlmann (1976) and Jozsa (1994), (8.4) is a natural generalization of (8.3) and represents maximum of (8.3) over all purifications $|\phi_i\rangle$ of ρ_i. More formally, if we define $F(\rho_1, \rho_2) = (Tr\sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}})^2$, then $F(\rho_1, \rho_2) = \max\{|\langle\phi_1|\phi_2\rangle|^2 \mid \phi_i \text{ is a purification of } \rho_i\}$.

Exercise 8.2.1 (Jozsa (1994)) *Show the following properties of the fidelity function defined above: (a)* $0 \leq F(\rho_1, \rho_2) \leq 1$*; (b) If* $\rho_1 = |\phi\rangle\langle\phi|$ *is a pure state, then* $F(\rho_1, \rho_2) = \langle\phi|\rho_i|\phi\rangle$*; if* $\rho_1, \rho_2 \geq 0$, $p_1 + p_2 = 1$*, then* $F(\rho, p_1\rho_1 + p_2\rho_2) \geq p_1F(\rho, \rho_1) + p_2F(\rho, \rho_2)$*; (c)* $F(\rho_1 \otimes \rho_2, \rho_3 \otimes \rho_4) = F(\rho_1, \rho_3)F(\rho_2, \rho_4)$*; (d)* $F(\rho_1, \rho_2)$ *is preserved by unitary transformations.*

A quantum channel will be considered as faithful if, in an appropriate limit, the expected fidelity of transmission tends to unity. In other words the chance to distinguish channel outputs from inputs by any quantum measurement should tend to zero.

Remark 8.2.2 It may happen that a quantum state to be transmitted is entangled with some other state, We may now be intertested that entanglement is preserved by the transmission. Because of that it was not clear how to define capacity of quantum channels. To deal with the problem of the quality of transmission of entanglement the concept of **entanglement fidelity** has been introduced by Barnum et al. (1998) and they have shown the equivalence of the concept of quantum capacity based on entanglement fidelity and transmission fidelity—to be discussed later.

8.2.2 Shannon's coding theorems

The concept of entropy $S(X)$ of a random classical source (variable) X allows to state in a simple and elegant way one of the basic results of the classical data compression, which solves the following fundamental transmission problem: how many bits are needed to transmit through a noiseless channel, n values of X. The answer is elegant: $nS(X)$ bits. (They are sufficient in the limit and one cannot do better.) More precisely:

Theorem 8.2.3 (Shannon's noiseless coding theorem) *If* $x_1, \ldots, x_n$ *are independently obtained values of a random source* X *and* x_i *is obtained with probability* p_i*, then any string* $x_1 \ldots x_n$ *can be compressed to a* $nS(p)$*-bit string,* $(p = (p_1, \ldots, p_n))$*—and no more–in the following asymptotic sense.*

For any $\varepsilon, \delta > 0$*, for sufficiently large* n*, for any* $\lambda(n) \geq n[S(p) + \delta]$, $\lambda(n) \in \{1, \ldots, n\}$*, there exists an encoding (compression method) that encodes (compresses) each binary string* $x_1 \ldots x_n$ *to* $y_1 \ldots y_{\lambda(n)}$*, in such a way that* $x_1 \ldots x_n$ *can be successfully recovered from* $y_1 \ldots y_{\lambda(n)}$*, with probability greater than* $1 - \varepsilon$*. Moreover, the above compression method produces the minimal compression in the sense that for any* $\varepsilon, \delta > 0$*, for sufficiently large* n*, and* $\lambda(n) \leq n[S(p) - \delta]$*, for any compression method that maps* $x_1 \ldots x_n$ *into* $y_1 \ldots y_{\lambda(n)}$*, the probability that* $x_1 \ldots x_n$ *can be successfully recovered from* $y_1 \ldots y_{\lambda(n)}$ *is less than* $1 - \varepsilon$*.*

Since one of the main aims of this section is to present an analogous result for quantum data compression, let us discuss the basic ideas behind the proof of Shannon's theorem for the case of binary variables (Steane, 1997).

Let us assume that Alice wants to communicate to Bob a sequence X of n values $x_1, \ldots, x_n$ of a binary variable X that takes the value 1 with probability p and the value 0 with the probability $1 - p$. The mean number of 1's in X is in such a case np.

If np is an integer, then the probability of obtaining a binary sequence of length n with np 1's is

$$p^{np}(1-p)^{n-np} = 2^{-nH(p)} \tag{8.5}$$

Exercise 8.2.4 *Show in detail the equality (8.5).*

The result in (8.5) naturally leads to the following encoding strategy. A sequence X of n binary values is said to be an ε-**typical sequence**, $\varepsilon > 0$, if its probability $p(X)$ satisfies the inequality

$$2^{-n(H(p)+\varepsilon)} \leq p(X) \leq 2^{-n(H(p)-\varepsilon)}.$$

It can be shown that the probability that the sequence Alice wants to send is an ε-typical sequence is greater than $1-\varepsilon$, for sufficiently large n, no matter how small ε is.

The above facts lead naturally to the following communication decision. Alice does not need to communicate X to Bob directly. It is enough that Alice sends Bob information which of the typical sequences she wants to send. (Of course, they have to agree beforehand how to number all typical sequences.) Alice therefore sends only the number of the typical sequence instead of the sequence itself (and any nontypical , unlikely sequence, directly).

How good is the method? It can be shown that all of $2^{nH(p)}$ typical sequences have the same probability. In order to communicate a number of one of them, it is sufficient to send $nH(p) < n$ bits. In addition, there is no way for Alice to do better because all typical sequences have the same probability.

Shannon's noiseless theorem provides theoretical limitations how well can (classical) data be compressed. The next important task is to develop simple methods that can either achieve optimality or to get close to it.

Example 8.2.5 (Huffman code) *For $n = 4$ and $p = \frac{1}{4}$ the best possible data compression, according to Shannon's theorem, requires sending on average $4H(\frac{1}{4}) = 3.245$ bits for communicating 4 values of a binary variable X. A simple and practical method known as Huffman code, see table 8.3, in which less probable sequences are encoded by longer words and more probable sequences by shorter words, requires sending on average 3.273 bits per message.*

message	code	message	code	message	code	message	code
0000	10	0100	010	1000	011	1100	11101
0001	000	0101	11001	1001	11011	1101	111110
0010	001	0110	11010	1010	1100	1110	111101
0011	11000	0111	1111000	1011	111111	1111	1111001

Figure 8.3: Huffman code for $n = 4$

Remark 8.2.6 Shannon's noiseless coding theorem also provides an interpretation of Shannon entropy $S(X)$ as the mean number of bits necessary to code the outputs of a random variable X using an ideal code.

8.2.3 Schumacher's noiseless coding theorem

The basic quantum data compression problem and its solution are quantum analogues of the problem discussed in the previous section. This time we have a quantum source X producing a mixed state with the density matrix ρ and the task is to find an achievable limit on the number of qubits needed to transmit n pure states produced by the source. The answer is that $nQS(\rho)$ qubits are needed and sufficient in average to transmit n states of the source (Schumacher (1995). Let us now discuss both the problem and its solution in more detail, adopting the proof and example of Jozsa and Schumacher (1994).

Let a source X produce pure states (as signals) $|\psi_1\rangle, \ldots, |\psi_n\rangle$ in H_n, with probabilties $p_1, \ldots, p_n$. In addition, let the corresponding density matrix of the source X be

$$\rho = \sum_{i=1}^{n} p_i |\psi_i\rangle\langle\psi_i|.$$

Example 8.2.7 *Let the source produce states $|\phi\rangle$ and $|\psi\rangle$ both with probability $\frac{1}{2}$. Let $|\phi\rangle = \sqrt{0.9}|a\rangle + \sqrt{0.1}|b\rangle$ and $|\psi\rangle = \sqrt{0.9}|a\rangle - \sqrt{0.1}|b\rangle$, where $\{|a\rangle, |b\rangle\}$ is an orthonormal basis of H_2 . The density matrix of the source is $\begin{pmatrix} 0.9 & 0 \\ 0 & 0.1 \end{pmatrix}$ and $S(\rho) = H(0.9) = 0.469$.*

The basic idea of the encoding-transmission-decoding process goes as follows: n outputs (signals) of the quantum source X with the density matrix ρ are encoded by a state in a Hilbert space of a smaller dimension (than that of n-signals), transmitted and decoded into n signals again—this time of another source X' with the density matrix ρ'. It may happen that some information is lost in the encoding process, or during the transmission, and therefore we have to expect that in general $\rho \neq \rho'$. For the whole encoding-transfer-decoding process to be successful we require that their transmission fidelity $F(\rho, \rho')$ is as close to 1 as required.

It can be shown that for any $0 < \varepsilon \leq 1$ there is an encoding/decoding scheme requiring for transmission slightly more than $nQS(\rho)$ qubits with transmission fidelity $1 - \varepsilon$. This can be done as follows (if we restrict ourselves to the case that the pure states of the source are qubits).

If n signals are to be transmitted, then the total composed state of n symbol is represented by a vector of 2^n-dimensional Hilbert space. However, in case $QS(\rho) < 1$, it is very likely (and the probability goes to 1 for $n \to \infty$), that in any given realization the total state of n signals actually falls in a **typical subspace** of the Hilbert space of the composed block of signals (an analogue of typical sequences from the classical case). It has been shown by Jozsa and Schumacher (1994), that the dimension of the typical subspace is $2^{nQS(\rho)}$. As a consequence only $nQS(\rho)$ qubits are required to represent quantum information in n signals.

Example 8.2.8 *For blocks $s_1 s_2 s_3$ of three signals of the source X from Example 8.2.7 let us encode s_1 and s_2 using the mapping $|a\rangle \to |0\rangle$, $|b\rangle \to |1\rangle$ and let us ignore s_3. During the decoding process let us decode s_1 and s_2 fully and s_3 always as $|a\rangle$. Encodings are therefore into a 4-dimensional space and the fidelity of this encoding/decoding scheme is* 0.9.

Theorem 8.2.9 (Schumacher's noiseless coding theorem) *For any quantum source X with the density matrix ρ and any $\varepsilon, \delta > 0$ it holds:*

(a) If $S(\rho) + \delta$ qubits are available per signal, then for large n, there exists a coding/decoding scheme of fidelity $F \geq 1 - \varepsilon$ for strings of the signals of length n.

(b) If $S(\rho) - \delta$ qubits are available per signal, then, for any encoding/decoding scheme, some strings of length n will be decoded with the fidelity $F < \varepsilon$ for n sufficiently large.

Proof. We assume a quantum source X producing pure states (sequences of pure states) $|\psi_1\rangle, \ldots, |\psi_m\rangle$ of a Hilbert space H_n with the probability p_i for $|\psi_i\rangle$. The density matrix ρ of X is $\rho = \sum_{i=1}^m p_i |\phi_i\rangle\langle\phi_i|$. In the proof we use two lemmas.

Lemma 8.2.10 *Let $|\phi_i\rangle\langle\phi_i| \longleftrightarrow W_i$, $1 \le i \le m$, be an association of density matrices to signals, where each W_i is a density matrix of a mixed state with pure states over a d-dimensional subspace D of H_n. Let the sum of d largest eigenvalues of ρ be θ. Then the fidelity of transmissions $|\phi_i\rangle \longrightarrow W_i$ is at most θ.*

Proof of Lemma 8.2.10. Since each W_i is "supported" by D there is, for each $1 \le i \le m$, an orthonormal basis $|\psi_1^{(i)}\rangle, \ldots, |\psi_d^{(i)}\rangle$ of D consisting of eigenvectors of W_i such that

$$W_i = \sum_{j=1}^{d} g_j^{(i)} |\psi_j^i\rangle\langle\psi_j^{(i)}|,$$

where $0 \le g_j^{(i)} \le 1$. Denote by Γ the projection into D. Hence for any $1 \le i \le m$, $\Gamma = \sum_{j=1}^d |\psi_j^{(i)}\rangle\langle\psi_j^{(i)}|$. In addition, for any $1 \le i \le m$, we have

$$\langle\phi_i|W_i|\phi_i\rangle = \sum_{i=1}^{d} g_j^{(i)} \langle\phi_i|\psi_j^{(i)}\rangle\langle\psi_j^{(i)}|\phi_i\rangle = Tr(|\phi_i\rangle\langle\phi_i|\Gamma).$$

Hence, for the fidelity F it holds

$$F \le \sum_{i=1}^{m} p_i \, Tr(|\phi_i\rangle\langle\phi_i|\Gamma) = Tr(\rho\Gamma).$$

Let now $|e_1\rangle, \ldots, |e_n\rangle$ be an orthonormal basis of eigenvectors of ρ with the corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Then

$$Tr(\rho\Gamma) = \sum_{i=1}^{n} \lambda_i \langle e_i|\Gamma|e_i\rangle,$$

where $0 \le \langle e_i|\Gamma|e_i\rangle \le 1$, and $\sum_{i=1}^n \langle e_i|\Gamma|e_i\rangle = Tr(\Gamma) = d$, see Exercises 9.2.24. From the minimax properties of eigenvalues (see Kato, 1978), this implies that $Tr(\rho\Gamma) \le \theta$.

Observe that the above upper bound for $Tr(\rho\Gamma)$ is obtained if D is the subspace spanned by d eigenvectors corresponding to d largest eigenvalues. ∎

The lemma claims that for a fixed d we get the highest fidelity by taking as D the space generated by eigenvectors corresponding to d largest eigenvalues. Let us denote by D_X such a subspace of H_n. We present now an encoding-decoding method based on the choice od D as D_X.

Let $|0\rangle$ be any fixed state in D_X and let $D_X^\perp$ be the orthogonal complement of D_X in H_n. For each $|\phi_i\rangle$ we construct W_i as follows:

If $|\phi_i\rangle = \alpha_i|l_i\rangle + \beta_i|r_i\rangle$, where $|l_i\rangle \in D, |r_i\rangle \in D_X^\perp$ are unit vectors, then

$$W_i = |\alpha_i|^2 |l_i\rangle\langle l_i| + |\beta_i|^2 |0\rangle\langle 0|. \tag{8.6}$$

Technically, W_i can be designed by applying observable P_{D_X} on $|\phi_i\rangle$ and, if the result is 0, i.e. $|\psi_i\rangle$ projects to $D_X^\perp$, $|0\rangle$ is taken as the post-measurement state.

Lemma 8.2.11 *Suppose that the sum of the d largest eigenvalues of ρ is greater than $1-\varepsilon$. Then the coding/decoding process defined by the mapping $|\phi_i\rangle \longleftrightarrow W_i$ has fidelity $F > 1-2\varepsilon$.*

Proof. On the base of equalities and inequalities

$$\langle\phi_i|W_i|\phi_i\rangle = |\alpha_i|^2\langle\phi_i|l_i\rangle^2 + |\beta_i|^2|\langle\phi_i|0\rangle|^2 \geq |\alpha_i|^2|\langle\phi_i|l_i\rangle|^2 = |\alpha_i|^4 \geq 2|\alpha_i|^2 - 1$$

we get

$$F = \sum_{i=1}^{m} p_i\langle\phi_i|W_i|\phi_i\rangle \geq 2\sum_{i=1}^{m}(p_i|\phi_i|^2) - 1 = 2\,Tr(\rho P_{D_X^\perp}) - 1 \geq 1 - 2\varepsilon,$$

if we make use of the fact that $\sum_{i=1}^{m} p_i|\alpha_i|^2 = Tr(\rho P_(D_X^\perp)$ is the sum of d largest eigenvalues and therefore it is larger than $1-\varepsilon$. ∎

To finish the proof of Schumacher's theorem let us formulate another of its ingredients—a quantum modification of the classical result on typical sequences mentioned in Section 8.2.2.

Let $\lambda = \{\lambda_1, \ldots, \lambda_m\}$ be probabilities of X and let for any $k \in \mathbf{N}$, Λ_k be the set of probabilities (if elements of k-tuples are multipied and taken to form a multiset) of X^k. Let the probability of any subset of X_k be the sum of probabilities of its members. The result on typical sequences we will use has the form (Jozsa and Schumacher, 1994):

Lemma 8.2.12 *Let $\varepsilon, \delta > 0$. (1) For sufficiently large k, the set X_k may be partitioned into a subset L of "likely sequences" with at most $2^{k(QS(X)+\delta)}$ elements, which has probability greater than $1-\varepsilon$, and into its complement of "unlikely sequences", which has the probability smaller than ε.*

(2) Any subset of X_k, which has less than $2^{k(QS(\rho)-\delta}$ elements, has probability smaller than $2^{k(QS(\rho)-\delta)}$.

Continuation of the proof of Schumacher's theorem. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of the density matrix ρ of X. $\lambda_1, \ldots, \lambda_n$ can be seen as forming a probability distribution Λ as considered above. In this case $QS(\rho) = S(p)$ for $p = (\lambda_1, \ldots, \lambda_n)$. The k-block version X^k has the density matrix $\rho_k = \bigotimes_{i=1}^{k} \rho$ and its eigenvalues are products of k-tuples of vectors from X^k.

Suppose now that $\varepsilon, \delta > 0$ are given. From the part (1) of Lemma 8.2.12, it follows that for all sufficient large k there is a set of $2^{k(QS(\rho)+\delta)}$ eigenvalues of ρ_k with the total sum greater than $1 - \frac{1}{2}\varepsilon$. Hence the sum of $2^{k(QS(X)+\delta)}$ largest eigenvalues is also greater than $1 - \frac{1}{2}\varepsilon$. By Lemma 8.2.11 there exists a coding scheme for X^k using $k(QS(\rho_k) + \delta)$-qubits per signal of X^k, i.e., $(QS(\rho) + \delta)$ qubits per signal of X, with fidelity $F > 1 - \varepsilon$, for k signals. This proves the part (a) of Schumacher's theorem.

From the point (2) of Lemma 8.2.12 it follows that for sufficiently large k any subset of X^k of the size less than $2^{k(QS(\rho)-\delta)}$ has probability less than ε; that is the sum of the $2^{k(QS(\rho)-\delta)}$ largest eigenvalues of ρ_k has the sum less than ε. By Lemma 8.2.10 every encoding/decoding scheme for X^k using $k(QS(\rho) - \delta)$ qubits per symbol of X^k, that is $QS(\rho) - \delta$ qubits per signal of X, will have fidelity less than ε for k signals. Since this holds for all sufficiently large k, this proves the part (b) of Schumacher's theorem. ∎

Example 8.2.13 *Let us consider again the source from Example 8.2.7 and let H_a be its 4-dimensional subspace spanned by the following set of three qubit states*

$\{|aaa\rangle, |aab\rangle, |aba\rangle, |baa\rangle\}$ *having majority of as. Let us use the following encoding/decoding scheme:*

Encoding. *Let the states in H_a be encoded as follows:*

$$|aaaa\rangle \to |00\rangle, \quad |aab\rangle \to |01\rangle, \quad |aba\rangle \to |10\rangle, \quad |baa\rangle \to |11\rangle \tag{8.7}$$

and states not in H_a as $|00\rangle$.

Decoding. *Each two qubit state is decoded as the corresponding state in H_a, following encodings (8.7).*

Analysis. *Observe that due to our encoding methods none of the actual triples $s_1s_2s_3$ of signals is encoded perfectly because none of them lies in H_a. In spite of that we get high fidelity for the overall encoding/decoding scheme. Let Γ be the projection into H_a. The density matrix ρ_a has a diagonal form with elements*

$$(0.9^3, 0.9^2 \cdot 0.1, 0.9^2 \cdot 0.1, 0.9 \cdot 0.1^2, 0.9 \cdot 0.1^2, 0.9 \cdot 0.1^2, 0.9 \cdot 0.1^2, 0.1^3)$$

and therefore $Tr(\rho\Gamma) = 0.9^3 + 3(0.9)^2 0.1 = 0.972 = 1 - 0.18$. By Lemma 8.2.10, this encoding/decoding scheme has fidelity at least $0.964 = 1 - 2 \cdot 0.18$.

Remark 8.2.14 The first attempts to develop quantum information theory were based on applying classical information theory to probabilities derived from the representations of quantum mixed states. Schumacher's theorem has actually been the beginning of a new approach, in which quantum information theory is based on concepts, measures and codings that are inherently quantum.

Remark 8.2.15 Observe that quantum data compression has the following remarkable property: it allows to compress and expand each of the 2^n of distinct sequences of nonorthogonal states with fidelity approaching 1 even though the sequences cannot be reliably distinguished from one another by any measurement.

8.2.4 Dense quantum coding

There is a special way, due to Ambainis et al. (1998), presented in the following, how to encode m bits into $n < m$ qubits, for n not too much smaller than m, in such a way, which has no classical analogue, that each bit, but not all, can be retrieved with significant probability.

The Holevo theorem (see Holevo, 1973) says that no more than m bits can be transmitted (faithfully) by transferring m qubits. Codings described below are therefore quite surprising, even if they do not contradict Holevo theorem because even if each of m bits can be retrieved from $n < m$ qubits with significant probability, not all can be retrieved. These codings make use of special properties of quantum projection measurements. Namely, the fact that projection measurements are not commutative and a particular measurement can destroy information that could be revealed by other measurement.

Encodings discussed in the following have three parameters, n, m and p (notation $m \xrightarrow{p} n$). They consist of mapping m bits into n qubits and, in addition, in providing m measurements $\mathcal{O}_i$, $1 \le i \le m$, such that using the measurement $\mathcal{O}_i$ the ith bit can be retrieved from encoded qubits with probability p. Since such encodings use a random number as a parameter, we talk about random access encodings.

Definition 8.2.16 *A* $m \xrightarrow{p} n$ **random access encoding** *is a function* $f : \{0,1\}^m \times \mathcal{R}_{\{0,1\}} \rightarrow H_n$ *such that for any* $1 \leq i \leq m$ *there is a projection measurement* $\mathcal{O}_i$ *which when applied to the value-state of* f *returns* 0 *or* 1 *and such that for any* $b \in \{0,1\}^m$, $Pr(\mathcal{O}_i(f(b,r)) = b_i) \geq p$, *where* $\mathcal{R}$ *is a source of random bits (*f *is called the encoding function and* $\mathcal{O}_i$ *are decoding mappings).*

Example 8.2.17 *(A* $2 \xrightarrow{0.85} 1$ *encoding.) Let* $u_0 = |0\rangle$, $u_1 = |1\rangle$, $|v_0\rangle = |0'\rangle$, $|v_1\rangle = |1'\rangle$ *and* $f(x_1, x_2) = \frac{1}{\sqrt{2+\sqrt{2}}}(|u_{x_1}\rangle + |v_{x_2}\rangle)$. *The mapping* f *has the desirable properties provided the first (second) bit is measured with respect to the standard (dual) basis. (This follows from the fact that the distance between the codeword and the right subspace is* $\frac{\pi}{8}$ *and the access probability is therefore* $\cos^2(\frac{\pi}{8}) \approx 0.853$.)

A classical bit-to-bit $m \xrightarrow{p} n$ encoding can be defined similarly and it has been proved by Ambainis et al. (1998) that no $2 \xrightarrow{p} 1$ classical random access encoding exists if $p > \frac{1}{2}$. The potential of classical $m \xrightarrow{p} n$ encodings is already well understood.

Lemma 8.2.18 *If* $\frac{1}{2} < p \leq 1$, *then* $n \geq (1 - H(p))m$ *for any classical* $m \xrightarrow{p} n$ *encoding.*

Proof. Let f be a (possibly probabilistic) encoding mapping randomly and uniformly chosen $X = X_1 \ldots X_m$ from $\{0,1\}^m$ into $Y = f(X) \in \{0,1\}^n$. In addition, let $Z = Z_1 \ldots Z_m$ be random variables with values in $\{0,1\}^m$ obtained using the decoding. Now the claim of the lemma follows from two general series of inequalities concerning Shannon entropy and mutual information.

$$n \geq S(Y) \geq I(X : Y) = S(X) - S(X|Y) = m - S(X|Y),$$

$$S(X|Y) \leq S(X|Z) \leq \sum_{i=1}^{m} S(X_i|Z) \leq \sum_{i=1}^{m} S(X_i|Z_i)$$

and from the inequality $S(X_i|Z_i) \leq H(p)$. ∎

It has been shown by Ambainis et al. (1998) that the bound in Lemma 8.2.18 is pretty good. Namely, there is a classical $m \xrightarrow{p} n$ encoding with $n = (1 - H(p))m + \mathcal{O}(\lg n)$ if $p > \frac{1}{2}$. In the quantum case it holds.

Theorem 8.2.19 *If there is a* $m \xrightarrow{p} n$ *quantum encoding where* $p > \frac{1}{2}$, *then* $n > \Omega(\frac{m}{\lg m})$.

Theorem 8.2.19 is a corollary of the following two lemmas:

Lemma 8.2.20 *If there is an* $m \xrightarrow{p} n$ *encoding for some* $p > \frac{1}{2}$, *then there is also an* $m \xrightarrow{1-\varepsilon} \mathcal{O}(n \lg \frac{1}{\varepsilon})$ *encoding for any* $\varepsilon > 0$.

Proof. Given an encoding $f : \{0,1\}^m \times \mathcal{R} \rightarrow H_n$ with decodings $\mathcal{O}_i$, $1 \leq i \leq m$ and the success probability $p > \frac{1}{2}$, let for $t \in \mathbf{N}$, $f_t\{0,1\}^m \times \mathcal{R}^t \rightarrow \bigoplus_{i=1}^n H_n$ be a new encoding defined by $f_t(x, r_1, \ldots, r_t) = \bigoplus_{i=1}^n f(x, r_i)$—the tensor product of t independent copies of the original code. Let the new decoding $\mathcal{O}_i'$ consist of applying $\mathcal{O}_i$ to each of t copies of the code and taking the majority of the outcomes as the final 0/1 result. By applying one of Chernoff's bounds we get that the error probability decreases exponentially with t and it is therefore at most ε if $t = \mathcal{O}(\lg \frac{1}{\varepsilon})$. ∎

Lemma 8.2.21 *If there is an* $m \stackrel{1-\varepsilon}{\to} n$ *quantum encoding with* $\varepsilon < \frac{1}{64m^2}$, *then* $n = \Omega(m)$.

For the proof, which is quite technical, see Ambainis et al. (1998).

Open problem 8.2.22 *How good is lower bound of Theorem 8.2.19?*

The existence of $n \stackrel{0.85}{\to} n$ encoding follows from Example 8.2.17. Encoding $3n \stackrel{0.79}{\to} n$ is discussed in the following exercise.

Exercise 8.2.23 *Design a* $3 \stackrel{0.79}{\to} 1$ *encoding. (Hint: make eight vertices of a maximal cube embedded in the Bloch sphere to encode* 3 *bits. For decoding use: standard, dual and circular basis.)*[5]

Lemma 8.2.18 and Theorem 8.2.19 show the gap between the classical and quantum random access encodings.

8.2.5 Quantum Noisy Channel Transmissions

For quantum information to be useful, it must be transmitted and processed reliably. The fundamental limits on amount of quantum information that can be transmitted reliably along a noisy channel are therefore of large importance.

Classical channel capacity is the greatest number C such that for any $R < C$ and any $\varepsilon > 0$, there exist m and n and an encoding (error-correcting code mapping) $\mathcal{E}$ of any m-bit strings into n bit strings, where $\frac{m}{n} > R$ such that every m bit string x can be received, after decoding $\mathcal{D}$, with error probability less than ε at the receiving end of the channel, from its output y. In short:

$$C = \limsup\{c \,|\, \forall_{R<c}\forall_{\varepsilon>0}\exists_{\mathcal{E},\mathcal{D}}\exists_{m,n}(\frac{m}{n} > R \wedge Pr(\mathcal{D}(y) \neq x) < \varepsilon)\}.$$

The above result about the capacity of the noisy classical channel is called Shannon's bound for the capacity of noisy channels.

The basic definition of the capacity $QC(\mathcal{N})$ of quantum channels parallel that for classical channels. Namely, $QC(\mathcal{N})$ is the largest number R such that for any $\varepsilon > 0$ there exists n such that any state of n qubits can be transmitted with fidelity at least $1 - \varepsilon$ by fewer that $\frac{n}{R}$ (forward) uses of the channel $\mathcal{N}$. n qubits are encoded into m channel inputs, then transmitted and, finally decoded.

The basic definition of the capacity of a quantum noisy channel $\mathcal{N}$ can be concisely expressed as follows (Lloyd, 1997, Barnum et al. 1997, 1997a, Bennett, 1998):

$$QC(\mathcal{N}) = \lim_{\varepsilon\to 0}\limsup_{n\to\infty}\{\frac{n}{m} \,|\, \exists_{m,\mathcal{E},\mathcal{D}}\forall_{\psi\in H_{2^n}(\langle\psi|\mathcal{D}\mathcal{N}^{\otimes m}\mathcal{E}(|\psi\rangle\langle\psi|}|\psi\rangle > 1-\varepsilon)\}.$$

where $\mathcal{E}$ is an encoding superoperator from n qubits to m quantum channels, see Figure 8.4, and $\mathcal{D}$ is a decoding superoperator from m channel outputs to n qubits.

A communication through a quantum channel can be accompanied by a communication through a classical channel. The above definition of the capacity of a quantum channel refers only to the transmission unassisted by classical communications. Two other measures

[5] Personal communication: I. Chuang → A. Ambainis, A. Nzyak, A. Ta Shma and U. Vazirani → the author.

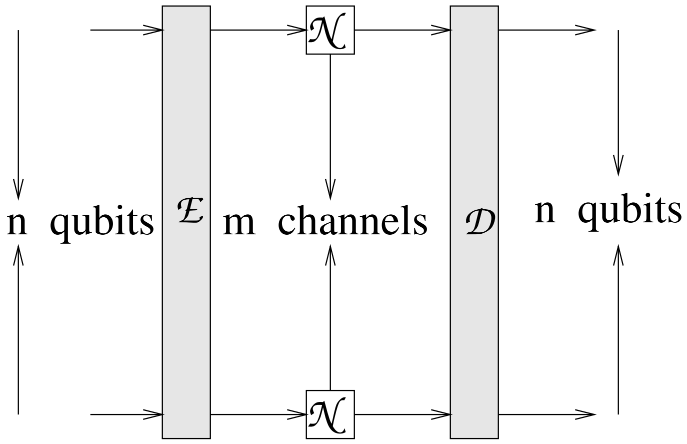


Figure 8.4: Channel transmission scheme

of quantum channel capacity, $QC_1(\mathcal{N})$ and $QC_2(\mathcal{N})$, are defined as asymptotic quantum capacities of the quantum channel $\mathcal{N}$ assisted by one-way classical communication (from the sender to the receiver) or by two-way classical communication, between the sender and the receiver (see Figure 8.5), where the superoperator $\mathcal{E}$ stands for the encoding and $\mathcal{D}$ for the decoding processes. It has been shown, by Bennett et al. (1996a), that one-way classical

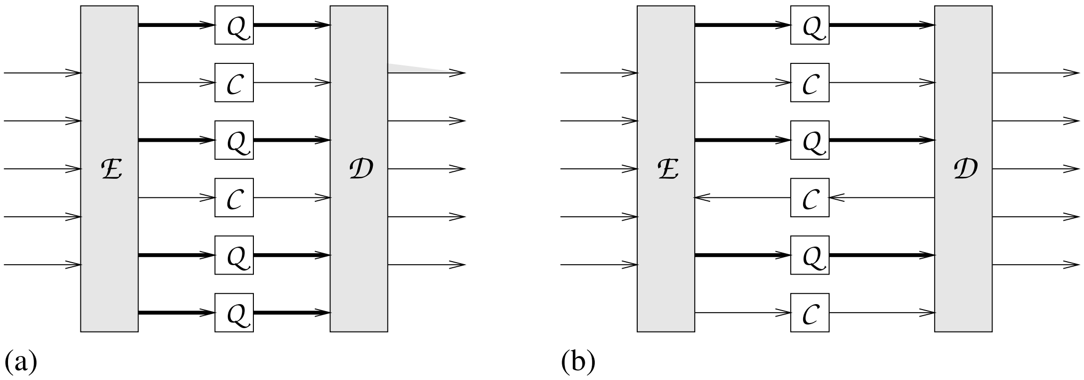


Figure 8.5: Communication through a quantum channel assisted by one-way and two-way classical communication. $\mathcal{Q}$ denotes a quantum channel and $\mathcal{C}$ a classical channel

communication does not help and therefore $QC_1(\mathcal{N}) = QC(\mathcal{N})$ for any quantum channel. On the other hand, two-way classical communication can increase the capacity of some noisy quantum channels.

The basic relations between these capacities can be summarized as follows (see Bennett et al. 1996a and Bennett, 1998b):

- $QC(\mathcal{N}) \leq C(\mathcal{N})$ for all quantum channels $\mathcal{N}$ and $QC(\mathcal{N}) < C(\theta)$ for some $\mathcal{N}$.
- $QC(\mathcal{N}) \leq QC_2(\mathcal{N})$ for all quantum channels $\mathcal{N}$ and $QC(\mathcal{N}) < QC_2(\mathcal{N})$ for some $\mathcal{N}$.
- $QC_2(\mathcal{N}) \leq C(\mathcal{N})$ for all known quantum channels $\mathcal{N}$ and $QC_2(\mathcal{N}) < C(\mathcal{N})$ for some.

The inequality $QC(\mathcal{N}) \leq QC_2(\mathcal{N})$, follows from the definition. From the fact that orthogonal quantum states can be used to transmit bits, it follows that $QC(\mathcal{N}) \leq C(\mathcal{N})$. No channel $\mathcal{N}$ is known for which $QC_2(\mathcal{N}) > C(\mathcal{N})$.

8.2.6 Capacities of erasure and depolarizing channels

Quantum capacities are known well only for few types of quantum channels. With two such cases we deal now: the erasure channel and the depolarizing channel. For most of other channels at most some upper and lower bounds are known.

The quantum **erasure channel** (QECH, see Grassl et al. 1996), is the channel in which every qubit sent through the channel either gets through fine, with probability $1-p$, or it gets totally random, with probability p and, in addition, it is known to the receiver, which case has occurred.

In the case of 2EPP, if Alice sends n EPR pairs through the erasure channel, then pn of them get destroyed and $(1-p)n$ will be intact. In addition, Bob knows which pair remain intact. He informs Alice and they discard the remaining pairs. The resulting rate is $1-p$ and clearly one cannot do better. Hence $QC_2(QECH) = 1-p$.

No cloning argument can be used to show (see Bennett et al. 1997a), that the erasure channel has the capacity at most $1-2p$ and therefore 0 if $p \geq \frac{1}{2}$. Hence $QC(QECH) = \max\{0, 1-2p\}$.

The quantum **depolarizing channel** (QDCH) is a very natural channel. Each qubit is left intact with probability $1-p$ and it is affected by each of the error operators $\sigma_x, \sigma_y, \sigma_z$ with the same probability $\frac{p}{3}$. The receiver does not know on which qubits randomization has been performed.

Another definition of the erasure channel is that the incoming qubit is, with probability p, replaced by an "erasure state" $|2\rangle$, orthogonal to the states $|0\rangle$ and $|1\rangle$ (thereby erasing the qubit), and informing the receiver that it has been erased.

For the depolarizing channel $\mathcal{D}$ the following bounds are known (see Bennett et al. (1996).

if $p < 0.25408$ **then** $QC(QDCH) > 0, QC_2(QDCH) > 0, C(QDCH) > 0$;

if $\frac{1}{3} < p < \frac{2}{3}$ **then** $QC_1(QDCH) = 0, QC_2(\mathcal{D}) > 0$, $C(QDCH) > 0$;

if $\frac{2}{3} \leq p \leq 1$ **then** : $QC_2(QDCH) = 0$, $C(QDCH) > 0$.

8.3 Quantum Entanglement

Quantum entanglement can be seen as the purest form of quantum information (see Bennett, 1998a). Quantitative and qualitative theory of entanglement as well as methods of creating, preserving, transmitting, and manipulating entanglement are therefore of great interest and importance for quantum information processing.

Entanglement is also an important communication resource. For example, for teleportation and dense coding. For that we mostly need pure entangled states (otherwise the EPR-channel gets noisy) and maximally entangled states—as for superdense coding.

Unfortunately, pure and maximally entangled states are not easy to create and to preserve. In practice we have therefore mainly noisy and week entanglement. An important problem is therefore whether there is an efficient way to purify noisy entanglement or, in other words, to distill pure entanglement out of the noisy one. Fortunately, the answer is positive.

8.3.1 Transformation and the partial order of entangled states.

The very basic questions concerning the existence and methods of transformations of one entangled state into another by local operations and classical communication are already quite well understood. An elegant necessary and sufficient condition for the existence of such a transformation is known, in the terms of the majorization of vectors of eigenvalues of the corresponding bipartite states, due to Nielsen (1998). The result implies the existence of different (incommensurate) types of entanglement. In order to formulate the main result the concept of "majorization" between vectors needs to be defined.

Definition 8.3.1 *A real-valued vector* $X = (x_1, \ldots, x_d)$ *is* **majorized** *by a real-valued vector* $Y = (y_1, \ldots, y_d)$*, notation* $X \prec Y$*, if, for* $1 \leq k \leq d$*,* $\sum_{j=1}^{k} x_j^{\downarrow} \leq \sum_{j=1}^{k} y_j^{\downarrow}$ *with equality holding for* $k = d$*, where* $x_i^{\downarrow}$ *denotes the* i*th largest element among* $x_1, \ldots, x_d$*.*

For a pure state $|\phi\rangle$ of a bipartite Hilbert space $H_A \otimes H_B$, let λ_ϕ denotes the vector of eigenvalues of the H_A-state $\rho_\phi = Tr_{H_B}(|\phi\rangle\langle\phi|)$.

Theorem 8.3.2 *(Nielsen (1998) If* $|\psi\rangle$ *and* $|\phi\rangle$ *are pure states of* $H_A \otimes H_B$*, then* $|\psi\rangle$ *can be transformed to* $|\phi\rangle$ *using local operations and classical communication if and only if* $\lambda_\psi \prec \lambda_\phi$*.*

Two states in $H_A \otimes H_b$ are called **incommensurate** if none can be obtained from another by local operations and classical communication.

Example 8.3.3 *Incommensurate are, for example, states*

$$|\psi\rangle = \sqrt{\frac{1}{2}}|11\rangle + \sqrt{\frac{2}{5}}|22\rangle + \sqrt{\frac{1}{10}}|33\rangle \text{ and } |\phi\rangle = \sqrt{\frac{3}{5}}|11\rangle + \sqrt{\frac{1}{5}}|22\rangle + \sqrt{\frac{1}{5}}|33\rangle$$

The proof of Theorem 8.3.2 also yields a method how to construct $|\psi\rangle$ out of $|\phi\rangle$ using at most $d-1$ bits of classical communication.

Open problem 8.3.4

1. *Find a necessary and sufficient condition that a mixed state* ρ_1 *can be transformed into a mixed state* ρ_2 *using only local operations and classical communication.*
2. *Find a necessary and sufficient condition to transform one three (or more) party entangled state into another one by local operations and classical communications of all involved parties.*

8.3.2 Entanglement purification/distillation

Starting with the pioneering paper by Bennett et al. (1996b), several purification protocols have been developed.

The idea behind the basic purification step is simple. Alice generates two pairs of entangled qubits (a_1, b_1) and (a_2, b_2) and sends the second of each pair to Bob through a noisy channel. Alice performs XOR on a_1 and a_2 and Bob on b_1 and b_2 and both measure target qubits—see Figure 8.6— a_2 and b_2, and communicate to each other the results of their measurements. If they agree they consider the first entangled pair (a_1, b_1) as good. If not they discard both pairs.

The overall purification process then goes as follows (see Figure 8.7). Alice generates pairs (a_1, b_1), (a_2, b_2), ... and sends the second particle of each pair to Bob (through a

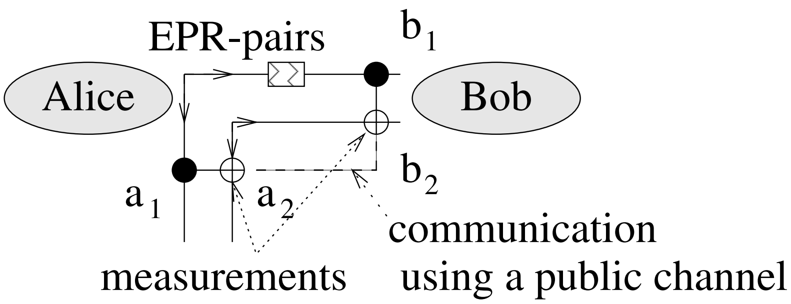


Figure 8.6: Basic step of the entanglement purification

noisy channel). To each pair of particles (a_{2i-1}, b_{2i-1}), (a_{2i}, b_{2i}) they perform the above purification step at the end of which they either keep the pair (a_{2i-1}, b_{2i-1}) or discard both pairs.

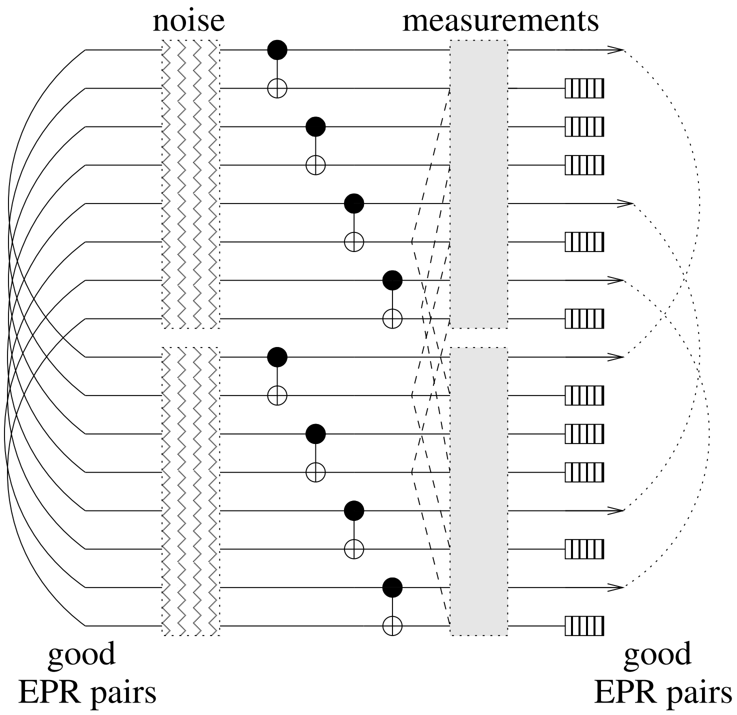


Figure 8.7: Entanglement purification with 2-way communication-80%

The input for the purification process do not have to be perfectly entangled pairs and such pairs can be produced by a source and get through the noisy channels on both ways to Alice and to Bob.

The above purification protocol can be modified and generalized by allowing both parties to perform first various local operations (superoperators) and to use some other post-selection procedures. Efficiency of purification is the main issue.

Example 8.3.5 *(Bennett et al. 1996b) Let us assume that a purification procedure is applied to a mixed state ρ and the aim is to obtain the singleton $|\Psi^-\rangle$. The purity of ρ, with respect to the goal $|\Psi^-\rangle$, can be expressed by the fidelity $F = \langle\Psi^-|\rho|\Psi^-\rangle$. In order to explore the impact of a given purification procedure instead of ρ we can consider the so-called* **Werner state**

$$W_F = F|\Psi^-\rangle\langle\Psi^-| + \frac{1-F}{3}(|\Psi^+\rangle\langle\Psi^+| + |\Phi^+\rangle\langle\Phi^+| + |\Phi^-\rangle\langle\Phi^-|), \tag{8.8}$$

which has the same fidelity as ρ.

In the simplified purification protocol given by Bennett et al. (1996b), at first Hadamard rotations are applied locally, on each of the two pairs and after the BXOR operation such rotations are applied again. They showed that in such a case new mixed state has with probability $\frac{1}{4}$ a better fidelity F', with respect to $|\Psi^-\rangle$, namely

$$F' = \frac{F^2 + \frac{1}{9}(1-F)^2}{F^2 + \frac{2}{3}F(1-F) + \frac{5}{9}(1-F)^2} > F.$$

Purification with one-way communication

The above model of purification is also called **purification with 2-way communication**, or **two-way entanglement purification protocol** (2EPP), because both parties have to communicate classically to reach the outcome.

Purification protocols with one-way classical communication (1EPP) have been developed by Bennett et al. (1996a). They have also proved that 2EPP are more powerful than 1EPP. For example, 1EPP cannot purify Werner states W_F if $F \leq \frac{5}{8}$, but there is a positive lower bound (0.00457) on the yield when $W_{\frac{5}{8}}$ is purified using 2EPP.

Entanglement purification and error correction

Entanglement purification protocols can be seen also as providing, in addition to QECC, another way of protecting quantum states from their interaction with the environment. Not only that, a close relation between quantum error-correcting codes and 1EPP can be shown.

The scenario we deal with is that Alice prepares a number of EPR pairs and sends one member of each pair to Bob. In a 1EPP Bob can receive quantum and classical information from Alice but cannot communicate to Alice.

At first we show how to make out of an quantum error correcting code a 1EPP. Alice generates EPR-pairs and encodes those qubits she wants to sent to Bob and sends them, encoded, to Bob. He performs error correction and decoding. The encoded qubits retain their entanglement with Alice's qubits. This way distillation works on 100%.

Secondly, let us have an 1EPP capable to distill k good EPR-pairs out of n noisy pairs. We show how to use such 1EPP to make a QECC. Alice will be the encoder and Bob the decoder.

Alice creates n EPR-pairs, sends them (the corresponding parts) to Bob and performs her part of 1EPP. In addition, she takes k qubits she wants to protect using her half of what will be the k good pairs and sends the classical information about the measurement results to Bob. Afterwards Bob completes the purification protocol to get k good pairs. Since they are good pairs he can complete the teleportation protocol. The whole process can be seen as encoding k qubits by n qubits.

8.3.3 Entanglement concentration and dilution

Entanglement purification requires classical public communication. Another technique, with similar aim, that requires no public communication is the so-called entanglement.

Entanglement concentration is a method to concentrate entanglement of a large number of partially entangled states into a smaller number of maximally entangled states. On the other hand, **entanglement dilution** is a method to dilute a small set of maximally entangled states into a larger set of partially entangled states (see Bennett et al. 1996).

8.3.4 Quantifying entanglement

As discussed in Section 2.2.2 the amount of entanglement $E(\phi)$ of a pure state $|\phi\rangle$ of a bipartite system $A \oplus B$ is defined, for $\rho = |\phi\rangle\langle\phi|$, through quantum entropy by

$$E(\phi) = -Tr\rho_A \lg \rho_A = -Tr\rho_B \lg \rho_B.$$

Using the Schmidt decomposition theorem, page 374, $|\phi\rangle$ can be expressed in the form

$$|\phi\rangle = \sum_{i=1}^{d} c_i|\alpha_i\rangle|\beta_i\rangle,$$

where c_i are positive and $\{\alpha_i\}_{i=1}^d$, $\{\beta_i\}_{i=1}^d$ are orthonormal sets of states in A and B. In such a case $\rho_A = T_B|\phi\rangle\langle\phi| = \sum_{i=1}^d c_i|\alpha_i\rangle\langle\alpha_i|$, and $E(\phi) = \sum_{i=1}^d -c_i^2 \lg c_i^2$.

A strong justification of the above definition of the entanglement of pure states is the following result (see Bennett et al. 1996b). Consider n entangled pairs of particles, each in the state $|\phi\rangle$. Let Alice hold one particle of each pair and Bob, spatially separated, holds the other. If $|\phi\rangle$ has E ebits of entanglement, then n pairs can be reversibly converted, by local operations and classical communications, into m singletons, where $\frac{m}{n}$ approaches E for large n and the fidelity approaches 1.

Example 8.3.6 *(Bennett et al. 1996a) If a state $|\phi\rangle$ has representation $|\phi\rangle = \sum_{i=1}^4 \alpha_i|m_i\rangle$, where $|m_i\rangle$ are states of the magic basis, then $E(\phi) = H(\frac{1}{2} + \frac{1}{2}\sqrt{1 - \sum_{i=1}^4 |\alpha_i|^2})$.*

Several measures of entanglement have been defined and investigated for mixed states and density matrices They are based on clear physical ideas. Each of them can be seen as a fundamental concept of quantum information theory.

Entanglement of formation (creation), E_f, (Bennett et al. 1996a), is defined for a density matrix ρ by

$$E_f(\rho) = \min\{\sum_i p_i E(\psi_i) \,|\, \rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|\},$$

where ρ is the density matrix for mixed states$\bigoplus_{i=1}^n (p_i, \psi_i)$. In other words, entanglement of formation, $E_f(\rho)$, is the least expected entanglement of any ensemble of pure states, with ρ as the density matrix.

Entanglement of formation is known exactly for many states, including all two-qubit states (see Hill and Wooters, 1997), having no more than two non-zero eigenvalues. In addition, Uhlmann (1997) has shown how to compute entanglement of formation for those states where E_f is not known to have a closed form.

Open problem 8.3.7 *Is entanglement of formation additive?*

Entanglement of distillation (distillable entanglement), E_d (see Bennett et al. 1996a), is intuitively, the maximum asymptotic yield of singleton states that can be produced from the given mixed state by local operations and classical communication.

Distillable entanglement is further classified by the type of classical communication allowed (no (E_{d0}), one-way (E_{d1}), and (E_{d2}) two-way classical communication) and depends also on the type of the local operations allowed.

Another way to define, informally, entanglement of distillation of a mixed state ρ is the maximum over all allowable protocols of the expected rate at which singletons can be obtained from a sequence of states ρ. For a rigorous treatment of the entanglement of distillation see Rains (1998).

Entanglement of assistance, E_a, (DiVincenzo et al. 1998), is in a sense dual to the entanglement of formation

$$E_(\rho) = \max\{\sum_i p_i E(\psi_i) \mid \rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|\},$$

Entanglement of relative-entropy, E_r**,** (Vedral and Plenio, 1997) defined by

$$E_r(\rho) = \min_{\rho' \in \mathcal{D}} (Tr\rho(\lg\rho - \lg\rho')),$$

where $\mathcal{D}$ denotes the set of unentangled mixed states.

For the entanglement of relative-entropy the following relation holds:

$$E_r(\rho_1 \otimes \rho_2) = E_r(\rho_1)E_r(\rho_2)$$

for all density matrices ρ_1 and ρ_2.

The following relations hold among the above entanglement measures (see Vedral and Plenio, 1997, and Bennett et al. 1996a):

$$E_{d0} \leq E_{d1} \leq E_{d2} \leq E_r \leq E_f.$$

All these entanglement measures are reduced to the entropy of entanglement for pure states.

By Vedral and Plenio (1997), a measure E of entanglement should satisfy the following conditions; (a) $E(\rho) = 0$ if and only if the state ρ is separable; (b) local unitary operations leave $E(\rho)$ invariant; (c) the expected entanglement cannot increase under local (general) measurements and classical communication.

8.3.5 Bound entanglement

A surprising discovery about the existence of two qualitatively different types of entanglement of mixed states was made by Horodecki et al. (1998). They showed that there is a

- "free", distillable entanglement;

and a

- "bound", enatnglement that is not distillable.

Horodecki et al. (1997) showed that any entangled mixed state of two qubits can be distilled to get the singleton. However, this is not true in general for entangled states. For example, there are entangled mixed states of two qutrits that cannot be distilled (and therefore they cannot be used for a quantum communication).

Example 8.3.8 *(Horodecki et al. 1998a) Two qutrit mixed states*

$$\sigma_+ = \frac{1}{3}(|0\rangle|1\rangle\langle 0|\langle 1| + |1\rangle|2\rangle\langle 1|\langle 2| + |2\rangle|0\rangle\langle 2|\langle 0|), \sigma_- = \frac{1}{3}(|1\rangle|0\rangle\langle 1|\langle 0| + |2\rangle|1\rangle\langle 2|\langle 1| + |0\rangle|2\rangle\langle 0|\langle 2|)$$

are separable. However, the state

$$\sigma_\alpha = \frac{2}{7}|\Psi^+\rangle\langle\Psi^+| + \frac{\alpha}{7}\sigma_+ + \frac{5-\alpha}{7}\sigma_-,$$

where $|\Psi^+\rangle = \frac{1}{\sqrt{3}}(|0\rangle|0\rangle + |1\rangle|1\rangle + |2\rangle|2\rangle)$, *is*

$$\begin{array}{rll} \textit{separable,} & \textit{if} & 2 \le \alpha \le 3 \\ \textit{bound entangled,} & \textit{if} & 3 < \alpha \le 4 \\ \textit{free entangled,} & \textit{if} & 4 > \alpha \le 5 \end{array}$$

A method to construct states with bound entanglement was developed by DiVincenzo et al. 1998a).

8.4 Quantum information processing principles and primitives

The development of concepts, methods, and results of importance for quantum information processing has to be accompanied by a study of the mosr fundamental principles and primitives of quantum information theory.

8.4.1 Search for quantum information principles

The thesis that information is physical and that the role of information in physics is analogous to that of energy in thermodynamics leads naturally to the search for informati on processing principles and laws. For example for principles and lawsındexlaw analogous to those in thermodynamics. It is only natural t hat quantum entanglement is expected to play the key role in suc h principles and laws. One such emerging principles seems to be **no-increasing of entanglement principle** (under local operations and communic ation).

Informally, this principle says that entanglement cannot inc rease under local operations and classical communication. More formally, it says that if E is a "proper" measure of entanglement and an input state ρ is transformed by quantum operations into the mixed state $\{(p_i, \rho_i)\}_{i=1}^k$, then $\sum_{i=1}^k p_i E(\rho_i) \le E(\rho)$.

This principle has been proven to hold for entanglement of f ormation (see Bennett et al. 1997a). A form of this principle has been proven to be equivalent with no-cloning theorem (see M. Horodecki and R. Horodecki, 1997). They also formulated the following implications of the principle: "The entanglement of the com pound system does not change under unitary processes in one of the subsystems.

8.4.2 Quantum information processing primitives

Identification and comparison of quantum information processing primitives as well as the study of reducibilities among them is one of the most fundamental goals of quantum information processing theory.

Two such primitives are, for example, qubits and ebits. Let us therefore summarize some of their basic relations, and properties.

Qubits are, from a point of view of transmissions, a directed channel resource (Bennett, 1998a). They can be sent from a sender to a receiver in a particular direction. On the other hand ebits are undirected resource shared between a sender and a receiver.

Ebits are a weaker communication resource than qubits in the following sense. One can create one ebit by transmitting one qubit. On the other hand, sharing any number of ebits does not suffice by itself to transmit a qubit in either direction.

Both qubits and ebits, as communication primitives, have to be considered in two settings: in communication through a noiseless channel and through a noisy channel,

A detailed analysis of quantum information processing primitives and reducibilities has been done by Bennett (1998).

Appendix

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

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

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

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

9.1 Quantum Theory

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

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

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

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

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

On the other hand, quantum theory is often seen as forcing us to accept, as the best we currently have, philosophically highly unsatisfying views of the world that do not square with our common sense perception of the functioning of the Universe.[1]

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

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

9.1.1 Pre-history of quantum theory

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

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

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

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

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

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

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

In 1900 Max Planck[2] came with a revolutionary theory that electromagnetic oscillations could carry energy only in "quantas", the energy E of which satisfies the relation

$$E = h\nu$$

to the frequency ν, where h is known today as Planck constant.[3] With this idea the black-body radiation problem could be dealt with.

In spite of Planck's ability to come up with a new theory in accordance with the experimental results, as in the case of the black body radiation problem, his theory did not receive too much attention until the next step occurred. Einstein came up with the theory that electromagnetic field also can exist only in discrete units and on this basis he was able to explain the photoelectric effect.[4] This implied, for example, that light itself must actually be particles, because it was known, due to the results of Maxwell, that light consists of the oscillations of the electromagnetic fields. However, it was also established experimentally that light sometimes behaves as waves. -

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

Niels Bohr[5] made an important use of the Planck relation in 1913. He discovered that the **angular momentum—spin**—of electrons in orbit about the nucleus can occur only in integer multipies of the number $\frac{h}{2\pi}$, for which Dirac introduced the symbol $\hbar$. The only permissible values of the spin of electrons are therefore

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

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

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

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

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

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

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

In 1923, Louis de Broglie proposed that anything which oscillates with some frequency ν can occur only in discrete units of mass $\frac{h\nu}{c^2}$ (which came from a combination of the Planck identity $E = h\nu$ with Einstein equation $E = mc^2$).[6] This implies that particles sometimes behave as waves and sometimes as fields and that particles and field oscillations are the same things. De Broglie stated the hypothesis that the wave–particle duality is a general property of microscopic objects. This brought an elegant and unified view of Nature but it could not be considered as a new consistent theory yet.

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

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

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

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

Computer science, and actually even the most theoretical parts of it, has been developed, and is still developed, basically as a servant of the computer industry and, consequently, with quite restricted and short-term goals. This is why it actually ignored in a sense its most basic scientific goal—to explore fully the potentials and limitations of computing and communication based on the laws of physics[7]. Theoretical computer science, especially complexity theory, found interest and importance to work with computer models far outside the framework of current classical computers, but these models, such as alternating Turing machines, were considered only as tools. Attempts to see the main scientific aims of fundamental computer science (informatics), in a broader context, as in Gruska (1993), have been and still are rare.[8]

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

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

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

On the other hand, physics has only very slowly developed an understanding that information is a physically important phenomenon, concept, and resource[9]. In addition, physics practically ignored an important fact that within computer science in general, and in complexity theory in particular, important new concepts, methods, results and insights have been developed that can be useful also for extending our understanding of the physical world. It is also worth noticing that one could hardly expect some deeper developments in quantum computing before complexity theory made significant advances and found ways to classify computational tasks in a deep way, and before such of its modern branches as randomized complexity theory were developed.

9.1.2 Heisenberg's uncertainty principle

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

In 1927 Heisenberg[10] discovered one of the fundamental principles and limitations of quantum mechanics: a bound on the accuracies with which certain pairs of observables are measured. The principle says that measuring the value of one observable more accurately makes the value of another, noncommutative, observable less certain. In addition, there is a certain intrinsic uncertainty with which values of two observables can be measured, and in general once a way of measurement is fixed this uncertainty, in general, increases.

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

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

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

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

Heisenberg's principle claims that the following lower bound on the uncertainties holds when two observables are measured independently.[11]

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

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

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

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

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

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

Many instances of Heisenberg's principle are well known. One of them, and historically the first one, considered already by Heisenberg, concerns the simultaneous measurement of the position and the momentum of particles and it is called Heisenberg's microscope. Classically (see Figure 9.1a), you can measure both, position and momentum. Quantum mechanically, if you measure the position precisely (Figure 9.1b), then for various possible

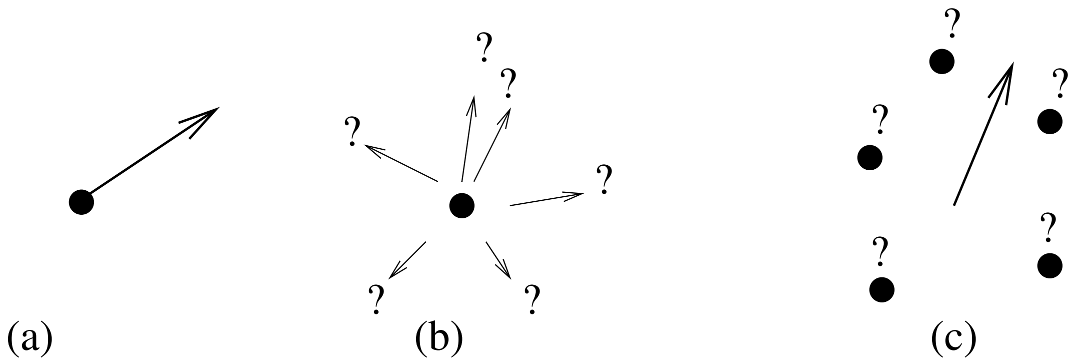


Figure 9.1: Measurement of the position and of the momentum of particles

values of the momentum only probabilities are given. Similarly, if you measure precisely momentum first (Figure 9.1c), then there are many options for positions and for each of them only the probability is known. More exactly, in this case it is assumed that two conjugate observables, the position and momentum, are measured simultaneously, which results in the fact that precision in the measurement of position is obtained at the expense of precision of the measurement of momentum.

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

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

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

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

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

Example 9.1.2 (Uncertainty principle for polarized photons) *Photons*[12] *are elec-*

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

tromagnetic waves and their electric and magnetic fields are perpendicular to the direction of propagation and also to each other. An important property of photons is polarization—it refers to the bias of the electric field in the electromagnetic field of the photon. If the electric field vector is always parallel to a fixed line we have **linear polarization** *(see Figure 9.2). If the electric field vector rotates about the direction of propagation forming a (right-) left-handed screw, we have (right) left* **eliptic polarization**. *If the rotating electric field vector inscribes a circle, we have (right) left* **circular polarization**.

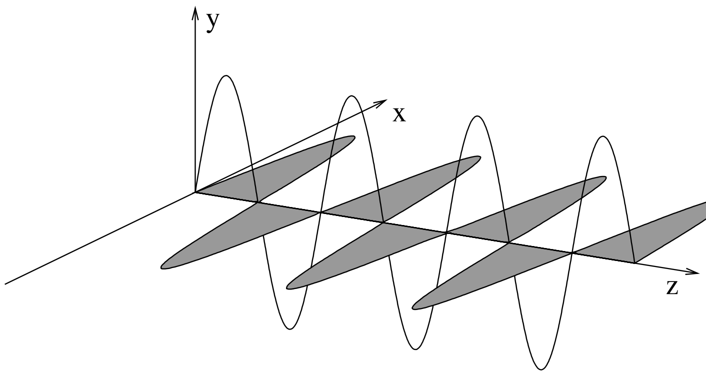


Figure 9.2: Electric and magnetic fields of a linearly polarized photon

Polarization[13] *is a property of photons with which one can demonstrate well what can and cannot be done in the quantum world. In addition, polarized photons are key transmission elements in quantum cryptography.*

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

Photons whose electronic fields oscillate in a plane at either $0°$ or $90°$ to some reference line are called usually **rectilinearly polarized** *and those whose electric field oscillates in a plane at $45°$ or $135°$ as* **diagonally polarized**. *Polarizers that produce only vertically or horizontally polarized photons are depicted in Figure 9.3a,b.*

Generation of orthogonally polarized photons. *For any two orthogonal polarizations (that differ by 90°) there are generators that produce photons of two given orthogonal polarizations. For example, a calcite crystal, properly oriented, can do the job. Figure 9.3c shows a calcite crystal that causes θ-polarized photons to be either horizontally polarized, with probability $\cos^2\theta$, or vertically polarized, with probability $\sin^2\theta$.*

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

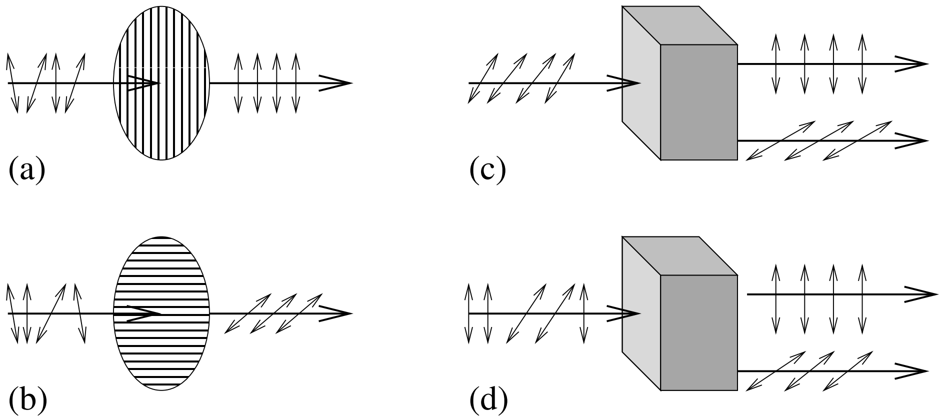


Figure 9.3: Photon polarizers and measuring devices-80%

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

Measuring of polarization. *A calcite crystal, which has the properties of birefringence, can be used to separate horizontally and vertically polarized photons (see Figure 9.3d). Another crystal can be used to distinguish two polarizations for diagonally polarized photons. Can we measure both of them at the same time?*

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

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

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

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

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

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

Since

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

these two observables do not commute.

9.1.3 Quantum theory versus physical reality

God is subtle, but malicious he is not.

A. Einstein

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

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

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

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

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

There are several reasons why current quantum theory is considered by some schools of thought in physics and in the philosophy of science not only as being far from complete but actually as a theory no one fully understands, and a theory that is not able to attach in a fully satisfactory way a definite physical meaning to its mathematical concepts.[14] For example

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

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

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

2. The existence of non-locality and non-local interactions[15] as pointed out by EPR Gedanken experiment discussed in Section 9.1.5.

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

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

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

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

9.1.4 Quantum measurements

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

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

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

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

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

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

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

Example 9.1.4 *(Peres, 1993) There is a way to produce photons with various polarizations, but there is no way to measure polarization of a particular photon.*[16] *The question "What is the polarization of that photon?" cannot be answered and is considered in quantum physics as having no meaning. The legitimate question that can be answered experimentally is whether or not a particular photon has a specified polarization. As a consequence, if Alice prepares a sequence of photons and sends them to Bob without disclosing their polarization, then there is no instrument whatsoever to sort them into bins for polarization* $0 - 30^\circ$, $30 - 60^\circ \dots$ *in the way that agrees with the polarization as produced by Alice. This fact is actually used in Section 6.2 to make a secure system for cryptographical key generation.*

3. It is not always possible to obtain by measurements full information about the unknown state of a quantum object. In addition, the unknown states of quantum objects are considered as having no definite value except when they are measured.[17]

4. There have been attempts to consider quantum measurement as the very basic concept of quantum physics (see Peres, 1993), and to use it to derive from it such concepts as quantum state. By Peres, a measurement consists of a *preparation*, a *test* and a *selection*, where preparations and tests are even more fundamental concepts. A quantum state is then defined as an equivalence class of preparations.[18]

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

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

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

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

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

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

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

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

- **Repeatable measurements.** These are measurements M such that if M is applied to the result x obtained by M, we get x again. Projections P (with the property $P = P^2$) are examples of repeatable measurements. As discussed in Peres (1993), not all meaningful measurements are repeatable, and repeatable measurements are more an exception than the rule. In spite of that the term "measurement" is generally used for repeatable measurements. The non-repeatable measurements are also of importance because they may provide more information than ideal repeatable ones.
- **Maximal or complete measurements.** These are measurements that produce the same number of outcomes as the dimension of the corresponding Hilbert space. Each maximal test is uniquely determined by an orthonormal basis.
- **YES-NO measurements**. They are specified by a subspace and its orthonormal complement. The projection into one of the subspaces is interpreted as YES and into the second as a NO answer. A special case of a **primitive measurement** is when one of the subspaces is spanned by just one vector.
- **POV measurements**. **P**ositive **o**perator **v**alued **m**easurements (see Bush et al. 1997, Peres, 1993, and Section 9.2.8), are different forms of measurement. For example, the number of outcomes they produce can be larger than the dimension of the underlying Hilbert space. Formally, a POV measurement is given by a set $\{O_i\}_{i=1}^k$ of positive Hermitian operators such that $\sum_{i=1}^k O_i = I$.

9.1.5 Quantum paradoxes

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

Design of paradoxes[19] has been an old methodology in physics to point out, in a irresistible way, some inconsistency of physical theories. Three such paradoxes, that have played an

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

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

Maxwell's demon

It seems that it was in connection with the attempts to explain Maxwell's[20] demon paradox, from 1867, that information processing considerations started to play a significant role in physics for the first time.

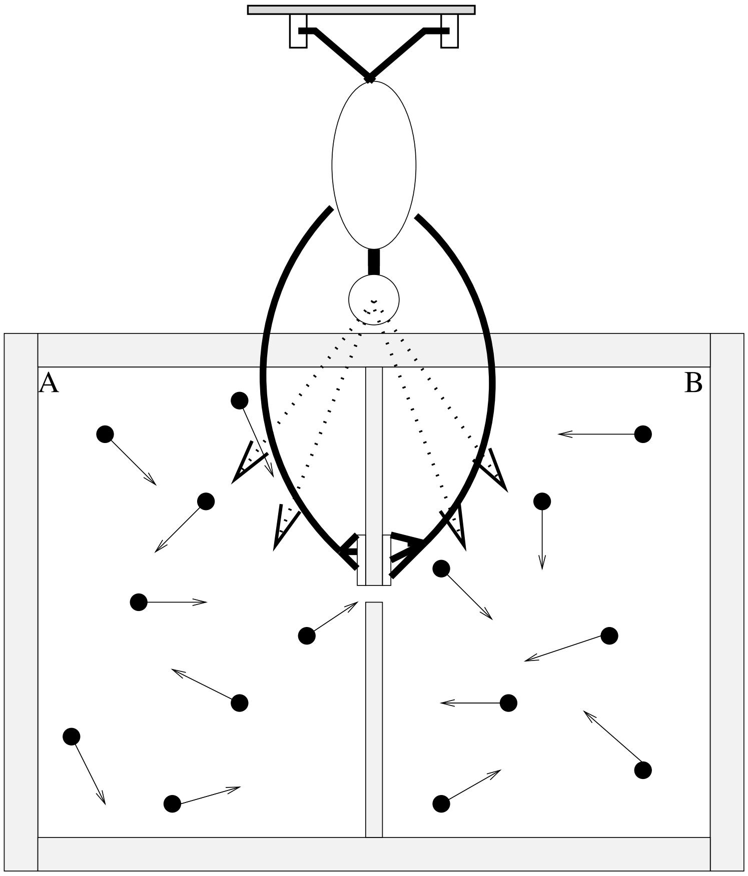


Figure 9.4: Maxwell's demon

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

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

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

Maxwell's demon paradox created enormous controversy among physicists.[21] The first explanation was offered by Szilard in 1929 and it was based on the belief that measurements that the demon has to perform, on the location and speed of molecules, increase its entropy and this compensate for the decrease of the entropy in the system.

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

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

Schrödinger cat

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

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

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

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

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

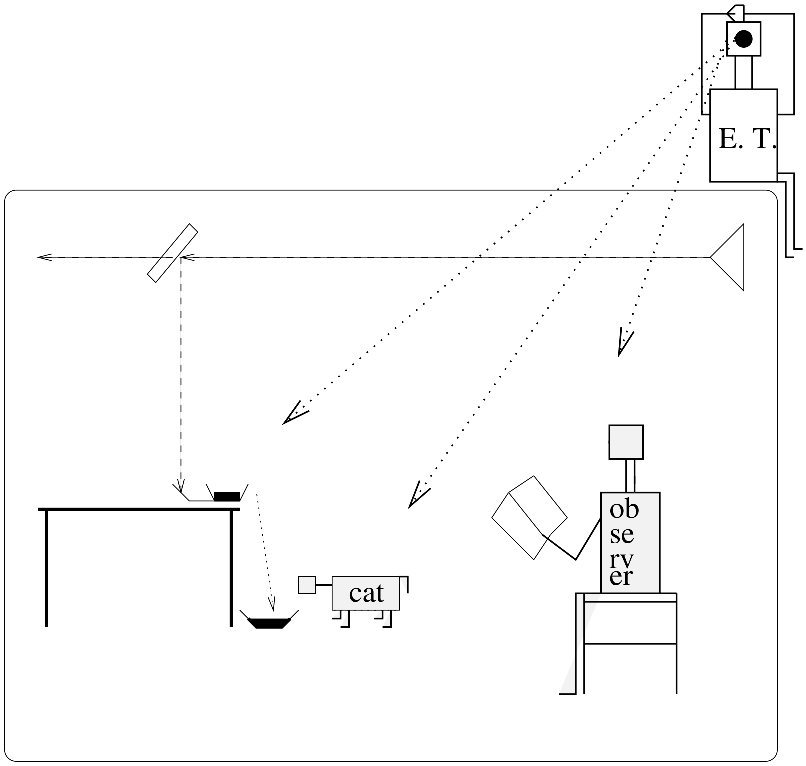


Figure 9.5: Schrödinger's cat

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

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

Schrödinger's Gedanken experiment led to the development of the so-called many-world interpretation of quantum mechanics (see Everett, 1957, 1977) and Section 9.1.4.[22]

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

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

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

EPR measurements

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

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

In order to discredit views of physics resulting from the Copenhagen interpretation, Albert Einstein,[23] Boris Podolsky and Nathan Rosen (EPR), in their famous paper "Can quantum mechanical description of reality be considered complete?", in 1935, developed an experimental set-up that helped to illustrate how strange consequences follow from quantum theory.

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

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

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

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

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

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

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

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

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

Einstein and his colleagues concluded from their Gedanken experiment that explanations of the real phenomena which the current quantum physics offers (namely, its Copenhagen interpretation) are not complete and suggested the way, a program, to fix it—how a proper fundamental theory of Nature should look like. The EPR program asked for *completeness* ("In a complete theory there is an element corresponding to each element of reality."), *locality* ("The real factual situation of system A is independent of what is done with system B, which is spatially separated from the former.") and *reality* ("If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity.").

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

9.1.6 The quantum paradox

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

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

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

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

A measurement is in the Copenhagen interpretation a step we have to take to get from the quantum world to the classical world, but the Copenhagen interpretation gives slightest explanation of how this step is performed. It only says what its outcomes are. There is nothing in quantum theory that explains or determines the exact mechanism of quantum measurement and the resulting state collapse. In particular, both measurements and state collapse are presented in such a way that they would not require involvement of forces of any kind.[24]

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

9.1.7 Interpretations of quantum theory

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

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

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

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

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

Quantum formalisms versus interpretations

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

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

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

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

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

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

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

Main interpretations

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

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

The main interpretations of linear quantum mechanics, in which evolution is described by the linear Schrödinger equation, are the following ones: Copenhagen interpretation[25], many-worlds interpretation (Everett, 1957, 1977), and the hidden variable interpretations, Bohm (1952), or its more modern version—an ontological interpretation of Bohm and Hiley (1993).

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

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

There are also various models of nonlinear quantum mechanics in which evolution is described by a nonlinear Schrödinger equation—obtained from linear Schrödinger equation by introducing nonlinear modifications. For example, models due to de Broglie (1956) and Weinberg (1989). No experiments supported yet the existence of nonlinear evolutions, but some physicists do not consider their existence as completely ruled out.[26]

The Copenhagen interpretation is based on the following two principles.

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

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

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

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

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

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

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

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

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

2. The Copenhagen interpretation of quantum mechanics relies on a notion called "measurement" which seemed to be beyond the abilities of quantum mechanics itself to explain.[27] Quantum mechanics seemed to be unable to explain by which physical processes a measurement act makes things to change from undetermined to fully determined ones.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

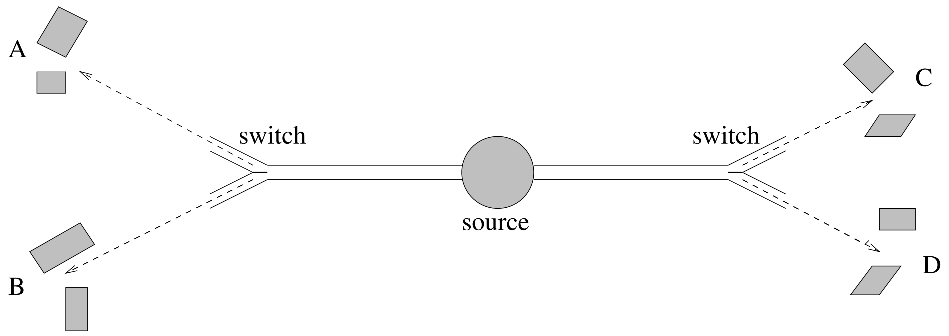


Figure 9.6: Aspect's experiment

its mathematical language is merely a useful approximation.

9.1.8 Incompleteness of quantum mechanics

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

Assume that a pair of electrons whose spins add up to zero is created by some source and sent off in two separate directions. In each electron's path we install two Stern–Gerlach magnets and a switching device that can send an electron randomly to one of the two magnets—see Figure 9.6. All four magnets are set at different angles.

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

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

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

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

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

Bell's theorem and one of Bell inequalities[28].)

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

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

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

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

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

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

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

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

9.2 Hilbert Space Framework for Quantum Computing

Truth is great and its effectivness endures.

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

Ptahhotpe, 24th century BC

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

9.2.1 Hilbert spaces

The basic underlying concept is that of a vector space.

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

- $a \cdot (\phi + \psi) = a \cdot \phi + a \cdot \psi$, $(a +_f b) \cdot \phi = a \cdot \phi + b \cdot \phi$ $\quad$ {*distributive laws*}
- $(a \cdot (b \cdot \phi)) = (a \times_f b) \cdot \phi$
- $1 \cdot \phi = \phi$.[29]

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

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

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

The inner product introduces on H the norm

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

and the metric[30]

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

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

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

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

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

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

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

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

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

An inner-product space with a carrier H is called **complete** if for any sequence $\{\phi_i\}_{i=1}^{\infty}$ with $\phi_i \in H$, and with the property $\lim_{i,j\rightarrow\infty} ||\phi_i - \phi_j|| = 0$ there is a unique element $\phi \in H$ such that $\lim_{i\rightarrow\infty} ||\phi - \phi_i|| = 0$. A complete inner-product space is called **Hilbert space**.[31]

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

Orthonormal bases of Hilbert spaces

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

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

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

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

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

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

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

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

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

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

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

Theorem 9.2.10 *For each closed subspace W of a Hilbert space H there exists a unique subspace $W^\perp = \{\phi \,|\, \langle\phi|\psi\rangle = 0$ for any $\psi \in W\}$ such that each $\phi \in H$ has a unique representation $\phi = \psi_1 + \psi_2$, $\psi_1 \in W$, $\psi_2 \in W^\perp$. In such a case we write $H = W \oplus W^\perp$.*[32]

Those orthonormal sets that are bases are of special importance.

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

It can be shown that all bases of a Hilbert space H have the same cardinality, called dimension of H. If the cardinality is finite we speak about the finite-dimensional Hilbert space.[33] Otherwise, we talk about the infinite-dimensional Hilbert space. For example, in the case of spin states of a particle the corresponding Hilbert space has finite dimension. Namely, in the case of a spin-$\frac{1}{2}n$ particle the corresponding Hilbert space is $(n+1)$-dimensional. Infinite-dimensional Hilbert space is in the case of states being positions of particles or configurations of space unbounded Turing machine.

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

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

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

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

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

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

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

Self-duality of Hilbert spaces

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

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

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

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

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

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

9.2.2 Linear operators

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

Definition 9.2.18 *By a (**linear**) **operator** of Hilbert space H we understand a linear mapping $A : H \rightarrow H$ (the domain of which forms a subspace of H).*[34] *A linear operator A is called **positive** or **semi-definite** (notation $A \geq 0$) if $\langle\psi|A\psi\rangle \in \mathbf{R}^{\geq 0}$ for every vector $\psi \in H$. If A_1, A_2 are linear operators, then we write $A_1 \geq A_2$ if and only if $A_1 - A_2 \geq 0$. Let A be an operator of a Hilbert space H such that the closure of the domain of A is H. The **adjoint operator** of A is an operator A^* the domain of which consists exactly of those $\psi \in H$ for which the linear mapping $\phi \rightarrow \langle\psi|A\phi\rangle$ is continuous on the domain of A, and for all $\psi \in$ domain(A^*), $\langle\psi|A\phi\rangle = \langle A^*\psi|\phi\rangle$. If $A = A^*$, then the operator A is called **self-adjoint** or **Hermitian**. By $\mathcal{L}(H)$ we denote the class of all self-adjoint operators of H.*

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

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

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

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

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

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

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

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

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

$$P_W\phi = \phi_W$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Another important case is when a linear operator is bounded.

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

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

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

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

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

All linear operators of finitely dimensional Hilbert spaces are bounded.

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

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

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

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

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

9.2.3 Mixed states and density matrices

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

If some eigenvalues of ρ are degenerate, then

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

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

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

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

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

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

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

Schmidt decomposition theorem

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

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

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

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

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

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

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

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

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

Superoperators

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

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

9.2.4 Probabilities and observables

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

9.2.5 Evolution of quantum states

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

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

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

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

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

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

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

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

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

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

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

9.2.6 Measurements

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

9.2.7 Tensor products and Hilbert spaces

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

9.2.8 Generalized measurements-POV measurements

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

Quantum information gain

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

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

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

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

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

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

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

The generalized measurements represent a way to increase information gain.

POV measurements

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

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

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

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

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

and therefore

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

where A_μ is defined by

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

and A_μ is an operator on H.

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

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

One difference between von Neumann's projection measurement and POVM is that in the later case the number of possible outcomes may be larger than the dimensionality of the underlying Hilbert space H. The probability of the outcome μ is given by $Tr(A_\mu \rho)$ instead of von Neumann's $Tr(P_\mu \rho)$.[35]

Information gain via POVM

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

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

It can be shown that that the average information gain is

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

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

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

9.3 Deterministic and Randomized Computing

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

9.3.1 Computing models

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

Deterministic models

The very basic model of the universal computer, invented in 1937 by A. M. Turing, is that of **one–tape Turing machine** (TM). A Turing machine $\mathcal{M}$, see Figure 9.7, consists of a **bi–infinite tape**, divided into an infinite number of cells in both directions, with one distinctive starting cell, or 0-cell. Cells of the tape can contain any symbol from a finite **tape alphabet** Γ, or a symbol $\sqcup$ representing the empty cell; a **read–write head** is positioned at any moment of the discrete time on a cell; a **finite control** unit is always in one of the states of a finite set Q of states and implements a transition function

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

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

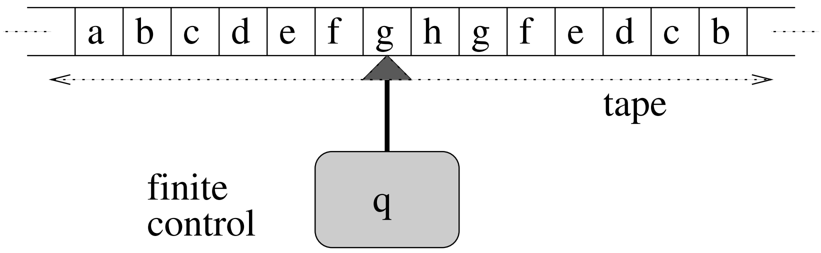


Figure 9.7: One-tape Turing machine

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

A natural and important generalization of one-tape Turing machines are **multitape Turing machines** (MTM)—see Figure 9.8. A k-tape Turing machine has k tapes, one of them is the input tape, and there is a head on each of the tapes. Formally, a k tape MTM is defined as $\mathcal{M} = \langle \Sigma, \Gamma, Q, q_0, \delta \rangle$, where

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

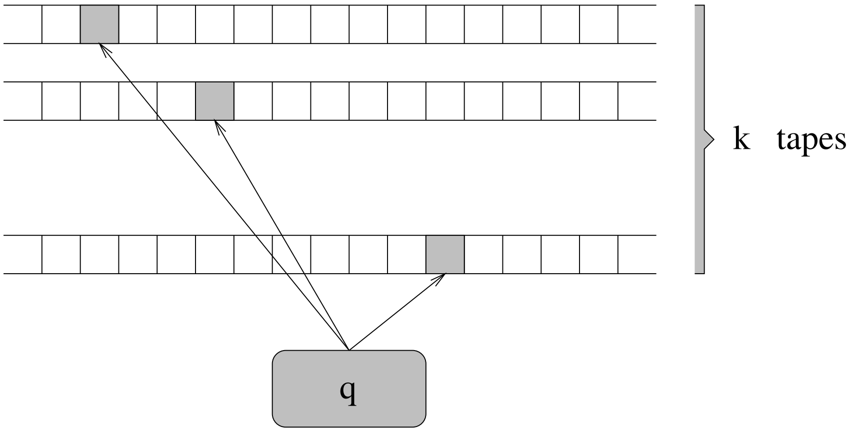


Figure 9.8: Multitape Turing machines

If, at the beginning of a computation step, $\mathcal{M}$ is in the state q, its head on the $i$the tape reads the symbol σ_i and $\delta(q, \sigma_1, \ldots, \sigma_k) = (q', \sigma'_1, \ldots, \sigma'_k, d_1, \ldots, d_k)$, then $\mathcal{M}$ moves, in one step, into the state q', the head on the ith tape replaces σ_i by σ'_i and moves in the direction d_i. One-tape TM can be seen as a special case of MTM.

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

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

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

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

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

For example,

$$\text{a transition } \delta(q_i, a_j) = (q_k, a_l, d) \text{ can be encoded as } 0^i 1 0^j 1 0^k 1 0^l 1 0^{d'},$$

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

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

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

Another very basic model of Turing machines is the oracle Turing machine—see Figure 9.9. This model is to capture the idea of computing with a help (of an oracle or a subroutine). Let us illustrate this concept again on one-tape oracle Turing machines.

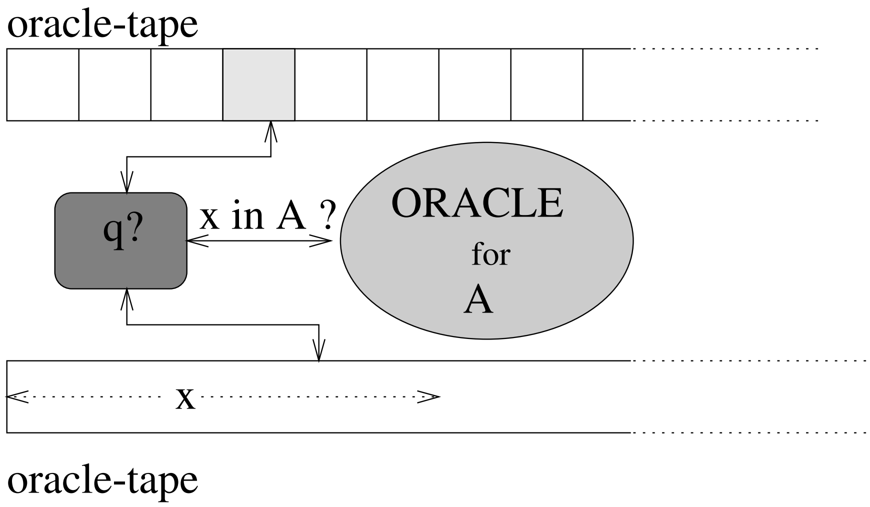


Figure 9.9: Oracle Turing machine

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

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

Nondeterministic model

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

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

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

This means that a step of an NTM is not in general uniquely determined and several alternative steps may be offered to choose (nondeterministically). As a consequence, a configuration C of an NTM may have several potential next configurations, and $\mathcal{M}$ can go, nondeterministically, from C to one of them. We can therefore see the overall computational process of an NTM not as a sequence (of subsequent) configurations, but as a tree of (branching) configurations—see Figure 9.10.

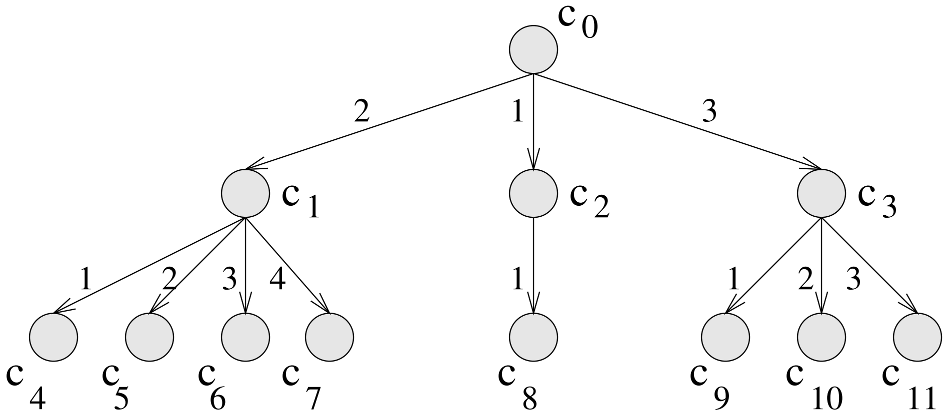


Figure 9.10: Tree of configurations of NTM

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

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

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

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

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

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

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

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

9.3.2 Randomized computations

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

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

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

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

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

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

A computation of a PTM $\mathcal{M}$ can be again seen as a configuration tree in which to each configuration transfer a probability is assigned—see Figure 9.11. From that one can derive the probability that a computation reaches a particular node—this is the product of probabilities assigned to all edges (transfers) on the path from the root to the particular node.

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

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

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

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

Algorithm 9.3.10 (RQUICKSORT)

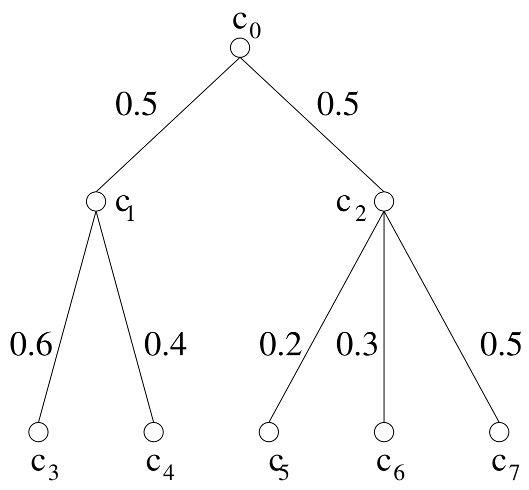


Figure 9.11: Configuration tree of a PTM-80%

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

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

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

9.3.3 Complexity classes

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

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

Complexity classes of deterministic computing

The very basic classes of classical deterministic computing are:

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

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

The basic relations between these complexity classes have the form

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

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

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

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

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

It holds:

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

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

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

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

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

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

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

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

Relativized complexity classes

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

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

Two main questions to consider are:

1. Are there oracles A and B such that $\mathcal{L}_1^A \neq \mathcal{L}_2^A$ and $\mathcal{L}_1^B = \mathcal{L}_2^B$?

2. Is it true that $\mathcal{L}_1^R \neq \mathcal{L}_2^R$ for a randomly chosen oracle R?

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

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

Randomized complexity classes

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

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

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

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

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

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

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

The first model (see Figure 9.12a), is that of a one-tape Turing machine enhanced by a special "randomized tape" and with the head which moves one square right after each reading from the random tape. Before any computation the randomized tape is assumed to be filled up with an infinite random binary string. At some steps of computation the symbol under the current position of the head on random tape is read—this corresponds to a "random coin-tossing" at such a step—and computation continues depending on the outcome of the reading.

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

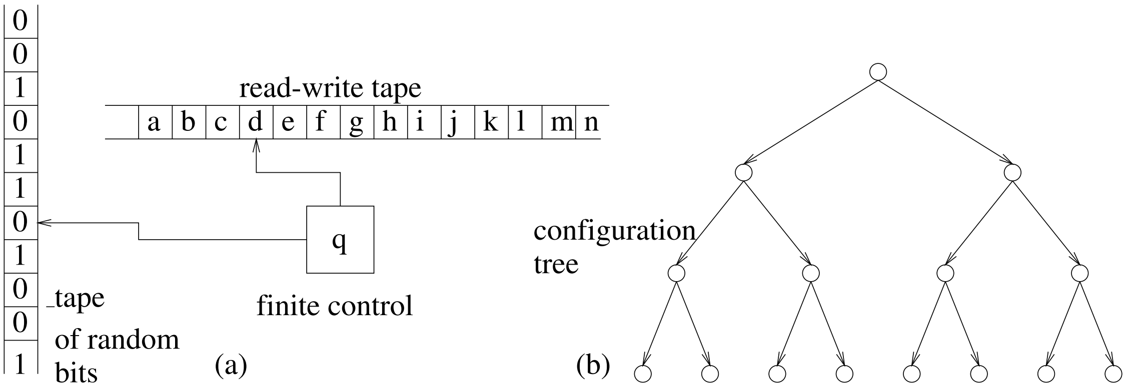


Figure 9.12: A Turing machine with a tape of random bits and a configuration tree of a special NTM

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

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

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

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

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

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

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

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

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

9.3.4 Computational theses

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

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

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

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

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

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

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

9.4 Exercises

1. Show that the following matrix is unitary:

$$A = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{i}{2} & -\frac{1}{2} & \frac{i}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{i}{2} \\ \frac{1}{2} & -\frac{i}{2} & -\frac{1}{2} & \frac{i}{2} \end{pmatrix}.$$

2. Determine under which conditions the following matrices are unitary:

$$\text{(a)} \quad \begin{pmatrix} e^{i\theta} & 0 & 0 & ie^{i\theta}\sin\theta_1 \\ 0 & \cos\theta_2 & i\sin\theta_2 & 0 \\ 0 & i\sin\theta_2 & \cos\theta_2 & 0 \\ ie^{i\theta}\sin\theta_1 & 0 & 0 & e^{i\theta}\cos\theta_1 \end{pmatrix}, \quad (b) \begin{pmatrix} a & c & c & b \\ c & a & b & c \\ c & b & a & c \\ b & c & c & a \end{pmatrix}.$$

3. Compute eigenvalues and eigenvectors all Pauli matrices.

4. Express, using rotation, phase shift and the Pauli matrix σ_x, matrices: (a) σ_y; (b) σ_z.

5. Show that Pauli matrices have the following properties: (a) $\sigma_i\sigma_k + \sigma_k\sigma_i = 2\delta_{ik}$, for $i, k \in \{x, y, z\}$; (b) $\sigma_y\sigma_z - \sigma_z\sigma_y = 2i\sigma_x$; (c) $\sigma_z\sigma_x - \sigma_x\sigma_z = 2i\sigma_y$.

6. Show how to make a NAND gate out of two Toffoli gates.

7. How many of n input/output gates are reversible for: (a) $n = 1$ and $n = 2$; (b) $n = 3$; (c) for an arbitrary n.

8. Compute eigenvalues and eigenvectors of the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

9. Determine eigenvalues and eigenvectors of the following matrices: (a) Hadamard matrix; (b) XOR matrix.

10. Determine when two non-orthogonal states $|\phi_1\rangle$ and $|\phi_2\rangle$ can be written, by a suitable choice of the basis as $|\phi_1\rangle = \cos\alpha|0\rangle + \sin\alpha|1\rangle$ and $|\phi_2\rangle = \sin\alpha|0\rangle + \cos\alpha|1\rangle$.

11. Show that for any two states ϕ, ψ of a Hilbert space H:

$$||\phi+\psi||^2 - ||\phi-\psi||^2 + i||\phi - i\psi||^2 - i||\phi + i\psi||^2 = 4\langle\phi|\psi\rangle.$$

12. Express the state $|\phi\rangle \otimes |\phi\rangle$, where $|\phi\rangle = \alpha|0\rangle + \beta|1\rangle$ in the Bell basis.

13. Design a matrix that maps Bell states in H_4 into the standard basis.

14. A family of bases of a Hilbert space H_n is called conjugate if each pair of the bases from the family is conjugate. (a) Find a family of three conjugate bases in H_2; (b) (Wiesner, 1983) Show that for any n there is an m such that the Hilbert space H_m has a family of n conjugate basis.

15. Find an example of an infinite matrix A such that A^2 is not well defined.

16. Show that if $|\phi\rangle$ and $|\psi\rangle$ are orthogonal vectors of a Hilbert space H, then there exists a Hermitian operator A in H such that $|\phi\rangle$ and $|\psi\rangle$ are eigenvectors of A.

17. Let A, B and C be Hermitian matrices such that $\mathbf{0} \leq B \leq C$. Show that $CA = \mathbf{0}$ implies $BA = \mathbf{0}$.

18. Show that the trace of a Hermitian or unitary matrix is equal to the sum of its eigenvalues, and the determinant of such a matrix equals the product of its eigenvalues.

19. Show that if a matrix U is unitary and H is Hermitian, then $U^T HU$ is Hermitian and if H is unitary, then $U^T HU$ is unitary too.

20. Let A be a (bounded) linear operator of a Hilbert space H. Show that the following conditions are equivalent: (1) A is positive; (2) $A = B^*B$ for some bounded linear operator B; (3) $A = A^*$ and all its eigenvalues are positive.

21. (a) Show that if A is a Hermitian matrix, then there is a unique Hermitian matrix B, denoted $\sqrt{A}$, such that $B^2 = A$. (b) Show that there are Hermitian matrices A, B such that $\sqrt{AB} \neq \sqrt{A}\sqrt{B}$.

22. Show that if A is a Hermitian matrix, then the matrix e^{iA} is unitary.

23. Let $\{\phi_i\}_{i=1}^n$ be n orthonormal states of a Hilbert space H_n and ϕ_0 an arbitrary state of H_n. Show that $\sum_{i=1}^n ||\phi_i - \phi_0||^2 \geq 2n - 2\sqrt{n}$.

24. Let S be a set of elements of an inner-product space H. Is the concept of the small subspace containing S well defined and unique?

25. (Jozsa, 1994) Denote $[a, b, c]$ the matrix from the Exercise 2b.; determine (a) $Tr[a, b, c]$; (b) Tr([a.b,c][d,e,f]); (c) show that any two matrices $[a, b, c]$ and $[d, e, f]$ commute.

26. Show that the sequence of functions

$$E = \{e_k\}_{k=-\infty}^{\infty}, \quad e_k(x) = \frac{e^{ikx}}{\sqrt{2\pi}},$$

forms an orthonormal set on $L_2([0, 2\pi))$. (One of the fundamental results in the theory of Fourier series is that E is a basis of $L_2([0, 2\pi))$.)

27. Design a quantum circuit consisting of elementary gates and realizing the mapping:

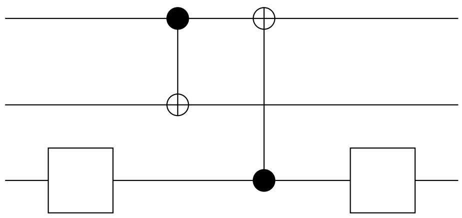
$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\phi} \end{pmatrix}.$$

28. Describe the mapping realized by the gates described by the matrices

$$\frac{1}{2}\begin{pmatrix} 1 + e^{i\pi\alpha} & 1 - e^{i\pi\alpha} \\ 1 - e^{i\pi\alpha} & 1 + e^{i\pi\alpha} \end{pmatrix},$$

depending on different values of α.

29. Let $x, y \in \{0, 1\}^n$. Determine $H_n(\frac{1}{\sqrt{2}}(|x\rangle + |y\rangle))$.



(a)

Figure 9.13: Examples of networks

30. Determine what kind of unitary transformation is performed by the circuit in Figure 9.13 provided the square gate represents: (a) Hadamard transformation; (b) transformation H'; (c) Pauli matrices.

31. Show how to design a network to transform the state $\frac{1}{\sqrt{2}}(|0000\rangle + |1111\rangle)$ into the four-qubit state $\frac{1}{\sqrt{8}}\sum_{i \text{ even}} |i\rangle$.

32. Design a 3-qubit input/output network to map states $\{|i\rangle \,|\, 0 \le i \le 7\}$ into the 3-particle Bell states: $\frac{1}{\sqrt{2}}(|000\rangle \pm |111\rangle)$, $\frac{1}{\sqrt{2}}(|001\rangle \pm |110\rangle)$, $\frac{1}{\sqrt{2}}(|010\rangle \pm |101\rangle)$, $\frac{1}{\sqrt{2}}(|100\rangle \pm |011\rangle)$.

33. (Barenco et al. 1995) Let
$$U = \begin{pmatrix} u_{00} & u_{01} \\ u_{10} & u_{11} \end{pmatrix}$$
be a unitary matrix. For $m \in \{0, 1, 2 \ldots\}$ let $\bigwedge_m(U)$ be an $(m+1)$-bit operator defined by
$$\bigwedge_m(U)(|x_1, \ldots, x_m, y\rangle) = \begin{cases} u_{y0}|x_1, \ldots, x_m, 0\rangle + u_{y1}|x_1, \ldots, x_m, 1\rangle & \text{if } \bigwedge_{k=1}^m x_k = 1 \\ |x_1, \ldots, x_m, y\rangle & \textit{otherwise} \end{cases}$$
for all $x_1, \ldots, x_m, y \in \{0, 1\}$. (a) Describe the matrix $\bigwedge_m(U)$; (b) $\bigwedge_m(U)$ is called the $(m+1)$-bit Toffoli gate if $U = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$—explain why.

34. Show that if $W \in U(2)$, then the $\bigwedge_i(W)$ gate can be simulated by the network from Figure 9.14, where A, B and C are in $SU(2)$, if and only if $W \in SU(2)$.

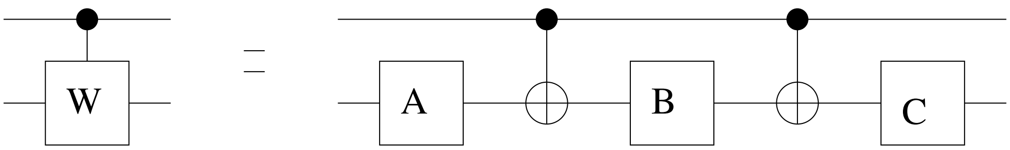


Figure 9.14: Decomposition of networks

35. Show that for any special unitary matrix $W \in SU(2)$ there exist matrices $A, B, C \in SU(2)$ such that $ABC = 1$ and $A\sigma_x\sigma_x C = W$.

36. "Majority voting". Design an m input/output reversible circuit, $m \geq 3$, as small as possible, which in the case of inputs $x, y, z, 0^{(m-3)}$ produces the output $u, u, u, x_1, \dots, x_{m-3}$ where $u = majority\{x, y, x\}$, for some $x_1, \dots, x_m$.

37. **Coin weighting problem** (Terhal and Smolin, 1997). There are n different coins of two types: good and bad. A set of n coins is represented by a binary string y of length n with $y_i = 1$ if and only if the ith coin is bad. Assume that by weighting we can determine which coins are good. Each weighting is represented by a query string x, where x_i specifies whether the ith coin is to be included in the set to be weighted. The result of weighting is $w(x, y) = x \cdot y$. The task is to retrieve y.

38. Given a function $f : \{0,1\}^n \rightarrow \{0,1\}^m$, $m \leq n$, see Cleve et al. (1998), of the form$f(x) = (A \cdot x) \oplus b$, where A is an $m \times n$ Boolean matrix, b is a binary m-tuple and $\oplus$ is applied bitwise. (a) Show that $f(x)$ has constant parity if $(1, \dots, 1) \cdot A = (0, \dots, 0)$ and balanced parity otherwise. (b) Design a quantum circuit to determine all entries of A by evaluating f only m times.

39. Design a 2QFA which in time $\mathcal{O}(n)$ accepts all words of length n from the alphabet $\{0, 1\}$ with the same number of 0s and 1s and in time $\mathcal{O}(kn)$ rejects all other words with probability at least $1 - 4^{-k}\binom{2k}{k}$.

40. Each state $|\phi\rangle = \sum_{i=0}^{2^n-1} \alpha_i |i\rangle$ induces a probability distribution $P_{|\phi\rangle}(i) = |\alpha_i|^2$ on the basis states. The *total variation* distance between two distributions $P_{|\phi\rangle}$ and $P_{|\psi\rangle}$ is defined by $||P_{|\phi\rangle} - P_{|\psi\rangle}||_1 = \sum_{i=0}^{2^n-1} |P_{|\phi\rangle}(i) - P_{|\psi\rangle}(i)|$. Show that $||P_{|\phi\rangle} - P_{|\psi\rangle}||_1 \leq 2|||\phi\rangle - |\psi\rangle||$.

41. (Jozsa, 1994) Let M, ρ_1, ρ_2 be Hermitian matrices of degree 2 with positive eigenvalues λ_1, λ_2 being eigenvalues of M. Show: (a) $Tr\sqrt{M} = \sqrt{\lambda_1} + \sqrt{\lambda_2}$; (b) $(Tr\sqrt{M})^2 = TrM + 2\sqrt{detM}$; (c) If $M = \sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}}$, then $TrM = Tr\rho_1\rho_2$ and $detM = det\rho_1 det\rho_2$. (d) $F(\rho_1, \rho_2) = Tr\rho_1\rho_2 + 2\sqrt{det\rho_1 det\rho_2}$.

9.5 Historical and Bibliographical References

Beginnings of quantum information processing can be traced back into the first attempts to investigate physical resources information processing needs.

In 1929 Leo Szilard, in his attempts to solve Maxwell's demon paradox, developed the concept of the bit of information and associated the entropy $kTln2$ with the acquisition of one bit. On this base, through von Neumann (see Burks, 1966), a belief started to be a part of computer folklore that at least $kTln2$ of energy is needed to process one bit of information at room temperature. An important new insight came form Landauer (1961), namely that not acquisition but erasure of information is of importance from physical point of view. Landauer showed that any erasure or overwriting of one bit requires at least $\mathrm{kTln}\,2$ of energy (about 3×19^{-21} Joule at room temperature, and that if a computation is reversible then, in principle, it requires no dissipation and no power expenditure. Landauer also pointed out that it is possible to eliminate erasures and overwriting of bits by keeping all intermediate results on a "scratchpad". For early papers on entropy and information, as well as on Maxwell's demon, see Leff and Fix (1990).

Landauer (1961) seems to be also the first to keep pointing out very explicitly and consistently physical nature of information. However, only after the first works on quantum computing by Feynman (1982) and Deutsch (1985), the fundamental connection between the laws of physics and computation started to be systematically and intensively investigated.

Lecerf (1963) was the first to show that any computation can be made reversible (and therefore in principle requires no energy computation).[36] However, Lecerf's result went unnoticed and it was Bennett (1973) who showed the same[37] and made the result well-known and initiated a broader research in the area. (Bennett showed that it is possible to design a reversible universal Turing machine capable to reset its tape to the initial state before ending its computation.) Next important steps were discoveries of universal gates for reversible classical computing by Petri (1967), Toffoli (1980, 1981) and Fredkin and Toffoli (1982). Very stimulating was also the so-called "billiard model" of reversible computing due to Fredkin and Toffoli (1982). For excellent accounts on the history of reversible computing see Bennett (1988). For a detailed presentation of various issues concerning reversibility see Feynman (1996).

The earliest specific results on quantum computing were obtained by Benioff (1982, 1982a) who showed that Turing machines can be simulated by unitary quantum processes.

Feynman (1982, 1985) seems to be the first to bring up explicitly and convincingly, the idea that quantum mechanics phenomena can improve efficiency of computations. He gave arguments/evidence why it seems to require exponential time to simulate quantum phenomena on classical computers He has also implicitly suggested to use a computer based on quantum mechanical principles to overcome the difficulties.

The concept of a fully quantum Turing machine was introduced by Deutsch (1985), who also showed the existence of a universal QTM. This result was not yet fully satisfactory because his universal QTM was not able to simulate any QTM in polynomial time. More formal and detailed investigations of QTM started with Bernstein and Vazirani (1993) and Yao (1993) (see also Hirvensalo, 1996, and especially Bernstein and Vazirani, 1997), where a detailed construction of universal quantum Turing machines is provided.

[36] Later (see Lecerf, 1963a), he showed the existence of the universal reversible Turing machine. His motivation was to demonstrate that the Post correspondence problem remains undecidable even if it is restricted to injective morphism. He did not pay attention that his universal reversible TM requires more than polynomial time for simulation. His ideas were extended and clarified by Ruohonen (1985).

[37] In addition, Bennett (1989) showed that reversibility does not cost too much in terms of time or space.

Investigation of quantum networks started by Deutsch (1989). He considered the so-called "quantum computational networks" and established some of their properties. These networks can be viewed as quantum analog of classical logical circuits with the feedback—sequential logical circuits. A quantum analog to the standard acyclic (Boolean) circuit model was introduced by Yao (1993) and since then this is the main model of quantum circuits. Yao has also shown that any function computable in polynomial time by a QTM has a polynomial size circuit.

The existence of 3-qubit universal gates was first shown by Deutsch (1995). Sleator and Weinfurter (1995), DiVincenzo (1995) and Barenco (1995) showed the existence of a two-qubit universal gate. Barenco (1995), Lloyd (1995) and Deutsch, Barenco and Ekert (1995) showed that almost any two-qubit gate is universal and the proofs are nonconstructive. Networks for addition and modular addition are due to Vedral, Barenco and Ekert (1995). Quantum superoperators circuits were investigated by Aharonov et al. (1998).

Deutsch's randomized algorithm was the first example of a quantum algorithm which could solve an algorithmic problem faster than any classical algorithm. Deterministic solution to Deutsch's problem was discovered independently by Tapp, as pointed out by Cleve et al. (1998). The first example of a problem with the expected polynomial time quantum algorithm, but with no polynomial tome randomized algorithm was due to Bernstein and Vazirani (1993).

The history story of attempts to solve the hidden subgroup problems parallel that of the design of quantum algorithms superpolynomially faster than their classical counterparts. The history starts with Simon (1994) who has solved the hidden subgroup problem for the group $\mathbf{Z}_2^n$ under the promise that the hidden subgroup is of order 2. Next important contributions were Shor's algorithms. Kitaev (1996) made a significant step to find a more general framework in terms of the Abelian group stabilizer problem. The observation that all these problems are special cases of the hidden subgroup problem has been formulated by several authors, for example by Høyer (1997). A complex treatment of the subject for the Abelian case is by Mosca and Ekert (1998). The first exact polynomial time algorithm for the group $\mathbf{Z}_2^n$ is due to Brassard and Høyer (1997).

Shor's polynomial time algorithms for factorization and discrete logarithm have appeared for the first time in Shor (1994) and an improved and extended version of his results is in Shor (1997). Of interest is also presentation of Shor's results by Ekert and Jozsa (1996). Fast quantum implementation of the QFT_q, the key component of these algorithms and of many others, was discovered for different q by Cleve (1994), Coopersmith (1994) and Deutsch (see Ekert and Jozsa, 1996). Griffiths and Niu (1996) have shown that DFT can be performed using only one-bit gates and measurements of single qubits. Kitaev (1995) developed another method to solve factorization and discrete logarithm in polynomial time. Shor's and Kitaev, methods have turned out to be equivalent.

The basic paper on $\mathcal{O}(\sqrt{n})$ searching in unordered database was due to Grover (1996), Lower bound $\Omega(\sqrt{n})$ for quantum searching in an unsorted database was shown first by Bennett, Bernstein, Brassard, Vazirani (1997), in 1994, sooner than Grover developed his algorithm. Bøyer et al. (1996) showed that Grover's algorithm is within 10% optimal. Zalka (1997) proved that Grover's algorithm is optimal in a well defined sense. A relatively simple proof that each search in unsorted database requires $0.707\sqrt{n}$ queries was given by Grover (1998a). The basic generalization of Grover's algorithm to the case that more searched for elements exist is due to Bøyer et al. (1996). A detailed analysis of Grover's algorithm and motivations behind is due to Jozsa (1999). A variety of generalizations and modifications of Grover's algorithm are due to Dürr and Høyer (1996), , Biron et al. (1998),, Brassard et al. (1998) and Grover (1996, 1997b, 1998b).

The first attempts to introduce a quantum algorithm speed-up or design methodology were due to Brassard et al. (1998a) and Grover (1998). Basic results on quantum counting are due to Brassard et al. (1998a) and Mosca (1998).

From other attempts to develop faster-than-classical quantum algorithms: oracle learning (van Dam, 1998), quantum strategies (Meyer, 1998), computational learning (Ventura and Martiner, 1998).

Results have also been reported to get some specific answers to the original Feynman's question about the possibilities of a more efficient simulation of quantum mechanics on quantum computers than on classical ones. Abrams and Lloyd (1997) showed that this is possible for quite a general class of discrete quantum systems, and Baghosian and Taylor IV (1998) showed that for Schrödinger equation for interacting many-body systems.

Particular results on limitations of quantum computing are due to Ozhigov (1998, 1998a) and Farhi et al. (1998). A general methodology to show lower bounds, using classical results how to make lower bounds on degrees of polynomials representing or approximating a Boolean function, are due to Beals et al. (1998).

The concept of two-way quantum finite automaton was introduced by Watrous (1997c). One-way quantum automata were introduced by Kondacs and Watrous (1997) and they explored their power comparing to the classical finite automata. Equivalence of reversible finite automata and one-way quantum automata that accepts with very high probability is due to Ambainis and Freivalds (1998). Basic results on space efficiency of 1QFA are due to Ambainis and Feivalds (1998) and Ambainis et al. (1998), For a variety of models of quantum grammars and automata see Moore and Crutchfield (1997).

There have been a variety attempts to introduce quantum cellular automata. Some of the first attempts, see Grössing and Zeilinger (1981) and the subsequent series of papers, have little common with the current view of QCA. Concerning other models see Biafore (1994)—discussing problems of synchronization—Margolus (1994)—a study of space-periodic quantum cellular automata and Lloyd (1993, 1994)—a discussion of possibilities to realize a special type of quantum cellular automata. Formally, linear quantum cellular automata were introduced by Watrous (1995). Efficiency of procedures to decide well-formedness and unitarity of simple LQCA has been investigated by Dürr, LěThanh and Santha (1996) as well as by Dürr and Santha (1996).

The first results on computational quantum time complexity were due to Deutsch (1985), Barthiaume and Brassard (1982, 1982a, 1982b, 1994), Bernstein and Vazirani (1993) and Simon (1994). For a more detailed treatment of the subjects see Bernsteain and Vazirani (1997), Bennett et al. (1997) and Adleman et al. (1997). In the last paper impacts of amplitudes started to be systematically investigated. The basic paper on quantum space complexity classes is due to Watrous (1997).

Yao (1993) made the first steps to develop quantum communication complexity . The first idea to use quantum entanglement to decrease communication complexity of problems was presented by Cleve and Buhrman (1997)—by one qubit. Asymptotical improvements, presented in Section 2.2, are due to van Dam et al. (1997). Basic paper on the relation between quantum computational and communication complexity is Buhrman et al. (1998).

The origin of quantum cryptography can be tracked back to Wiesner around 1970, who suggested how to use differently and randomly polarized photons "for transmitting two messages either but not both of which may be recovered", and a way how one could have "counterfeit-proof money". These interesting ideas were not published and therefore did not get much attention. Additional reason for that was the fact that the idea of "counterfeit-proof money" required facilities to store quantum states for a longer time, what was seen as unfeasible at that time. However, in earlies 1980s Bennett started to realize the potential of Wiesner's ideas for cryptography and the basic observations have been published by Bennett at al. (1982). Wiesner himself published his original paper in 1983. Of large importance for the development of quantum cryptography was the first conference on quantum cryptography organized by A. K. Ekert in 1993. The first cryptographical protocol, known as BB84, was published by Bennett and Brassard (1984). Bennet and his collaborators made also the first steps to verify QKD ideas by experiments and in 1989, using polarized photons, demonstrated transmissions of photons for the distance of 32cm. In 1991 Ekert published his ideas of QKD based on entanglement and in 1992 Bennett published a "minimal" QKD protocol, known as B92 and suggested that it could be implemented using single-photon interference with photons propagating over optical fiber for a long distance. Since then several significant experiments have been performed with signal propagating over the distance of several (tens) of kilometres—see Marand and Townsend (1995), Muller et al. (1995), and Hughes et al. (1996). Muller et al. (1995) described sending quantum signals from Nyon to Geneva using as quantum channel an optical fibre deployed beneath lake Geneva. Impressive progress has also be made in open-space quantum key generation. From successful experiment for the distance 32cm in 1989 to almost one kilometer (Buttler et al). For more details on the history of quantum cryptography see Brassard and Crépeau (1996) and Brassard (1988, 1994).

The history of attempts to develop QCP and to prove their (unconditional) security is short but reach on interesting, deep and surprising developments.

The first quantum coin-flipping protocol was developed by Bennett and Brassard (1984), and they also showed its unsecurity. Crépeau and Killian (1988) developed a protocol for oblivious transfer which was secure provided neither party could store photons for long time and only projection measurements were allowed. It was later shown (see Mayers, 1996) for the last step) that a secure quantum bit commitment is sufficient to get secure quantum oblivious transfer. A protocol for quantum bit commitment, so called BCJL protocol, was developed by Brassard, Crépau, Jozsa and Langlois (1993) and it was claimed that it is unconditionally secure. (Actually BCJL protocol was a modification of the protocol developed by Brassard and Crépau (1991).) However, Lo and Chau (1997) and Mayers (1998) have shown a way to cheat, using an "EPR-attack" and by delaying the measurement. (The cheating strategy requires quantum computer!)

Bennett et al. (1993) is the basic reference for teleportation. Teleportation circuit is due to Brassard (1996). The idea of superdense coding is due to Bennett and Wiesner (1992).

The first analysis of errors in quantum computation in the initial state (data), called software errors, and in the Hamiltonian, called hardware errors, was performed by Zurek (1984). Measurement (or readout) errors were first analyzed by Peres (1985).

The discovery of quantum error-correcting codes and of fault-tolerant techniques for quantum computation have been further surprising and important steps toward the feasibility of quantum computing.

Quantum error correction was assumed to be impossible till 1995. A break-through came with pioneering papers of Shor (1995) and Steane (1996). Shor exhibited a way to encode one qubit by an entangled state of nine qubits in such a way that a single bit or sign error could be corrected. Steane was able to encode one qubit by seven qubits. An optimal way to encode one qubit, by an entangled state of five qubits, was first shown by Laflamme et al. (1996) and by Bennettet al. (1996).

The first attempts to develop a methodology for design of quantum error correcting codes were due to Calderbank and Shor (1996), Steane (1996a) and Calderbank et al (1996)—to make a quantum error-correcting code out of one or two classical error correcting codes. A more general approach to stabilizer or additive codes was due to Gottesman (1996) and Calderbank and Shor (1997). First example of a nonadditive code has been due to Rains et al. (1997).

A necessary and sufficient conditions for a subspace of a Hilbert space to form a quantum error-correcting code was derived by Bennett et al. (1996a) and Knill and Laflamme (1997). Quantum versions of the Hamming bound and of the Gilbert-Varshamov bound were derived by Ekert and Macchiavello (1996). An infinite class of codes satisfying the quantum Hamming bound was developed by Gottesman (1996). Of large importance was also the establishment of the relation of the quantum error-correcting codes to the preservation of quantum entanglement in a noisy environment by Bennett et al. (1996a 1996b). For a survey on quantum error-correction codes see Gottesman (1997), and Steane (1998).

The method of stabilization of quantum computation and memory by symmetrization, presented in Section 7.3, is due to Barenco et al.(1997). Its origin goes back to Deutsch, in 1993, and Berthiaume et al. (1994).

Pioneering papers in the area of quantum fault-tolerant computations were due to Shor (1996) and Zurek and Laflamme (1996). Shor has shown that there exist fault-tolerant implementations for universal set of gates. DiVincenzo and Shor (1996) showed a systematic way how to design fault-tolerant syndrom computation circuits for additive codes. A general approach to fault-tolerant computation based on stabilizer codes was developed by Gottesman (1997) and Steane (1998b)

Concatenated codes, as a way to make reliably long transmissions of quantum information, are due to Knill and Laflamme (1996). The first paper pointing out a possibility to have fault-tolerant gates was due to Kitaev (1997). For a survey on fault-tolerant computing see Preskill (1998). The idea of quantum repeaters is due to Briegel et al.

The concept of quantum entropy is due to von Neumann (1932) and an intensive survey of quantum entropy results is due to Wehrl (1978). For classical results on communication theory see Shannon and Weaver (1969).. For “classical results” on information theory for quantum systems see Levitin (1987). Schumacher’s (1995) noiseless quantum coding theorem has been a basis of a new approach to quantum information theory based on inherently quantum concepts. In addition, the basic definition of the capacity of quantum channels is also due to Schumacher (1995). Another proof of the basic Schumacher’s theorem was done by Jozsa and Schumacher (1994). An explicit algorithm for performing Schumacher’s noiseless compression of quantum bits was given by Cleve and DiVincenzo (1996). For a quantum analogue of Shannon’s bound and code for the noisy channel transmission see Lloyd (1997). Basic results on dense coding are due to Ambainis et al. (1998). Transmission fidelities were defined and explored by Jozsa and Schumacher (1994) and Jozsa (1994). For an attempt to develop quantum analogies to basic concepts of classical information theory see Cerf and Adami (1996).

For basic results concerning quantum purification see Bennett et al. (1996a, 1996b). For the relation between entanglement purification and quantum error-correcting codes see Bennett et al. (1996a). Basic results concerning quantum concentration are due to Bennett et al. (1996) and Lo with Popescu (1997). Measures of entanglement are due to Bennett et al. (1996a), Vedral and Plenio (1997), Vedral et al. and DiVincenzo et al. (1998). Bennett (1998,1998a) provides both the survey of current quantum information theory concepts and results and presents also directions for further research. For a survey of results on quantum information primitives and reducibilities see Bennett (1998).

For popular presentation of basic problems of quantum mechanics see Penrose (1990, 1994) and Lindley (1996). Introductions to quantum mechanics can be found in von Neumann (1932), Dirac (1947) and Feynman, Leighton and Sandes (1964).. One of the main references to quantum mechanics is Peres (1993)—a graduate level textbook and reference book. For POV measurements see Bush, Grabowski and Lahti (1997). For an introduction to Hilbert spaces and their relation to quantum mechanics see Cohen (1989) and Packel (1974). For the history and philosophy of quantum mechanicsss Jammer (1974) and Bub (1997).

There are quite a few of excellent introductory papers, surveys and tutorials on quantum computing and communication: Ekert (1995), DiVincenzo (1995a, 1996), Svozil (1995), Brassard (1995, 1996). Berthiaume (1997), Steane (1997), Aharonov (1998), Rieffel and Polak (1998), Bennett (1998). Good surveys are also in PhD theses of Barenco (1996), Hirvensalo (1996), Gottesman (1997). See also papers in a special issue of Physics Today (March 1998).

The first two books on quantum computing are due to Williams and Clearwater (1997), with various quantum simulations on CD-ROM and Berman et al. (1998), oriented mainly to physicists and on a systematic and detailed treatment of various implementation issues.

An excellent source of papers on quantum computing is the following WWW site:

http://xxx.lanl.gov/archive/quant-ph

of the Quantum Physics Archive at Los Alamos National Laboratory with the following average number of papers per month and year: 1995 (26), 1996 (38), 1997 (54), 1998 (85).

Bibliography

[1] D. S. Abrams and Seth Lloyd. Nonlinear quantum mechanics implies polynomial-time solution for NP-complete and #P complete problems. Technical report, quant-ph/9801041, 1998.

[2] Leonard M. Adleman, Jonathan DeMarrais, and Ming-Deh A. Huang Quantum computability. *SIAM Journal of Computing*, 26(5):1524–1540, 1997.

[3] Dorit Aharonov. Quantum computation. Technical report, quant-ph/9812037, 1998.

[4] Dorit Aharonov, Alexei Kitaev, and Noam Nisan. Quantum circuits with mixed states. In *Proceedings of* 30*th ACM STOC*, pages 20–30, 1998. quant-ph/9806029.

[5] Y. Aharonov and J. Anandan Meaning of the density matrices. Technical report, quant-ph/9803018, 1998.

[6] Eric Allender and Mitsunori Ogihara. Relationship among PL, $\#L$ and the determinant. *RAIRO*, 30:1–21, 1996.

[7] L. Allen and Joseph H. Eberly *Optical resonance and two-level atoms*. Dover Publications, 1975.

[8] Masami Amano and Kazuo Iwama. Undecidability on quantum finite automata. In *Proceedings of 31st ACM STOC*, page to be published, 1999.

[9] Andris Ambainis and Rūsiņš Freivalds. 1-way quantum finite automata: strengths, weaknesses and generalizations. In *Proceedings of* 39*th IEEE FOCS*, pages 332–341, 1998. quant-ph/9802062.

[10] Andris Ambainis, Ashwin Nayak, Amnon Ta-Shma , and Umesh Vazirani. Dense quantum coding and a lower bound for 1-way quantum finite automata. Technical report, quant-ph/9804043, 1998.

[11] Andris Ambainis, Leonard Schulman, Amnon Ta-Shma, Umesh Vazirani, and Avi Wigderson. The quantum communication complexity of sampling. In *Proceedings of* 39*th IEEE FOCS*, pages 342–351, 1998.

[12] Serafino Amoroso and Yale N. Patt Decision procedures for surjectivity and injectivity of parallel maps for tesselation structures. *Journal of Computer and System Sciences*, 6:448–464, 1972.

[13] Mohammad Ardehali, Gilles Brassard, H. F. Chau and, and Hoi-Kwong Lo. Efficient quantum key distribution. Technical report, quant-ph/9803007, 1998.

[14] Alain Aspect, Jean Dalibard, and Gérard Roger. Experimental tests of Bell's inequalities using time-varying analyzers. *Physical Review Letters*, 49:1804–1807, 1982.

[15] Leslie E. Ballentine *Quantum mechanics*. Prentice Hall, 1970.

[16] Adriano Barenco. Dense coding based on quantum entanglement. *Journal of Modern Optics*, 42:1253–1259, 1995.

[17] Adriano Barenco. A universal two-bit gate for quantum computation. *Proceedings of Royal Society London A*, 449:679–683, 1995.

[18] Adriano Barenco. *Quantum computation*. PhD thesis, University of Oxford, 1996.

[19] Adriano Barenco, Charles H. Bennett, Richard Cleve, David P. DiVincenzo, Norman Margolus, Peter W. Shor, Tycho Sleator, John A. Smolin, and Harald Weinfurter. Elementary gates of quantum computation. *Physical Review A*, 52(5):3457–3467, 1995.

[20] Adriano Barenco, André Berthiaume, David Deutsch, Artur K. Ekert, Richard Jozsa, and Chiarra Macchiavello. Stabilization of quantum computations by symmetrization. *SIAM Journal of Computing*, 26(5):1541–1557, 1997. quant-ph/9604028.

[21] Adriano Barenco, David Deutsch, Artur K. Ekert, and Richard Jozsa. Conditional quantum dynamics and logic gates. *Physical Review Letters*, 74:4083–4086, 1995a.

[22] Adriano Barenco and Artur K. Ekert Quantum computation. *Acta Physica Slovaca*, 45(3):205–216, 1995. In Proceedings of 3rd Central-European Workshop on Quantum Optics.

[23] Adriano Barenco, Artur K. Ekert, Kalle-Anntti Suominen, and Päivi Törmä. Approximate quantum Fourier transform and decoherence. *Physical Review A*, 54(1):139–146, 1996b.

[24] Stephen M. Barnett and Simon J. D. Phoenix Bell's inequality and rejected-data protocols for quantum cryptography. *Journal of Modern Optics*, 40(8):1443–1448, 1993.

[25] Howard Barnum, C. M. Caves, Christopher. A. Fuchs, Richard Jozsa, and Benjamin W. Schumacher Noncomuting mixed states cannot be broadcast. *Physical Review Letters*, 76(15):2828–2831, 1996. quant-ph 9511010.

[26] Howard Barnum, Emanuel H. Knill, and Michael A. Nielsen On quantum fidelities and channel capacities. Technical report, quant-ph/9809010, 1998.

[27] Howard Barnum, Michael A. Nielsen, and Benjamin W. Schumacher Information transmission through noisy quantum channels. *Physical Review A*, 57:4153–4175, 1997.

[28] Howard Barnum, John Smolin, and Barbara Terhal Results on quantum channel capacity. Technical report, quant-ph/9711032, 1997a.

[29] Robert Beals. Quantum computation of Fourier transforms over symmetric groups. In *Proceedings of 29th ACM STOC*, pages 48–53, 1997.

[30] Robert Beals, Harry Buhrman, Richard Cleve, and Michele Mosca. Quantum lower bounds for polynomials. In *Proceedings of 39th IEEE FOCS*, pages 352–361, 1998.

[31] David Beckman, Amalovoyal N. Chari, Srikrishna Devabhaktuni, and John Preskill. Efficient networks for quantum factoring. *Physical Review A*, 54:1034–1063, 1996.

[32] John S. Bell On the Einstein-Podolsky-Rosen paradox. *Physics 1*, pages 195–200, 1964. Reprinted in Quantum Theory and Measurement, (eds): J. A. Wheeler and W. H. Zurek, 403-408.

[33] Paul A. Benioff The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines. *Journal of Statistical Physics*, 22:563–591, 1980.

[34] Paul A. Benioff Quantum mechanical Hamiltonian models of discrete processes that erase their own histories: application to Turing machines. *International Journal of Theoretical Physics*, 21(3/4):177–202, 1982.

[35] Paul A. Benioff Quantum mechanical Hamiltonian models of Turing machines that dissipate no energy. *Physical Review Letters*, 48:1581–1585, 1982a.

[36] Paul A. Benioff Models of quantum Turing machines. *Fortschritte der Physics*, 46(4-5):423–441, 1998.

[37] Charles H. Bennett Logical reversibility of computation. *IBM Journal of Research and Development*, 17:525–532, 1973.

[38] Charles H. Bennett Notes on the history of reversible computation. *IBM Journal of Research and Development*, 32(1):16–23, 1988.

[39] Charles H. Bennett Logical depth and physical complexity. In *The Universal Turing Machines—a Half Century Survey*, pages 227–257. Kammer & Univerzagt, Hamburg, 1988d.

[40] Charles H. Bennett Time/space trade-offs for reversible computation. *SIAM Journal of Computing*, 18:766–776, 1989.

[41] Charles H. Bennett Quantum cryptography using any two nonorthogonal states. *Physical Review Letters*, 68(21):3121–3124, 1992.

[42] Charles H. Bennett Future directions for quantum information. In *Introduction to quantum computation and information*, page to be published. World Scientific, 1998.

[43] Charles H. Bennett Quantum information. *Physica Scripta*, T76:210–217, 1998a.

[44] Charles H. Bennett Quantum information. Technical report, Tutorial, MFCS'98, 1998b.

[45] Charles H. Bennett, Ethan Bernstein, Gilles Brassard, and Umesh Vazirani. Strength and weaknesses of quantum computation. *SIAM Journal on Computing*, 26(5):1510–1523, 1997.

[46] Charles H. Bennett, Herbert J. Bernstein, Sandu Popescu, and Benjamin W. Schumacher Concentrating partial entanglement by local operations. *Physical Review A*, 53:2046–2053, 1996.

[47] Charles H. Bennett, François Bessette, Gilles Brassard, Louis Salvail, and John A. Smolin Experimental quantum cryptography. *Journal of Cryptology*, 5(1):3–28, 1992.

[48] Charles H. Bennett and Gilles Brassard. Quantum cryptography: public key distribution and coin tossing. In *Proceedings of IEEE Conference on Computers, Systems and Signal processing, Bangalore (India)*, pages 175–179, 1984.

[49] Charles H. Bennett and Gilles Brassard. The dawn of a new era for quantum cryptography. The experimental prototype is working! *SIGACT News*, 20(4):78–82, 1989.

[50] Charles H. Bennett, Gilles Brassard, S. Briedbart, and Stephen J. Wiesner. Quantum cryptology, or unforgeable subway tokens. In *Proceedings of Crypto'82*, pages 267–275, 1982.

[51] Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70:1895–1899, 1993.

[52] Charles H. Bennett, Gilles Brassard, Claude Crépeau, and Marie-Hélène Skubiszewska. Practical quantum oblivious transfer. In *Advances in Cryptography, Proceedings of Crypto'91, LNCS 576, Springer Verlag*, pages 351–366, 1991.

[53] Charles H. Bennett, Gilles Brassard, and N. David Merman. Quantum cryptography without Bell's theorem. *Physical Review Letters*, 68(5):557–559, 1992a.

[54] Charles H. Bennett, Gilles Brassard, Sandu Popescu, Benjamin W. Schumacher, John A. Smolin, and William K. Wootters Purification of noisy entaglement and faithful teleportation via noisy channels. *Physical Review Letters*, 76(5):722–725, 1996b.

[55] Charles H. Bennett, Gilles Brassard, and J.-M. Robert Privacy amplification by public discussion. *SIAM Journal of Computing*, 17(2):210–229, 1988.

[56] Charles H. Bennett, David P. DiVincenzo, Christopher A. Fuchs, Tal Moric, Eric M. Rains, Peter W. Shor, and John A. Smolin Quantifying non-locality without entanglement. Technical report, quant-ph/9804053, 1998.

[57] Charles H. Bennett, David P. DiVincenzo, and John A. Smolin Capacities of quantum erasure channels. Technical report, quant-ph/9701015, 1997a.

[58] Charles H. Bennett, David P. DiVincenzo, John A. Smolin, and William K. Wootters Mixed state entaglement and quantum error correction. *Physical Review A*, 54:3824–3851, 1996a. quant-ph/0604024.

[59] Charles H. Bennett, Christopher A. Fuchs, and John A. Smolin *Quantum Communication, Computing and Measurement*, chapter Entanglement-enhanced classical communication on a noisy quantum channel, pages 79–88. Plenum, New York, 1997b.

[60] Charles H. Bennett and Stephen J. Wiesner Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Physical Review Letters*, 69(20):2881–2884, 1992.

[61] Gennady P. Berman, Gary D. Doolen, Ronnie Mainieri, and Vladimir I. Tsifrinovich *Introduction to quantum computing*. World Scientific, 1998.

[62] Ethan Bernstein. *Quantum complexity theory.* PhD thesis, University of California, Berkeley, 1997.

[63] Ethan Bernstein and Umesh Vazirani. Quantum complexity theory. In *Proceedings of 25th ACM STOC*, pages 11–20, 1993.

[64] Ethan Bernstein and Umesh Vazirani. Quantum complexity theory. *SIAM Journal of Computing*, 26(5):1411–1473, 1997.

[65] André Berthiaume. *Complexity theory retrospectives II*, chapter Quantum computation, pages 23–51. Springer-Verlag, 1997.

[66] André Berthiaume and Gilles Brassard. Oracle quantum computing. In *Proceedings of the Workshop on Physics of Computation*, pages 195–199, 1992.

[67] André Berthiaume and Gilles Brassard. Oracle quantum computing. In *Proceedings of 7th IEEE Conference on Structure in Complexity Theory*, pages 132–137, 1992a.

[68] André Berthiaume and Gilles Brassard. The quantum challenge to structural complexity. In *Proceedings of Structure in Complexity Conference*, pages 132–137, 1992b.

[69] André Berthiaume and Gilles Brassard. Oracle quantum computing. *Journal of Modern Optics*, 41(12):2521–2535, 1994.

[70] André Berthiaume, David Deutsch, and Richard Jozsa. The stabilization of quantum computation. In *Proceedings of the Third Workshop on Physics and Computation*, pages 60–62. IEEE Computer Society Press, 1994.

[71] Michael Biafore. Can computers have simple Hamiltonians? In *Proceedings of Physics and Computation, PhysComp'94*, pages 63–69, 1994.

[72] Ivo Bialynicki-Birula and J. Mycielski Uncertainty relations for information entropy in wave mechanics. *Commen. Math. Phys*, 44:129–132, 1975.

[73] Eli Biham, Ofer Biham, David Biron, Markus Grassl, and Daniel A. Lidar Exact solution of Grover's quantum search algorithm for arbitrary initial amplitude distribution. Technical report, quant-ph/9807027, 1998.

[74] Eli Biham, Bruno Huttner, and Tal Mor. Quantum cryptographic network based on quantum memories. *Physical Review A*, 54(4):2651–2658, 1996.

[75] David Biron, Ofer Biham, Markus Grassl, and Daniel A. Lidar Generalized Grover search algorithm for arbitrary initial amplitude distribution. Technical report, quant-ph/9801066, 1998.

[76] Manuel Blum. Coin flipping by telephon. A protocol for solving impossible problems. In *Proceedings of the 24th IEEE FOCS*, pages 133–137, 1982.

[77] Bruce M. Boghosian and Washington Taylor IV Simulating quantum mechanics on a quantum computers. *Physica D*, 120:30–42, 1997. quant-ph/9701019.

[78] David Bohm. A suggested interpretation of the quantum theory in terms of "hidden" variables I and II. *Physics Review*, 85:166–193, 1952. also in "Quantum theory and measurements (ed. J. A. Wheeler and W. H. Zurek), Princeton University Press, 1983.

[79] David Bohm and Basil J. Hiley *The undivided universe—an onthological interpretation of quantum theory*. Routledge, 1994.

[80] D. Boschi, S. Branca, F. DeMartini, L. Hardy, and Sandu Popescu. Experimental realization of teleporting an unknown pure quantum state via dual classical Einstein–Podolsky–Rosen channels. *Physical Review Letters*, 80:1121–1125, 1998.

[81] Sugato Bose, Martin B. Plenio, and Vlatko Vedral. Mixed state dense coding and its relation to entanglement measures. Technical report, quant-ph/9810025, 1998.

[82] Dik Bouwmeester, Jean-Wei Pan, Matthew Daniel, Harald Weinfurter, and Anton Zeilinger. Observation of three-photon Greenberer-Horner-Zeilinger entanglement. Technical report, quant-ph/9810035, 1998.

[83] Dik Bouwmeester, Jean-Wei Pan, K. Mattle, M. Eibl, Harald Weinfurter, and Anton Zeilinger. Experimental quantum teleportation. *Nature*, 390(6660):575–579, December 1997.

[84] Michel Boyer, Gilles Brassard, Peter Høyer, and Alain Tapp. Tight bounds on quantum searching. In *Fourth Workshop on Physics and Computation, Ed. T. Toffoli, M. Biaford, J. Lean*, pages 36–43. New England Complex System Institute, 1996. See also Fortschritte der Physics, 46, N4-5, 493-505, 1998, and quantum-ph/9605034.

[85] Gilles Brassard. *Modern cryptography. A tutorial*. Springer-Verlag, LNCS 325, 1988.

[86] Gilles Brassard. Cryptology column—quantum computing: the end of classical cryptography? *SIGACT News*, 25(4):15–21, 1994.

[87] Gilles Brassard. *Computer Science Today*, chapter A quantum jump in Computer Science, pages 1–14. Springer-Verlag, LNCS 1000, 1995.

[88] Gilles Brassard. New trends in quantum computing. In *Proceedings of STAC'96*, pages 3–10, 1996.

[89] Gilles Brassard. Teleportation as quantum computation. *Proceedings of PhysComp-96, Physica D*, 120:43–47, 1998.

[90] Gilles Brassard and Claude Crépeau. Quantum bit commitment and coin tossing protocols. In *Adanvecs in cryptology—CRYPTO'90*, pages 49–61. LNCS 537, Springer-Verlag, 1991.

[91] Gilles Brassard and Claude Crépeau. Cryptology column—25 years of quantum cryptography? *SIGACT News*, 27(3):13–24, 1996.

[92] Gilles Brassard, Claude Crépeau, Richard Jozsa, and Denis Langlois. A quantum bit commitment scheme provably unbreakable by both parties. In *Proceedings of 34th IEEE FOCS*, pages 362–371, 1993.

[93] Gilles Brassard, Claude Crépeau, Dominic C. Mayers, and Louis Salvail. A brief review on the impossibility of quantum bit commitment. Technical report, quant-ph/9712023, 1997.

[94] Gilles Brassard, Claude Crépeau, Dominic C. Mayers, and Louis Salvail. Defeating classical bit commitments with a quantum computer. Technical report, quant-ph/9806031, 1998b.

[95] Gilles Brassard and Peter Høyer. On the power of exact quantum polynomial time. Technical report, quant-ph/9612017, 1996.

[96] Gilles Brassard and Peter Høyer. An exact quantum polynomial-time algorithm for Simon's problem. In *Proceedings of Israeli Symposium on Theory of Computing and systems*, pages 12–23, 1997.

[97] Gilles Brassard, Peter Høyer, and Alain Tapp. Quantum cryptanalysis of hash and claw-free functions. In *Proceedings of LATIN'98*, pages 163–169. LNCS 1380, Springer-Verlag, 1998. quant-ph/9705002.

[98] Gilles Brassard, Peter Høyer, and Alain Tapp. Quantum counting. Technical report, quant-ph/9805082, 1998a.

[99] H.-J. Briegel, W. Dür, Juan I. Cirac, and Peter Zoller. Quantum repeaters for communication. Technical report, quant-ph/9803056, 1998.

[100] Dagmar Bruß. Optimal eavesdropping in quantum cryptography with six states. Technical report, quant-ph-98050019, 1998.

[101] Jeffrey Bub. *Interpreting the quantum world*. Cambridge University Press, 1997.

[102] Harry Buhrman, Richard Cleve, and Avi Wigderson. Quantum vs. classical communication and computation. In *Proceedings of* 30*th ACM STOC*, pages 63–68, 1998. quant-ph/9802040.

[103] Harry Buhrman and Ronald de Wolf. Lower bounds for quantum search and derandomization. Technical report, quant-ph/9811046, 1998.

[104] Guido Burkard, Daniel Loss, and David P. DiVincenzo Coupled quantum dots as quantum gates. Technical report, quant-ph/9808026, 1998.

[105] Arthur W. Burks *J. von Neumann: Theory of self-reproducing automata*. University Illinois Press, 1966.

[106] Paul Bush, Marian Grabowski, and Pekka J. Lahti *Operational quantum physics*. Springer, 1997.

[107] W. T. Buttler, Richard J. Hughes, Paul G. Kwiat, S. K. Lamoreaux, G. G. Luther, G. L. Morgan, J. E. Nordholt, C. G. Peterson, and C. M. Simmons Practical free-space quantum key distribution over 1 km. Technical report, quant-ph/9805071, 1998.

[108] Vladimír Bužek, Mark Hillery, and Reinhard F. Werner Optimal manipulations with qubits: universal NOT gate. Technical report, quant-ph/9901053, 1999.

[109] Vladimír Bužek, Samuel L. Braunstein, Mark Hillery, and Dagmar Bruß. Quantum copying: A network. *Physical Review A*, 56(5):3446–3452, 1997.

[110] Vladimír Bužek and Mark Hillery. Quantum copying: beyond the no-clonning theorem. *Physical Review A*, 54:1844–1852, 1996.

[111] A. Rob Calderbank, Eric M. Rains, Peter W. Shor, and Neil J. A. Sloane Quantum error correction via codes over GF(4). Technical report, quant-ph/9608006, 1996.

[112] A. Rob Calderbank, Eric M. Rains, Peter W. Shor, and Neil J. A. Sloane Quantum error correction and orthogonal geometry. *Physical Review Letters*, 78(3):405–408, 1997.

[113] A. Rob Calderbank and Peter W. Shor Good quantum error-correcting codes exist. *Physical Review A*, 54:1098–1105, 1996. quant-ph/9512032.

[114] Nicolas J. Cerf and Chris Adami. Quantum information theory of entanglement and measurement. In J. Leão T. Toffoli, M. Biafore, editor, *Proceedings of Physics and Computation, PhysComp'96*, pages 65–71. New England Complex Systems Institute, 1996. see also Physica D, 120, 1998, 62-81 and quant-ph/9605039.

[115] Vladimír Černý. Quantum computers and intractable (NP-complete) computing problems. *Physical Review A*, 48(1):116–119, 1993.

[116] H. F. Chau and Hoi-Kwong Lo. Making an empty promise with a quantum computer. *Fortschritte für Physics*, 46(4-5):507–519, 1998.

[117] Isaac L. Chuang, Neil A. Gershenfeld, and Mark Kubinec. Experimental implementation of fast quantum searching. *Physical Review Letters*, 80(15):3408–3411, 1998.

[118] Isaac L. Chuang, Lieven M. K. Vandersypen, Xinlan Zhou, Debbie W. Leung, and Seth Llyod. Experimental realization of a quantum algorithm. *Nature*, 393:143–146, 1998a. quant-ph/9801037.

[119] Isaac L. Chuang and Yoshihisa Yamamoto. A simple quantum computer. *Physical Review A*, 52(5):3489–3496, 1995.

[120] Juan I. Cirac and Peter Zoller. Quantum computations with cold trapped ions. *Physical Review Letters*, 74:4091–4094, 1995.

[121] John F. Clauser, Richard A. Holt, Michael A. Horne, and Abner Shimony. A proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23:880–884, 1969.

[122] Richard Cleve. A note on computing Fourier transforms by quantum programs. Technical report, http://www.cpsc.ucalgary.ca/˜cleve, 1994.

[123] Richard Cleve and Harry Buhrman. Substituting quantum entanglement for communication. *Physical Review A*, 56(2):1201–1204, 1997.

[124] Richard Cleve and David P. DiVincenzo Schumacher's quantum data compression as a quantum computation. *Physical Review A*, 54(4):2636–2650, 1996. quant-ph/9603009.

[125] Richard Cleve, Artur K. Ekert, Chiarra Macchiavello, and Michele Mosca. Quantum algorithms revisited. *Proceedings of Royal Society London, A*, 454:339–354, 1998.

[126] Richard Cleve, Daniel Gottesman, and Hoi-Kwong Lo. How to share a quantum secret. Technical report, quant-ph/9901025, 1999.

[127] Richard Cleve, Wim van Dam, Michael A. Nielsen, and Alain Tapp. Quantum entanglement and the communication complexity of the inner product function. Technical report, quant-ph/9708019, 1997.

[128] David W. Cohen *An introduction to Hilbert space and quantum logic.* Springer-Verlag, 1989.

[129] Don Coppersmith. An approximate Fourier transform useful in quantum factoring. Technical report, IBM, 1994. IBM Research report RC 19642.

[130] David G. Cory, Amr Fahmy, and Timothy F. Havel Nuclear magnetic resonance spectrocopy: an experimentally accessible paradigm for quantum computing. In *Proceedings of Physics and Computation, PhysComp96*, pages 87–91, 1996.

[131] David G. Cory, W. Mass, M. Price, Emanuel H. Knill, Raymond Laflamme, Wojciech H. Zurek, Timothy F. Havel, and S. S. Somaroo Experimental quantum error correction. *Physical Review Letters*, 81:2152–2155, 1998.

[132] Claude Crépeau. Equivalence between two flavours of oblivious transfers. In *Proceedings of CRYPTO'87, LNCS 293, Springer-Verlag*, pages 350–354, 1987.

[133] Claude Crépeau. Quantum oblivious transfer. *Journal of Modern Optics*, 41(12):2445–2454, 1994.

[134] Claude Crépeau and Joe Kilian. Achieving oblivious transfer using weakend security assumptions. In *Proceedings of 29th IEEE FOCS*, pages 42–52, 1988.

[135] Marek Czachor. Notes on nonlinear quantum algorithms. Technical report, quant-ph/9802051, 1998.

[136] Edward B. Davies *Quantum theory of open systems*. Academic Press, 1976.

[137] Louis de Broglie. *Une tentative d'intérpretation causale et nonlinéaire de la mecanique ondulatoire*. Gouthier-Villars, 1956.

[138] David Deutsch. Uncertainty in quantum measurement. *Physical Review Letters*, 50:631–633, 1983.

[139] David Deutsch. Quantum theory, the Church-Turing principle and the universal quantum computer. *Proceedings of the Royal Society London A*, 400:97–117, 1985.

[140] David Deutsch. Quantum computational networks. *Proceedings of Royal Society of London A*, 425:73–90, 1989.

[141] David Deutsch. *The fabric of reality*. Allen Lane, The Penguin Press, 1997.

[142] David Deutsch, Adriano Barenco, and Artur K. Ekert Universality in quantum computation. *Proceedings of Royal Society London A*, 449:669–677, 1995. quant-ph/9505018.

[143] David Deutsch, Artur K. Ekert, Richard Jozsa, Chiarra Macchiavello, Sandu Popescu, and Anna Sanpera. Quantum privacy amplification and the security of quantum cryptography over noisy channels. *Physical Review Letters*, 77(13):2818–2821, 1996.

[144] David Deutsch and Richard Jozsa. Rapid solution of problems by quantum computation. *Proceedings of the Royal Society London A*, 439:553–558, 1992.

[145] D. Dieks. Communication by EPR devices. *Physical Letters A*, 92:271–272, 1982.

[146] Paul A. M. Dirac *The principles of quantum mechanics*. Oxford University Press, 1947.

[147] David P. DiVincenzo Two-bit gates are universal for quantum computation. *Physical Review A*, 51:1015–1022, 1995.

[148] David P. DiVincenzo Quantum computation. *Science*, 270:255–261, 1995a.

[149] David P. DiVincenzo Quantum gates and circuits. *Proceedings of the Royal Society of London, Series A*, 454:261–276, 1998b. quant-ph/9705009.

[150] David P. DiVincenzo, Christopher A. Fuchs, Hideo Mabuchi, John A. Smolin, Ashish Thapliyal, and Armin Uhlmann. Entanglement of assistance. Technical report, quant-ph/9803033, 1998.

[151] David P. DiVincenzo and Daniel Loss. Quantum information is physical. *Superlattices and Microstructures*, 23:419–432, 1997. cond-mat/9710259.

[152] David P. DiVincenzo, Tal Mor, Peter W. Shor, John A. Smolin, and Barbara M. Terhal Unextendible product bases and bound entanglement. Technical report, quant-ph/9808030, 1998a.

[153] David P. DiVincenzo and Asher Peres. Quantum code words contradict local realism. *Physical Review A*, 55(6):4089–4092, 1997.

[154] David P. DiVincenzo and Peter W. Shor Fault-tolerant error correction with efficient quantum codes. *Physical Review Letters*, 77:3260–3263, 1996.

[155] David P. DiVincenzo and Barbara M. Terhal Decoherence: the obstacle to quantum computing. *Physics World*, pages 53–57, March 1998.

[156] Peter Domokos, Jean-Michel Raimond, Michel Brune, and Serge Haroche. Simple cavity-QED two-bit universal quantum logic gate: the principle and expected performance. *Physical Review A*, 52(5):3554–3559, 1995.

[157] Lu-Ming Duan and Guang-Can Guo. Reducing decoherence in quantum computer memory with all quantum bits coupling to the same environment. Technical report, quant-ph/9612003, 1996.

[158] Jean-Christophe Dubacq. How to simulate Turing machines by invertible one-dimensional cellular automata. *International Journal of Foundations of Computer Science*, 6:395–402, 1995.

[159] Christoph Dürr. *Automates cellulaires quantiques finis*. PhD thesis, LRI, Université Paris-Süd, 1997.

[160] Christoph Dürr and Peter Høyer. A quantum algorithm for finding the minimum. Technical report, Université Paris-Süd-Odense, quant-ph/9607014, 1996.

[161] Christoph Dürr, Huong LêThanh, and Miklos Santha. A decision procedure for well-formed linear quantum cellular automata. *Random Structures and Algorithms*, 11(4):381–394, 1996. Preliminary version in Proceedings of 13th STACS'96, 281-292.

[162] Christoph Dürr and Miklos Santha. A decision procedure for unitary linear quantum cellular automata. In *Proceedings of 37th IEEE FOCS*, pages 38–45, 1996. To appear in SIAM Journal of Computing, quant-ph/9604007.

[163] Cynthia Dwork and Larry Stockmeyer. On the power of 2-way probabilistic finite automata. In *Proceedings of 30th IEEE FOCS*, pages 480–485, 1989.

[164] Albert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum mechanical description of physics reality be considered complete? *Physical Review*, 47:777–780, 1935. Also in *quantum theory and measurement*, ed. J. A. Wheeler, Wojciech H. Zurek, Princeton University Press, 1983.

[165] Artur K. Ekert Quantum cryptography based on Bell's theorem. *Physical Review Letters*, 67(6):661–663, 1991.

[166] Artur K. Ekert Quantum computation. In *Atomic Physics 14 - 14th International Conference on Atomic Physics, Boulder, 1994, AIP Proceedings 323 (AIP Press, New York)*, pages 450–4??, 1995.

[167] Artur K. Ekert From quantum code-making to quantum code-breaking. Technical report, quant-ph/9703035, 1997.

[168] Artur K. Ekert and Richard Jozsa. Quantum computation and Shor's quantum factoring algorithm. *Review of Modern Physics*, 68:733–753, 1996.

[169] Artur K. Ekert and Richard Jozsa. Quantum algorithms entanglement-enhanced information processing. *Philosophical Transactions of Royal Society*, A356:1762–1782, 1998. quant-ph/9803072.

[170] Artur K. Ekert and Chiarra Macchiavello. Quantum error correction for communication. *Physical Review Letters*, 77:2585–2588, 1996. quant-ph/9602022.

[171] Artur K. Ekert, John Rarity, Paul Tapster, and G. Palma Practical quantum cryptography based on two-photon interferometry. *Physical Review Letters*, 69:1293–1295, 1992.

[172] Mark Ettinger and Peter Høyer. On quantum algorithms for noncommutative subgroups. Technical report, quant-ph/9807029, 1998.

[173] Mark Ettinger and Peter Høyer. Hidden subgroup states are almost orthogonal. Technical report, quant-ph/9901034, 1999.

[174] Mark Ettinger and Peter Høyer. A quantum observable for the graph isomorphism problem. Technical report, quant-ph/9901029, 1999a.

[175] Shimon Even, Oded Goldreich, and A. Lempel A randomized protocol for signing contracts. In *Proceedings of Crypto'82, Plenum Press*, pages 205–210, 1983.

[176] Hugh Everett. The theory of the universal wave function. In *The many-world interpretation of quantum mechanics*, pages 1–140. Princeton University Press, 1977. Reprinted PhD Thesis.

[177] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. A limit on the speed of quantum computation in determining parity. Technical report, quant-ph/9802045, 1998.

[178] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. A limit on the speed of quantum computation for insertion into an ordered list. Technical report, quant-ph/9812057, 1998a.

[179] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. How many functions can be distinguished with k quantum queries? Technical report, quant-ph/9901012, 1999.

[180] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. Invariant quantum algorithms for insertion into an ordered list. Technical report, quant-ph/9901059, 1999.

[181] Stephen Fenner, Frederic Green, Steven Homer, and Randall Pruim. Quantum **NP** is hard for **PH**. In *Proceedings of 6th Italian Conference on Theoretical Computer Science*, pages 241–252, 1998. quant-ph/9812056.

[182] Richard P. Feynman Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6/7):467–488, 1982.

[183] Richard P. Feynman Quantum mechanical computers. *Foundations of physics*, pages 507–531, 1986. Originally appeared in Optics News, February 1985, p 11-20.

[184] Richard P. Feynman *Feynman lectures on computation*. Addison-Wesley, 1995. Edited by Anthony J. G. Hey and Robin W. Allen.

[185] Richard P. Feynman, Robert B. Leighton, and Matthew Sands. *The Feynman lectures on physics, Volume 3*. Addison-Wesley, 1964.

[186] Lance Fortnow and John Rogers. Complexity limitations on quantum computation. In *Proceedings of 13th IEEE Conference on Computational Complexity*, pages 202–209, 1998.

[187] Edward Fredkin. Digital mechanics. *Physica D*, 45:254–270, 1990.

[188] Edward Fredkin and Tommaso Toffoli. Conservative logic. *International Journal of Theoretical Physics*, 21(3-4):219–253, 1982.

[189] Rūsiņš Freivalds. Probabilistic two-way machines. In *Proceedings of MFCS'81, LNCS 118*, pages 33–45, 1981.

[190] Christopher A. Fuchs and Asher Peres. Quantum-state disturbance versus information gain: uncertainty relation for quantum information. *Physival Review A*, 53(4):2038–2045, 1996.

[191] Neil A. Gershenfeld and Isaac L. Chuang Bulk spin-resonance quantum computation. *Science*, 275:350–356, 1997.

[192] Neil A. Gershenfeld, Isaac L. Chuang, and Seth Lloyd. Bulk quantum computation. In *Proceedings of the 4th Workshop on Physics and Computation. Phys96*, pages 134–134, 1996.

[193] James Glanz. Measurements are the only reality, say quantum tests. *Science*, 270:1439–1440, 1995.

[194] R. J. Glaubner *Frontiers in Quantum Optics*, chapter ?????, pages ???–??? Adam Helger, Bristol, 1988.

[195] Daniel Gottesman. Class of quantum error correcting codes saturating the quantum Hamming bound. *Physical Review A*, 54:1862–1868, 1996. Preprint quant-ph/9604038.

[196] Daniel Gottesman. *Stabilizator codes and quantum error correction*. PhD thesis, California Institute of Technology, 1997. quant-ph/9705052.

[197] Daniel Gottesman. A theory of fault-tolerant quantum computation. *Physics Review A*, 57:127–137, 1997a. quant-ph/9702029.

[198] Daniel Gottesman. The Heisenberg representation of quantum computers. Technical report, quant-ph/9807006, 1998.

[199] Markus Grassl, Thomas Beth, and Thomas Pellizzari Codes for quantum erasure channels. *Physical Review A*, 56:33–38, 1997. quant-ph/9610042.

[200] Daniel M. Greenberger, Michael A. Horne, and Anton Zeilinger. *Bell's theorem and the conception of the universe*, chapter Going beyond Bell's theorem, pages 69–. Kluwer Academic, Dordrecht, 1989.

[201] Robert B. Griffiths and Chi-Sheng Niu. Semiclassical Fourier transform for quantum computation. *Physical Review Letters*, 76(17):3228–3231, 1996.

[202] Gerhard Grössing and Anton Zeilinger. Quantum cellular automata. *Complex systems*, 2:197–208, 1988. Originally appeared in Proceedings of Conference on Cellular Automata, Cambridge, MA, 1986.

[203] Lov K. Grover A fast quantum mechanical algorithm for estimating a median. Technical report, quant-ph/9607024, 1996.

[204] Lov K. Grover A fast quantum mechanical algorithm for database search. In *Proceedings of 28th ACM STOC*, pages 212–219, 1996a.

[205] Lov K. Grover Quantum computer can search arbitrarily large databases by a single querry. *Physical Review Letters*, 79:4709–4712, 1997.

[206] Lov K. Grover Quantum mechanics helps in searching for a needle in a haystack. *Physical Review Letters*, 78:325–328, 1997a. quant-ph/9605043.

[207] Lov K. Grover Quantum telecomputation. Technical report, quant-ph/9704012, 1997b.

[208] Lov K. Grover A framework for fast quantum mechanical algorithms. In *Proceedings of 30th ACM STOCS*, pages 53–62, 1998.

[209] Lov K. Grover How fast can a quantum computer serach? Technical report, quant-ph/9809029, 1998a.

[210] Lov K. Grover Quantum search on structured problems. Technical report, quant-ph/9802035, 1998b.

[211] Jozef Gruska. Why we should not any longer only repair, polish and iron computer science education? *Education and Computers*, 8:303–330, 1993.

[212] Jozef Gruska. *Foundations of computing.* Thomson International Computer Press, 1997.

[213] Alfréd Haar. Zur Theorie der orthogonalen Funktionssysteme. *Mathematische Annalen*, LXIX:331–371, 1910.

[214] E. Hagley, X. Maltre, G. Nogues, C. Wunderlich, M. Brune, Jean-Michel Raimond, and Serge Haroche. Generation of Einstein-Podolsky-Rosen pairs of atom. *Physical Review Letters*, 79(1):1–5, 1997.

[215] Lisa Hales and Sean Hallgren. Sampling Fourier transforms on different domains. Technical report, quant-ph/9812060, 1998.

[216] Paul Hausladen, Richard Jozsa, Benjamin Schumacher, Michael Westmoreland, and William K.Wootters Classical information capacity of a quantum channel. *Physical Review A*, 54:1869–1876, 1996.

[217] Mark Hillery, Vladimír Bužek, and André Berthiaume. Quantum secret sharing. Technical report, quant-ph/9806063, 1998.

[218] Raymond Hill. *A first course in coding theory.* Claredon Press, Oxford, 1986.

[219] Scott Hill and William K. Wootters Entanglement of a pair of quantum bits. Technical report, quant-ph/9703041, 1997.

[220] Mika Hirvensalo. *On quantum computation.* PhD thesis, Turku Center for Computer Science, 1997.

[221] D. G. Hoffman, D. A. Leonard, C. C. Linder, K. T. Phelps, C. A. Rodger, and J. R. Wall *Coding theory, the essentials*. Marcel Dekker Inc., 1991.

[222] Tad Hogg. Highly structured searches with quantum computers. *Physical Review Letters*, 80:2473–2476, 1998. quant-ph/9508012.

[223] Tad Hogg, Carlos Mochon, Wolfgang Polak, and Eleanor G. Rieffel Tools for quantum algorithms. Technical report, quant-ph/9811073, 1998.

[224] Alexander S. Holevo Bounds for the quality of information transmitted by quantum communication channel. *Problemy Peredachi information*, 9(3):3–11, 1973. English translation: Problems of Information Transmission, V9, 1973, 177-183.

[225] Michal Horodecki, Pawel Horodecki, and Ryszard Horodecki. Inseparable two spin-$\frac{1}{2}$ density matrices can be distilled to a singleton form. *Physical Review Letters*, 78:574–577, 1997.

[226] Michal Horodecki, Pawel Horodecki, and Ryszard Horodecki. Mixed-state entanglement and distillation: is there a "bound" entanglement in nature? Technical report, quant-ph/9801069, 1998.

[227] Michal Horodecki, Pawel Horodecki, and Ryszard Horodecki. Bound entanglement can be activated. Technical report, quant-ph/9806058, 1998a.

[228] Michal Horodecki and Ryszard Horodecki. Are there basic laws of quantum information processing? Technical report, quant-ph/9705003, 1997.

[229] Ryszard Horodecki. Unitary transformation ether and its possible applications. *Annalen der Physi (Leipzig)k*, 48:479–488, 1991.

[230] Peter Høyer. Efficient quantum transforms. Technical report, quant-ph/9702028, 1997.

[231] Juraj Hromkovič. *Communication complexity and parallel computing.* Springer Verlag, 1997.

[232] Richard J. Hughes Cryptography, quantum computation and trapped ions. Technical report, Los Alamos National Laboratory, 1997. preprint LA-UR-97-4986.

[233] Richard J. Hughes, D. M. Alde, P. Dyer, G. G. Luther, G. L. Morgan, and M. M. Schauer Quantum cryptography. *Contemporary physics*, 36(3):149–163, 1995.

[234] Richard J. Hughes, Daniel F. V. James, J. J. Gomez, M. S. Gulley, M. H. Holzscheiter, Paul G. Kwiat, S. K. Lamoreaux, C. G. Peterson, V. D. Sandberg, M. M. Schauer, C. M. Simmons, C. E. Thorburn, D. Tupa, P. Z. Wang, and A. G. Whille The Los Alamos trapped ion quantum generator computer experiment. *Fortschritte Physics*, 46:329–361, 1998.

[235] Richard J. Hughes, Daniel F. V. James, Emanuel H. Knill, Raymond Laflamme, and Albert G. Petschek Decoherence bounds on quantum computation with trapped ions. *Physical Review Letters*, 77:3240–3243, 1996a.

[236] Richard J. Hughes, G. G. Luther, G. L. Morgan, C. G. Peterson, and C. M. Simmons Quantum cryptography over underground optical fibers. In *Advances in Cryptology, Proccedings of CRYPTO'96*, pages 329–342. LNCS 1109, Springer-Verlag, 1996a.

[237] L. P. Hugston, Richard Jozsa, and William K. Wootters A complete classification of quantum ensembles having a given density matrix. *Physics Letters A*, 183:14–18, 1993.

[238] Russe Impagliazzo and Steven Rudich. Limits on the provable consequances of one-way permutations. In *Proceedings of 21st ACM STOC*, pages 44–61, 1989.

[239] Max Jammer. *The philosophy of quantum mechanics.* John Wiley & Sons, 1974.

[240] Jonathan A. Jones and Michele Mosca. Implementation of a quantum algorithm to solve Deutsch's problem on a nuclear magnetic resonance quantum computers. *Journal of Chemical Physics*, 109(5):1648–1653, 1998. quant-ph/9801027, 9808056.

[241] T. F. Jordan *Linear operators for quantum mechanics.* Wiley, 1969.

[242] Richard Jozsa. Fidelity for mixed quantum states. *Journal of Modern Optics*, 41(12):2315–2323, 1994.

[243] Richard Jozsa. Entanglement and quantum computation. In S. Huggett, L. Mason, K. P. Todd, S. T. Tsou, and N. M. J. Woodhouse, editors, *The Geometric Universe*, pages 369–379. Oxford University Press, 1997.

[244] Richard Jozsa. Quantum algorithms and the Fourier transform. *Proceedings of Royal Society London A*, 454:323–337, 1997a. quant-ph/9707033.

[245] Richard Jozsa. Searching in Grover's algorithm. Technical report, quant-ph/9901021, 1999.

[246] Richard Jozsa and Benjamin W. Schumacher A new proof of quantum noiseless coding theorem. *Journal of Modern Optics*, 41(12):2343–2349, 1994.

[247] Jānis Kaņeps and Rūsiņš Freivalds. Running-time to recognize nonregular languages by 2-way probabilistic automata. In *Proceedings of MFCS'91, LNCS 510*, pages 174–185, 1991.

[248] B. E. Kane A Si–based nuclear spin quantum computer. *Nature*, 393:133–13?, 1998.

[249] Jarkko Kari. Reversibility of 2D cellular automata is undecidable. *Physica D*, 45:379–385, 1990.

[250] T. Kato *Perturbation theory for linear operators*. Springer-Verlag, 1976.

[251] Robert W. Keyes Miniaturization of electronics and its bounds. *IBM Journal of Research and Development*, 32:24–28, 1988.

[252] Joe Kilian. Founding cryptography on oblivious transfer. In *Proceedings of 20th ACM STOC*, pages 20–31, 1988.

[253] Alexei Yu. Kitaev Quantum measurements and Abelian stabilizer problem. Technical report, quant-ph/9511026, 1995.

[254] Alexei Yu. Kitaev *Quantum communication, computing and measurement*, chapter Quantum error correction with imperfect gates, pages 181–188. Plenum Press, New York, 1996.

[255] Alexei Yu. Kitaev Fault-tolerant quantum computation by anyons. Technical report, 1997.

[256] Emanuel H. Knill and Raymond Laflamme. Concatenated quantum codes. Technical report, quant-ph/9608012, 1996.

[257] Emanuel H. Knill and Raymond Laflamme. Theory of quantum error-correcting codes. *Physical Review A*, 55(2):900–911, 1997. quant-ph/9604034.

[258] Emanuel H. Knill, Raymond Laflamme, and Wojciech H. Zurek Accuracy threshold for quantum computation. Technical report, quant-ph/9610011, 1996.

[259] Emanuel H. Knill, Raymond Laflamme, and Wojciech H. Zurek Resilent quantum computation: error models and thresholds. Technical report, quant-ph/9702058, 1997.

[260] Attila Kondacs and John Watrous. On the power of finite state automata. In *Proceedings of 36th IEEE FOCS*, pages 66–75, 1997.

[261] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, 1997.

[262] Paul G. Kwiat, K. Mattle, Harald Weinfurter, Anton Zeilinger, A. V. Sergienko, and Y. M. Shih New high-intenzity source of polarization—entangled photon pairs. *Physics Review Letters*, 75:4337–4341, 1995.

[263] Paul G. Kwiat and Harald Weinfurter. Embedded Bell-state analysis. *Physical Review A*, 58(4):2623–2626, 1998.

[264] Raymond Laflamme, Emanuel H. Knill, Wojciech H. Zurek/indexZurek, Wojciech H., P. Catasti, and S. U. S. Hanappen. NMR GHZ. Technical report, quant-ph/9709025, 1997.

[265] Raymond Laflamme, Cesar Miquel, Juan Paz, and Wojciech H. Zurek Perfect quantum error correction code. *Physical Review Letters*, 77:198–201, 1996. quant-ph/9602019.

[266] Ralf Landauer. Information is physical. *Physics Today*, 44:23–29, 1991.

[267] Ralf Landauer. Is quantum mechanically coherent computation useful? In D. H. Fing and B.-L. Hu, editors, *Proceedings of the Drexel-4 Symposium on Quantum Nonintegrability—Quantum Classical Correspondence*, pages 37–55. International Press, 1994.

[268] Ralf Landauer. Is quantum mechanics useful? *Philosophical Transactions of Royal Society of London A*, 353:367–376, 1995.

[269] Klaus-Jörn Lange, Pierre McKenzie, and Alain Tapp. Reversible space equals deterministic space. In *Proceedings of 12th Annual IEEE Conference on Computational Complexity*, pages 45–50, 1997.

[270] Yves Lecerf. Récursive insolubilité de l'equation genérale de diagonalisation de deux monomorphismes de monoides libres $\phi x =^* x$. *Comptes Rendus de l'Académie des Sciences*, 257:2940–2943, 1963.

[271] Yves Lecerf. Machines de Turing réversibles. récursive insolubilité en $n \in \mathbf{N}$ de léquation $u = \theta^n u$, où θ est un isomorphisme de codes. *Comptes Rendus de l'Académie des Sciences*, 257:2597–2600, 1963a.

[272] Harvey S. Leff and Andrew R. Fex *Maxwell's demon: entropy, information*. Adam Hilger, 1990.

[273] Lev B. Levitin ????????? In *Proceedings of the 4th All-Union Conference on Information and Coding Theory, (Moscow-Tashkent)*, pages ???–???, 1969.

[274] Lev B. Levitin *Information complexity and control in quantum physics*, chapter Information theory for quantum systems, pages 15–?? Springer, 1987.

[275] Noah Linden and Sandu Popescu. The halting problem for quantum computers. Technical report, quant-ph/9806054, 1998.

[276] David Lindley. *Where does the weirdness go?* BasicBooks, 1996.

[277] Seth Lloyd. A potentially realizable quantum computer. *Science*, 261:1569–1571, 1993.

[278] Seth Lloyd. Envision a quantum computer. *Science*, 263:695, 1994.

[279] Seth Lloyd. Almost any quantum gate is universal. *Physical Review Letters*, 75:346–349, 1995.

[280] Seth Lloyd. Capacity of the noisy quantum channel. *Physical Review A*, 55(3):1613–1622, 1997.

[281] Hoi-Kwong Lo. Insecurity of quantum secure computations. *Physical Review A*, 56(2):1154–1172, 1997.

[282] Hoi-Kwong Lo and H. F. Chau Is quantum bit commitment really possible? *Physical Review Letters*, 78(17):3410–3413, 1997.

[283] Hoi-Kwong Lo and H. F. Chau Why quantum bit commitment and ideal quantum coin tossing are impossible? Technical report, quant-ph/9711065, 1997a.

[284] Hoi-Kwong Lo and Sandu Popescu. Concentrating entanglement by local actions—beyond mean values. Technical report, quant-ph/9707038, 1997.

[285] Samuel J. Lomonaco A quick glance at quantum cryptography. Technical report, quant-ph/9811056, 1998.

[286] F. J. MacWilliams and Neil J. A. Sloane *The theory of error-correcting codes*. North-Holland, Amsterdam, 1977.

[287] Yurii Manin. *Solvable and unsolvable*. Soviet radio, 1980. In Russian.

[288] C. Marand and Paul D. Townsend Quantum key distribution over the distance as long as 30 km. *Optics Letters*, 20:1695–1697, 1995.

[289] Norman Margolus. *Complexity, entropy and physics*, chapter Parallel quantum computation, pages 273–. Addison-Wesley, 1994.

[290] K. Mattle, Harald Weinfurter, Paul G. Kwiat, and Anton Zeilinger. Dense coding in experimental quantum communication. *Physics Review Letters*, 76:4656–4659, 1996.

[291] Dominic C. Mayers Quantum key distribution and string oblivious transfer in noisy channels. Technical report, quant-ph/9606003, 1996.

[292] Dominic C. Mayers Unconditinaly secure quantum bit commitment is impossible. *Physical Review Letters*, 78:3414–3417, 1998.

[293] Dominic C. Mayers and Andrew C.-C. Yao Unconditional security in quantum cryptography. Technical report, quant-ph/9802025, 1998.

[294] Dominic C. Mayers and Andrew C.-C. Yao Quantum cryptography with imperfect apparatus. Technical report, quant-ph/9809039, 1998a.

[295] David A. Meyer Quantum strategies. Technical report, quant-ph/9804010, 1998.

[296] Christoper Monroe, Dan M. Meeckhof , B. E. King, W. M. Itano, and Daniel J. Wineland Demonstration of a fundamental quantum logic gate. *Physical Review Letters*, 75(25):4714–4717, 1995.

[297] Cristopher Moore and James P. Crutchfield Quantum automata and quantum grammars. Technical report, Santa Fe, 1997.

[298] Cristopher Moore and Martin Nilsson. Parallel quantum computation and quantum codes. Technical report, quant-ph/9808027, 1998.

[299] Cristopher Moore and Martin Nilsson. Some notes on parallel computation. Technical report, quant-ph/9804034, 1998a.

[300] Kenichi Morita and Masateru Harao. Computational universality of one-dimensional reversible (injective) cellular automata. *Transactions of the IEICE*, E72:758–762, 1989.

[301] Michele Mosca. Quantum searching, counting and amplitude modification by eigenvector-analysis. In *Proceedings of MFCS'98 Workshop on Randomized Algorithms*, pages 90–100, 1998.

[302] Michele Mosca. Counting on a quantum computer. Technical report, MFCS'98, 1998b. Talk at MFCS'98 Workshop on randomization.

[303] Michele Mosca and Artur K. Ekert The hidden subgroup problem and eigenvalue estimation on a quantum computer. In *Proceedings of* 1*st NASA International Conference on Quantum Computing* & *Quantum Communication*, page ??? LNCS 1504, Springer-Verlag, 1998.

[304] Rajeev Motwani and Prabhakar Raghavan. *Randomized algorithms*. Cambridge University Press, 1995.

[305] Antoine Muller, Hugo Zbinden, and Nicolas Gisin. Underwater quantum coding. *Nature*, 378:449–449, 1995.

[306] Michael Nielsen A partial order on the entangled states. Technical report, quant-ph/9811053, 1998.

[307] Michael A. Nielsen, Emanuel H. Knill, and Raymond Laflamme. Complete quantum teleportation by nuclear magnetic resonance. Technical report, quant-ph-9811020, 1998.

[308] Noam Nisan. CREW PRAMs and decision trees. *SIAM Journal of computing*, 20(6):999–1007, 1991.

[309] Noam Nisan and Mario Szegedy. On the degree of Boolean functions as real polynomials. In *Proceedings of* 24*th ACM STOC*, pages 462–467, 1992.

[310] Masanao Ozawa. Quantum Turing machines: local transitions, preparation, measurement, and halting problem. Technical report, quant-ph/9809038, 1998.

[311] Masanao Ozawa and Harumichi Nishimura. Local transition functions of quantum Turing machines. Technical report, quant-ph/9811069, 1998.

[312] Yuri Ozhigov. Fast quantum verification for formulas of predicate calculus. Technical report, quant-ph/9809015, 1998.

[313] Yuri Ozhigov. Quantum computers speed up classical with probability zero. Technical report, quant-ph/9803064, 1998a.

[314] Edward W. Packel *Functional analysis, a short course*. Indext Educational Publishers, 1974.

[315] Ramamohan Paturi. On the degree of polynomials that appriximate Boolean functions. In *Proceedings of* 24*th ACM FOCS*, pages 468–474, 1992.

[316] Roger Penrose. *The emperor's new mind*. Vintage, 1990.

[317] Roger Penrose. *Shadows of the mind*. Oxford University Press, 1994.

[318] Asher Peres. Reversible logic and quantum computers. *Physical Review A*, 32(6):3266–3276, 1985.

[319] Asher Peres. *Quantum theory: Concepts and methods*. Kluwer Academic Publisher, 1993.

[320] Asher Peres. Error symmetrization in quantum computers. Technical report, quant-ph/9605009, 1996.

[321] Asher Peres and William K. Wootters Optimal detection of quantum information. *Physics Review Letters*, 66:1119–1122, 1991.

[322] Carl Adam Petri. Gründsätzliches zur Beschreibung diskreter Prozesse. In *Proceedings 3. Colloquium über Automatentheorie (Hannover), 1965*, pages 121–140, 1967.

[323] Simon J. D. Phoenix, Stephen M. Barnett, Paul D. Townsend, and Keith J. Blow Multi-user quantum cryptography on optical networks. *Journal of modern optics*, 42:1155–1163, 1995.

[324] Simon J. D. Phoenix and Paul D. Townsend Quantum cryptography: how to beat the code breakers using quantum mechanics. *Contemporary Physics*, 36(3):165–195, 1995.

[325] Jean-Erie Pin. On the languages accepted by finite reversible automata. In *Proceedings of 14th ICALP*, pages 237–249. LNCS 267, Springer-Verlag, 1987.

[326] Martin B. Plenio and Peter L. Knight Realistic lower bounds for the factorization time of large numbers on a quantum computer. *Physical Review A*, 53(5):2986–2990, 1996.

[327] Martin B. Plenio, Vlatko Vedral, and Peter L. Knight Quantum error correction in the presence of spontaneous emission. Technical report, quamt-ph/9603022, 1996.

[328] John Preskill. Fault tolerant quantum computation. Technical report, quant-ph/9712048, 1997. To appear in "Introduction to quantum computation", edited by H.-K. Lo, S. Popescu and T. P. Spiller.

[329] John Preskill. Quantum computing. Technical report, CALTECH, 1998. http://www.theory.caltech.edu/ preskill/ph229.

[330] Markus Püschel, Martin Röteller, and Thomas Beth. Fast Fourier transforms for a class of non-abelian groups. Technical report, quant-ph/9807064, 1998.

[331] Michael O. Rabin Probabilistic automata. *Information and Control*, 6(3):230–244, 1963.

[332] Michael O. Rabin How to exchange secrets by oblivious transfer. Technical report, Technical Memo TR-81, Aiken Computational Laboratory, Harward University, 1981.

[333] Eric M. Rains A rigorous treatment of distillable entanglement. Technical report, quant-ph/9809078, 1998.

[334] Eric M. Rains, R. H. Hardin, Peter W. Shor, and Neil J. A. Sloane A nonadditive quantum code. *Physical Review Letters*, 79:953–954, 1997.

[335] Eleanor G. Rieffel and Wolfgang Polak. An introduction to quantum computing for non-physicists. Technical report, quant-ph/9809016, 1998.

[336] Yurii Roghozin. On the notion of universality and small universal Turing machines. *Theoretical Computer Science*, 168:215–240, 1996.

[337] Hein Röhrig. An upper bound for searching an unordered list. Technical report, quant-ph/9812061, 1998.

[338] D. A. Ross A modification of Grover's algorithm as a fast database search. Technical report, quant-ph/9807078, 1998.

[339] Martin Rötteler and Thomas Beth. Polynomial-time solution to the hidden subgroup problem for a class of non-abelian groups. Technical report, quant-ph/9812070, 1998.

[340] Keijo Ruohonen. Reversible machines and Post's correspondence problem for biprefix morphisms. *EIK*, 21(12):579–595, 1995.

[341] Bruce Schneier. *Applied cryptography*. Wiley, New York, 1996.

[342] Leonard Schulman and Umesh Vazirani. Scalable NMR quantum computation. Technical report, quant-ph/9804060, 1998.

[343] Benjamin W. Schumacher Quantum coding. *Physical Review A*, 51(4):2738–2747, 1995.

[344] Benjamin W. Schumacher and Michael A. Nielsen Quantum data processing and error correction. *Physical Review A*, 54(4):2629–2635, 1996.

[345] Claude E. Shannon and Warren Weaver. *The mathematical theory of communication*. University of Illinois Press, Urbana, 1949.

[346] Peter W. Shor Algorithms for quantum computation: discrete log and factoring. In *Proceedings of 35th IEEE FOCS*, pages 124–134, 1994.

[347] Peter W. Shor Scheme for reducing decoherence in quantum computer memory. *Physical Review A*, 52:2493–2496, 1995.

[348] Peter W. Shor Fault-tolerant quantum computation. In *Proceedings of 37th IEEE FOCS*, pages 56–65, 1996.

[349] Peter W. Shor Polynomial time algorithms for prime factorization and discrete logarithms on quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, 1997.

[350] Daniel R. Simon On the power of quantum computation. In *Proceedings of 35th IEEE FOCS*, pages 116–123, 1994. See also SIAM Journal of Computing, V26, N5, 1474-1483, 1997.

[351] Tycho Sleator and Harald Weinfurter. Realizable universal quantum logic gates. *Physical Review Letters*, 74(20):4087–4090, 1995.

[352] Henry P. Stapp The Copenhagen interpretation. *American Journal of Physics*, 40:1098–1116, 1972.

[353] Andrew M. Steane Error correcting codes in quantum theory. *Physics Review Letters*, 77(5):793–797, 1996.

[354] Andrew M. Steane Multiple particle interference and quantum error correction. *Proceedings of Royal Society London*, A 452:2551–2577, 1996a.

[355] Andrew M. Steane Quantum Reed-Muller code. Technical report, quant-ph/9608026, 1996b.

[356] Andrew M. Steane Simple quantum error-correcting codes. *Physical Review A*, 54:4741–4751, 1996c. quant-ph/9605021.

[357] Andrew M. Steane Quantum computing. Technical report, quant-ph/9708022, 1997.

[358] Andrew M. Steane Quantum error correction. In *Introduction to quantum computation*, page to be published. World Scientific, 1998.

[359] Andrew M. Steane Enlargement of Calderbank Shor Steane quantum codes. Technical report, quant-ph/9802061, 1998a.

[360] Andrew M. Steane Efficient fault-tolerant computing. Technical report, quant-ph/9809034, 1998b.

[361] Karl Svozil. Quantum computation and complexity theory I and II. *Bulletin of EATCS*, 55 and 56:170–207, 116–136, 1995.

[362] Max Tegmark. The interpretation of quantum mechanics: many worlds or many words. Technical report, quant-ph/9709032, 1996.

[363] W. Teich, K. Obermayer, and G. Mahler Structural basis of multidisciplinary quantum systems II. Effective few particle dynamics. *Physical Review B*, 37(4):8111–8120, 1988.

[364] Barbara M. Terhal and John A. Smolin Single quantum querying of a database. *Physical Review A*, 58:1822–1826, 1997. quant-ph/9705041.

[365] Wolfgang Tittel, J. Brendel, Hugo Zbinden, and Nicolas Gisin. Experimental demonstration of quantum correlation over more than 10km. *Physical Review A*, 57(5):3229–3232, 1998. Preprint quant-ph/9806043.

[366] Wolfgang Tittel, Gregoire Robordy, and Nicolas Gisin. Quantum cryptography. *Physics World*, pages 41–45, 1998a.

[367] Tommaso Toffoli. Computation and construction universality of reversible cellular automata. *Journal of Computer and System Sciences*, 15:213–231, 1977.

[368] Tommaso Toffoli. Reversible computing. In *Proceedings of ICALP'80*, pages 632–644. LNCS 84, Springer-Verlag, 1980.

[369] Tommaso Toffoli. Bicontinuous extensions of invertible combinatorial functions. *Mathematical Systems Theory*, 14:13–23, 1981.

[370] Q. A. Turchette, C. J. Hood, W. Lange, Hideo Mabuchi, and H. J. Kimble. Measurement of conditional phase shifts for quantum logic. *Physical Review Letters*, 75:4710–4713, 1995.

[371] Armin Uhlmann. The "transition probability" in the state space of a *-algebra. *Reports Mathematical Physics*, 9:273–279, 1976.

[372] Armin Uhlmann. Entropy and optimal decomposition of states relative to a maximal commutative algebra. Technical report, quant-ph/9704017, 1997.

[373] Wim van Dam. A universal quantum cellular automaton. In M. Biafore J. Leão T. Toffoli, editor, *Proceedings of Physics and Computation, PhysComp'96*, pages 323–331. New England Complex Systems Institute, 1996.

[374] Wim van Dam. Quantum oracle interrogation. Technical report, quant-ph/9805006, 1998.

[375] Wim van Dam, Peter Høyer, and Alain Tapp. Multiparty quantum communication complexity. Technical report, quant-ph/9710054, 1997.

[376] Wim van Dam. Quantum oracle interrogation: getting all information for almost half price. In *Proceedings of* 39*th IEEE FOCS*, pages 362–367, 1998.

[377] S. J. van Enk, Juan I. Cirac, and Peter Zoller. Ideal quantum communication over noisy channels: a quantum optical implementation. *Physical Review Letters*, 78:4293–4296, 1997.

[378] J. H. van Lint *Introduction to coding theory*. North Holand, 1995.

[379] Umesh Vazirani. Quantum computing. Technical report, http://www.cs.berkeley.edu/ṽazirani, 1997.

[380] Vlatko Vedral, Adriano Barenco, and Artur K. Ekert Quantum networks for lementary operations. *Physical Review A*, 54(1):147–153, 1996. quant-ph/9511018.

[381] Vlatko Vedral and Martin B. Plenio Entanglement measures and purification procedures. Technical report, quant-ph/9707035, 1997.

[382] Vlatko Vedral and Martin B. Plenio Basics of quantum computations. Technical report, quant-ph/9802065, 1998.

[383] Vlatko Vedral, Martin B. Plenio, Mark A. Ripin, and Peter L. Knight Quantifying entanglement. Technical report, quant-ph/9702027, 1997.

[384] Dan Ventura and Tony Martinez. A quantum computational learning algorithm. Technical report, quant-ph/9807052, 1998.

[385] John von Neumann. Thermodynamik quantenmechanischer Gesamtheiten. *Göttinger Nachrichten*, pages 273–291, 1927.

[386] John von Neumann. *Mathematische Grunglagen der Quantenmechanik*. Springer-Verlag, 1932. English translation: Mathematical Foundations of Quantum mechanics, Princeton University, 1955.

[387] Joachim von zur Gathen and James R. Rucke Polynomials with two values. *Combinatorica*, 17(3):345–362, 1997.

[388] Klaus Wagner. The complexity of combinatorial problems with succinct input representation. *Acta Informatica*, 23:325–356, 1986.

[389] John Watrous. On one-dimensional cellular automata. In *Proceedings of 36th IEEE FOCS*, pages 528–537, 1995.

[390] John Watrous. Notes on a quantum cellular automaton illustrating the EPR paradox. Technical report, University of Wisconsin, 1997.

[391] John Watrous. On the power of 2-way quantum finite automata. Technical report, University of Wisconsin, 1997.

[392] John Watrous. Relationships between classical and quantum space bounded complexity classes. In *Proceedings of 13th IEEE Conference on Computational Complexity*, pages 210–227, 1997a.

[393] John Watrous. Quantum simulations of classical random walks and undirected graph connectivity. Technical report, quant-ph/9812012, 1998.

[394] Alfred Wehrl. General properties of entropy. *Rewiews of Modern Physics*, 50:221–260, 1978.

[395] Steve Weinberg. Testing quantum mechanics. *Annals of Physics*, 194:336–386, 1989.

[396] Stephen J. Wiesner Conjugate coding. *SIGACT News*, 15:78–88, 1983.

[397] Collin P. Williams and Scott H. Clearwater *Explorations in quantum computing*. Springer-Verlag, 1998.

[398] Daniel J. Wineland, Christoper Monroe, W. M. Itano, D. Kielpinski, B. E. King, C. J. Myatt, Q. A. Turchette, and C. S. Wood Quantum computation, spectroscopy of trapped ions and Schrödinger cat. Technical report, quant-ph/9809028, 1998a.

[399] Daniel J. Wineland, Christoper Monroe, W. M. Itano, B. E. King, D. Leibfried, Dan M. Meeckhof, C. J. Myatt, and C. S. Wood Experimental processor on trapped ion quantum computer. *Fortschritte Physics*, 46(4-5):363–390, 1998.

[400] William K. Wootters Entanglement of formation of an arbitrary state of two qubits. Technical report, quant-ph/9709029, 1997.

[401] William K. Wootters and Wojciech H. Zurek A single quantum cannot be cloned. *Nature*, 299:802–803, 1982.

[402] C. S. Wu and I. Shaknov The angular correlation of scattered annihilation radiation. *Physical Review*, 77(1):136–136, 1950.

[403] Tomoyuki Yamakami and Andrew C.-C. Yao $\mathbf{NQP} = \mathbf{co\text{-}C_{=}P}$. Technical report, quant-ph/9812032, 1998.

[404] Andrew C.-C. Yao Quantum circuit complexity. In *Proceedings of 34th IEEE FOCS*, pages 352–361, 1993.

[405] Christof Zalka. Grover's quantum searching algorithm is optimal. Technical report, quant-ph/9711070, 1997.

[406] Christof Zalka. Fast version of Shor's quantum factoring algorithm. Technical report, quant-ph/9806084, 1998.

[407] Anton Zeilinger, Michael A. Horne, Harald Weinfurter, and Marek Zukowski. Three particle entanglements from two entangled pairs. *Physical Review Letters*, 78(16):3031–3034, 1997.

[408] Lin Zhang and Ian Fuss. Quantum Reed-Muller codes. Technical report, quant-ph/9703045, 1997.

[409] Wojciech H. Zurek Reversibility and stability of information processing systems. *Physical Review Letters*, 53:391–394, 1984.

[410] Wojciech H. Zurek and Raymond Laflamme. Quantum logical operations on encoded qubits. *Physical Review Letters*, 77:4683–4686, 1996. quant-ph/9605013.

[411] Konrad Zuse. *Der Computer—Mein Lebenswerk*. Springer-Verlag, 1984.