Introduction to Quantum Computing (M743)

Zdzisław Meglicki

February 5, 2002

Document version

$Id: M743.tex,v 1.87 2002/02/05 21:18:42 gustav Exp $

Contents

Chapter 1

Preliminaries

This course aims to cover the newly emerging discipline of quantum computing. It is addressed to graduate students, senior undergraduate students, and academics who would like to learn more about this research area.

The course is based on a variety of materials including "Quantum Computation and Quantum Information" by Michael L. Nielsen and Isaac L. Chuang [80], "Lecture Notes for Physics 229: Quantum Information and Computation" by John Preskill, Caltech, September 1998 [87], "Introduction to Quantum Computation and Information", edited by Lo, Popescu and Spiller [70], "The Physics of Quantum Information", edited by Bouwmeester, Ekert and Zeilinger [17], "Explorations in Quantum Computing" by Colin P. Williams and Scott H. Clearwater, Springer Verlag, New York, 1998 [104], our previous notes for B679, research papers, and various other sources.

Quantum computing is *not* about computational quantum physics, or quantum chemistry as it is done *today*. It is *not* about modelling quantum systems using classical computers, although we will attempt to model a very simple two-gate quantum computer using a classical computer in section 4.5.7, which talks about the Feynman Quantum Computer.

Quantum computing is about computing *with* quantum systems, called *quantum computers*. Once you have a system like this, in principle you can model various quantum phenomena with it far more efficiently than with a classical computer. It is also possible to use quantum computers to solve otherwise intractable computational problems in other areas, e.g., break unbreakable codes.

Quantum computing is rooted, as the name suggests, in quantum mechanics. The logic it is based on is that of quantum logic, not classical logic. Quantum logic differs from classical logic in many important respects. To begin with there is no distribution axiom in quantum logic. Instead of operating on bits, we will operate on *qubits*, which, for most practical purposes can be thought of as spin-$\frac{1}{2}$ systems or polarization states of a photon. If you have ever encountered spin-$\frac{1}{2}$ systems you will notice some similarities to binary logic: to begin with spin-$\frac{1}{2}$ systems can exist in two states: $|\uparrow\rangle$ or $|\downarrow\rangle$, like a classical bit, which can be either 1 or 0. But unlike classical bits, qubits can also exist in a superposition

of $|\uparrow\rangle$ and $|\downarrow\rangle$. They can be *entangled* with other qubits too. Superposition and entanglement enrich quantum logic (and quantum physics in general) immensely. In the so called thermodynamic limit many of those quantum riches average away, they disappear, and we are left with classical physics, classical bits, and classical computers. It is possible to use a quantum computer like a classical probabilistic computer, but this is not a very interesting use of quantum systems.

One of the holy grails of the newly emerging discipline of quantum computing is to find what more can be done with quantum computers and if those truly peculiar quantum features associated with qubits can be utlized in their full quantumness to accomplish tasks that may be too hard for classical computers.

The other holy grail is to build usable and programmable quantum computers. The systems that quantum computer scientists play with today, such as heteropolymers, nuclear magnetic resonance machines, quantum electrodynamic cavities, quantum dots and Josephson junctions demonstrate that quantum computing is possible, but we're here still only on the level of just a few qubits, and people get very excited if they can demonstrate something very basic such as a single qubit control, or a very simple error correction procedure.

For a physicist, chemist and a philosopher quantum computers are very interesting for a yet another reason: you can turn the table around and use a quantum computer in order to investigate the very fundamental principles of quantum mechanics. You can look at quantum interference effects, at quantum entanglement, at quantum teleportation. All this, possibly, on a single chip with quantum dots and Josephson junctions. Einstein Podolsky Rosen paradox, Schrödinger cat paradox, wave function collapse – these and many other puzzling effects can be studied experimentally using quantum computers and quantum computing.

1.1 The Aims of the Course

The main purpose of the course is to provide an introduction to and a fairly thorough review of quantum computing as it stands today. If you work through the notes diligently, at the end of the day you should be able to follow most quantum computing literature published in Nature, Science, Physical Review A, or on the LANL Archive. Special emphasis is placed on experimental accomplishments in quantum computing and related technologies. These are quoted whenever available. Whatever physics and mathematics are introduced and discussed serve primarily as a background needed to design and analyze quantum circuits and to discuss their implementation and related laboratory experiments. Axiomatic and formalistic approaches to quantum mechanics are neither mentioned nor followed. Like Peres, I strongly believe that quantum mechanics does not happen in Hilbert space – it happens in the laboratory.

1.2 When and Where?

There will be two lectures a week in the Fall semester of 2001.

Classes will be held on Wednesdays and Fridays, 2:30PM to 3:45PM, in SW218.

1.3 Required Background

How much quantum mechanics do you need to know to attend the lecture?

I will attempt to explain enough of it to make the lecture comprehensive even if you have never studied quantum mechanics before. The lecture is aimed at senior undergraduate and graduate students majoring in Computer Science, Chemistry, Molecular Biology, Electronic Engineering, Mathematics, and Physics. With such a broad background I cannot assume much and so the course is going to provide whatever is required to follow the material.

Nevertheless quantum mechanics being a very difficult and a very rich subject cannot be easily compressed to just one or two lectures. Consequently the more you know about it the better. An introductory quantum mechanics course, e.g., similar to Feynman's volume III [35] but *not* to the Berkeley Physics Course, is going to be helpful. The reason why Berkeley Course is less useful than Feynman in our context is because we will work almost exclusively with mechanics of simple discrete systems, and will stay away from the Schrödinger Wave Equation. Yet most traditional Quantum Mechanics courses dwell more on the latter than on the former.

The course will also require a certain level of mathematical skills. You need to have *some* background in algebra and analysis: vector spaces, complex numbers, complex functions, ordinary differential equations – roughly at the level that corresponds to what you end up with after a second year mathematics course. It is good to know a difference between a form and a vector. It is good to know about linear operators, eigenvalues, and stuff like this.

1.4 Recommended Reading

This section recapitulates what I have already said above. The books listed below are *recommended* reading. None of them are *required*. Quite detailed lecture notes are going to be provided on-line (this document) and they should be sufficient to follow the course.

The notes will change and evolve as the course develops. Although the B679 Notes of 1999 are a starting point for us, we covered a lot of additional material in the Spring 2001, and only some of it found its way to the on-line notes. This material as well as other updates will be added to the notes as we plough on.

Apart from the notes I recommended the following three texts, on which the notes are largely based:

- M. A. Nielsen and Isaac L. Chuang, "Quantum Computation and Quantum Information", Cambridge University Press, 2000, ISBN 0521635039, 700 pages [80]

  This is the most recommended reading of all. Chuang from IBM and Nielsen from Caltech and the University of Queensland (Australia) are amongst the most experienced researchers in the field. Their book is probably the first comprehensive introduction to the ideas and techniques of quantum computation and information.

  The book is divided into three parts.

  The first part provides introduction to quantum mechanics and to computer science – concepts and methods needed to follow the remainder of the test.

  The second part discusses quantum circuits, quantum algorithms and realizations of various quantum computers.

  The last part is dedicated to quantum information theory.

  To find more about the book connect to

  http://www.squint.org/qci

  or to

  http://www.cup.org

  or to

  http://uk.cambridge.org

  To view Mike Nielsen's home page go to

  http://www.theory.caltech.edu/~mnielsen/

  To view Isaac Chuang's home page go to

  http://www.almaden.ibm.com/cs/people/ichuang

  You can buy this book on-line from `amazon.com`. *There is a link that points directly to the appropriate shopping page at* `www.squint.org`.

- John Preskill and Alexei Kitaev, "Lecture Notes for Physics 229, Quantum Information and Computation" [87]

  http://www.theory.caltech.edu/people/preskill/ph229

  This is a rather high flying and ambitious text. Perhaps even the best I have seen in this area so far. The course is given twice weekly, 90 minutes at a time, over two semesters. Provided notes are divided into 6 large chapters, and my impression is that the authors still didn't manage to cram all they wanted to convey to their audience in the time available, so that only the last chapter, 6th, is dedicated to Quantum Computing, with the preceding chapters providing background. But the background

> *they provide is very thorough. The course is addressed to (Caltech) physicists, mathematicians, computer scientists and engineers. It is a highly recommended reading for those, who would like to study the subject in some depth.*

- Colin P. Williams and Scott H. Clearwater, "Explorations in Quantum Computing", Springer Verlag, 1998, IBSN 0-387-94768-X, 307 pp., CDROM included [104]

  > *This neat little text provides a very basic introduction to Quantum Computing. Quantum Mechanics material itself is reduced to an absolute minimum. Computer Science and computational aspects, on the other hand are treated in more depth. The book provides some simulation examples and illustrations that make its rather difficult subject quite palatable. There are some errors in the book though, and their effect is such that some "derivations" accompanied by a fair amount of hand-waving are totally incomprehensible. It is at times like these that you learn to appreciate more rigorous approach.*

In addition students who need to catch up on quantum mechanics should read

- Richard P. Feynman, Robert B. Leighton and Matthew L. Sands, "The Feynman Lectures on Physics", Addison-Wesley, 1989, Volume 3, "Quantum Mechanics" [35]

  > *This last volume of* Feynman Lectures *is seldom used or even recommended in introductory physics courses. One of the reasons is that it introduces the Schrödinger wave equation pretty late and instead devotes something like 2/3rd of its initial material to study of simple finite quantum mechanical systems, or, in other words to what some people call* matrix mechanics. *But this is just what we need for our course, so if you haven't studied quantum mechanics in the past at all, this is a very good place to start. If you have, but haven't progressed much beyond scattering of a Schrödinger wave against a rectangular barrier or a well, this is again a place to go to. If your course was very heavy on functional analysis, linear combinations of atomic orbitals, and the like, but quite lightweight on basic foundations, you may find Feynman's volume 3 quite an eye opener too.*

The following are books, which I found useful, insightful and interesting in many ways and which are going to enrich students' understanding of the topic:

- Arno Böhm, "Quantum Mechanics", Springer-Verlag, 1979, 522 pp. [15]

  > *A beautiful, precise and quite thorough text/monograph about quantum mechanics, based on a course that the author taught at*

the University of Texas at Austin. Not for the beginners. Covers foundations, oscillators, rotators, angular momenta and the Wigner-Eckart theorem, Kepler problem, perturbation theory, spin, multielectron systems, time evolution, the Stern-Gerlach experiment and the measurement theory, transitions, elastic and inelastic scattering, resonances, time reversal, and decay of unstable systems.

- Asher Peres, "Quantum Theory: Concepts and Methods", Kluwer Academic Publishers, 1993, ISBN 0-7923-2549-4, 446 pp. [86]

  This is a splendid text for those who want to understand quantum theory better, as opposed to just manipulating its mathematical formulas thoughtlessly, which does *happen more often than you would think.*

  The book introduces formal tools of quantum mechanics with great precision but without being excessively abstract. The physical interpretation is rigorous: no use is made of the uncertainty principle and other ill-defined notions. *The book provides one of the best discussions of Bell's theorem and then goes on to some of the most interesting topics of current research: spacetime symmetries, quantum thermodynamics, quantum information theory, irreversibility, quantum chaos and measurement theory.*

- Josef M. Jauch, "Foundations of Quantum Mechanics", Addison-Wesley, 1968, 299 pp. [52]

  This text concentrates on logical foundations of quantum mechanics and on quantum logic. As such it should appeal to logicians and computer scientists interested in quantum computing. On the other hand it predates our present day interest in quantum computing and even the very concept of a qubit. Consequently, quantum logic as presented by Jauch is not immediately transferable onto the much more precise and simpler world of qubits. Nevertheless the book contains a number of important insights and may prove to be of some assistance.

- David Bohm and B. J. Hiley, "The Undivided Universe: An Ontological Interpretation of Quantum Theory", Routledge, 1993, 397 pp. [14]

  Like Giordano Bruno, David Bohm was a heretic and a martyr. Following de Broglie and Schrödinger he concocted a very interesting and insightful interpretation of Quantum Mechanics that attracted wrath and ire of orthodox physicists, whose usual approach to mystery and weirdness of Quantum Mechanics used to be to sweep it all under the carpet and resort to a dogma.

> *Luckily in 1964 a British physicist, John Stewart Bell, demonstrated that some of the most intriguing questions about the interpretation of Quantum Mechanics can be investigated experimentally. Then in 1982 Aspect, Dalibart and Roger carried on experiments suggested by Bell that demonstrated* non-locality *of quantum physics, and quantum physics was never the same again.*
>
> *Today quite serious folks, for example John Preskill (see above), ask insightful questions about "where do quantum probabilities come from" and "why does a quantum measurement select a single basis state from a superposition thereof", and old Bohm laughs in his grave and says "I told you so".*
>
> *In short some of Bohm's insights bring home the bacon, and for this reason I'm going to bring them to this lecture occasionally, although like Bohm himself I'm quite ready to acknowledge that not all in the presented picture may be quite right.*
>
> *There are two reasons why some of this stuff is of relevance to quantum computing. The first one is quantum teleportation, and the second one is that in quantum computing information is stored on a multi-qubit system, and a system like that is described in terms of a tensor product on Hilbert space. This, in turn, implies non-local and* anti-relativistic *interactions, and those bring us right back to "The Undivided Universe".*

The following texts provide a standard introduction to present day electronics. Since quantum computing is currently at the stage where elementary electronics was 40 or even 50 years ago, these texts should help students understand the basic differences between quantum computing and conventional computing technologies, as well as gain a better understanding of how computing is actually done in real life – and hence, what really matters as opposed to what may be just a fancy decoration.

- Richard Dalven, "Introduction to Applied Solid State Physics", Plenum Press, 1981, 330 pp. [27]

  > *This book provides a good discussion of physics behind present day computing: pn-junctions, MIS junctions, MIS devices. But for us the most important is chapter 8 that talks about Josephson junctions and devices based thereon. Josephson junction emerges as a leading technology for the HTMT Petaflops Computer, see*
  >
  > *http://www.sc99.org/proceedings/invtalk.htm#sterling*
  >
  > *for some background, and a leading technology for solid state implementation of a qubit, see Nakamura* et al. *[79]*

- James J. Brophy, "Basic Electronics for Scientists", McGraw-Hill Kogakusha, 1977, 430 pp. [21]

This text will lead you all the way from the basic principles of solid state device physics to simple computational circuitry: flip-flops, counters and registers. It should be helpful at this very early stage in quantum computing in figuring out and understanding how to assemble simple quantum logic circuitry. The text will also bring home the fact that there is plenty of quantum mechanics in present day classical computing too.

1.5 The Notes

These lecture notes are *not* a book. They are what you would write down in your own notebook if you were to attend a traditional lecture of this nature. For your (and our) convenience I have placed these notes on-line.

The notes are based on numerous sources. And so

- Chapters 2 and 3 are based on Dalven [27], Brophy [21] and Williams and Clearwater [104].
- The first 3 sections of chapter 4 are based on Feynman [35] with some minor modifications of our own (we replace the Stern-Gerlach apparatus with a birefringent crystal), whereas the section about Berry phase is based on Berry's original paper [12].
- The section about multiparticle systems is based initially on Feynman [35], then it briefly switches to Nielsen and Chuang [80]. Section about stopping and reversing time is based on paper by Leung, Chuang, Yamaguchi and Yamamoto [67]. The discussion of Feynman computer is taken from Williams and Clearwater [104] as is the section about nonlocality, which is further enhanced by material taken from a paper by Pan, Bouwmeester, Daniell, Weinfurter and Zeilinger [82].
- The sections about the measurement and interaction with the environment are based on Preskill and Kitaev [87], but their content is really condensed to the bare minimum compared to what you'll find in the original source. The part that talks about NMR measurement is based on Nielsen and Chuang [80].
- Chapter 5 is based mostly on Preskill and Kitaev [87] with some borrowings from Nielsen and Chuang [80], and chapter 6 is based mostly on Nielsen and Chuang [80], though the introduction to the chapter is based on Williams and Clearwater [104]. The brief discussion of decoherence-free spaces is based on a paper by Kielpinski, Meyer, Rowe, Sackett, Itano, Monroe and Wineland [59].
- Various asides in the notes are derived from all over the place, with the main sources being perhaps Bohm and Hiley [14], and Peres [86], followed by recent papers and even press clippings.

1.6 Computer Simulations

You will find a number of quantum computing and quantum teleportation simulations, as well as other auxiliary codes designed for Mathematica in

/afs/ovpit.indiana.edu/common/mathematica/quantum_computing

This area is restricted only to users with valid AFS tokens in the AFS cell `ovpit.indiana.edu`. The material contained therein is copyrighted [104]. There are three directories there: `mac`, `unix`, and `windows`, which contain the following Mathematica notebooks for Mathematica Version 2.2 and Mathematica Version 3.0

BraKet Defines basic operations on bras and kets, direct product, nuts and bolts QM

ErrorCo Simulation of Quantum Error Correction

Feynman Simulation of Feynman's Quantum Computer

Interfer Graphical Illustration of Interference Effect

OTPExamp Example of a Provably Secure Cryptosystem

QBugs Code for Monte Carlo Analysis of Error Propagation

QCdata Database of References in Quantum Computing

QCrypt Simulation of Quantum Cryptography

RSAExamp Example of RSA Public-Key Cryptography

SearchEn A Search Engine for our Quantum Computing Database

ShorFact Simulation of Shor's Algorithm

Schroe2D Animation of Particle Impinging a Double Slit (Interference)

Teleport Simulation of Quantum Teleportation

TimingFa Illustration that Factoring is a Hard Problem Classically

Indiana University Distributed Storage Systems Group can provide AFS clients for all commonly used UNIXes, Linuxes, Windows NT and 2000, and for MacOSX.

1.7 Conference Announcements

April 22-25, 2002 2002 Applied Computational Research Society Joint Meeting, Computational Micro and Nano Technology, San Juan, Puerto Rico

April 1-5, 2002 Quantum Computing III - Part of SPIE's 16th Annual International Symposium on AeroSense, Orlando, Florida

September 25 - October 2, 2002 Perspectives in Decoherence Control and Quantum Computing, Center for Theoretical Physics, Ann Arbor, Michigan (Workshop)

January 14-17, 2002 QIP 2002, The Fifth Workshop on Quantum Information Processing, IBM, T. J. Watson Research Center, Yorktown Heights, New York

1.8 Scholarship Announcements

The National Science Foundation has funded graduate fellowships at Indiana University for students interested in quantum computing. Each three year fellowship carries $18,000 stipend, a tuition waiver, health insurance, and no teaching duties. The fellows will be working with professor Zhenghan Wang on applications of topology and knot theory to theoretical computer science. The ultimate goals of quantum computing research are developing quantum algorithms that are exponentially faster than the algorithms of classical computer science and building scalable quantum computers. To apply for this fellowship submit a regular application to the Department of Mathematics at Indiana University and in your personal statement indicate your interest in this kind of research. For more information send e-mail to *zhewang@indiana.edu*.

Chapter 2

Physics of Computation

2.1 From Counters to Qubits

Consider a simple 4-bit counter shown in Figure 2.1.

The basic computational elements in the counter are JK flip-flops. A typical JK flip-flop is based on an RS master-slave flip-flop with *clear* and *set* and comprises 9 NAND gates, as shown in Figure 2.2.

Perhaps the simplest way to make a NAND gate is to use 2 transistors as shown in Figure 2.3.

Now, what's inside a transistor? Figure 2.4 shows a cross-section of a *Metal-Oxide-Semiconductor Field Effect Transistor*, or MOSFET for short.

The idea here is to modulate the flow of electrons from the *Source* to the *Drain* by applying a signal to the *Gate*. In particular a MOSFET can be used like a switch. A negative potential can be applied to the gate that will repel all electrons from the region just under the oxide. Without carriers there can be no current. Alternatively if a sufficiently high positive potential is applied to the gate it will attract electrons from the bulk to the region under the oxide.

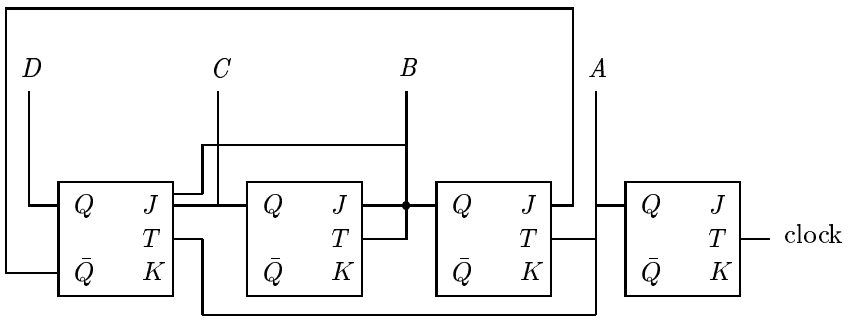


Figure 2.1: An asynchronous binary-coded decimal 4-bit counter (from Brophy [21]).

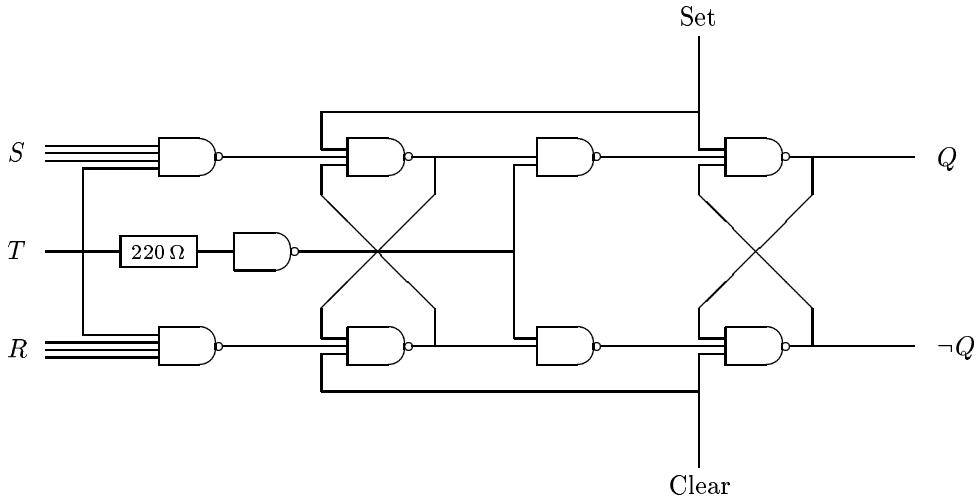


Figure 2.2: RS master-slave flip-flop (from Brophy [21]). Connecting $\neg Q$ to one of the S inputs and Q to one of the R inputs results in a toggle action. The remaining S inputs are renamed to J and the remaining R inputs are renamed to K. The resulting configuration is then called a *JK master-slave flip-flop*.

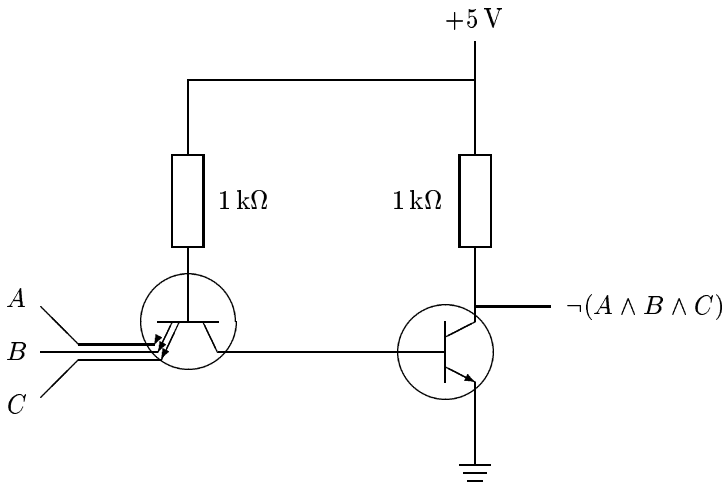


Figure 2.3: Transistor-transistor logic implementation of a NAND gate (from Brophy [21]).

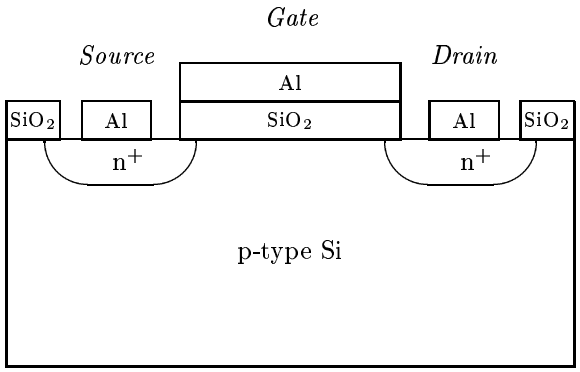


Figure 2.4: Schematic view of a MOSFET (from Dalven [27]).

A very thin electron rich *inversion layer* will form there and a current will flow from the source to the drain.

This device is based on the notion that in a single crystal of Silicon you can have regions with different types of conductivity. The n-type regions under the source and the drain are obtained by diffusing or implanting Phosphorus in Silicon. The p-type bulk is Boron rich. What this does to conductivity is that in the n-type region current is conducted by electrons, whereas in the p-type region current is conducted by holes. Holes and electrons in crystals are very similar to positrons and electrons in vacuum. The mathematics that describes the former and the latter is very similar.

How does all this come about?

A relatively simple quantum mechanical computation tells us that as you bring together more and more atoms, their energy levels split more and more finely. Ultimately in a crystal, where atoms of Silicon are only about 5 Å apart, you end up not with energy levels, but with energy bands. In particular there are two bands there of special importance: the valence band and the conductivity band. Electrons in the valence band are still attached to their atoms. But electrons in the conductivity band are like ionized electrons in vacuum: they can move freely around the crystal.

But quantum mechanics gives us more than the bands. Quantum mechanics gives us also a very peculiar distribution function that describes the proportion of electrons at each energy in a crystal. This distribution function is due to Fermi and Dirac and it looks as follows:

$$f(E) = \frac{1}{e^{(E-E_F)/(kT)} + 1} \tag{2.1}$$

For energies that are much higher than kT and E_F, which is referred to as the *Fermi level*

$$f(E) \approx e^{-E/(kT)} \tag{2.2}$$

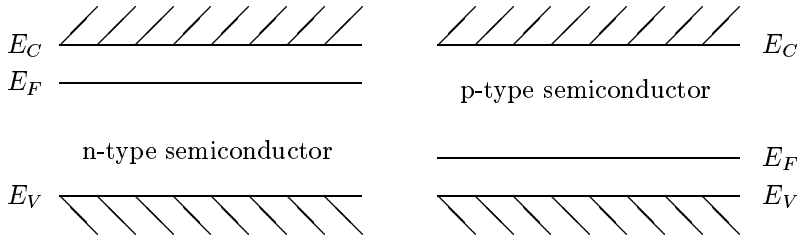


Figure 2.5: A schematic diagram of energy bands in n- and p-type semiconductors (from Dalven [27]).

which is the Boltzmann distribution. But for energies that are much lower than E_F the Fermi-Dirac distribution attains a constant value.

If the Fermi level happens to be positioned in the conductivity band, then it means that the conductivity band is full of electrons. Such materials conduct current very well. All metals belong in this category.

If the Fermi level is positioned in the middle of the energy band gap, and the gap between the valence band and the conductivity band is very wide, i.e., wider than several kT, then the material is an insulator.

If the Fermi level is positioned in the valence band, then the valence band is full of holes. A material like that conducts current very well too, although there are no metals in this category. A silicon wafer very heavily doped with Boron may behave like that.

But if the gap is relatively narrow compared to kT and the Fermi level is positioned within the gap, then depending on whether it is closer to the conductivity band or to the valence band, you will find either electrons contributed to the conductivity band (in an n-type semiconductor) or electrons *withdrawn* from the valence band (in a p-type semiconductor), which leaves holes.

Figure 2.5 shows the band structure and the Fermi level in a semiconductor.

Holes like electrons can move freely within the crystal. Both holes and electrons have an *effective* mass, which is usually different from the electron mass in vacuum. Furthermore effective mass and mobility may differ for electrons and holes in the same material, which implies differing conductivities in n- and p-type regions.

All this stuff is quantum mechanical. And yet, devices that are based on the *diffusion* of electrons or holes within a crystal implement classical logic. How come? The reason for this is that in a typical semiconductor material there is a very large number of electrons and holes involved in transmitting and storing information. A typical electron density in an n^+ region may be $10^{18}\,\mathrm{cm}^{-3}$, sometimes even $10^{19}\,\mathrm{cm}^{-3}$ (such semiconductors are said to be *degenerate*). Quantum systems that comprise a very large number of constituents and interacting frequently with the environment and with each other lose their distinct quantum character and all that we end up with is a classical diffusive flow of electrons or holes from the source to the drain.

What are typical dimensions in a MOSFET today? You may often hear expressions such as "0.18 micron technology". What this means is that the tiniest details that can be produced on a wafer are $0.18\,\mu$m wide. In the $0.18\,\mu$m technology the length of the channel that links the source with the drain is going to be somewhat more than $0.18\,\mu$m. Perhaps twice that. This level of miniaturization is currently encountered in only the best and newest devices, for example the Intel 1 GHz Coppermine chip, which was demonstrated in February 2000, featured $0.18\,\mu$m interconnects [44]. More commonly it is going to be something like $0.32\,\mu$m or $0.28\,\mu$m. $0.18\,\mu$m is $1,800\,$Å, which is 360 Si lattice constants (Si lattice constant is $\approx 5\,$Å). The thickness of the silicon dioxide layer under the gate is somewhat less, e.g., $200\,$Å (40 Si lattice constants). It cannot be made too thin for two reasons:

First If the oxide is too thin, electrons may tunnel through it, thus shortcircuiting the gate.

Second The capacitance of a two plate capacitor, and a gate in a MOSFET is an example of such, is inversely proportional to the thickness of the gap between the plates of the capacitor, i.e., the thinner the gap, the greater the capacitance:

$$C \propto \frac{1}{d} \tag{2.3}$$

But the time it takes for the capacitor to discharge is proportional to the square root of the capacitance:

$$\tau \propto \sqrt{C} \propto \sqrt{\frac{1}{d}} \tag{2.4}$$

This means that the thinner the oxide, the slower the device.

The capacitor discharge time also depends on the length and the width of the gate: the smaller the area, the shorter the discharge time. Also the shorter the channel, the less time electrons need in order to cross it.

There is a thin charge depletion layer between the n^+ and p-type regions. The width of the depletion layer is given by:

$$W = \sqrt{\frac{\epsilon kT}{2\pi e^2}\left(\frac{N_a + N_d}{N_a N_d}\right)\log\left(\frac{N_a N_d}{n_i^2}\right)} \tag{2.5}$$

where ϵ is a dilectric constant of Si, e is the elementary charge, kT is ambient temperature in Joules, N_a and N_d are fixed charge densities in the depletion layer, and n_i is the intrinsic charge density. For very pure silicon with N_a and N_d of the order of $10^{15}\,\mathrm{cm}^{-3}$, $W \approx 1\,\mu$m, and it can be made narrower with increased doping. In a typical modern MOSFET the depletion layer would normally stretch all the way from the source to the gate.

How much can we shrink the gate and still expect a classical operation? As the gate gets shorter, the first thing we'll notice is that instead of diffusing

electrons may begin to move ballistically between the source and the drain. This is actually quite good, because ballistic transport is bound to be faster than diffusive transport. At the International Electron Devices Meeting in December 1999 Bell Laboratory researchers demonstrated a MOSFET with a 500 Å gate (100 Si lattice spacings), although 248 nm (2,480 Å) lithography was used to create most of the device structures. For a gate that short, ballistic effects were clearly visible, and the transistor demonstrated performance much better than would be the case with diffusive electron transport only. The oxide deployed in the device was extremely thin. The researchers tried 13 Å and then 16 Å oxide. Oxide that is so thin leaks because of quantum tunneling. Consequently the transistors were not very efficient. However at 16 Å the leakage across the oxide was still tolerable [43].

But if the gate shrinks by a yet another order of magnitude, e.g., to 50 Å, electrons may begin to tunnel between the drain and the source. At this stage the device ends up leaking in every direction, and ceases to operate as a switch. There is a physical limit to how far you can shrink those devices and still expect them to work according to classical physics.

> The limits that plague MOSFETs can be overcome by switching to quite different technology, while still preserving the basically classical functioning of a transistor. On the 27th of August 2001 IBM demonstrated a carbon nanotube based flip-flop [25]. Carbon nanotubes are only 10-atoms wide. But they can be quite long. They are relatively easy to produce too. Certain chemical processes result in nanotube precipitation. This precipitation is then captured onto soapy surfaces. The trick is in assembling them to form useful electronic circuits. Once perfected this technology should yield at least a 10,000-fold improvement over the best electronic circuitry of today. Consequently, if today's state of the art PC is a 2GFLOPS machine with about a GB memory and, say, 80GB hard drive, a nanotube based PC should deliver 20TFLOPS, it should have 10TB memory and 160TB disk space. Do not laugh. We may well see this technology sooner than you'd expect!

On the other hand, going down into the quantum domain, and changing the computational paradigm, at least hardwarewise (e.g., to hell with MOSFET) may open new and entirely different ways of doing computing. And this is exactly what this lecture is going to be about.

> There is a difference between what is currently called *molecular computing* and quantum computing, even though both rely on using individual molecules or atoms to accomplish various operations. Molecular computing as understood today is usually classical computing, where MOSFET switches are replaced with molecular switches.
>
> An example of a very simple molecular computer is a catenane, an organic molecule composed of two interlocking rings. It is possible to make one of the rings move between two different states, for example a different angle with respect to the other ring. The trigger for the change can be

either electric or optical. This device is a simple molecular switch that can replace a MOSFET. Catenanes were invented by J. Fraser Stoddart from the University of California Los Angeles (UCLA). The demonstration of switching in catenanes was performed by Fraser Stoddart and James Heath in late August 2000 [48].

While such devices accomplish ultimate in miniaturization, the logic that they implement is still classical logic, and no advantage is taken of quantum parallelizm, about which more below.

A good review and comparison of molecular, quantum and membrane computing can be found in "Computing with Cells and Atoms: An Introduction to Quantum, DNA and Membrane Computing" by Cristian S. Calude and Gheorghe Paun [22].

Another novel approach to computation is called *spintronics*. Spintronics can be classical or quantum. Here the idea is that instead of associating information with electric charge, as is the case for *electronics*, we could associate it with electron spin instead [24]. Once loaded with information spin packets could be moved between semiconductors, much the same as we move charge packets today in order to transmit information. That this is possible has been demonstrated recently by four researchers from Santa Barbara, who demonstrated a persistent spin conduction in biased semiconductor heterostructures [73].

There is a laboratory device called Molecular Beam Epitaxial Reactor ("MBE" for short). Using this device it is possible to grow extremely thin and very highly controlled layers of various materials on semiconductor wafers. These layers can be only one lattice spacing thick, and fully conforming to the underlying crystal structure. This works especially well with GaAs and AlGaAs. The former is a semiconductor with a fixed energy band gap, but depending on the proportion of Al to Ga in the latter the width of the energy band gap in AlGaAs alloy can be varied continuously. By alternating GaAs and AlGaAs, it is possible to grow a very thin layer of GaAs sandwiched between two layers of AlGaAs (see Figure 2.6). The GaAs layer may be only one lattice spacing thick. The insertion of that layer between AlGaAs regions produces a very thin, sheet-like potential well also shown in Figure 2.6. That well can be so thin that only one electron may fit into it vertically. The resulting structure is a two dimensional electron sheet. Electrons in such a sheet have a very high mobility. It is possible to make field effect transistors, in which the two dimensional quantum sheet of electrons is placed right under the gate instead of an induced channel (as in MOSFETs). Such devices are called High Electron Mobility Transistors (HEMTs) and are used in microwave electronics (which is one to two orders of magnitude faster than digital microelectronics).

Such 2-dimensional electron sheets have many other marvellous properties. For example tiny electron-fluid vortices may form in them if the sheets are immersed in very high magnetic fields and cooled. Localized electrons can sometimes become trapped in those vortices: they sit right in the middle of a vortex, where normally there shouldn't be any electrons at all. When this happens the combined electron-vortex system acquires bosonic properties – such systems are referred to as *composite fermions*

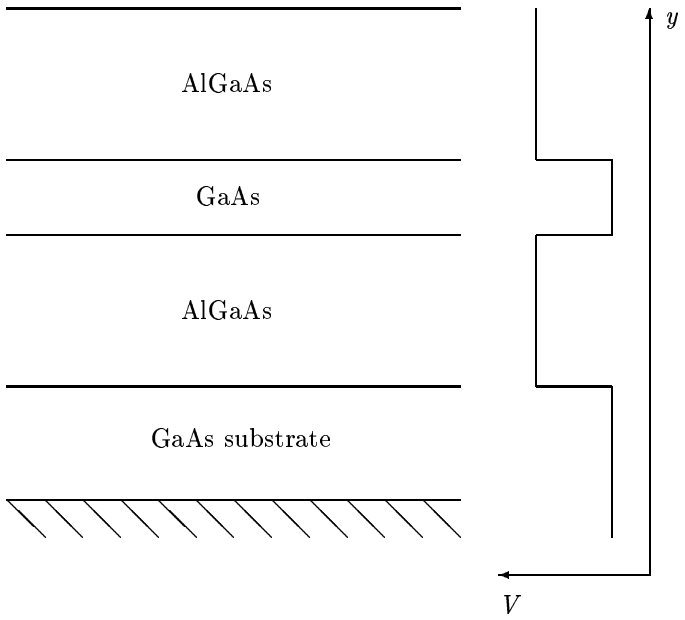


Figure 2.6: A quantum sheet structure built of GaAs and AlGaAs. The GaAs layer sandwiched between the AlGaAs layers may be only one lattice spacing thick. The plot on the right shows the potential well formed by the sandwich.

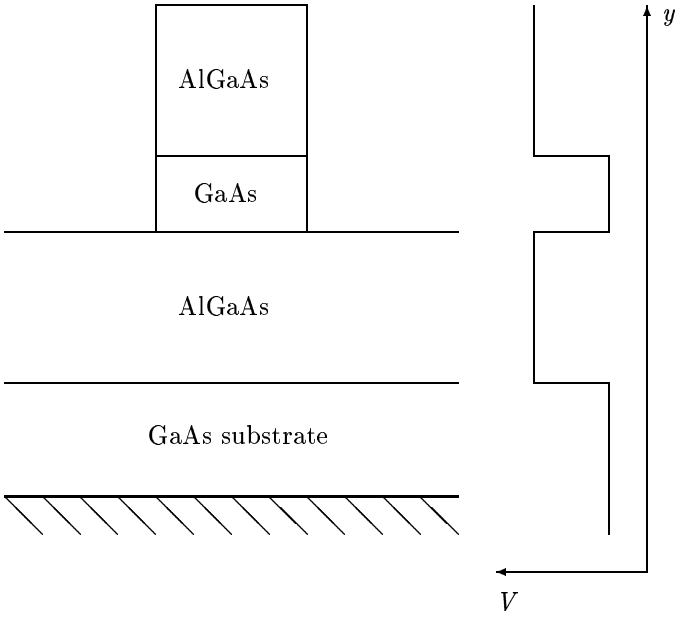


Figure 2.7: A quantum-line or a quantum-dot structure implemented in GaAs and AlGaAs.

> – and a peculiar kind of superconductivity may arise as the result [47]. Quantum sheets with vortices are of great interest to some quantum computing researchers, who believe that the topological character of the vortices may be more suitable for quantum computing devices that dynamic systems such as spins or photons [36], [37].
>
> Recently a group of Japanese researchers published a paper in Nature where they claimed to have resolved and imaged individual vortices of this type in $Bi_2\,Sr_2\,Ca\,Cu_2\,O_8$ thin films. The vortices were trapped by columnar defects produced by irradiation of the film by heavy ions [99].

Using electron-beam or X-ray lithography, we can pattern the quantum sheet shown in Figure 2.6 and produce either a quantum line or a quantum dot (see Figure 2.7). A single electron may become trapped in a quantum-dot and manipulated in various ways. Such devices are of great interest to physicists. They can be used to test the most basic assumptions of quantum mechanics in ways that the fathers of quantum mechanics themselves could only dream and hypothesize about.

But a quantum dot can be also used for quantum computation, because information can be stored on the single electron trapped within the quantum dot and manipulated together with the electron. Such a single isolated electron is called a *qubit*.

In practice a slightly more complex configuration of quantum dots has to be

used to implement a single qubit. Daniel Loss and David P. DiVincenzo from the Institute for Theoretical Physics of the University of California, Santa Barbara, demonstrated recently how a universal set of one- and two-quantum-bit gates can be constructed from quantum dots [71].

This lecture course will not progress much beyond what you can do with just a couple of qubits. Quantum computing today is at the point where classical computing was 70 years ago. We ponder on how to assemble very simple systems that can perform some very simple computations. We ponder on the nature of those computations. And we ponder on how to extract computational results from such a system. Reading and interpreting results of quantum computations is a much more complex process than is the case with classical computers.

Why should we climb this mountain at all? A simple answer is "because it is there". Very simple quantum computers can be made today and they *have* been made already. A small number of intriguing quantum algorithms was demonstrated. The potential applicability of quantum computers to attack quantum physics and quantum chemistry problems is undeniable: it derives right from physics itself. But most importantly, this is a completely new, open, and still quite unploughed research area. What else can be done with quantum computers? How can practical quantum computers be built? These are all open questions.

> It is possible to imitate certain important features of Quantum Computing with classical light beams. In May this year (2001) Ian Walmsley, professor of optics at the University of Rochester, demonstrated that "... if you have a quantum computer that is based entirely on quantum interference, we can build you a computer that is equally efficient, based entirely on light interference. And light is a whole lot easier to manipulate than quantum systems." What really distinguishes Quantum Computing is *not* superposition and interference, but entanglement. It turns out that some algorithms, e.g., database searches, which were thought of as quantum, do not rely on entanglement, and can therefore be run just as efficiently on a classical optical computer.

2.2 Thermodynamics of Computation

Computers are machines and like all machines they are subject to thermodynamic constraints. It costs a lot of time and energy to shape, maintain, and then move around a digital signal. This is why analog devices can be much faster, while generating less heat, than digital devices, and why analog computers, in particular, can be a lot faster than digital computers.

So why don't we use analog computers? Speed, as it turns out, is not everything. Errors and inaccuracies, for example, can be controlled within a digital computer much more precisely than in an analog computer. Digital computers are much easier to program too.

Can digital computers be improved so as to minimize production of heat?

Could heat generation be eliminated entirely? It is clear that there is something very fundamental missing from the Turing Machine model, because it says nothing about it.

It turns out that it is possible to think (in agreement with the laws of physics) of an *ideal* computer that would be somehow capable of shaping, maintaining, and moving around digital signals without any heat generation. There is one place, however, where heat *must* be produced. Whenever information is *erased* the phase space associated with the system that stores that information must shrink.

Information must be written somehow on a physical medium: whatever that medium may be. Processing information occurs by the means of manipulating the system on which the information has been written. Erasing a single bit of information reduces entropy of the system that stored that information by at least $\Delta S = k \log 2$. This reduction of entropy results in heat transfer to the environment. This result is due to Rolf Landauer, 1961 (also see Landauer's paper in Physics Today, "Information is Physical", [64]).

But this, apparently, is the only place, where an ideal computer wastes energy. So if we were to construct a computer that does not erase any information, such a computer could work without generating any heat at all. Of course, in reality the computer would still generate a lot of heat. Electric pulses moving along copper wires would have to work their way against resistance. Electrons diffusing from a source would still collide with crystal imperfections and with electrons in the drain, again, generating heat. But, at least in a fantasy world, you could replace all wires with superconductors, you could use ideal crystals without any imperfections and then the only place where you would still have to generate some heat would be whenever and wherever you erase information.

NOR, AND, NAND, and XOR gates are all irreversible: they must generate heat. Amount of information on the right hand side of

$$(a, b) \to \neg(a \wedge b) \tag{2.6}$$

is less than the amount of information on the left hand side.

But it is possible to perform any computation using only reversible steps (this result is due to Charles Bennett, "Logical Reversibility of Computation", 1973, [9]). Special gates have been conceived and fabricated (e.g., Toffoli gates, [98]), which maintain all information that is passed to them, so that the computation can be run forward and backward. The computation results in a very large amount of junk, because every intermediate step is remembered, but heat generation is eliminated while the computation goes on. After the computation is finished the computation can be run backwards in order to restore the initial state of the computer and avoid its spontaneous combustion.

Amongst the reasons why quantum computing attracted so much interest recently is that quantum computation is reversible (but not the read-out of the results of that computation). It is therefore possible, at least in principle, to carry out quantum computation without generating heat.

2.3 Quantum versus Classical Information

Information and logic, which is the art of managing information, cannot exist in detachment from the laws of nature. In particular information must be written on some physical substance, be it neural connections in our brain, or paper, or magnetic media, or electrons trapped in quantum dots, or a beam of photons. Physics of the media that is used to store information determines logic that can be used to manipulate that information. Classical media, e.g., transistors, magnetic domains, or paper determine classical logic as the means to manipulate that information. Electric charge either is stored in a particular CCD location or isn't. A magnetic domain on the drive either is aligned with the direction of the head or isn't. Things are either true or false, and any memory location can be read without destroying that memory location.

Not so in the world of quantum physics.

1. Quantum information cannot be copied with perfect fidelity. The latter was demonstrated by William Wootters (Williams College, Massachusetts), Wojtek Zurek and Dennis Dieks in 1982 [87]. This implies that quantum information cannot be *read* in its entirety either, because the process of *reading* is the same as the process of *copying*.

2. Quantum information can be *transferred* with perfect fidelity, but in the process the original must be destroyed. Quantum teleportation was first described by Bennett, Brassard, Crepeau, Jozsa, Peres, and Wootters in 1993 [10]. It was demonstrated experimentally in February 1998 by Nielsen, Knill, and LaFlamme [81].

3. Any measurement performed on a quantum system destroys most of the information contained in that system leaving it in one of the so called basis states. The discarded information is unrecoverable.

4. Although in some cases it is possible to predict with certainty which basis state a quantum system will end up in after the measurement, in general probabilistic predictions can be made only.

5. Certain observables cannot simultaneously have precisely defined values. This affects our ability to define initial conditions at the beginning of a computation, and further affects our ability to read the results.

6. Quantum information can be encoded (and usually is) in non-local correlations between the different parts of a physical system.

2.4 Quantum Communication

In a classical computer information is moved between processing elements (including memory and peripherals) using metal wires. Those wires are often printed on plastic boards or on oxidized silicon wafers. But, by and large,

they're the same wires that have been used for electric telegraphy since its invention by Sir William Fothergill Cooke, Sir Charles Wheatstone and Prof. Samuel F. B. Morse of the University of the City of New York in 1837. All three submitted patents for an electric telegraph in that year.

> The British telegraph, invented by Cooke and Wheatstone utilized 6 wires which actuated five needle pointers attached to five galvanoscopes at the receiver end. It was a very clumsy device and, not surprisingly, was never really deployed "in production". The Morse telegraph, however, was very practical. Soon after its invention it attracted attention of the US government, which financed its experimental deployment between Washington DC and Columbia in 1843. The telegraph was commissioned on the 24th of May 1844. The first message transmitted over the telegraph said "What hath God wrought!".
>
> The principles behind the Morse telegraph are very similar to the principles employed in digital communications. The Morse code was effectively the first binary code and it was used for the next 100 years all around the world in a very much the same way we transmit data over the Internet today.
>
> We can well say then that Internet was invented 163 years ago by Samuel Morse, and not by Al Gore. What we have seen since have been numerous and gradual improvements to the original idea and hardware, but the basic principle remains unchanged: information is transmitted over a wire as a sequence of digital pulses.

Long haul communications and sometimes high bandwidth local communications between computers or computer components also utilize optical waveguides, e.g., optic fibres. This technology dates back to 1950s. It was used originally for endoscopy, i.e., looking at stomach, lungs, and things at the other end. Optic fibres were first used for telecommunication in 1966 in Britain by Kao and Hockham (two electrical engineers working for British Telecom). Today nearly all long haul, i.e., interstate and intercontinental, data transfers take place over optics.

There is a substantial physics difference between moving data over copper wires by the means of diffusing massive electrically charged particles such as electrons and moving data by the means of massless and electrically neutral photons. Yet, once the information enters an information processing system at the receiver, it is transferred from photons to electrons, and from this point onwards it is moved around the old fashioned way, i.e., over telegraph wires. Purely optical information processing is in *statu nascendi* and hasn't left research laboratories yet.

But how do you move *quantum* information, which is contained in the wave function of a quantum system, not in a charge or in a pulse, between processing elements of a quantum circuit?

The answer is *teleportation*.

In the November 25th issue of Nature (1999), pp. 390–393, Daniel Gottesman and Isaac Chuang showed that a universal quantum computer can be built

by combining single qubit operations, teleportation, and Greenberger-Horne-Zeilinger states [41]. Moreover, they also present a systematic construction for an infinite class of *fault tolerant* quantum gates.

We will talk about a teleportation circuit (suggested by Brassard in 1996 [18]) in Chapter 4.5.8.

The simulation code for the circuit lives in

/afs/ovpit.indiana.edu/common/mathematica/quantum_computing/unix/nbooks3/teleport.nb

So, what is quantum teleportation?

I have already remarked that multiparticle quantum systems are described in terms of tensor products on Hilbert space, and that this implies existence of non-local interactions between components of a quantum system. In principle the whole universe is *entangled* and you cannot take a chunk of it and isolate it from the rest. Even particles that are on two opposite sides of the universe are connected by entanglement. This interaction is not mediated by any conventional field known to physics such as electromagnetism or gravity. It is instantaneous and in clear violation of special relativity. The latter can be restored for the so called "expectation values", i.e., measured quantities, but the anti-relativistic correlations are still there. Bohm [14] demonstrates that these non-local interactions can be described in terms of a very special anti-relativistic *quantum information field* that does not diminish with distance and that binds together the whole universe. This field is not physically measurable though and the only way in which it manifests itself is through the non-local correlations. So it is, at least at this stage, a matter of religion whether you want to believe in it or not. But the idea is interesting and derivable *entirely* from the Schrödinger equation.

Anyhow, whether you want to describe the non-local correlations in terms of Bohm's field or magic, they are there. Their existence was demonstrated experimentally by Aspect, Dalibard and Roger in 1982 [1], and predicted by Bell in 1964 [7]. But they manifest themselves also in superconductivity, superfluidity, and even in the Bohm-Aharonov effect. The first two are macroscopic phenomena, and in the Aspect, Dalibard and Roger experiment, the existence of non-local correlations is demonstrated over a distance of 12 m. More recently non-local correlations were demonstrated over a distance of some 20 km.

Quantum teleportation uses this non-local interaction, combined with a classical information channel (e.g., telephone wires) in order to transfer a quantum state, intact, from one location to another one. How exactly this is done will be explained in section 4.5.8.

2.5 Faster, Smaller, and Cheaper: Quantum Computers

2.5.1 Heteropolymers

The first heteropolymer based quantum computer was designed and built in 1988 by Teich, Obermayer and Mahler from Institut für Theoretische Physik at the University of Stuttgart [96], and then improved by Lloyd in 1993 [69].

In a heteropolymer computer a linear array of atoms is used as memory cells. Information is stored on a cell by pumping the corresponding atom into an excited state. Instructions are transmitted to the heteropolymer by laser pulses of appropriately tuned frequencies.

Because of dipole interactions between atoms in a heteropolymer energy levels of different atoms in the molecule are slightly different. It is therefore possible to send a laser pulse addressed to a specific atom or a group of atoms in the heteropolymer chain.

By sending a very carefully designed sequence of pulses of various colours, shapes, and durations it is possible to accomplish feats such as transmission of a single classical bit (or a qubit state, which is not in a superposition), from one atom to another one along the chain. But superpositions of multiatom states can be constructed too, and at this stage our heteropolymer becomes a true quantum computer.

The nature of the computation that is performed on selected atoms is determined by the shape and the duration of the pulse.

After the computation has completed answers are read by the means of resonance fluorescence.

2.5.2 Ion Traps

An ion trap quantum computer was first proposed by Cirac and Zoller in 1995 [26], and implemented first by Monroe and colaborators in 1995 [77], and then by Schwarzschild in 1996 [89].

The ion trap computer encodes data in energy states of ions and in vibrational modes between the ions. Conceptually each ion is operated by a separate laser. Ions are coupled to each other by electrostatic repulsion. A centre of mass of the system represents non-local entanglement and a "bus" to carry out manipulations on individual qubits of the entanglement.

Using a single berillium ion it is possible to implement a two input controlled NOT gate, by combining energy levels of an electron in the ion with energy levels of the ion in the harmonic trap. Initial measurements demonstrated a correct performance about 90% of the time and a decoherence time of about a millisecond.

A preliminary analysis demonstrated that Fourier transforms can be evaluated with the ion trap computer. This, in turn, leads to Shor's factoring algorithm, which is based on Fourier transforms.

Computation in an ion trap computer is steered by sending laser pulses to selected atoms. It turns out that there is a limit on the maximum number of bits that can be factored with such a computer and it depends on Z – the degree of ionization, L – number of trapped ions, τ – the decoherence time, A – the atomic number of an ion, λ – the wavelength of a laser and F – a focusing capability of the laser:

$$N_{\max} = \frac{2Z\tau}{L^{1.84}A^{1/2}F^{3/2}\lambda^{3/2}} \tag{2.7}$$

It is also possible to predict a total number of pulses needed to factor an integer of a certain size too. Because of the dependence of $N_{\max}$ on A and τ, some ions are better than others. For example an ion of ytterbium has a transition with a lifetime of 1533 days. If we were to use ytterbium, the maximum number of bits that can be factored is 385. The number of laser pulses required to factor the integer is 30 billion. The number of ytterbium ions required to perform the computation is 1926 [50]

2.5.3 QED Cavity

A quantum electrodynamic (QED) cavity computer was demonstrated by Turchette, Hood, Lange, Mabuchi, and Kimble in 1995 [100]. The computer consists of a QED cavity filled with some cesium atoms and an arrangement of lasers, phase shift detectors, polarisers and mirrors. Kimble's group implemented a 2-qubit quantum XOR gate. Control and target bits of the gate are two photons of differing polarization and colour, which pass through the cavity. A very low intensity of the beam results in there being only two photons in the cavity at a time, i.e., the target and the control photon. The control bit is a circularly polarized photon, and a target bit is a linearly polarized one. The dimensions of the cavity have been tuned to resonate with a specific transition in the cesium atom and target and control photons. On read-out rotation of the target photon polarization is measured in function of the control photon intensity. The computer implements a classical XOR function, i.e., the target bit is flipped depending on the state of the control bit. But the set up is also a true quantum computer, because it can create, manipulate, and preserve superpositions and entanglements.

2.5.4 Nuclear Magnetic Resonance

A Nuclear Magnetic Resonance (NMR) computer consists of an ampule filled with a liquid and an NMR machine. Simple! Each molecule in the liquid is an independent quantum memory register. Computation proceeds by sending radio pulses to the sample and reading its response. Qubits are implemented as spin states of the nuclei of atoms comprising the molecules. There is an Avogadro number, $\approx 6 \times 10^{23}$, of computers per mol (or $\approx 2.7 \times 10^{19}$ computers per cubic centimeter).

In an NMR computer the readout of the memory register is accomplished by a measurement performed on a statistical ensemble of, say, 2.7×10^{19} molecules. In this an NMR computer differs from some quantum computers considered so far, for example, a QED-cavity computer, or an ion-trap computer, in which a single isolated quantum system was used. It has been shown that an NMR computer can solve NP complete problems in polynomial time. Because an NMR computer is based on a very well understood and highly sophisticated technology most practical accomplishments in quantum computing so far have been achieved using NMR.

NMR works by looking at transitions associated with spin flips in nuclei comprising an investigated molecule. The ampule with the liquid is immersed in a very strong magnetic field, between 9 T and 15 T. Because there is a magnetic moment associated with the nuclon spin, also in case of neutrons, the energy of a nucleus varies depending on whether the spin is aligned with the direction of the magnetic field or counter-aligned. For spin I there can be $2I+1$ values that the component of the spin in the direction of the magnetic field can assume. As the spin of the nucleus flips between those values, photons of appropriate wavelength, usually within the radio part of the electromagnetic spectrum are emitted. Flipping of the spin can be induced by massaging the sample with a radio pulse, in which the direction of the magnetic field is perpendicular to the direction of the spin, and thus the direction of the original static magnetic field.

As the nucleus interacts with the surrounding electron cloud and with other nuclei within the molecule the energy levels that correspond to those $2I + 1$ values shift. This way by looking at NMR spectra it is possible to infer the chemical composition and structure of a molecule. That's the normal use of NMR.

But, of course, you can reverse this process too, and *knowing* the chemical structure of a molecule and its NMR spectrum you can manipulate its nuclear spins by bathing the ampule in radio pulses of appropriately tuned frequencies.

An so, if you have, say, a molecule with 3 magnetically active nuclei, you can use it to build a 3-qubit Toffoli gate. Given current state of NMR technology it should be possible to cook 73 qubit registers. For larger registers the size of the samples would have to be increased exponentially.

> The most powerful NMR computer demonstrated so far is a 5-qubit system built by Isaac Chuang and his colleagues from IBM Almaden Research Center, Stanford University and the University of Calgary in early August 2000 [46], [102]. The computer uses a specially designed molecule with five fluorine atoms. ^{19}F magnetically active isotope of fluorine is used for this purpose. The computation, which solves the order-finding problem, comprises 200 logical steps (quantum gates), which are executed in about 0.3 s.
>
> But NMR systems can be used also for more than "conventional" spin-flipping quantum computers. In February 2000 Jones, Vedral, Ekert and Castagnoli demonstrated an NMR experiment in which a controlled phase shift gate was implemented by the means of a conditional Berry phase [53]. The point here is that although phase shifts are usually dynamic in origin,

they may also arise as a result of geometric operations, as, for example, in the Bohm-Aharonov effect. In particular those Berry phases depend only on the geometry of the path executed, and gates based on this concept should therefore be resilient to errors, or at least some types of errors. If this pans out then intrinsically fault-tolerant quantum computing may be possible.

2.5.5 Quantum Dots

The four quantum computer implementations discussed in previous sections all suffer from the same malady: compared to present day notebooks they're quite impractical. They require a bulky and very expensive equipment – lasers, very strong magnets, cryogenics – and tremendous experimental, theoretical and mathematical skills to boot. On the one hand this is exactly what makes quantum computing such fun, on the other, this is exactly what keeps it away from our desktops and computer rooms.

It is a common opinion that quantum computers need to be implemented as solid state devices before they can be used for anything other than entertainment and doctoral dissertations. This opinion may reflect old-fashioned and outmoded XXth century thinking, so you shouldn't treat it like a gospel, but it reflects our experience in these matters so far.

There are two quite realistic possibilities here. The first one is based on the quantum dot technology, the second one on the Josephson junction technology.

Two promising quantum dot based schemes were proposed recently by Loss and DiVincenzo [71] and by Sherwin, Imamoglu and Montroy [91]. Both groups are associated with the Center for Quantum Computation and Coherence in Nanostructures of the Quantum Institute of the University of California at Santa Barbara.

The first scheme is based on a concept of an array of quantum dots, in which the dots are connected with their nearest neighbours by the means of gated tunneling barriers. Such gated tunneling barriers are very difficult to make (a tunneling barrier should not be thicker than about 100 Å, so how can you attach a gate electrode to it?), but such barriers have already been achieved experimentally using a split-gate technique by Waugh, Berry, Crouch, Livermore, Mar, Wetervelt, Campman, and Gossard in 1996 (two teams participated in this work: one from Harvard and one from Santa Barbara).

This scheme has one great advantage: the qubits are controlled electrically. In their paper Loss and DiVincenzo demonstrated how single qubit and the so called square root of the *swap* gates can be made from quantum dots. The quantum XOR gate can, in turn, be made of these, and once you have the quantum XOR gate and single-qubit gates, you can do *any* quantum computation with it (this result is due to Barenco, Bennett, Cleve, DiVincenzo, Margolus, Shor, Sleator, Smolin and Weinfurter, 1995 [6]). The disadvantage of this architecture is that quantum dots can communicate with their nearest neighbours only. Data read-out is quite difficult too. The authors propose that electrons trapped in quantum dots could be transferred via tunneling into an array of supercooled

paramagnetic dots, which would nucleate a formation of a ferromagnetic domain whose magnetization direction could then be measured by conventional means. Another possiblity would be to use a spin-dependent switchable "spin valve" tunnel barrier. An initial condition could be set by cooling the chip in a uniform externally applied magnetic field down to cryogenic temperatures, and then by pushing electrons into the dots through the spin-valve tunnel barriers.

Well, it looks like there is just no escape from cryogenics and magentic fields – even if solid state devices are used in place of molecules or trapped ions.

The second scheme is very similar to the QED cavity computer discussed in section 2.5.3. The difference this time is that instead of using an atom of cesium trapped in an electrodynamic cavity we use a single electron trapped in a quantum dot, which is a QED cavity in its own right. Information is stored in the lowest energy levels of the dot, with the energy levels themselves controlled by a dedicated gate electrode. Then instead of using external optical lasers we use infrared semiconductor lasers in the THz regime. The whole system could probably be implemented on a single GaAs wafer or, at worst, as a hybrid GaAs chip. The laser could provide connectivity between several quantum dots, not necessarily adjacent ones, thus performing the function of a data bus. The computation is steered by applying a sequence of adiabatic voltage pulses to individual quantum dots.

The lifetime of a cavity photon must be sufficiently long to enable multiple quantum gate operations with high fidelity. The cavity itself would thus have to have a very high figure of merit (or a very low loss). Metal based cavities are too lossy, so they cannot be used in this context. The authors proposed using a dielectric cavity made of ultrapure silicon instead.

> There are many other very interesting things that can be done with quantum dots.
>
> For example, in August 2000 Michler, Imamoglu, Mason, Carson, Strouse, and Buratto from the University of California at Santa Barbara observed photon antibunching from a single cadmium selenide quantum dot at room temperature [49], [75]. This demonstrates that a quantum dot acts like an artificial atom with a discrete anharmonic spectrum. The result is very surprising, because, first, it shows that manipulations and observations pertaining to a single quantum dot can be performed and single photon emission events observed with great precision, and that all these observations can be carried out at room temperature. The latter brings great encouragement to people who work on dot based quantum computers.
>
> Quantum dots can be also used for classical computing, but in this case computing is done quite differently than with MOSFETs. Craig Lent and his co-workers at the University of Notre Dame use electron configurations rather than electron currents to implement various logical operations [38], [94]. Every dot in their system contains two electrons. Electrostatic repulsion pushes the electrons into the opposite corners of a quantum dot. If the dot is additionally flat there are only 4 corners to choose from, so the electron pair can either connect south-west and north-

east corners or north-west and south-east corners. If two dots are put next to each other electrons in these dots will assume the same configuration. If an electron pair in a dot that is a part of a chain is flipped forcibly, the change may then propagate along the chain like in a domino, flipping every other pair. It is possible to construct quantum dot wires, inverters, gates, fanouts and various other circuit elements based on this elegant principle.

2.5.6 The Kane Computer

This computer looks a little like a quantum dot computer, but in other ways it is more like an NMR computer, so it really forms a category of its own.

The Kane computer was proposed by B. E. Kane from Cambridge, UK, in a brief paper in Nature in 1998 [56]. His idea was to embed a single magnetically active nucleus of ^{31}P in a crystal of isotopically clean magnetically inactive ^{28}Si. The sample would then be placed in a very strong magnetic field in order to set the spin of ^{31}P parallel or antiparallel with the direction of the field. The spin of the ^{31}P nucleus can then be manipulated by applying a radio frequency pulse to a control electrode, called A-gate, adjacent to the nucleus. Electron mediated interaction between spins could in turn be manipulated by applying voltage to electrodes, called J-gates, placed between the ^{31}P nuclei.

It is not impossible to build a computer like this even today. Single atoms of phosphorus can be placed on an ideally polished silicon surface with atomic precision, every 200 Å or so, using *atomic force microscopes* (AFM) and *scanning tunneling microscopes* (STM). The atoms can then be burried under an ultrapure epitaxial layer grown either in an MBE or even just in a CVD reactor. It should be possible to ensure isotopic purity of silicon and phosphorus too, though this may not be cheap.

However, it is not obvious why a computer like that should be any better than, say, a molecule based NMR computer. It certainly wouldn't be any faster, because its speed would be ultimately determined by the required frequency of the RF pulses. Also the electrodes attached to the device could melt when high electric currents are induced in them when a very strong external mangetic field is turned on. Furthermore it is not clear that interactions with a very large number of surrounding silicon atoms would not destroy the coherence of any quantum configuration very quickly.

Nevertheless it is an interesting idea, and it is currently being pursued actively at the Semiconductor Nanofabrication Facility of the University of New South Wales in Sydney, Australia [90].

2.5.7 Josephson Junctions

The Josephson junction quantum computer was demonstrated in April 1999 by Nakamura, Pashkin and Tsai of NEC Fundamental Research Laboratories in Tsukuba, Japan [79]. So a Josephson junction quantum computer is no longer a stuff of theoretical considerations, but a practical reality. In the same

month, only about one week earlier, Ioffe, Geshkenbein, Feigel'man, Fauchère and Blatter, independently, described just such a computer in Nature [51].

Nakamura, Pashkin and Tsai's computer is built around a *Cooper pair box*, which is a small superconducting island electrode weakly coupled to a bulk superconductor. Weak coupling between the superconductors creates a Josephson junction between them. Like most other junctions, the Josephson junction is also a capacitor, which is charged by the current that flows through it. A gate voltage is applied between the two superconducting electrodes. If the Cooper box is sufficiently small, e.g., as small as a quantum dot, the charging current breaks into discrete transfer of individual Cooper pairs, so that ultimately it is possible to just transfer a single Cooper pair across the junction. The effectiveness of the Cooper pair transfer depends on the energy difference between the box and the bulk and a maximum is reached when a voltage is applied, which equalizes this energy difference. This leads to resonance and *observable* coherent quantum oscillations [2].

This contraption, like the quantum dot computer of Loss and Vincenzo [71], has the advantage that it is controlled electrically. Unlike Loss and Vincenzo's computer, this one actually exists in the laboratory. Nakamura, Pashkin and Tsai did not perform any computations with it though. At this stage it was enough of an art to observe the coherence for about 6 cycles of the Cooper pair oscillations, while the chip was cooled to about 30 mK and carefully shielded from external electromagnetic radiation.

Further progress is likely to be difficult. Nevertheless NEC executives were justifiably very proud of this remarkable accomplishment and announced that a practical quantum computer may be only a decade or so away from reality [42]. Tsai was more cautious, but emphasized the importance of having mastered the control of quantum states in the Cooper pair box.

Because of the importance of Josephson junction in quantum computing and in the HTMT computer [95] I am going to spend some more time on it in future.

2.5.8 Topological Quantum Computer

The so called *topological quantum computer* is a computer in which qubits are encoded into a system of anyons. Anyons are quasiparticles in 2-dimensional media. One can show that in such media particle statistics are neither strictly fermionic nor bosonic. But in a way anyons are still closer to fermions, because a fermion-like repulsion exists between them, for example the trajectories of two anyons cannot cross. What makes anyons very unusual though is that this fermion-like interaction depends on how they move with respect to each other. Their movements are described by the so called *braid* group.

The existence and properties of anyons were predicted by Leinaas and Myrheim in 1977 [66]. And only a few years later, Laughlin [65] proposed that some unusual effects observed in 2-dimensional electron sheets cooled to nearly an absolute zero and immersed in a horrendous magnetic field could be explained by assuming that in these conditions collective excitations of electron fluid can form, which have anyon properties.

The idea behind the topological quantum computer is to make use of the *braid group* properties that describe the motion of anyons in order to carry out quantum computations. This idea was proposed by Freedman, Kitaev and Wang in 2000 [36], who also claimed that such a computer should be invulnerable to quantum errors, because of the topological stability of anyons.

In a paper published in October 2001, Averin and Goldman [3] proposed a device based on these ideas. In their model anyons group around *anti-dot* holes of $0.2\,\mu$m diameter made in a 2D electron sheet. The holes are separated by $0.01\,\mu$m wide gates. Individual anyons are then moved between the anti-dots in a way that allows for controlled braiding. They demonstrated how a two-qubit controlled-NOT gate and single qubit gates could be constructed this way, thus showing that their scheme implemented universal computation.

They also discussed the decoherence mechanisms that would affect their model and provided some estimates of the dissipation and decoherence rate, which showed that the device would not be significantly better, in this respect, than, say, schemes based on quantum dots. The advantage of the topological quantum computer, however, is in the ease with which controlled entanglements can be created between qubits. The other advantage would be in the stability of anyon-encoded qubits against depolarization errors.

Chapter 3

An Abstract Quantum Computer

3.1 Quantum Turing Machine

The aim of mathematical theory of computation is to discuss and model computation in abstraction from any particular implementation of a computer. As technology develops computers change and evolve. Furthermore at any given time there are many different computer architectures around. In order to be useful mathematics has to abstract all those superficial differences away and concentrate on what really constitutes computation.

Early in XXth century Gödel, Church and Turing proposed 3 different models of computation:

- general recursive functions of Gödel
- lambda expressions of Church (Lisp and other functional programming languages are based on those)
- the Turing machine

and somewhat later it turned out that they were very much the same. But later still, towards the end of XXth century, it turned out that certain physical assumptions, which may not necessarily correspond to how certain computations can be done, were smuggled into all three models. In particular quantum computation, the subject of this lecture, is not modelled correctly by any of the above. But there are even some aspects of classical computation, which are not adequately accounted for by the Turing machine and equivalent models, e.g., the thermodynamics of computation.

The Turing machine was invented by Alan Turing in 1936 [101] in order to address Hilbert's *Entscheidungsproblem*. It is sufficiently simple so that various mathematical theorems can be proven about it. In particular by using this model Turing was able to show that the Entscheidungsproblem must be answered in

negative, i.e., there is no mechanical procedure, in general, which can be used to decide a theoremhood of an arbitrary statement in mathematics. This, in combination with Gödel theorem, came as a bit of a surprise to mathematicians. On the other hand, it merely demonstrated what the greatest mathematicians always knew and practiced, namely that mathematics is an art. Also, that some of the best mathematics is constructed by broadening a particular theory that is an object of some assertion and attacking the problem from a higher level. For example, algebraic problems can often be tackled with a surprising efficacy by rolling out the apparatus of complex analysis.

The original purpose of Turing machine was to model a formalist mathematical reasoning, the way Hilbert wanted it to be. And Hilbert wanted a mathematician to forget about the meaning of various mathematical constructs and, instead, just operate on symbols. In order to do that the mathematician had to

- invoke some symbol transformation rules
- record each step on paper
- go back and forth over the proof and sometimes combine earlier inferences with the later ones
- have some mechanism for deciding about which transformation rule to use.

Turing simplified this whole procedure by

- replacing symbols with sequences of 1s and 0s,
- replacing a writing pad with a 1-dimensional paper tape divided into cells which would accommodate either 1 or 0,
- inventing a read/write head, which could go back and forth over the tape,
- allowing the head to exist in various *states* that would define a *context* within which read/write operations would take place.

How does Turing machine go about its business?

The computation begins with the program and the initial data being written on the tape, which is otherwise empty. The head is put in a *state*, which tells it: read the program. The program is read, program instructions are intepreted, i.e., the head moves up and down, reading some data, writing some other data, and, in general, using the rest of the tape as a scratch pad.

Soon enough it turned out that Turing machine was universal, that is, that every *classical* computation could be performed with it. It also turned out that, in a way, all classical computers are the same: be it a PC, or a Fujitsu VPP, or a Mac – the most essential difference between them is the colour, transparency and the size of the box they live in. If a computation can be done on a Fujitsu VPP, it can be done on a PC too, even if it's going to take longer.

The original Turing machine was deterministic (DTM): the head would be always in a single state, which would uniquely determine which direction it would go into and how far. There is a variant of the Turing machine, which is not deterministic. The head may be in a state, which gives the machine certain choices as to the direction and length of the next traverse. The choices are then made by throwing dice and possibly applying some weights to the outcome. A machine like that is called a probabilistic Turing machine (PTM), and it turns out that it is more powerful than the deterministic Turing machine in the sense that anything computable with DTM is also computable with PTM and usually faster.

But both PTM and DTM are based on classical physics: the states of the tape and of the head are always readable and writable, data can be always copied, everything is uniquely defined.

A mathematical theory of computation that is based on quantum physics is bound to be different. As you move from classical physics to quantum physics there is a qualitative change in concepts that has profound ramifications.

So here's a brief history of how quantum Turing machine came about.

1973 Bennett demonstrates that a reversible Turing machine is possible [9].

1980 Benioff observes that since quantum mechanics is reversible a computer based on quantum mechanical principles should be reversible too [8].

1982 Richard Feynman shows that no classical Turing machine can simulate quantum phenomena without an exponential slow down, and then observes that a *universal quantum simulator* can [34].

1985 David Deutsch of Oxford University, UK, describes the first true quantum Turing machine (QTM) [28]

In the quantum Turing machine read, write, and shift operations are all accomplished by quantum interactions. The tape itself exists in a quantum state as does the head. In particular in place of the Turing cell on the tape, that could hold either 0 or 1, in quantum Turing machine there is a *qubit*, which can hold a quantum superposition of 0 and 1. The quantum Turing machine can encode many inputs to a problem simultaneously, and then it can perform calculations on all the inputs at the same time. This is called *quantum parallelism.*

A qubit is often represented graphically by a sphere with an arrow in it, see figure 3.1.

Arrow up corresponds to a classical 1. Arrow down corresponds to a classical 0. Arrow in between corresponds to a superposition of 1 and 0. Additionally the arrow may be rotated about the vertical axis, as shown in figure 3.1, which corresponds to the phase of the qubit.

The tape of the quantum Turing machine can now be drawn as shown in figure 3.2.

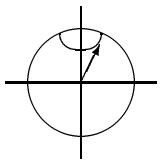


Figure 3.1: A graphical representation of a qubit

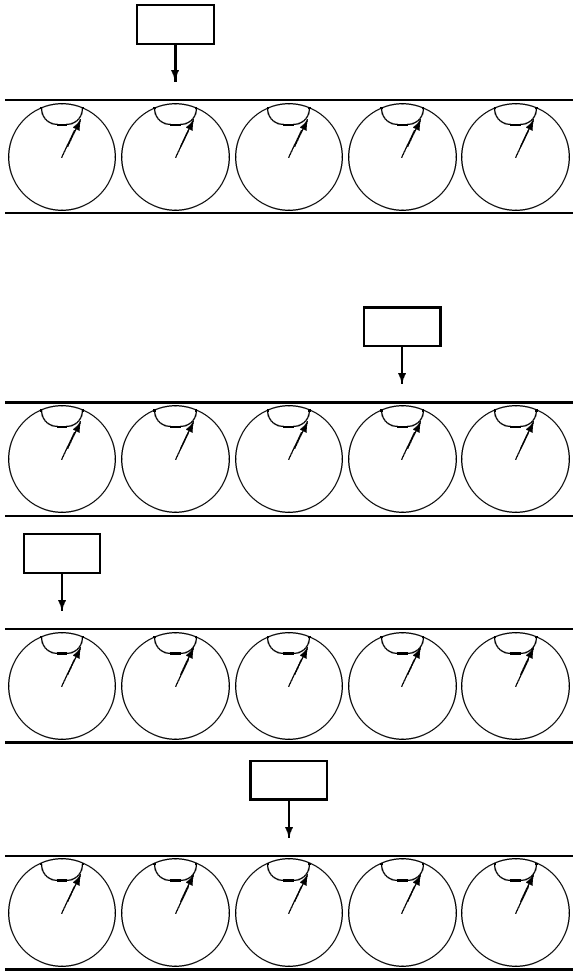


Figure 3.2: A graphical representation of a quantum Turing machine and its evolution. The top row represents an initial condition. As the machine evolves, the head moves simultaneously in three different directions – the state of the machine becomes a superposition of the three states illustrated below the top row.

The machine evolves in many different directions simultaneously. After some time t its state is a superposition of all states that can be reached from the initial condition in that time.

This model, like the classical Turing machine was sufficiently simple and at the same time universal to prove various theorems about quantum computation.

3.2 Quantum Computability

There is a connection between physics and computability. To begin with all computers are physical devices. But, it turns out, that it's the other way round too. Simple physical systems can be made into computers.

In 1990 Christopher Moore from Cornell University showed that a single classical particle moving in a three-dimensional potential well made of a finite number of parabolic mirrors is equivalent to a Turing machine, and hence is capable of universal computation [78].

> This very interesting result is far from trivial.
>
> The reason for this is that a motion of a single classical particle in a "normal" potential well is described in its entirety by Newton's equations of motion. These equations are deterministic and in great many cases easily integrable. As the result the state of the particle can be predicted for all times ahead, or, in other words, any question about the state of the particle at some time in the future can be easily decided. So a system like this cannot map onto a Turing machine, which is undecidable.
>
> Since early 60s (of the XXth century!) scientists have been investigating some classical and quantum systems, which, while deterministic in principle, turned out to be totally chaotic in practice. Damped pendula, accelerator beams, spin-orbit coupling, and many other similar systems, which are usually referred to as chaotic Hamiltonian systems, fall in this category.
>
> However, these systems still cannot be mapped onto the Universal Turing Machine. Because the systems are fully deterministic in principle, it is possible to answer with precision many questions about them, for example escape rates, Lyapunov exponents, periodic sequences, all can be easily determined.
>
> In his brief paper Moore discussed a new type of "maps", i.e., transformations that a dynamical system is subjected to as it evolves in time. The maps are called *Generalized Shift Maps*. Moore showed that there was a simple correspondence between the maps and the Universal Turing Machine: the latter could be easily implemented using a particular form of the Generalized Shift Map. This immediately implied that as the Universal Turing Machine is undecidable then so was the dynamics governed by the Generalized Shift Maps.
>
> The next question that Moore answered was if this type of dynamics could be found in physical systems. A simple example of a system like that was a classical particle (or a family of classical particles) trapped between a finite number of parabolic mirrors, which had a property of expanding and contracting congruences of particle trajectories.

> The resulting motion was unpredictable in a *qualitatively stronger* way than that of chaotic Hamiltonian systems. It was, in fact, as unpredictable as the Universal Turing Machine. The trapped particles *were* in effect the Universal Turing Machine.

Consequently, a demonstration that the universal Turing machine is not capable of something or other implies that the corresponding physics systems (simple or complex) are similarly not capable of some equivalent result.

It is also possible to demonstrate a physics equivalent of the Gödel's theorem. In 1985 Asher Peres argued that it was possible to make certain true statements about a physical system, which could not be confirmed by a measurement [85], sic!

These observations led Peres and Wojtek Zurek to state that quantum mechanics, being a formal system, could not be closed: some truths pertaining to quantum mechanical systems had to exist that could not be analyzed within the body of quantum mechanics [84]. This, argued Peres in "Quantum Theory: Concepts and Methods" [86],

> was not a flaw of quantum theory, but a logical necessity in a theory which is self-referential and describes its own means of verification. This, continued Peres, reminds of Gödel's undecidability theorem: the consistency of a system of axioms cannot be verified because there are mathematical statements that can neither be proved nor disproved by the formal rules of the theory; but they may nonetheless be verified by methamathematical reasoning.

The latter clearly suggests, although this is probably quite unintentional, that as was the case, for example, with the Fermat conjecture, which had ultimately been proven some 358 years later by Andrew Wiles from Princeton University and Richard Taylor from Cambridge University, who went far beyond elementary number theory and used tricks from algebraic geometry to deliver the proof, similarly those unanswerable questions in Quantum Mechanics may be answered by going beyond Quantum Mechanics – perhaps also some 358 years later.

This is a very telling example of how one can *turn the tables* and use theory of computation in order to say something profound about physics.

Quantum Turing machine can be used to simulate the classical Turing machine and the probabilistic Turing machine too. But quantum Turing machine can do more than that. For example it can generate *truly* random numbers, something that classical Turing machines cannot do.

Quantum parallelism is not easy to harness though. On measurement of final results the wave function of the computer must collapse, so that only a single result is delivered. On the other hand, it turns out that it is possible to measure certain *joint* properties of all the outputs.

The unique features of a quantum computer pose the following paradox: imagine that the computer is used to prove automatically a mathematical theorem. Classical computer programs that do just that exist and have delivered a

number of genuine proofs of nontrivial theorems. But in a quantum computer the details of the reasoning cannot be followed. An attempt to do that converts the quantum computer into a classical computer. The situation is exactly the same as in the Feynman double-slit Gedankenexperiment. The moment you insert an apparatus that can tell you which way the particle goes, the quantum intereference image vanishes and you're left with a classical distribution and a classical trajectory. This led some authors, e.g., Williams and Clearwater [104], to ponder a situation whereupon a quantum computer would be able to tell you if your theorem is true or false, but it would not be possible to extract the proof.

This may indeed be the case, but it does not imply that a classical proof of that theorem does not exist or that it cannot be found. One can demonstrate easily that the solution of the Schrödinger wave equation that describes the double-slit Gedankenexperiment represents a congruence of classical trajectories [14]. Whether there is a classical particle that follows those trajectories or not is highly debatable. But this is a matter of *interpretation*. From a strictly mathematical point of view a congruence of trajectories is there in the solution of the Schrödinger equation.

Translating this result onto our quantum theorem prover tells us that if we were able to somehow measure the whole wave function of the computer as it goes through the proof, and it may be possible to do that by running the job repetitively and measuring distributions, then it should be possible to extract a "classical trajectory" from that function that represents a classical proof of our theorem. The wave function will, in fact, deliver a whole congruence of proofs of numerous theorems, of which ours will be but one.

3.3 Quantum Complexity and Quantum Algorithms

Complexity is a measure of how efficiently can solvable problems be solved. What is of special interest is the growth of solution time and memory requirements with the size of the problem. The idea is to scrutinize a problem itself and deliver the assessment of its *complexity* abstracted away from any particular computer architecture. So, like Turing machine, complexity is an abstract concept. But complexity is related to the model of computation. There are special complexity classes for a deterministic Turing machine, special classes for a probabilistic Turing machine and then special classes for a quantum Turing machine.

The most important class is P: a class of problems that grow polynomially, as opposed to an exponential growth. Problems that grow exponentially are considered intractable in general. An example of such an intractable problem is large integer factoring which is exponential in the number of bits needed to represent an integer number. For example to factor a 200 digit number would take nearly 3 billion years on a machine that runs a million instructions per second.

P	problem can be solved by a UTM in polynomial time at worst	QP	problem can certainly be solved by a QTM in polynomial time at worst	P $\subset$ QP [13]
ZPP	problem can certainly be solved by a PTM in polynomial time on average	ZQP	problem can be solved by a QTM without errors in polynomial time	ZPP $\subset$ ZQP
BPP	problem can be solved by a PTM in polynomial time with probability greater than 2/3	BQP	problem can be solved by a QTM in polynomial time at worst with probability greater than 2/3, i.e., in 1/3 cases the computer may return an erroneous result	BPP $\subseteq$ BQP

Table 3.1: Classical and quantum complexity classes.

The most important classes are shown in table 3.1

Apart from the classes shown in the table there are quite a few more, distinguished by ever finer criteria that specify them. An important class is NP. This is a class of problems that are known to be intractable. But imagine that you can guess a solution to such a problem. Having guessed it you need to check that this indeed is a solution. But how easy it is to check a solution to such a problem. This is a problem in its own right. If this problem is of class P, then the class of the original problems, i.e., the intractable ones is called NP.

A relatively small number of quantum algorithms is known. Amongst them are:

- find a true statement in a list of two statements, Deutsch and Jozsa 1992 [29]
- integer factoring, Simon 1994 [93], Shor 1994 [92], Kitaev 1995 [60]
- database search, Grover 1996 [39]
- median estimation, Grover 1996 [40]
- mean estimation, Durr and Hoyer 1996 [32]

In most cases they rely on a quantum version of the Fourier transform. The quantum Fourier transform algorithm itself was discovered by Bernstein and Vazirani in 1993 [11].

3.4 Quantum Circuits

For practical purposes, especially for the design and analysis of quantum computers and algorithms an abstraction of a quantum circuit is going to be more useful than a quantum Turing machine. In 1993 Andrew Yao demonstrated the

equivalence of a quantum Turing machine and quantum circuits for the QP class of problems [105]. This *does not* prove the equivalence in general, but since the Turing machine itself is not a very useful concept in practical applications, there hasn't been much more development in this direction.

Some people believe that practical quantum computers will implement specific functions as quantum circuits and that they are not going to be programmable devices. It is difficult to see where this assertion comes from because NMR computers of today are perfectly programmable – programs are conveyed to molecules as judiciously tuned and shaped sequences of radio-frequency pulses. But semantically speaking, the only programming language for quantum computing today is the language of quantum circuits. It may be quite entertaining to think of higher level languages that would somehow encapsulate quantum logic and that could be compiled to the language of quantum gates.

Quantum circuitry comprises quantum gates in a sequence as they are applied to quantum registers. A quantum register is a collection of individually addressable qubits with individually controllable couplings and is represented by horizontal lines in circuit diagrams. Vertical lines in circuit diagrams illustrate couplings between various qubits of the register. You can see a typical example of a quantum circuit in Figure 4.14, section 4.5.8, page 141. Each quantum gate is reversible with a transformation between inputs and outputs described by a unitary operator. An operator $\boldsymbol{U}$ is said to be unitary when

$$\boldsymbol{U}^{\dagger}\boldsymbol{U} = \mathbf{1} \tag{3.1}$$

Quantum circuits may also contain other non-unitary operations such as measurements, which may be performed towards the end of a computation or in the middle of a computation on a part of the register. For example, CSS circuits discussed in section 6.3, page 235, all perform measurements of error syndromes prior to application of error correction procedures. Measurements are *not* unitary operations. But measurements, unitary transformations, and some other procedures and natural processes can be described together using a formalism of *quantum operations*, to which you will be introduced in section 4.7.2.

> *A quantum circuit can therefore be defined as a quantum register, to which a finite number of quantum operations is applied.*

There is another subtlety involved here. A quantum register will always evolve by itself following whatever free Hamiltonian describes its structure. The notion of a quantum circuit *assumes* that this natural evolution is *halted* and replaced with a steered evolution in the form of the sequence of quantum operations. This may not always be possible. Whether it is possible depends on the dynamics of the system chosen as the register. In section 4.5.6 you will see how this can be done for an NMR computer.

> The definition of a quantum circuit as a finite sequence of quantum operations is quite general and encompasses all circuits we are going to

study in this course. You will also find a narrower definition in literature, in which a quantum circuit is defined as a finite sequence of *unitary* operations terminated with a measurement. It turns out that for certain purposes a circuit with a measurement in the middle of its sequence of operations can be transformed into a circuit with all measurements at the end.

In the Brassard circuit (Figure 4.14, page 141) a measurement is performed in the middle of the computation. In the actual experiment, about which you will learn in section 4.5.8, this is done by allowing quantum states on two carbon nuclei in a molecule of trichloroethylene to decohere naturally. The states that are left on the carbon nuclei after this operation are then used in following computations. We therefore have a genuine non-unitary operation carried out in the middle of the computation with a causal relationship between quantum states on the carbon nuclei immediately after the measurement and just a little later when the computation restarts. The states are, as a matter of fact, the same and the algorithm depends on them being the same.

But if we were to *ignore* this causal relationship this whole operation could be modeled by assuming that the point at which the decoherence is allowed to take place is the *final* point in this part of the circuit. Then the point at which the computation is restarted can be thought of as an *entry point* to the circuit where *two additional qubits* are endowed with certain initial conditions, which, *incidentally only* happen to be what the original states on the carbon nuclei decohered to.

In this way the circuit can be thought of as comprising unitary operations within and non-unitary operations, i.e., the imposition of initial conditions and the measurements, at the left and right edges of the circuit only.

Although such a transformation does not correspond necessarily to how the operations are actually performed, it may be useful for mathematical purposes, e.g., when proving theorems about quantum circuits in general.

Of course, one can object that ignoring causal relationships between parts of the circuit and redrawing it for the sake of sweeping non-unitary operations to the edges produces a different circuit that is not going to behave *always* in the same way as the original circuit. For example, the two carbon nuclei in the molecule of trichloroethlene may decohere to $|\,0\rangle$ and $|\,1\rangle$. In the original circuit the computation will then restart from $|\,0\rangle$ and $|\,1\rangle$ on the carbon nuclei. But nothing stops us from assigning $|\,1\rangle$ and $|\,1\rangle$ to the two additional qubits in the redrawn circuit, which would lead to a quite different final result. The redrawn circuit is therefore not quite equivalent to the original circuit. It imitates the behaviour of the original circuit for certain assignments of initial conditions, as long as these correlate with whatever the measurement is going to return on the top two lines, but it also adds other possibilities, which the original circuit would not realize.

Chapter 4

A Brief Rehash of Quantum Mechanics

4.1 Probability Amplitudes

4.1.1 The Interference Experiment

Consider the experiment shown in Figure 4.1. We start with a strong beam of coherent monochromatic light, e.g., a laser beam, which emerges from source s. The beam then passes through two narrow slits and, as you should remember from school, an interference pattern results on the screen. The pattern is registered by a photographic emulsion, or by a highly sensitive light detector, which as is shown in Figure 4.1, we can move along the x axis.

Next we turn the intensity of light down until... something quite amazing happens. At very low light intensities light becomes granular. The detector begins to "tick" (or flash) audibly (or visibly) as it receives light packets. The packets all transmit the same amount of energy, which is related to the colour of light by the well known Planck-Einstein relation $E = \hbar\omega$, where E is energy, ω is the angular frequency of light and $\hbar$ is $h/2\pi$, where h is the Planck Constant.

What happens to the interference fringes we've seen on the screen when the light intensity was high? The packets arrive at random locations. It is quite impossible to predict where a given packet is going to hit. Yet, as the counts accumulate, we see the inteference pattern emerging. What used to be light intensity in classical physics, here becomes probability density. A probability that a photon emitted from s is going to arrive at some point x if there are two slits in between.

The image would be quite different if there were three slits or if there were two walls with slits in between. The packets of arriving light seem very well localized. You can make the detector as small as you like, even as small as a single atom, and it is still going to detect the arrival of the whole packet of light.

Light intensity in classical physics is proportional to E^2, where $\boldsymbol{E}$ is electric

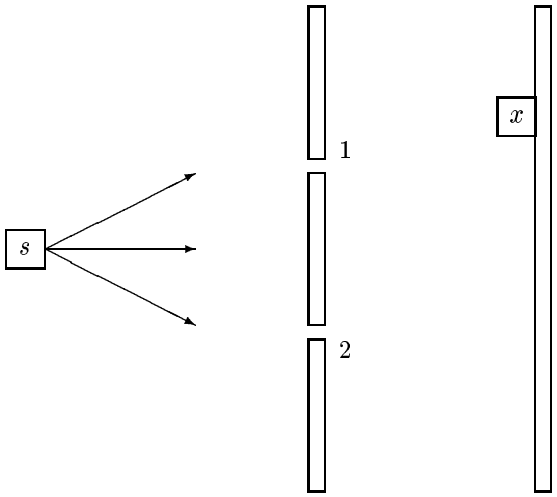


Figure 4.1: The interference experiment with photons: s – source, x – a position of the detector, 1 – slit 1, 2 – slit 2.

field vector. The interference pattern arises on the level of $\boldsymbol{E}$ and is caused by the vectors associated with light beams arriving from the two slits adding with varying phase.

In quantum mechanics a similar thing happens. In order to derive an expression for probability density associated with arrival of photons in the detector located at some x we need to have some *amplitudes* (equivalents of $\boldsymbol{E}$), which we could add with different phases, so that the inteference pattern would emerge.

We call these *probability* amplitudes and we use the following symbol to describe a probability amplitude that a photon emitted from source s is going to arrive at the detector location x:

$$\text{probability amplitude} = \langle x \mid s \rangle \tag{4.1}$$

The probability density is then obtained by squaring the amplitude, which, in general, is a complex number.

$$\text{probability} = \langle x \mid s \rangle \langle x \mid s \rangle^* = |\langle x \mid s \rangle|^2 \tag{4.2}$$

The probability amplitude that an electron emitted from source s is going to travel through 1 and then from 1 to the detector located at x is:

$$\langle x \mid s \rangle_{\text{via 1}} = \langle x \mid 1 \rangle \langle 1 \mid s \rangle \tag{4.3}$$

The interference pattern is obtained by adding probability amplitudes for the two alternative paths, one through slit 1 and the other through slit 2.

$$\langle x \mid s \rangle_{\text{via 1 or 2}} = \langle x \mid 1 \rangle \langle 1 \mid s \rangle + \langle x \mid 2 \rangle \langle 2 \mid s \rangle \tag{4.4}$$

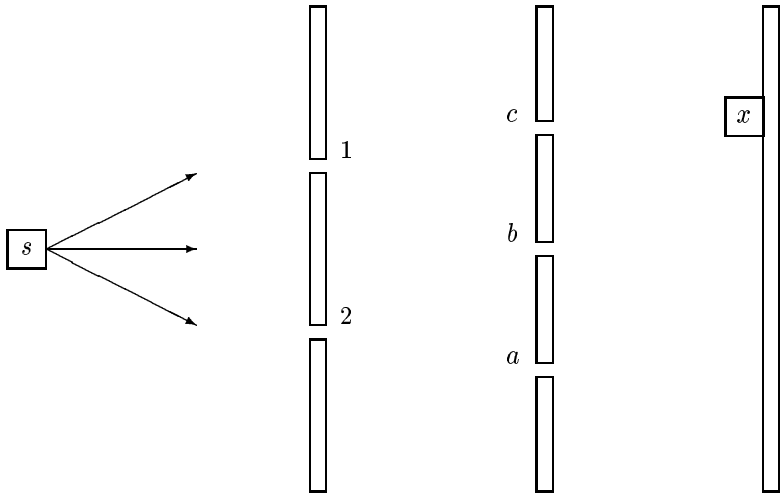


Figure 4.2: A more complex interference experiment: s – source, x – a position of the detector, 1 and 2 are slits in the first wall, a, b, and c are 3 slits in the second wall.

Depending on the position of x the two components in the above sum are added with different phases and so the interference pattern emerges when the amplitude $\langle x \mid s\rangle_{\text{via 1 or 2}}$ is squared. The mechanics of this are much the same as in the classical theory of light. The difference being basically that of interpretation. What in classical physics was light intensity, here becomes probability density. What in classical physics was a field amplitude, here becomes probability amplitude.

The same machinery can be used for more complex situations, for example, for an experiment with two screens and five slits shown in Figure 4.2.

The probability amplitude that a photon which passed through slit 1 arrives at location x is given by

$$\langle x \mid 1\rangle_{\text{via a, b, or c}} = \langle x \mid a\rangle\langle a \mid 1\rangle + \langle x \mid b\rangle\langle b \mid 1\rangle + \langle x \mid c\rangle\langle c \mid 1\rangle \tag{4.5}$$

and similarly for a photon which passed through slit 2. The probability amplitude for a photon that leaves source s and arrives at the detector located at x is now the sum of probability amplitudes for all possible trajectories:

$$\langle x \mid s\rangle_{\substack{\text{via 1 or 2, and then}\\ \text{via a, b, or c}}} = \sum_{\substack{i=1,2\\ j=a,b,c}} \langle x \mid j\rangle\langle j \mid i\rangle\langle i \mid s\rangle \tag{4.6}$$

In order to convert these expressions into numbers we need an expression for $\langle \boldsymbol{r}_2 \mid \boldsymbol{r}_1\rangle$, i.e., a probability amplitude that a photon that had passed through $\boldsymbol{r}_1$ arrived at $\boldsymbol{r}_2$. This is given by

$$\langle \boldsymbol{r}_2 \mid \boldsymbol{r}_1\rangle = \frac{e^{i\boldsymbol{p}\cdot\boldsymbol{r}_{12}/\hbar}}{r_{12}}, \tag{4.7}$$

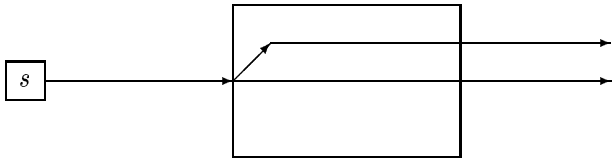


Figure 4.3: A birefringent crystal splitting the beam of incident light into two perpendicularly polarized beams.

where $\boldsymbol{p}$ is the momentum of the photon:

$$\boldsymbol{p} = \hbar \boldsymbol{k} \tag{4.8}$$

where $\boldsymbol{k}$ is the wave number of the photon.

A truly amazing discovery of early XXth century was that this formalism describes not only photons, but also electrons and all other elementary particles. For relativistic particles the momentum $\boldsymbol{p}$ is given by:

$$p^2c^2 = E^2 - \left(m_0c^2\right)^2 \tag{4.9}$$

and for slow Newtonian particles, we can use the following expression instead

$$\frac{p^2}{2m} = E_k \tag{4.10}$$

4.1.2 Experiments with Polarized Light

Apart from having intensity and colour light can be also polarized. This can be seen best by transmitting a beam of light through a birefringent crystal. Two beams emerge: one is polarized horizontally, the other one vertically. Figure 4.3 illustrates this experiment. We can now place two birefringent crystals back-to-back and merge the two beams back into one. If the original beam was not polarized, then the merged beam will not be polarized either. This is shown in Figure 4.4.

We can also insert a block between the two crystals, as shown in Figure 4.5 and filter one of the beams away, so that the beam that emerges has a well defined polarization.

This apparatus can now be used to prepare an initial state of the beam. We can then send the beam through another apparatus like this one, although in this second apparatus we may choose to filter away the top beam instead of the bottom one. If we filter away the bottom beam in the first apparatus and then we insert a filter in the top beam trajectory in the second apparatus, we will filter away all light in the beam and end up with nothing.

Figure 4.4: A beam is first split into two polarized beams, which then merge back into one beam.

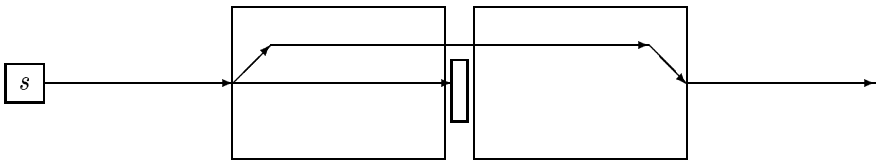


Figure 4.5: A beam is first split into two polarized beams, then one beam is filtered away, so that light that emerges from the apparatus is fully polarized.

Suppose that light in the top beam is polarized vertically and light in the bottom beam is polarized horizontally. We will use symbols v and h to denote polarization states of the beams. Now we are going to repeat our procedure with turning down light intensity until we can see individual photons, and what we used to call light inensity in classical physics now become probability. Photons carry polarization like classical beams.

We can describe our experiment in which we have extinghished the whole beam by writing the following formula:

$$\langle h \mid v \rangle = 0 \tag{4.11}$$

The probability amplitude that vertically polarized photon can be found in the horizontally polarized state is zero. Similarly, by blocking the top beam first and the bottom beam next we arrive at:

$$\langle v \mid h \rangle = 0 \tag{4.12}$$

Then we can block the bottom beam in both apparatuses, and of course, the top beam will emerge at the end undisturbed. The probability that a photon that is in vertical polarization to begin with is found again in the same state is 1. Hence the corresponding amplitude is 1 too:

$$\langle v \mid v \rangle = 1 \tag{4.13}$$

and similarly

$$\langle h \mid h \rangle = 1 \tag{4.14}$$

Now suppose the second apparatus has been rotated about the beam axis by some angle. Now the situation becomes really interesting. From classical physics we know that if a vertically polarized beam strikes a polarizer, whose axis has been rotated away from vertical, the intensity of the transmitted beam will be attenuated, until, the polarizer has been rotated by 90 degrees at which stage the beam becomes completely extinguished.

Switching to photons we discover that the relations we wrote above will no longer hold. Namely:

$$\begin{aligned} \langle h' \mid v \rangle &\neq 0 \\ \langle v' \mid h \rangle &\neq 0 \\ \langle h' \mid h \rangle &\neq 1 \\ \langle v' \mid v \rangle &\neq 1 \end{aligned}$$

These probability amplitudes are a function of the angle of rotation of the second apparatus. We could figure out how they change by, again, making analogies with classical physics (and we know what the solution to this problem is there), but we can make some general statements without it. And these statements are as follows:

1. For every polarization h or v the probability that a photon will emerge from the second apparatus as either h' or v' is 1:

$$\langle h' \mid v\rangle\langle h' \mid v\rangle^* + \langle v' \mid v\rangle\langle v' \mid v\rangle^* = 1 \tag{4.15}$$

and

$$\langle h' \mid h\rangle\langle h' \mid h\rangle^* + \langle v' \mid h\rangle\langle v' \mid h\rangle^* = 1 \tag{4.16}$$

These two relations hold for *every* angle subtended between the two apparatuses.

2. Consider an experiment with three apparatuses. The first and the third one are not rotated with respect to each other and the middle one is rotated by an angle α about the beam axis. Now suppose that there are no filters at all in the middle apparatus. This means that the beam that enters this apparatus exits unchanged. Hence the probability that a photon polarized vertically in the first apparatus is registered as still polarized vertically by the third apparatus is 1, and the same holds also for a photon polarized horizontally. But when you write down what happens inside the middle apparatus the following emerges:

$$\begin{aligned} \langle v \mid v'\rangle\langle v' \mid v\rangle + \langle v \mid h'\rangle\langle h' \mid v\rangle &= 1 \\ \langle h \mid v'\rangle\langle v' \mid h\rangle + \langle h \mid h'\rangle\langle h' \mid h\rangle &= 1 \end{aligned}$$

Comparing relations obtained in these two experiments shows that we must have:

$$\begin{aligned} \langle h' \mid v\rangle\langle h' \mid v\rangle^* + \langle v' \mid v\rangle\langle v' \mid v\rangle^* &= 1 \\ \langle v \mid h'\rangle\langle h' \mid v\rangle + \langle v \mid v'\rangle\langle v' \mid v\rangle &= 1 \end{aligned}$$

and

$$\begin{aligned} \langle h' \mid h\rangle\langle h' \mid h\rangle^* + \langle v' \mid h\rangle\langle v' \mid h\rangle^* &= 1 \\ \langle h \mid h'\rangle\langle h' \mid h\rangle + \langle h \mid v'\rangle\langle v' \mid h\rangle &= 1 \end{aligned}$$

Observe that these relations must hold for *all* angles α. This can be ensured only if

$$\begin{aligned} \langle h' \mid h\rangle^* &= \langle h \mid h'\rangle \\ \langle v' \mid h\rangle^* &= \langle h \mid v'\rangle \\ \langle h' \mid v\rangle^* &= \langle v \mid h'\rangle \\ \langle v' \mid v\rangle^* &= \langle v \mid v'\rangle \end{aligned}$$

Now let us consider an even more complex experiment. Suppose we have an apparatus S, followed by an apparatus T, followed by some very complex apparatus A made of various filters and birefringent crystals, followed again by

T and by S. We can describe polarization states generated by apparatus S by S_i, where i is either v or h. Similarly for apparatus T. The first S apparatus defines the state of the beam and the last S aparatus measures the state of the emerging beam. The probability amplitude that an initial state χ, which can be either S_v or S_h, will emerge as some other state ϕ is described by an amplitude

$$\langle \phi \mid A \mid \chi \rangle \tag{4.17}$$

which thanks to the apparatus T can be described as follows:

$$\langle \chi \mid A \mid \phi \rangle = \sum_{ij} \langle \chi \mid T_i \rangle \langle T_i \mid A \mid T_j \rangle \langle T_j \mid \phi \rangle \tag{4.18}$$

Matrix $\langle T_i \mid A \mid T_j \rangle$ thoroughly describes apparatus A.

4.1.3 Dirac Notation and Hilbert Space

In the previous section we have demonstrated the following rules for probability amplitudes describing polarization states χ and ϕ of a photon:

$$\langle \chi \mid \phi \rangle = \sum_{i=h}^{v} \langle \chi \mid i \rangle \langle i \mid \phi \rangle \tag{4.19}$$

$$\langle \chi \mid \phi \rangle = \langle \phi \mid \chi \rangle^* \tag{4.20}$$

$$\langle i \mid j \rangle = \delta_{ij} \tag{4.21}$$

Dirac observed that you can drop $\langle \chi$ from these formulas and instead work with:

$$\mid \phi \rangle = \sum_i \mid i \rangle \langle i \mid \phi \rangle \tag{4.22}$$

Those "half-amplitude" objects, $\mid \phi \rangle$ are called *quantum states*, or, sometimes *kets* (because they are the second half of a *bracket*). The complementary objects such as $\langle \chi \mid$ are called *bras*.

The language of quantum states or *bras* and *kets* is uncannily similar to the language of forms and vectors. Here we have:

$$\boldsymbol{v} = \sum_i v^i \boldsymbol{e}_i = \sum_i \boldsymbol{e}_i \langle \boldsymbol{\omega}^i, \boldsymbol{v} \rangle, \tag{4.23}$$

where $\boldsymbol{e}_i$ are basis vectors and $\boldsymbol{\omega}^i$ are basis forms conjugate to $\boldsymbol{e}_i$, i.e.,

$$\langle \boldsymbol{\omega}^i, \boldsymbol{e}_j \rangle = \delta^i{}_j \tag{4.24}$$

This, of course, is not accidental. *Kets* can be thought of as vectors, and *bras* can be thought of as forms. The vector space of quantum mechanics conceived thusly is in every respect a kosher vector space with some enhancements:

1. vector components are complex numbers

2. because
$$\langle \chi \mid \phi \rangle = \langle \phi \mid \chi \rangle^*$$
every vector can be upgraded to a form and vice versa. This implies that the vector space of quantum mechanics has a naturally defined scalar product: $\langle \chi \mid \phi \rangle$. This scalar product satisfies additional conditions:
   (a) $\langle \phi \mid \phi \rangle > 0$ for $\mid \phi \rangle \neq 0$ and 0 otherwise
   (b) $\langle \chi \mid a\phi_1 + b\phi_2 \rangle = a\langle \chi \mid \phi_1 \rangle + b\langle \chi \mid \phi_2 \rangle$, where a and b are complex scalars.
3. the space is *complete* in the norm $\|\phi\| = \sqrt{\langle \phi \mid \phi \rangle}$, this means that every Cauchy series in this norm is convergent to an element of the space

A vector space with these properties is called a *Hilbert space*.

The last condition is often broken by admission of states with well defined momentum, energy, or both, for example

$$\Psi(x) = e^{ipx/\hbar} \tag{4.25}$$

Such states are not kosher, because their norm is infinite, i.e., they are spread over the entire infinite universe. But they are used for computational and conceptual convenience. If you admit these states to the Hilbert space, the resulting space is called a *rigged* Hilbert space.

4.1.4 The Copenhagen Interpretation

The interpretation that an expression such as $\langle i \mid \phi \rangle$ represents a *probability amplitude* that a quantum system originally in state $\mid \phi \rangle$ is found, when measured, in state $\mid i \rangle$ and that
$$\mid \phi \rangle = \sum_i \mid i \rangle \langle i \mid \phi \rangle$$
yields a complete description of a quantum system in terms of probability amplitudes of finding that system in any of the *basis* states is often called the *Copenhagen Interpretation*. However, even experts disagree on what exactly is meant by Copenhagen Interpretation. As observed by Asher Peres[86]:

> Ballentine gives this name to the claim that "a pure state provides a complete and exhaustive description of a single system." The latter approach is called by Stapp the "absolute-ψ interpretation." Stapp insists that "critics often confuse the Copenhagen interpretation, which is basically pragmatic, with the diametrically opposed absolute-ψ interpretation ... In the Copenhagen interpretation, the notion of absolute wave function representing the world itself is unequivocally rejected." There is therefore no real conflict between Ballentine and Stapp, except that one of them calls Copenhagen interpretation what the other considers as the exact opposite of the Copenhagen interpretation.

4.2 Quantum Evolution

4.2.1 A Historical Note

Although we promised to stay away from Schrödinger equation it is important, as we embark on defining equation that describes quantum evolution, and for historical reasons too, that we bring some of the early picture of quantum mechanics here.

Planck and Einstein It all began when Planck derived correct equations describing black body radiation by assuming that matter can absorb and emit light in portions of $E = h\nu = \hbar\omega$. Einstain took it then one step further in his paper about the photoelectric effect, where he assumed that light propagates as particles, which he called photons, and the relation between light colour (frequency) and photon energy was $\omega = E/\hbar$

de Broglie Then de Broglie advanced the following intriguing hypothesis. He said that if light, which until then has been thought of as a wave process, can sometimes behave as particles, then particles such as electrons can perhaps behave sometimes as waves. He then proceeded to explain quantization of energy levels in an atom of hydrogen as corresponding to standing electron waves that form around a nucleus. The energy of the electron and the frequency of the wave were linked by the Planck relationship.

stationary state In a stationary state the electron wave would be described by $\Psi(\boldsymbol{r}, t) = ae^{-i\omega t} = ae^{-i(E/\hbar)t}$ This is a flat wave that fills the entire universe. It is OK, because here we have a particle with a well defined momentum (which is zero). By the Heisenberg uncertainty principle, which says that $\Delta x \Delta p_x \approx h$ we must have $\Delta x \approx \infty$.

uniform motion Now let us change the system of reference so that the particle gains some momentum The energy-momentum of the particle, $(E, \boldsymbol{p})$ and its location in space-time, $(t, \boldsymbol{r})$, are 4-vectors. Remember that $Et - \boldsymbol{p} \cdot \boldsymbol{r}$ is a Lorentz invariant. Therefore in the new system of reference de Broglie's wave must have the following form: $\Psi(\boldsymbol{r}, t) = ae^{-i(Et - \boldsymbol{p}\cdot\boldsymbol{r})/\hbar}$, from which we derive, by comparing it with an expression for a traveling wave, $\Psi(\boldsymbol{r}, t) = ae^{-i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})}$ that $E/\hbar = \omega$ and $\boldsymbol{p}/\hbar = \boldsymbol{k}$

Schrödinger (from [57]) Ernest Schrödinger then took de Broglie's relation on board and reasoned as follows. If you take a partial derivative of Ψ with respect to time you get $\frac{\partial \Psi(\boldsymbol{r},t)}{\partial t} = \frac{-i}{\hbar} E\Psi(\boldsymbol{r}, t)$, or $i\hbar\frac{\partial \Psi}{\partial t} = E\Psi$. So, he speculated, the time derivative operator must correspond to energy: $i\hbar\frac{\partial}{\partial t} = \hat{E}$

When you take a gradient of de Broglie's wave function you get $\frac{\partial \Psi(\boldsymbol{r},t)}{\partial \boldsymbol{r}} =$

$\frac{i}{\hbar}\boldsymbol{p}\Psi(\boldsymbol{r},t)$, or $-i\hbar\frac{\partial\Psi}{\partial\boldsymbol{r}} = \boldsymbol{p}\Psi$. So, he speculated, the gradient operator must correspond to momentum: $-i\hbar\frac{\partial}{\partial\boldsymbol{r}} = \hat{\boldsymbol{p}}$

Then he speculated further that if he combined the energy and momentum operators in the same way that energy and momentum combine in classical physics, $E = \frac{p^2}{2m} + V$, i.e., $\hat{E} = \frac{\hat{p}^2}{2m} + \hat{V}$, he would end up with a partial differential equation that should describe propagation of de Broglie waves:

$$i\hbar\frac{\partial\Psi(\boldsymbol{r},t)}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi(\boldsymbol{r},t) + V(\boldsymbol{r},t)\Psi(\boldsymbol{r},t) \tag{4.26}$$

and so he arrived at his celebrated Schrödinger equation. It was a lucky and educated guess. His equation, of course, cannot be derived from classical physics.

This was still before people performed experiments with photons like the one we have described above, so it wasn't clear back then what the function $\Psi(\boldsymbol{r},t)$ really represented. The Copenhagen interpretation came much later, and not everybody was happy with it. In particular neither Planck, nor Einstein, nor de Broglie, nor Schrödinger even ever subscribed to it, sic! De Broglie and Schrödinger developed their own alternative interpretation for this function, which was further developed by Bohm. We are going to talk about this interpretation a little in the last sub-section of this section.

Using the Schrödinger equation it is possible to describe quite easily many interesting quantum effects such as discretization of electron energy levels in atoms, tunneling across a potential barrier and many others. These you will find covered in every standard book about quantum mechanics, and we won't dwell on them here. If at any stage we'll need them, we'll use them like "off the shelf physics" and refer you to literature.

4.2.2 The Hamiltonian Matrix

The passage of time and its effect on a quantum state can be thought of as an *apparatus* for moving the quantum state in time: $\boldsymbol{U}(t_2,t_1)$

$$\langle\chi \mid \boldsymbol{U}(t_2,t_1) \mid \phi\rangle = \sum_{ij}\langle\chi \mid i\rangle\langle i \mid \boldsymbol{U}(t_2,t_1) \mid j\rangle\langle j \mid \phi\rangle$$

where $\mid i\rangle$ and $\mid j\rangle$ are some basis states, which can be polarization states of a photon, or spin states of an electron, or energy levels of electron in an atom of hydrogen.

The apparatus for moving the quantum state in time must be transitive:

$$\boldsymbol{U}(t_3,t_1) = \boldsymbol{U}(t_3,t_2)\boldsymbol{U}(t_2,t_1)$$

Consider a small increment in time and the effect this increment has on the quantum state:

$$\mid \Psi(t+\Delta t)\rangle = \boldsymbol{U}(t+\Delta t,t) \mid \Psi(t)\rangle$$

Since $\boldsymbol{U}(t,t)$ must be an identity, we can expand $\boldsymbol{U}(t+\Delta t,t)$ in a Taylor series around $\boldsymbol{U}(t,t)$ thusly:

$$\boldsymbol{U}(t+\Delta t,t) = \mathbf{1} - \frac{i}{\hbar}\boldsymbol{H}\Delta t$$

The $-i/\hbar$ factor is here for *historical reasons* only: we'll see soon enough why.

The evolution of $\mid \Psi\rangle$ over time Δt now looks as follows:

$$\mid \Psi(t+\Delta t)\rangle = \left(\mathbf{1} - \frac{i}{\hbar}\boldsymbol{H}\Delta t\right) \mid \Psi(t)\rangle$$

Moving $\mid \Psi\rangle$ from the left hand side to the right hand side and dividing both sides by Δt yields:

$$\frac{\mid \Psi(t+\Delta t)\rangle - \mid \Psi(t)\rangle}{\Delta t} = -\frac{i}{\hbar}\boldsymbol{H} \mid \Psi(t)\rangle$$

$$i\hbar\frac{\partial \mid \Psi(t)\rangle}{\partial t} = \boldsymbol{H} \mid \Psi(t)\rangle \tag{4.27}$$

The $\boldsymbol{H}$ matrix in this equation is called the Hamiltonian matrix, because of the equation's similarity to equations of Hamiltonian formalism in classical dynamics.

Now Compare this equation with the Schrödinger equation 4.26

$$i\hbar\frac{\partial\Psi(\boldsymbol{r},t)}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi(\boldsymbol{r},t) + V(\boldsymbol{r},t)\Psi(\boldsymbol{r},t)$$

We get the following:

- $\Psi(\boldsymbol{r},t) = \mid \Psi(t)\rangle$
- $\boldsymbol{H} = -\frac{\hbar^2}{2m}\nabla^2 + V(\boldsymbol{r},t)$

Here $\Psi(\boldsymbol{r},t)$ is a probability amplitude of finding a particle in location $\boldsymbol{r}$ at time t, and $-\frac{\hbar^2}{2m}\nabla^2 + V(\boldsymbol{r},t)$ is the Hamiltonian matrix.

You can see that $\Psi(\boldsymbol{r},t)$ is a vector if you identify each $\boldsymbol{r}$ with an index. The vector is infinitely dimensional and the number of dimensions is $\aleph^3$. The Hamiltonian matrix in this case is an infinitely dimensional matrix that acts on vector $\Psi(\boldsymbol{r},t)$.

Such vector spaces are very difficult to work with: the relevant area of mathematics is called *functional analysis*. In this course we'll stay away from it. But it's good to know what's what, and, in particular, you'll see that even in finite systems, e.g., 2-dimensional ones, $\hbar\omega$ is still energy, as in the continuum case.

If $\boldsymbol{H}$ does not depend on time, the solution to equation 4.27 is:

$$\mid \Psi(t)\rangle = e^{-i\boldsymbol{H}t/\hbar} \mid \Psi(0)\rangle \tag{4.28}$$

so that:

$$\boldsymbol{U}(t,0) = e^{-i\boldsymbol{H}t/\hbar} \tag{4.29}$$

This expression can be understood easier in terms of its Taylor expansion:

$$e^{-i\boldsymbol{H}t/\hbar} = \mathbf{1} - \frac{i}{\hbar}\boldsymbol{H}t - \frac{1}{2!\hbar^2}\boldsymbol{H}\cdot\boldsymbol{H}t^2 + \cdots \tag{4.30}$$

4.2.3 Unitarity

Quantum states are normalized. Mathematically this means that $\langle\Psi \mid \Psi\rangle = 1$. Physically this means that if a given quantum system is in state $\mid \Psi\rangle$ then the probability of finding this system in state $\mid \Psi\rangle$ is 1.

Quantum evolution needs to preserve this condition, i.e., if a state vector evolves, then its length must not change. It should remain 1: $\langle\Psi\boldsymbol{U}^\dagger \mid \boldsymbol{U}\Psi\rangle = \langle\Psi \mid \boldsymbol{U}^\dagger\boldsymbol{U} \mid \Psi\rangle = 1$ This implies that operators describing quantum evolution must satisfy this condition: $\boldsymbol{U}^\dagger\boldsymbol{U} = \mathbf{1}$ Such operators are called *unitary*.

There is a geometric picture which conveys the same. Quantum states are described by normalized vectors in Hilbert space, i.e., they are points on a sphere of radius 1. A quantum evolution moves those points on the sphere, i.e., it *rotates* them. The condition $\boldsymbol{U}^\dagger\boldsymbol{U} = \mathbf{1}$ is a Hilbert space equivalent of matrix orthogonality, i.e., it says that $\boldsymbol{U}$ is a "rotation" matrix.

Equation $\boldsymbol{U}(t,0) = e^{-i\boldsymbol{H}t/\hbar}$ says that matrix $\boldsymbol{H}$ is one of the the *generators* of group of *unitary* matrices $\boldsymbol{U}$.

Unitary matrices form a group, which can be parametrized, e.g., by giving "angles" of rotations. Such parametrization is *smooth*, which means that unitary matrices form a manifold. A manifold, which is at the same time a group, is called a Lie group. So unitary matrices form a Lie group. The generators of a Lie group are in fact "vectors" that are tangent to the Lie group at some point. Such vectors form what is called a Lie algebra.

In summary:

- Evolution of a quantum state in Hilbert space is accomplished by a linear unitary operator. Such operators form a Lie group.
- An infinitesimal evolution of a quantum state in Hilbert space is accomplished by a generator of the unitary group. That generator is called a Hamiltonian matrix. Hamiltonian matrices belong to the Lie algebra of the Lie group of unitary transformations.

These statements aren't really as profound as they may appear to be. All they express is the fact that $\langle\Psi \mid \Psi\rangle$ is a probability that a quantum particle that is in state Ψ can be found in that state, and that probability is, of course, 1. And, that, as we evolve the system, its quantum state changes to some other quantum state, but the same tautology still applies to it.

Quantum mechanics would become much more interesting if some other, possibly non-probabilistic interpretation could be given to Ψ. We would no longer be restricted to a sphere of radius 1 then, and evolution operators would no longer have to be unitary.

Quantum mechanical equations of motion are quite unsatisfactory for another reason. In both the wave and in the discrete Schrödinger equations we have time, t, as the major player. In the wave equation we also have the guiding vector, $\boldsymbol{x}$, which describes points in a continuous 3-dimensional space. Yet both time and space are macroscopic concepts. You have to use classical measuring rods and clocks, the way it's done in Relativity classes, to define them. But there are no clocks and measuring rods in the quantum domain. So the Schrödinger equation is flawed from the very beginning because it mixes macroscopic and microscopic concepts, or, to put it in other words, it describes the evolution of a microscopic system from a macroscopic point of view, where time and space are well defined notions.

Little wonder then that the resulting picture tends to hark back to the macroscopic observer all the time. One could well say that quantum mechanics describes interactions of microscopic systems with a macroscopic world, rather than the microscopic systems on their own. This may be the root of various conceptual difficulties in quantum mechanics itself, as well as the root of sickly divergencies in quantum field theory.

Is a description of quantum systems, which does not rely on macroscopic variables, at all possible? The answer is *yes*. There is an example of a theory, which does not introduce the concepts of space and time at the microscopic level at all. Through various manipulations the theory then demonstrates how space and time arise in the thermodynamic limit from quantum mechanical interactions.

The theory of "Spin Networks" was originally conceived by Roger Penrose and an introductory paper about it published in "Quantum Theory and Beyond" published by Ted Bastin as far back as 1971 [5]. The theory progressed a little since then, though not by much. Some very interesting spin network results were shown recently for the Chern-Simons Theory (I wish I could find a reference to this paper...).

What is so interesting about Quantum Computing is that here we don't use time and space in describing and discussing quantum circuits either. Instead the workings of the circuits are derived from mutual couplings between qubits and from sequential ordering of the circuits.

Perhaps one day we will arrive at a new perspective from which we will view quantum circuits as fundamental and the Schrödinger equation as a derived semi-macroscopic expression.

4.2.4 Schrödinger and Hamilton-Jacobi Equations

Schrödinger equations, both the wave and discrete versions, operate on complex valued functions of time and space. Yet we are perhaps justifiably reluctant to use such functions in context of physics, where all measurable quantities are real. In classical physics complex valued functions are used as auxiliary devices in the theory of vibrating systems, and it is always possible to drop them in favour of real-valued functions only. Is this possible in quantum mechanics too?

The Schrödinger wave equation can be thought of as two coupled equations for real-valued functions. And when you do this then quite unexpected picture

emerges.

Recall the Schrödinger wave equation:

$$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi + V\Psi$$

Since Ψ is a complex valued function, we can always replace it with the polar form

$$\Psi = Re^{iS/\hbar} \tag{4.31}$$

where R and S are two real-valued functions. Substituting this polar form in the Schrödinger equation yields a complex differential equation, which can be readily split into real and imaginary parts. This results in two coupled equations, which no longer involve any complex numbers or complex valued functions:

$$\frac{\partial S}{\partial t} + \frac{(\boldsymbol{\nabla} S)^2}{2m} + V - \frac{\hbar^2}{2m}\frac{\nabla^2 R}{R} = 0 \tag{4.32}$$

$$\frac{\partial R^2}{\partial t} + \boldsymbol{\nabla}\cdot\left(R^2\frac{\boldsymbol{\nabla} S}{m}\right) = 0 \tag{4.33}$$

The top equation looks exactly like a classical Hamilton-Jacobi equation where apart from the normal potential V we have an additional term:

$$Q = -\frac{\hbar^2}{2m}\frac{\nabla^2 R}{R} \tag{4.34}$$

This term is called a *quantum potential* to distinguish it from the normal classical potential V.

The quantum Hamilton-Jacobi equation can be solved for a family of trajectories, satisfying

$$m\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = -\boldsymbol{\nabla} V - \boldsymbol{\nabla} Q \tag{4.35}$$

Of course, in order to solve this equation, we need to solve a normal Schrödinger equation first, so that we can find R, and from R also Q.

If you carry out this procedure for a double-slit experiment, you end up with a family of trajectories which get extremely close to each other within the slits themselves, but then fan out. The trajectories are not straight lines. They are kind of wiggly and when they eventually hit the screen they get again closer to each other in places where we expect the interference fringes to emerge, and they get sparse in the troughs of the interference pattern.

This picture provides a *classical* interpretation of the Schrödinger equation, which is consistent with all that we know about quantum mechanics. Even very small differences in positions that a classical particle can assume within the slot result in widely different locations when the particle hits the detector.

The quantum potential Q has some very unusual properties. To begin with it does not diminish with distance. It can be thought of as a non-local field, which the particle uses to sniff the whole universe as it moves from the source to the detector. In particular using this field the particle is aware of the presence

of the other slot, and conditions therein, when it traverses through its own slot. The presence of the other slot affects the particle's future trajectory so that eventually the interference pattern emerges, when a sufficiently large statistical ensemble of particles is emitted by the source.

This is the essence of the deBroglie-Schrödinger-Bohm interpretation of quantum mechanics. The quantum potential field Q was called a *pilot wave* by deBroglie. Later Bohm called it an *information field*.

This is a very interesting picture and it is a pity that very few physicists know about it. This often leads to incorrect statements along the lines that "a classical interpretation of quantum mechanics is not possible". Von Neumann even proved some "theorems" to this effect. His "theorems" were later shown by Bell to have been based on incorrect assumptions about classical physics.

As you have just seen, it is possible and even straightforward to concoct a classical interpretation of quantum mechanics. Whether it is a *correct* interpretation is another matter altogether, and from my own remarks in the previous section, you can probably guess that I expect quantum world to be much weirder than this classical picture. Because this interpretation is based on classical notions of space and time and because it derives entirely from the Schrödinger wave equation, it suffers from the same sin of mixing macroscopic notions of space and time with the microscopic world of quantum physics, where these notions become quite questionable.

But there is a great value in the deBroglie-Schrödinger-Bohm interpretation. The value is in the observation that the quantum potential Q, which is essential in this picture, is a non-local entity. This comes out even more profoundly, when similar manipulations are applied to many-body quantum systems. The non-locality, the entanglement of all quantum components are the most striking and unusual features of quantum physics. It is here that quantum physics differs from classical physics most.

4.3 Two-state Systems

In this section we are going to have a close look at a qubit and its dynamics. This time we have enough quantum mechanics background to do it in detail. Towards the end of this section, you will learn that although qubits can be implemented in various types of quantum hardware, e.g., they can be associated with electron spin, or with photon polarization, or with selected energy levels of electrons in an atom, they are all subjected to identical dynamics. This dynamics derives from there being just two states between which the qubit can switch.

4.3.1 The Ammonia Maser: a Quantum NOT Gate

Our first qubit example is an ammonia molecule, NH_3. The molecule looks like a little pyramid with a triangular base, formed by hydrogen atoms and with the nitrogen atom at the apex. There is an electric dipole moment associated with this molecule, which points along the direction that connects the apex of the

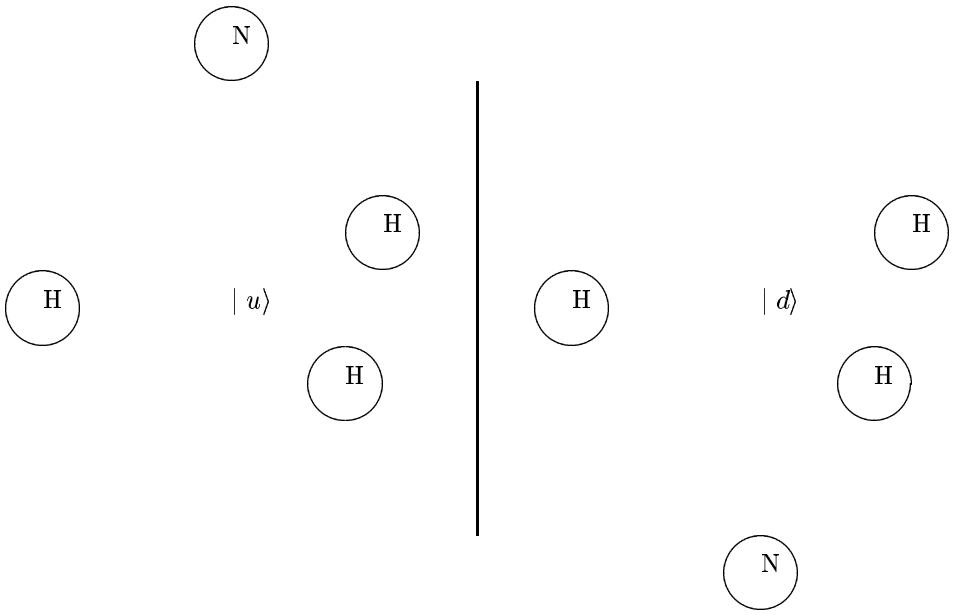


Figure 4.6: An ammonia molecule NH_3 in two possible states: $|\, u\rangle$ and $|\, d\rangle$.

pyramid with the center of its base. If you place the molecule in an electric field, the torque will align its dipole with the direction of the field. The molecule can now exist in two forms. The atom of nitrogen can be either above the plane of carbon atoms, or below. In other words, the pyramid can stand on its base or it can be upside down. Such two different arrangements of atoms are called isomers. In this case, the isomers of the molecule of ammonia are so similar, that you can see the difference between them only in the presence of an electric field. Figure 4.6 illustrates these two possible arrangements.

We are going to associate two different quantum states with these two configurations. If the atom of nitrogen is above the hydrogen plane, we're going to call this configuration *up* or $|\, u\rangle$ using the Dirac notation. If the atom of nitrogen is below the plane, we're going to call this configuration *down* or $|\, d\rangle$.

The two quantum states, $|\, u\rangle$ and $|\, d\rangle$ form a basis of a 2-dimensional Hilbert space. What you're going to see in the following subsection is that these two states do not form a good basis for a qubit, because the molecule fluctuates between them all the time. Even if no forces are applied to it. The fluctuations derive from the fact that there is a small possibility of quantum tunneling from the *up* to the *down* position, across a potential barrier formed by the base of three hydrogen atoms. And the thing about quantum systems is that *if they can tunnel, they will.*

A molecule of ammonia has other degrees of freedom too. It can rotate about various axes, it can move in various directions, its atoms can vibrate around the

positions of equilibrium. So the picture we're presenting here, in which we ignore all the other possible motions, is highly idealized. Nevertheless, the transition between $\mid u\rangle$ and $\mid d\rangle$ in ammonia molecules can be easily observed and separated from other transitions. Also a device can be built, which relies on this transition alone to emit coherent microwave radiation. This device is called an ammonia maser.

What makes a good qubit?

What should be the Hamiltonian for this system? Let us start with a representation of a general vector describing the state of ammonia molecule in the $\mid u\rangle$ and $\mid d\rangle$ basis:

$$\mid \Psi\rangle = C_u \mid u\rangle + C_d \mid d\rangle \tag{4.36}$$

Then let us start from a very simple diagonal Hamiltonian which looks like this:

$$\boldsymbol{H} = \begin{pmatrix} E_u & 0 \\ 0 & E_d \end{pmatrix} \tag{4.37}$$

Substituting this into the Schrödinger equation yields

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} C_u(t) \\ C_d(t) \end{pmatrix} = \begin{pmatrix} E_u & 0 \\ 0 & E_d \end{pmatrix} \cdot \begin{pmatrix} C_u(t) \\ C_d(t) \end{pmatrix} \tag{4.38}$$

This equation yields two independent ordinary differential equations:

$$\frac{\mathrm{d}}{\mathrm{d}t} C_u(t) = -i \frac{E_u}{\hbar} C_u(t) \tag{4.39}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} C_d(t) = -i \frac{E_d}{\hbar} C_d(t) \tag{4.40}$$

with the following solutions

$$C_u(t) = C_u(0) e^{-iE_u t/\hbar} \tag{4.41}$$

$$C_d(t) = C_d(0) e^{-iE_d t/\hbar} \tag{4.42}$$

If we now set $C_u(0) = 1$ and $C_d(0) = 0$ we see that our molecule just stays in the $\mid u\rangle$ state and there is this quantum vibration with frequency $E_u/\hbar$ associated with it.

Similarly, if we set $C_u(0) = 0$ and $C_d(0) = 1$ the molecule stays in the $\mid d\rangle$ state and there is the quantum vibration with frequency $E_d/\hbar$ associated with this state.

What is this "quantum vibration"? What vibrates, or rotates, is the phase of the qubit, even though the qubit itself just stays put in either the *up* or the *down* state. This vibration is related to the energy of the state, through the Planck-Einstein relation $E = \hbar\omega$. So the symbols E_u and E_d we used in the Hamiltonian stand for energies of states $\mid u\rangle$ and $\mid d\rangle$.

But if there is no electric field in the picture, then there is nothing physical to diffrentiate between $\mid u\rangle$ and $\mid d\rangle$ and therefore the two phase rotation frequencies and their corresponding energies should be identical, which yields

$$E_u = E_d = E_0 \tag{4.43}$$

If there was nothing else in the Hamiltonian, this system would make a nice qubit. You will see further down, how to use vibrating electric fields to flip the qubit from the *up* to the *down* configuration.

But, as we said, there is a small possibility that the molecule can *tunnel* between the two configurations. This possibility is described by adding two off-diagonal terms to our Hamiltonian. Because the two configurations are identical, the terms that describe tunneling from *up* to *down* and from *down* to *up* should be identical. And so our Hamiltonian assumes a new form:

$$\boldsymbol{H} = \begin{pmatrix} E_0 & A \\ A & E_0 \end{pmatrix} \tag{4.44}$$

and the Schrödinger equation for the system now looks as follows:

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} C_u(t) \\ C_d(t) \end{pmatrix} = \begin{pmatrix} E_0 & A \\ A & E_0 \end{pmatrix} \cdot \begin{pmatrix} C_u(t) \\ C_d(t) \end{pmatrix} \tag{4.45}$$

or

$$\begin{aligned} i\hbar \frac{\mathrm{d}}{\mathrm{d}t} C_u(t) &= E_0 C_u(t) + A C_d(t) \\ i\hbar \frac{\mathrm{d}}{\mathrm{d}t} C_d(t) &= A C_u(t) + E_0 C_d(t) \end{aligned}$$

These equations are no longer independent.

There is a standard method to solve such equations and it goes like this: first you diagonalize the matrix on the right. Having done this you end up with independent equations, which can be solved very easily. Then you switch back to the original basis from the eigen-basis and write down the solution. The eigenvalues of the Hamiltonian correspond to energies associated with eigenstates. This is true for all other so called *quantum observables* of quantum mechanics. Various physics quantities such as momentum, angular momentum, charge, can be associated with matrices, like the Hamiltonian matrix, and the values of the quantities that can be observed in experiments, correspond to the eigenvalues of the matrices.

But here it is instructive to solve these equations without the eigen-machinery, especially since the equations are very simple. If we multiply both equations by $1/\sqrt{2}$ and add them we get

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} \left(C_u + C_d\right)/\sqrt{2} = \left(E_0 + A\right)\left(C_u + C_d\right)/\sqrt{2} \tag{4.46}$$

And if we again multiply both equations by $1/\sqrt{2}$ and subtract the second equation from the first one, we end up with

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} \left(C_u - C_d\right)/\sqrt{2} = \left(E_0 - A\right)\left(C_u - C_d\right)/\sqrt{2} \tag{4.47}$$

These are two independent equations for two functions of time. The first function is $C_+(t) = (C_u(t) + C_d(t))/\sqrt{2}$ and the second function is $C_-(t) = (C_u(t) - C_d(t))/\sqrt{2}$. The solutions to these equations can now be readily written:

$$C_+(t) = C_+(0)e^{-i(E_0+A)t/\hbar} \quad (4.48)$$

$$C_-(t) = C_-(0)e^{-i(E_0-A)t/\hbar} \quad (4.49)$$

The original amplitudes $C_u(t)$ and $C_d(t)$ can be expressed in terms of $C_+(t)$ and $C_-(t)$ as follows:

$$\begin{aligned} C_u(t) &= \frac{1}{\sqrt{2}}(C_+(t) + C_-(t)) \\ &= \frac{1}{\sqrt{2}}\left(C_+(0)e^{-i(E_0+A)t/\hbar} + C_-(0)e^{-i(E_0-A)t/\hbar}\right) \quad (4.50) \\ C_d(t) &= \frac{1}{\sqrt{2}}(C_+(t) - C_-(t)) \\ &= \frac{1}{\sqrt{2}}\left(C_+(0)e^{-i(E_0+A)t/\hbar} - C_-(0)e^{-i(E_0-A)t/\hbar}\right) \quad (4.51) \\ & \quad (4.52) \end{aligned}$$

Suppose that at $t = 0$ $C_u(0) = 1$ and $C_d(0) = 0$. This means that at $t = 0$ $C_+(0) = 1/\sqrt{2}$ and $C_-(0) = 1/\sqrt{2}$ too. The solution for $C_u(t)$ then becomes

$$\begin{aligned} C_u(t) &= \frac{1}{2}e^{-iE_0t/\hbar}\left(e^{-iAt/\hbar} + e^{iAt/\hbar}\right) \\ &= e^{-iE_0t/\hbar}\cos\frac{At}{\hbar} \end{aligned}$$

and

$$\begin{aligned} C_d(t) &= \frac{1}{2}e^{-iE_0t/\hbar}\left(e^{-iAt/\hbar} - e^{iAt/\hbar}\right) \\ &= -ie^{-iE_0t/\hbar}\sin\frac{At}{\hbar} \end{aligned}$$

The probabilities of finding our molecule of ammonia in the *up* or in the *down* states are given by $C_uC_u^*$ and $C_dC_d^*$ respectively. These are:

$$P_u(t) = \cos^2\frac{At}{\hbar} \quad (4.53)$$

$$P_d(t) = \sin^2\frac{At}{\hbar} \quad (4.54)$$

First observe that $P_u(t) + P_d(t) = 1$, i.e., the molecule is *certain* to be found either in the *up* or in the *down* state at any time. But if the molecule starts in the *up* state the probability of its staying there will decay with time like $\cos^2(At/\hbar)$, until at time $t = \pi\hbar/(2A)$ it becomes zero. At the same time

the probability of the molecule being found in the *down* state will grow like $\sin^2(At/\hbar)$, until at time $t = \pi\hbar/(2A)$ it becomes one.

The molecule is going to fluctuate between the *up* and *down* configurations. Even though there is no force present that would keep flipping it between these two isomers. Because of this states $\mid u\rangle$ and $\mid d\rangle$ are not good candidates for a qubit.

But what can be said about quantum states that correspond to amplitudes C_+ and C_-. Can vectors be found in Hilbert space, which these amplitudes would correspond to? In order to answer this question we should retrace what we did when we added and subtracted amplitudes C_u and C_d using vector notation.

And so our initial vector had the form:

$$\mid \Psi\rangle = C_u \mid u\rangle + C_d \mid d\rangle \tag{4.55}$$

where $C_u C_u^* + C_d C_d^* = 1$. Now let us introduce a new pair of vectors:

$$\mid +\rangle = \frac{1}{\sqrt{2}}(\mid u\rangle + \mid d\rangle) \tag{4.56}$$

$$\mid -\rangle = \frac{1}{\sqrt{2}}(\mid u\rangle - \mid d\rangle) \tag{4.57}$$

These relationships can be inverted to obtain:

$$\mid u\rangle = \frac{1}{\sqrt{2}}(\mid +\rangle + \mid -\rangle) \tag{4.58}$$

$$\mid d\rangle = \frac{1}{\sqrt{2}}(\mid +\rangle - \mid -\rangle) \tag{4.59}$$

Substituting the latter two equations into Equation 4.55 yields

$$\begin{aligned} \mid \Psi\rangle &= C_u \frac{1}{\sqrt{2}}(\mid +\rangle + \mid -\rangle) + C_d \frac{1}{\sqrt{2}}(\mid +\rangle - \mid -\rangle) \\ &= \frac{1}{\sqrt{2}}(C_u + C_d) \mid +\rangle + \frac{1}{\sqrt{2}}(C_u - C_d) \mid -\rangle \end{aligned}$$

and so we discover that our amplitudes C_+ and C_- correspond to vectors $\mid +\rangle$ and $\mid -\rangle$.

Invoking once more our expressions for C_+ and C_-:

$$\begin{aligned} C_+(t) &= C_+(0)e^{-i(E_0+A)t/\hbar} \\ C_-(t) &= C_-(0)e^{-i(E_0-A)t/\hbar} \end{aligned}$$

we find that if the molecule of ammonia has been placed in state $\mid +\rangle$, whose energy is $E_0 + A$, initially, it will stay in this state and its phase is going to rotate with frequency $(E_0 + A)/\hbar$. If the molecule has been placed in state $\mid -\rangle$, whose energy is $E_0 - A$, initially, it will stay in this state and its phase is going to rotate with frequency $(E_0 - A)/\hbar$.

These two states, $|+\rangle$ and $|-\rangle$ are good candidates for a qubit. We can attempt to carry out computations using these states. States $|+\rangle$ and $|-\rangle$ are *eigenstates* of Hamiltonian 4.44. The two corresponding values of energy, $E_0 + A$ and $E_0 - A$ are the *eigenvalues* of the Hamiltonian. In basis $\{|+\rangle, |-\rangle\}$ Hamiltonian 4.44 assumes the following form:

$$\boldsymbol{H} = \begin{pmatrix} E_0 + A & 0 \\ 0 & E_0 - A \end{pmatrix} \tag{4.60}$$

In the next section we are going to perform our first quantum computation. We are going to force the change from $|+\rangle$ to $|-\rangle$ thus implementing a quantum NOT gate.

Ammonia molecule in a maser cavity

Consider now a Hamiltonian for an NH_3 molecule in an electric field $\mathcal{E}$ aligned with the direction of the molecule's dipole moment μ. The Hamiltonian is going to look much the same as before, but this time we add $\mu\mathcal{E}$ to the energy in the *up* configuration and subtract it from the energy in the *down* configuration:

$$\boldsymbol{H} = \begin{pmatrix} E_0 + \mu\mathcal{E} & A \\ A & E_0 - \mu\mathcal{E} \end{pmatrix} \tag{4.61}$$

The resulting equations of motion for amplitudes C_u and C_d now assume the following form:

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}C_u(t) = (E_0 + \mu\mathcal{E})\,C_u(t) + AC_d(t) \tag{4.62}$$

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}C_d(t) = AC_u(t) + (E_0 - \mu\mathcal{E})\,C_d(t) \tag{4.63}$$

Let us rewrite these equations in terms of C_+ and C_-. To do this we proceed as in the previous section.

$$\begin{aligned} i\hbar\frac{\mathrm{d}}{\mathrm{d}t}\left(C_u + C_d\right)/\sqrt{2} &= (E_0 + A)\left(C_u + C_d\right)/\sqrt{2} + \mu\mathcal{E}\left(C_u - C_d\right)/\sqrt{2} \\ i\hbar\frac{\mathrm{d}}{\mathrm{d}t}\left(C_u - C_d\right)/\sqrt{2} &= (E_0 - A)\left(C_u - C_d\right)/\sqrt{2} + \mu\mathcal{E}\left(C_u + C_d\right)/\sqrt{2} \end{aligned}$$

which yields

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}C_+ = E_+C_+ + \mu\mathcal{E}C_-, \quad \text{where } E_+ = E_0 + A \tag{4.64}$$

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}C_- = E_-C_- + \mu\mathcal{E}C_+, \quad \text{where } E_- = E_0 - A \tag{4.65}$$

Now substitute:

$$C_+(t) = \gamma_+(t)e^{-iE_+t/\hbar} \tag{4.66}$$

$$C_-(t) = \gamma_-(t)e^{-iE_-t/\hbar} \tag{4.67}$$

which should result in

$$\begin{aligned}\left(\gamma_+ E_+ + i\hbar\frac{\mathrm{d}\gamma_+}{\mathrm{d}t}\right)e^{-iE_+t/\hbar} &= E_+\gamma_+ e^{-iE_+t/\hbar} + \mu\mathcal{E}\gamma_- e^{-iE_-t/\hbar}\\ \left(\gamma_- E_- + i\hbar\frac{\mathrm{d}\gamma_-}{\mathrm{d}t}\right)e^{-iE_-t/\hbar} &= E_-\gamma_- e^{-iE_-t/\hbar} + \mu\mathcal{E}\gamma_+ e^{-iE_+t/\hbar}\end{aligned}$$

These equations can be further simplified by performing the following operations:

1. multiply the first equation by $e^{iE_+t/\hbar}$ and the second one by $e^{iE_-t/\hbar}$
2. subtract the $E_+\gamma_+$ term from both sides of the first equation and $E_-\gamma_-$ from both sides of the second equation
3. replace $(E_+ - E_-)/\hbar$ with $\omega_0 = 2A/\hbar$

which finally yields:

$$i\hbar\frac{\mathrm{d}\gamma_+}{\mathrm{d}t} = \mu\mathcal{E}\gamma_- e^{i\omega_0 t} \tag{4.68}$$

$$i\hbar\frac{\mathrm{d}\gamma_-}{\mathrm{d}t} = \mu\mathcal{E}\gamma_+ e^{-i\omega_0 t} \tag{4.69}$$

These equations look just about as nasty as what we started with and to make things worse there is now the harmonic term, $e^{\pm i\omega_0 t}$, to boot. But this harmonic term is going to be our salvation.

Observe that so far we haven't specified how $\mathcal{E}$ depends on time, so now we fill this gap. $\mathcal{E}$ represents a harmonic wave described by

$$\mathcal{E}(t) = 2\mathcal{E}_0\cos\omega t = \mathcal{E}_0\left(e^{i\omega t} + e^{-i\omega t}\right) \tag{4.70}$$

Now let us substitute this into our equations for γ_+ and γ_-

$$\begin{aligned}i\hbar\frac{\mathrm{d}\gamma_+}{\mathrm{d}t} &= \mu\mathcal{E}_0\gamma_-\left(e^{i(\omega+\omega_0)t} + e^{-i(\omega-\omega_0)t}\right)\\ i\hbar\frac{\mathrm{d}\gamma_-}{\mathrm{d}t} &= \mu\mathcal{E}_0\gamma_+\left(e^{i(\omega-\omega_0)t} + e^{-i(\omega+\omega_0)t}\right)\end{aligned}$$

Our problem now can be split into two distinct cases. The first one, which is very easy, is transitions at resonance, i.e., when $\omega \approx \omega_0$. The second one, somewhat harder describes a situation, when we're away from the resonance.

Transitions at resonance

In the vicinity of the resonance we make the following two approximations

1. When $\omega \approx \omega_0$, the term with $\omega + \omega_0 \approx 2\omega_0$ is going to oscillate rapidly. The rapid oscillations are going to blur into zero, if our system's inertia will not let it follow the oscillations. So we are going to drop this term altogether.

2. At the same time $e^{\pm i(\omega-\omega_0)t} \approx e^{\pm i(\omega_0-\omega_0)t} = 1$

As the result our equations simplify dramatically:

$$\begin{aligned} \frac{\mathrm{d}\gamma_+}{\mathrm{d}t} &\approx \frac{-i}{\hbar}\mu\mathcal{E}_0\gamma_- \\ \frac{\mathrm{d}\gamma_-}{\mathrm{d}t} &\approx \frac{-i}{\hbar}\mu\mathcal{E}_0\gamma_+ \end{aligned}$$

These two equations are now very easy to solve. Take $\frac{\mathrm{d}}{\mathrm{d}t}$ of both sides:

$$\begin{aligned} \frac{\mathrm{d}^2\gamma_+}{\mathrm{d}t^2} &= \frac{-i}{\hbar}\mu\mathcal{E}_0\frac{\mathrm{d}\gamma_-}{\mathrm{d}t} = -\left(\frac{\mu\mathcal{E}_0}{\hbar}\right)^2\gamma_+ \\ \frac{\mathrm{d}^2\gamma_-}{\mathrm{d}t^2} &= \frac{-i}{\hbar}\mu\mathcal{E}_0\frac{\mathrm{d}\gamma_+}{\mathrm{d}t} = -\left(\frac{\mu\mathcal{E}_0}{\hbar}\right)^2\gamma_- \end{aligned}$$

These are two independent harmonic oscillators, which admit the following solution:

$$\gamma_+(t) = a\cos\left(\frac{\mu\mathcal{E}_0}{\hbar}\right)t + b\sin\left(\frac{\mu\mathcal{E}_0}{\hbar}\right)t \tag{4.71}$$

$$\gamma_-(t) = i\left(b\cos\left(\frac{\mu\mathcal{E}_0}{\hbar}\right)t - a\sin\left(\frac{\mu\mathcal{E}_0}{\hbar}\right)t\right) \tag{4.72}$$

The reason why the solution for γ_- is a little different than the solution for γ_+ is that once we decide on a form for γ_+ then γ_- is automatically given by:

$$\gamma_- = \frac{i\hbar}{\mu\mathcal{E}_0}\frac{\mathrm{d}\gamma_+}{\mathrm{d}t} \tag{4.73}$$

Now let us remind you that

$$\gamma_+ = C_+e^{-iE_+t/\hbar} \tag{4.74}$$

$$\gamma_- = C_-e^{-iE_-t/\hbar} \tag{4.75}$$

hence

$$P_+(t) = |C_+(t)|^2 = |\gamma_+(t)|^2 \tag{4.76}$$

$$P_-(t) = |C_-(t)|^2 = |\gamma_-(t)|^2 \tag{4.77}$$

Suppose that for $t = 0$ $\gamma_+ = 1$ and $\gamma_- = 0$, i.e., the molecule is in the higher energy state $|+\rangle$. This implies that $a = 1$ and $b = 0$, and

$$\gamma_+ = \cos\left(\frac{\mu\mathcal{E}_0}{\hbar}\right)t \tag{4.78}$$

$$\gamma_- = -i\sin\left(\frac{\mu\mathcal{E}_0}{\hbar}\right) \tag{4.79}$$

The corresponding probabilities are now going to be:

$$P_+(t) = \cos^2\left(\frac{\mu\mathcal{E}_0}{\hbar}\right)t \tag{4.80}$$

$$P_-(t) = \sin^2\left(\frac{\mu\mathcal{E}_0}{\hbar}\right)t \tag{4.81}$$

The probability of finding the molecule in state $|+\rangle$ falls down like $\cos^2$ and becomes 0 at $t = \pi\hbar/(2\mu\mathcal{E}_0)$. On the other hand the probability of finding the molecule in state $|-\rangle$ grows like $\sin^2$ and becomes 1 at $t = \pi\hbar/(2\mu\mathcal{E}_0)$.

Consequently all we need to do in order to switch the molecule from the $|+\rangle$ state to the $|-\rangle$ state is to irradiate it with a harmonic electric wave of amplitude $\mathcal{E}_0$, the direction of the field parallel to the direction of the dipole, and frequency $\omega_0 = 2A$ for $\pi\hbar/(2\mu\mathcal{E}_0)$ seconds!

It is also easy to see that if the molecule starts in the $|-\rangle$ state then the same operation will switch it to the $|+\rangle$ state.

The maser cavity as a NOT gate

If the molecule of NH_3 is used to encode a qubit, e.g., $|+\rangle = 1$ and $|-\rangle = 0$ (remember that these are stationary states outside the cavity, so this is a good encoding) then the microwave cavity with an electric field oscillating with frequency $\omega_0 = 2A$ becomes a NOT gate for this one qubit register. The negation operation takes exactly $\frac{\pi\hbar}{2\mu\mathcal{E}_0}$ seconds to accomplish.

There are two ways to implement this gate. One would be to use a beam of moving ammonia molecules. We could use a variety of techniques to put some of the molecules in the $|+\rangle$ state and then filter away molecules in other states. Then we would send the $|+\rangle$ beam through the cavity, in which the field would vibrate continuously. We would have to ensure that the dimensions of the cavity and the speed of the molecules in the beam are such that the molecules spend exactly $\pi\hbar/(2\mu\mathcal{E}_0)$ seconds in the cavity as they pass through it.

The other way would be to bring the radiation pulse to a stationary molecule, which could be trapped in laser tweezers or simply put together with an Avogadro number of other ammonia molecules in a small vial.

The latter is a more common approach in quantum computing.

In classical computing logical gates are physical devices that are affixed to a silicon wafer. Data, in the form of electron pulses, moves through the gates as it is processed. In quantum computing data is loaded into a fixed register and gates are then brought to the register in the form of electro-magnetic radiation pulses of various polarization, duration, amplitude and frequency.

This different way of doing computation is also more flexible from the programming point of view. Every quantum computing program can be compiled all the way down to an optimized hardware level, since hardware in this case is simply a series of EM pulses. The generation and shaping of the pulses can be controlled by a classical computer.

Transitions off resonance

In order to analyze the behaviour of the system away from the resonance, we have to change our approximations. Let us return to the equations

$$\begin{aligned} i\hbar\frac{\mathrm{d}\gamma_-}{\mathrm{d}t} &= \mu\mathcal{E}_0\gamma_+\left(e^{i(\omega-\omega_0)t}+e^{-i(\omega+\omega_0)t}\right) \\ i\hbar\frac{\mathrm{d}\gamma_+}{\mathrm{d}t} &= \mu\mathcal{E}_0\gamma_-\left(e^{i(\omega+\omega_0)t}+e^{-i(\omega-\omega_0)t}\right) \end{aligned}$$

Assume that $\mathcal{E}_0$ is small and that t is small too. In this case the time is too short and the field is too weak to flip the molecule. We still assume that the term $e^{-i(\omega+\omega_0)t}$ oscillates so fast that it averages to zero, so that the effective equations of motion are:

$$\begin{aligned} i\hbar\frac{\mathrm{d}\gamma_-}{\mathrm{d}t} &= \mu\mathcal{E}_0\gamma_+e^{i(\omega-\omega_0)t} \\ i\hbar\frac{\mathrm{d}\gamma_+}{\mathrm{d}t} &= \mu\mathcal{E}_0\gamma_-e^{-i(\omega-\omega_0)t} \end{aligned}$$

Assume now that initially

$$\begin{aligned} \gamma_+ &= 1 \\ \gamma_- &\ll 1 \end{aligned}$$

Substituting this into the first equation above yields an approximate solution for γ_-:

$$\begin{aligned} & i\hbar\frac{\mathrm{d}\gamma_-}{\mathrm{d}t} = \mu\mathcal{E}_0e^{i(\omega-\omega_0)t} \\ \Rightarrow \quad & \mathrm{d}\gamma_- = \frac{\mu\mathcal{E}_0}{i\hbar}e^{i(\omega-\omega_0)t}\mathrm{d}t \\ \Rightarrow \quad & \gamma_- = \frac{\mu\mathcal{E}_0}{i\hbar}\frac{1}{i(\omega-\omega_0)}e^{i(\omega-\omega_0)t}+\text{const} \\ \Rightarrow \quad & \gamma_- = \frac{\mu\mathcal{E}_0}{\hbar(\omega-\omega_0)}\left(1-e^{i(\omega-\omega_0)t}\right) \end{aligned}$$

Transition probability during this time is given by

$$\begin{aligned} |\gamma_-|^2 &= \left(\frac{\mu\mathcal{E}_0}{\hbar(\omega-\omega_0)}\right)^2\left(1-e^{i(\omega-\omega_0)t}\right)\left(1-e^{-i(\omega-\omega_0)t}\right) \\ &= \left(\frac{\mu\mathcal{E}_0}{\hbar(\omega-\omega_0)}\right)^2\left(2-e^{i(\omega-\omega_0)t}-e^{-i(\omega-\omega_0)t}\right) \\ &= 2\left(\frac{\mu\mathcal{E}_0}{\hbar(\omega-\omega_0)}\right)^2\left(1-\cos\left(\omega-\omega_0\right)t\right) \end{aligned}$$

Now we are going to use the following trigonometric identity:

$$\cos 2\alpha = \cos^2\alpha - \sin^2\alpha$$

$$\Rightarrow \qquad \cos\alpha = \cos^2\frac{\alpha}{2} - \sin^2\frac{\alpha}{2}$$
$$\Rightarrow \qquad 1 - \cos\alpha = \cos^2\frac{\alpha}{2} + \sin^2\frac{\alpha}{2} - \cos^2\frac{\alpha}{2} + \sin^2\frac{\alpha}{2} = 2\sin^2\frac{\alpha}{2}$$

So that the solution can be now written in the following form:

$$\begin{aligned} |\gamma_-|^2 &= \left(\frac{2\mu\mathcal{E}_0}{\hbar(\omega-\omega_0)}\right)^2 \sin^2\frac{(\omega-\omega_0)\,t}{2} \\ &= \left(\frac{\mu\mathcal{E}_0 t}{\hbar}\right)^2 \frac{\sin^2\left((\omega-\omega_0)\,t/2\right)}{\left((\omega-\omega_0)\,t/2\right)^2} \end{aligned}$$

If we were to treat this expression as a function of ω and fix t, its shape would be bell-like with half-width of the order of π. We can therefore write the following restriction:

$$\omega - \omega_0 \lesssim \pi/t$$

or

$$\frac{f - f_0}{f_0} \lesssim \frac{1}{2tf_0}$$

Now assume that the length of the pulse $t \approx 1\,\text{ms}$. The resonance frequency for the ammonia molecule is $f_0 = 24\,\text{GHz}$. The above condition then yields:

$$\frac{f - f_0}{f_0} \approx \frac{1}{2\times 10^{-3}\,\text{s}\times 24\times 10^9\,\text{s}^{-1}} \approx 0.02\times 10^{-6} \approx 2\times 10^{-8}$$

In order to get a significant transition probability the frequency in the maser cavity must be within two parts per 100 million off the resonance frequency.

4.3.2 General Solution for a 2-State System

In previous sections we had a look at a specific very simple 2-state system. But it turns out that our system is quite representative. The basic reason for this is that there aren't that many ways in a 2-dimensional Hilbert space to do things differently. So all 2-state quantum system are quite similar. Once you've studied one qubit in depth, you've studied them all.

In order to see this more clearly we're going to find a general solution to the Schrödinger equation

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}\mid\Psi\rangle = \boldsymbol{H}\mid\Psi\rangle \tag{4.82}$$

with a general 2×2 Hamiltonian

$$\boldsymbol{H} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}$$

To find the solution we are going to seek such vectors $\mid\Psi\rangle$ that

$$\boldsymbol{H}\mid\Psi\rangle = E\mid\Psi\rangle \tag{4.83}$$

For such vectors the matrix equation (4.82) simplifies to two identical independent equations for the vector components, so that a simple solution can be written:

$$| \Psi(t)\rangle = | \Psi_0\rangle e^{-iEt/\hbar}$$

In general there will be 2 such directions in the Hilbert space, $| \Psi_1\rangle$ and $| \Psi_2\rangle$. They are called *eigenvectors* of $\boldsymbol{H}$. They may be associated with two different values of E: E_1 and E_2. These are called *eigenvalues* of $\boldsymbol{H}$. Eigenvectors of $\boldsymbol{H}$ form a basis in the Hilbert space, so all other solutions of equation (4.82) can be expressed in terms of linear combinations of the eigenvectors of $\boldsymbol{H}$.

In general eigenvalues of $\boldsymbol{H}$ are complex numbers, but if $\boldsymbol{H}$ is Hermitian, i.e., $\boldsymbol{H}^\dagger = \boldsymbol{H}$, then its eigenvalues are guaranteed to be real.

Equation (4.83) has a nontrivial solution if

$$\begin{aligned}
\det(\boldsymbol{H} - E\mathbf{1}) &= \det\begin{pmatrix} H_{11}-E & H_{12} \\ H_{21} & H_{22}-E \end{pmatrix} \\
&= (H_{11}-E)(H_{22}-E) - H_{12}H_{21} \\
&= E^2 - (H_{11}+H_{22})E + H_{11}H_{22} - H_{12}H_{21} \\
&= 0
\end{aligned}$$

The determinant $\Delta = b^2 - 4ac$ of this quadratic ($ax^2 + bx + c = 0$) equation in E is

$$\begin{aligned}
\Delta &= (H_{11}+H_{22})^2 - 4(H_{11}H_{22} - H_{12}H_{21}) \\
&= H_{11}^2 + 2H_{11}H_{22} + H_{22}^2 - 4H_{11}H_{22} + 4H_{12}H_{21} \\
&= H_{11}^2 - 2H_{11}H_{22} + H_{22}^2 + 4H_{12}H_{21} \\
&= (H_{11}-H_{22})^2 + 4H_{12}H_{21}
\end{aligned}$$

And now the solution ($x_\pm = \frac{-b \pm \sqrt{\Delta}}{2a}$) is

$$\begin{aligned}
E_\pm &= \frac{H_{11}+H_{22} \pm \sqrt{(H_{11}-H_{22})^2 + 4H_{12}H_{21}}}{2} \\
&= \frac{H_{11}+H_{22}}{2} \pm \sqrt{\frac{(H_{11}-H_{22})^2}{4} + H_{12}H_{21}}
\end{aligned}$$

And we can use symbols $| +\rangle$ and $| -\rangle$ for the corresponding eigenvectors. We will not attempt to find them explicitly. It is usually easier to find them for specific systems as required.

Observe that we can easily extract a solution for the ammonia molecule from the above. Substitute

$$\begin{aligned}
H_{11} &= H_{22} = E_0 \\
H_{12} &= H_{21} = A
\end{aligned}$$

then

$$E_\pm = E_0 \pm \sqrt{A^2} = E_0 \pm A \tag{4.84}$$

4.3.3 Spin $\frac{1}{2}$ System

Photons are not the only quantum particles that can be polarized. Almost all other particles can be polarized too, but in most cases their polarization is quite peculiar. We refer to it by the name of *spin*. Spin in quantum mechanics can be either an integer multiple of $\hbar$, i.e., 0, $\pm\hbar$, $\pm 2\hbar$, and such particles are called *bosons* or it can be $\pm\hbar/2$, $\pm 3\hbar/2$, $\pm 5\hbar/2$ and so on, and such particles are called *fermions*. In particular electrons, neutrinos, muons, protons, neutrons, and quarks are all fermions, whereas photons, mezons, gravitons (if they exist), alpha particles and Cooper pairs are bosons.

In this section we're going to have a look at Fermions, because they're very good candidates for qubits. Simple fermions like an electron or a proton, can have their spin up $+\hbar/2$ or down $-\hbar/2$. Unlike the Ammonia molecule we studied above, once you put a fermion in an up or a down state, they tend to stay there, unless you do something special to them and flip the spin. The spin of a Fermion cannot be rotated like an arrow in classical physics, because it is not an arrow. But a fermion can be put in a superposition of up and down states. A Hilbert space that describes spin states of a single Fermion is a $2x2$ Hilbert space, for which we have just found a general solution. The reason why up and down spin states are stable is because they are, as you will see later, eigenstates of the spin Hamiltonian, unlike the up and down states of the Ammonia molecule.

Most fermions, though not all, carry electric charge. The ones that do are electron, muon, proton, and quarks. The ones that don't are neutrinos and neutrons. All charged fermions have a magnetic momentum associated with their charge and with their spin, $\boldsymbol{\mu}$, which couples to an external magnetic field $\boldsymbol{B}$ the same way classical magnetic momentum does, which yields the following expression for energy of such a particle in the magnetic field

$$E = -\boldsymbol{\mu} \cdot \boldsymbol{B} \tag{4.85}$$

Because in quantum mechanics spin, and therefore also magnetic momentum, are *always* parallel to the direction of the magnetic field, we can write this expression in a scalar form:

$$E = -\mu B \tag{4.86}$$

Consider a charged spin-1/2 particle in a static magnetic field, which points in the z direction, so that it only has a z component B_z. The particle can have two energy states: $-\mu B_z$ for the magnetic moment parallel to the direction of the magnetic field and μB_z for the magnetic moment anti-parallel to the direction of the magnetic field. Using a vector notation $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ for the former and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ for the latter we can write the Hamiltonian matrix for this system quite readily as

$$\boldsymbol{H} = -\mu \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} B_z \tag{4.87}$$

But there is nothing special about the z direction, which is a matter of our convention. If we rotate a system of coordinates, so that the magnetic field acquires both B_x and B_y coordinates, the energy of the spin particle in the system is still going to be the same:

$$E_{\pm} = \pm\mu\sqrt{B_x^2 + B_y^2 + B_z^2} \tag{4.88}$$

What should the Hamiltonian matrix look like in this case? We can answer this question by making comparisons with our general solution for 2-dimensional quantum systems.

1. First observe that the energy is split equally around 0. This implies that $\frac{H_{11}+H_{22}}{2} = 0$, hence $H_{11} = -H_{22}$
2. This leaves:
$$\frac{H_{11} - H_{22}}{4} + H_{12}H_{21} = \mu^2\left(B_x^2 + B_y^2 + B_z^2\right) \tag{4.89}$$
3. If $\boldsymbol{B} = B_z\boldsymbol{e}_z$ then $H_{11} = -\mu B_z = -H_{22}$, hence $(H_{11} - H_{22})^2/4 = 4\mu^2 B_z^2/4 = \mu^2 B_z^2$ and this implies that in this case $H_{12}H_{21} = 0$, therefore H_{12} and H_{21} cannot have terms in B_z.
4. Since the B_z terms live entirely in H_{11} and H_{22} we must have $H_{12}H_{21} = \mu^2\left(B_x^2 + B_y^2\right)$
5. We also postulate that B_x and B_y appear in H_{12} and H_{21} *linearly*.
6. The solution is $H_{12} = \mu\left(B_x \mp iB_y\right)$ and $H_{21} = H_{12}^* = \mu\left(B_x \pm iB_y\right)$

The resulting Hamiltonian for a charge fermion in a magnetic field looks as follows:

$$\boldsymbol{H} = -\mu\begin{pmatrix} B_z & B_x - iB_y \\ B_x + iB_y & -B_z \end{pmatrix} \tag{4.90}$$

or

$$\boldsymbol{H} = -\mu\left(\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}B_z + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}B_x + \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}B_y\right) \tag{4.91}$$

The matrices multiplying B_x, B_y and B_z are called Pauli matrices:

$$\boldsymbol{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \boldsymbol{\sigma}_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \boldsymbol{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{4.92}$$

So that:

$$\boldsymbol{H} = -\mu\left(\boldsymbol{\sigma}_x B_x + \boldsymbol{\sigma}_y B_y + \boldsymbol{\sigma}_z B_z\right) = -\mu\boldsymbol{\sigma}\cdot\boldsymbol{B} \tag{4.93}$$

where $\boldsymbol{\sigma}$ is a "vector" whose x, y, and z components are the corresponding Pauli matrices.

Pauli matrices have the following properties:

$$\begin{aligned}
\boldsymbol{\sigma}_x^2 &= \boldsymbol{\sigma}_y^2 = \boldsymbol{\sigma}_z^2 = \mathbf{1} && (4.94)\\
\boldsymbol{\sigma}_x \cdot \boldsymbol{\sigma}_y &= -\boldsymbol{\sigma}_y \cdot \boldsymbol{\sigma}_x = i\boldsymbol{\sigma}_z && (4.95)\\
\boldsymbol{\sigma}_y \cdot \boldsymbol{\sigma}_z &= -\boldsymbol{\sigma}_z \cdot \boldsymbol{\sigma}_y = i\boldsymbol{\sigma}_x && (4.96)\\
\boldsymbol{\sigma}_z \cdot \boldsymbol{\sigma}_x &= -\boldsymbol{\sigma}_x \cdot \boldsymbol{\sigma}_z = i\boldsymbol{\sigma}_y && (4.97)\\
& && (4.98)
\end{aligned}$$

Observe that if you make the following identification:

$$\boldsymbol{i} = -i\boldsymbol{\sigma}_x \qquad \boldsymbol{j} = -i\boldsymbol{\sigma}_y \qquad \boldsymbol{k} = -i\boldsymbol{\sigma}_z \tag{4.99}$$

then

$$\begin{aligned}
\boldsymbol{i}^2 &= \boldsymbol{j}^2 = \boldsymbol{k}^2 = -\mathbf{1} && (4.100)\\
\boldsymbol{ij} &= -\boldsymbol{ji} = \boldsymbol{k} \qquad \boldsymbol{jk} = -\boldsymbol{kj} = \boldsymbol{i} \qquad \boldsymbol{ki} = -\boldsymbol{ik} = \boldsymbol{j} && (4.101)\\
\boldsymbol{ijk} &= -\mathbf{1} && (4.102)
\end{aligned}$$

The field of numbers spanned by $(\mathbf{1}, \boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k})$ is called the field of *quaternions*. So Pauli matrices combined with an identity matrix (and from this point onwards whenever we say "Pauli matrices" we will include the identity matrix in them) are isomorphic with quaternions.

You can build various interesting groups of complex 2×2 matrices out of Pauli matrices. But what groups are there to be built?

$\mathrm{GL}(2, \mathbb{C})$ This is a general group of 2×2 matrices with complex coefficients. You can build that group if you multiply Pauli matrices by complex coefficients.

$\mathrm{U}(2)$ These are unitary 2×2 matrices. They also have complex coefficients so we have that $\mathrm{U}(2) \subset \mathrm{GL}(2, \mathbb{C})$. For every element $\boldsymbol{U} \in \mathrm{U}(2)$ of this group the following holds:

$$\begin{aligned}
\boldsymbol{U} \cdot \boldsymbol{U}^\dagger &= \mathbf{1} && (4.103)\\
\det \boldsymbol{U} \, (\det \boldsymbol{U})^* &= 1 && (4.104)\\
\det \boldsymbol{U} &= e^{i\phi}, \quad \phi \in \mathbb{R} && (4.105)
\end{aligned}$$

$\mathrm{SU}(2)$ These are unitary 2×2 matrices, whose determinant is 1. So for these matrices we have $\phi = 0$ and $\mathrm{SU}(2) \subset \mathrm{U}(2) \subset \mathrm{GL}(2, \mathbb{C})$. It can be shown that $\mathrm{SU}(2)$ is the unit sphere in the quaternions, i.e.,

$$\mathrm{SU}(2) = \left\{ a + b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k} : a, b, c, d \in \mathbb{R}, a^2 + b^2 + c^2 + d^2 = 1 \right\} \tag{4.106}$$

$\mathrm{SL}(2, \mathbb{C})$ These are 2×2 matrices with complex coefficients and with determinant 1. Consequently, $\mathrm{SU}(2) \subset \mathrm{SL}(2, \mathbb{C})$, but $\mathrm{U}(2)$ is *not* a subset of $\mathrm{SL}(2, \mathbb{C})$. $\mathrm{SL}(2, \mathbb{C})$ maps onto the Lorentz group. The mapping is very faithful: in more precise terms $\mathrm{SL}(2, \mathbb{C})$ is a *double cover* of the connected

Lorentz group. You can see this quite easily as follows. Consider a 2×2 complex matrix

$$\boldsymbol{X} = t\mathbf{1} + x\boldsymbol{\sigma}_x + y\boldsymbol{\sigma}_y + z\boldsymbol{\sigma}_z \tag{4.107}$$

where (t, x, y, z) are spacetime coordinates of an event. It is easy to show that

$$\det \boldsymbol{X} = t^2 - x^2 - y^2 - z^2 \tag{4.108}$$

Now take any other matrix $\boldsymbol{L} \in \mathrm{SL}(2, \mathbb{C})$ and perform the following similarity operation on $\boldsymbol{X}$:

$$\boldsymbol{X}' = \boldsymbol{L} \cdot \boldsymbol{X} \cdot \boldsymbol{L}^\dagger \tag{4.109}$$

Taking determinants of both sides we get:

$$\det \boldsymbol{X}' = \det \boldsymbol{X} \tag{4.110}$$

So this operation leaves the Lorentz invariant intact, and, therefore, is a Lorentz transformation.

$\mathrm{SL}(2, \mathbb{C})$ is a very simple and a very basic group.

> *The fact that such a basic group is so closely related to the structure of spacetime, can only serve as a challenge to our understanding of physics. Is this a coincidence or a clue that we have still not fully understood? [4]*

Because Pauli matrices are Hermitian

> *every 2×2 Hamiltonian can be expressed as a linear combination of real numbers multiplying Pauli matrices.*

This implies that whatever we have learnt about specific 2×2 systems so far maps directly onto fermions and other 2×2 systems as well. For example, consider again the NH_3 molecule and its Hamiltonian in the $(|+\rangle, |-\rangle)$ basis, describing interaction of the molecule with an electric field $\mathcal{E}$:

$$\boldsymbol{H} = \begin{pmatrix} E_0 + A & \mu\mathcal{E} \\ \mu\mathcal{E} & E_0 - A \end{pmatrix} \tag{4.111}$$

We can always shift our energy scale so that $E_0 = 0$ then

$$\boldsymbol{H} = A\boldsymbol{\sigma}_z + \mu\mathcal{E}\boldsymbol{\sigma}_x \tag{4.112}$$

We can now translate the results of our Ammonia maser model to the world of, say, nuclear spins. In order to flip a spin pointing in the z direction, we need to pass the nucleus through a cavity with an oscillating magnetic field aligned in the x direction. The physics is somewhat different here than it was for the Ammonia molecule, because there the field had to be polarized in the same direction as the molecular dipole. But the equations are the same, even if their physical meaning differs somewhat.

Quaternions were invented by Rowan Hamilton in 1843. Hamilton, as you see, was only a step away from having invented Special Relativity and spin $\frac{1}{2}$ systems. Alas, it is easy to see things like this in retrospect. To know that this direction was worth pursuing at the time, Hamilton would have to know about Michelson-Morley experiment of 1887 and about Stern-Gerlach experiments of the early 1920s. And even then he would have to make the connection.

The concept of electron spin was introduced by Samuel Abraham Goudsmit and George Eugene Uhlenbeck, two graduate students at the University of Leiden in Netherlands, in 1925.

4.3.4 Nuclear Magnetic Resonance

Nuclear Magnetic Resonance is a very mature and precise technique for direct manipulation and detection of nuclear spin states using radiowaves. The technique was developed primary for chemical applications. Using NMR it is possible to infer the structure of a molecule. But NMR is also used in medicine and in material science to image internal structure of tissues and solids.

As we have already remarked in the introductory chapter, NMR experiments involve a very large number of molecules. The Avogadro number is 6×10^{23} of molecules per mole. A mole of an ideal gas occupies 22.4 litres, which is $22400\,\mathrm{cm}^3$. So for an ideal gas we have about 2.7×10^{19} molecules per cm^3. An NMR sample tube usually contains a magnetically active chemical diluted in a magnetically inactive solvent. This dilution is important, because it breaks molecule-molecule interaction, which would affect the readout. So the resulting number of molecules of interest per cm^3 is likely to be less than the Avogadro number. But it is still going to be a number well above a 100 million. For concentrations less than 100 million per sample, the signal is usually too weak to detect.

A typical molecule used in an NMR measurement would comprise a number of protons, all of which are magnetically active and produce an NMR signal at about 500 MHz in a magnetic field of about 12 T. Frequencies of different nuclei in a molecule usually differ by between a few kHz to a few hundred kHz depending on the position within the molecule. These differences are called *chemical shifts* and are caused by the presence of local magnetic fields generated by electron shells within the molecule. The fields vary from place to place, which how we can say something about the structure of a molecule by looking at its NMR spectrum.

In most quantum computing experiments active nuclei other than protons are used. This is because $^{13}\mathrm{C}$, $^{19}\mathrm{F}$, $^{15}\mathrm{N}$ and $^{31}\mathrm{P}$ yield spectral lines, which can be easier separated.

The heart of an NMR is a superconducting magnet, which generates a very uniform magnetic field in the z direction within a rather small region of about $1\,\mathrm{cm}^3$. Our sample must fit within this space to ensure that all molecules are placed within the field of the same direction and strength. The uniformity of the field can be ensured to within 1 part per billion. Helmholtz coils are used

to generate small oscillating magnetic fields in the x and y directions, which, as by now you know, can flip the spins between the *up* and *down* configurations. The fields can be pulsed very rapidly. But to ensure the homogeneity of this radio frequency field is extremely difficult, because the Helmholtz coils are much smaller than the superconducting coils that generate the background field. The same Helmoltz coils are also used to pick up radio frequency fields generated by precessing nuclei.

The computation usually begins with a waiting period of a few minutes, to let molecules thermalize. Radio frequency pulses of various frequencies polarizations and duration are then applied under the control of a normal computer (e.g., a PC). Immediately after the sequence of pulses has been applied, the high power pulse amplifiers are switched off, and a highly sensitive pre-amplifier is turn on, so that the final state of the spins can be measured. The measured signal is then run through Fast Fourier Transform to obtain a frequency spectrum.

Single spin dynamics in an NMR experiment

Although we have, by now, a good idea about how spins are flipped, let us consider the problem of spin interaction with a magnetic field given by

$$\boldsymbol{B} = B_0 \boldsymbol{e}_z + B_1 \left(\boldsymbol{e}_x \cos \omega t + \boldsymbol{e}_y \cos \omega t \right) \tag{4.113}$$

where $\boldsymbol{e}_x$, $\boldsymbol{e}_y$, and $\boldsymbol{e}_z$ are unit vectors pointing in the x, y, and z directions. The Hamiltonian matrix for this system looks as follows:

$$\boldsymbol{H} = -\mu \left(\boldsymbol{\sigma}_x B_1 \cos \omega t + \boldsymbol{\sigma}_y B_1 \sin \omega t + \boldsymbol{\sigma}_z B_0 \right) \tag{4.114}$$

The resonance frequency corresponds to the spin flip and is equal to

$$\omega_0 = 2\mu B_0 / \hbar \tag{4.115}$$

Also, we usually sweep $B_1 \mu / \hbar$ into a phenomenological coefficient g and drop the minus sign, because the sign depends on the charge anyway, so that the Hamiltonian becomes:

$$\boldsymbol{H}/\hbar = \frac{\omega_0}{2} \boldsymbol{\sigma}_z + g \left(\boldsymbol{\sigma}_x \cos \omega t + \boldsymbol{\sigma}_y \sin \omega t \right) \tag{4.116}$$

The Schrödinger equation of motion for this system is

$$i \frac{\mathrm{d}}{\mathrm{d}t} \mid \Psi(t) \rangle = \boldsymbol{H} \mid \Psi(t) \rangle / \hbar = \left(\frac{\omega_0}{2} \boldsymbol{\sigma}_z + g \left(\boldsymbol{\sigma}_x \cos \omega t + \boldsymbol{\sigma}_y \sin \omega t \right) \right) \mid \Psi(t) \rangle \tag{4.117}$$

Now substitute the following solution

$$\mid \Psi(t) \rangle = e^{-i \omega t \boldsymbol{\sigma}_z / 2} \mid \phi(t) \rangle \tag{4.118}$$

This yields the following:

$$i \left(-i \frac{\omega \boldsymbol{\sigma}_z}{2} e^{-i \omega t \boldsymbol{\sigma}_z / 2} \mid \phi(t) \rangle + e^{-i \omega t \boldsymbol{\sigma}_z / 2} \frac{\mathrm{d}}{\mathrm{d}t} \mid \phi(t) \rangle \right) = \boldsymbol{H} e^{-i \omega t \boldsymbol{\sigma}_z / 2} \mid \phi(t) \rangle \tag{4.119}$$

Multiplying both sides from the left by $e^{i\omega t\boldsymbol{\sigma}_z/2}$ and moving the first term to the right hand side yields

$$i\frac{\mathrm{d}}{\mathrm{d}t} \mid \phi(t)\rangle = \left(e^{i\omega t\boldsymbol{\sigma}_z/2}\boldsymbol{H}e^{-i\omega t\boldsymbol{\sigma}_z/2} - \frac{\omega\boldsymbol{\sigma}_z}{2}\right) \mid \phi(t)\rangle \tag{4.120}$$

This equation can be rewritten making use of the following identities:

$$e^{i\omega\boldsymbol{\sigma}_z t/2}\boldsymbol{\sigma}_z e^{-i\omega\boldsymbol{\sigma}_z t/2} = \boldsymbol{\sigma}_z \tag{4.121}$$

$$e^{i\omega\boldsymbol{\sigma}_z t/2}\boldsymbol{\sigma}_x e^{-i\omega\boldsymbol{\sigma}_z t/2} = \boldsymbol{\sigma}_x\cos\omega t - \boldsymbol{\sigma}_y\sin\omega t \tag{4.122}$$

$$e^{i\omega\boldsymbol{\sigma}_z t/2}\boldsymbol{\sigma}_y e^{-i\omega\boldsymbol{\sigma}_z t/2} = \boldsymbol{\sigma}_y\cos\omega t + \boldsymbol{\sigma}_x\sin\omega t \tag{4.123}$$

which you are going to prove laboriously in the exercise below. Making use of these yields:

$$i\frac{\mathrm{d}}{\mathrm{d}t} \mid \phi(t)\rangle = \left(\frac{\omega_0-\omega}{2}\boldsymbol{\sigma}_z + g\boldsymbol{\sigma}_x\right) \mid \phi(t)\rangle \tag{4.124}$$

EXERCISE: Prove these identities.
Hints:

1. First prove that
$$e^{i\alpha\boldsymbol{\sigma}_z} = \mathbf{1}\cos\alpha + i\boldsymbol{\sigma}_z\sin\alpha \tag{4.125}$$
This follows from the fact that $\boldsymbol{\sigma}_z^2 = \mathbf{1}$ and from the Taylor expansion of $e^{i\alpha\boldsymbol{\sigma}_z}$.
2. Use the above expression from both sides of $\boldsymbol{\sigma}_x$, and then use the identities
$$\begin{aligned}\boldsymbol{\sigma}_z\cdot\boldsymbol{\sigma}_x &= -\boldsymbol{\sigma}_x\cdot\boldsymbol{\sigma}_z = i\boldsymbol{\sigma}_y\\ \boldsymbol{\sigma}_y\cdot\boldsymbol{\sigma}_z &= -\boldsymbol{\sigma}_z\cdot\boldsymbol{\sigma}_y = i\boldsymbol{\sigma}_x\end{aligned}$$
3. Recall that
$$\begin{aligned}\sin 2\alpha &= 2\sin\alpha\cos\alpha\\ \cos 2\alpha &= \cos^2\alpha - \sin^2\alpha\end{aligned}$$
4. Repeat the procedure for $\boldsymbol{\sigma}_y$ and $\boldsymbol{\sigma}_z$ in the middle.

The solution to equation (4.124) is

$$\mid \phi(t)\rangle = e^{i((\omega_0-\omega)\boldsymbol{\sigma}_z/2+g\boldsymbol{\sigma}_x)t} \mid \phi(0)\rangle \tag{4.126}$$

Looking at this solution we can infer some modes of behaviour for the spin. For $\omega \approx \omega_0$ the term proportional to σ_z in the exponens vanishes. What we are then left with is a rotation of the spin about the x axis, which after a while results in spin flip. On the other hand, far from the resonance, the first term in the exponens is dominant, because g is very small. In this case the behaviour of the spin is dominated by the background magnetic field.

What do we mean by "the spin rotates"? We said all along that a spin does not rotate. That all it can do is to flip from up to down and vice versa. But spin can also exist in a superposition of up and down and on top of that it can have a phase. This we illustrate graphically by drawing an arrow, which represents a spin, inside a sphere. The tilt of the arrow away from the vertical axis (the lattitude) depends on the proportion of up to down. The longitude corresponds to the spin's phase. This picture is called a Bloch sphere representation of a spin.

There is an operator, which is defined as follows:

$$\boldsymbol{R}(\boldsymbol{n},\theta) = e^{-i\theta\boldsymbol{n}\cdot\boldsymbol{\sigma}/2} = \cos\frac{\theta}{2}\mathbf{1} - i\sin\frac{\theta}{2}\boldsymbol{n}\cdot\boldsymbol{\sigma} \tag{4.127}$$

The effect of applying this operator to the state that is represented by some Bloch vector $\boldsymbol{v}$, i.e., a vector with its beginning in the centre of the Bloch sphere and its sharp end on the sphere's surface, is to rotate it by an angle θ about the direction of the unit vector $\boldsymbol{n}$.

So this rotation does not take place in a real physical space. It is a rotation on the surface of the Bloch sphere. But it has some relation to rotations in the real physical space in the sense that it affects the proportion of up and down in the state of a spin.

Equation (4.127) is known as the Hamilton quaternion formula. It has this nice property that if we have two rotations described by $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ then their composition is simply $\boldsymbol{R}_1 \cdot \boldsymbol{R}_2$. Although Hamilton discovered this formula independently and was the first to publish it, it was already known to Gauss, who discovered it in 1819, though never published, and even to Euler who discovered its simplified form in 1776.

There are various ways to show that it indeed corresponds to rotations. A nice and very readable discussion of this formula, as well as a more general presentation of spin properties can be found in [76].

Comparing equations (4.127) and (4.126) shows that our solution to the NMR equation represents a single spin rotation about the axis

$$\boldsymbol{n} = \frac{\boldsymbol{e}_z + \frac{2g}{\omega_0-\omega}\boldsymbol{e}_x}{\sqrt{1+\left(\frac{2g}{\omega_0-\omega}\right)^2}} \tag{4.128}$$

by an angle

$$\theta = t\sqrt{\left(\frac{\omega_0-\omega}{2}\right)^2 + g^2} \tag{4.129}$$

4.3.5 A Classical Picture of Spin 1/2

A naive classical model of electron is a rotating ball of charge held together by Poincare braces. The braces are essential in order to hold the charge together and also in order to account for the Einstein relation $E = mc^2$. Without the braces we would only have $E = \frac{3}{4}mc^2$. You can evaluate magnetic moment that such an object should have and you'll find that it should be

$$\boldsymbol{\mu} = -\frac{q_e}{2m}L \tag{4.130}$$

where q_e is an elementary charge, m is the mass of the ball, and L is it internal angular momentum or spin.

A less naive analysis, which involves special relativistic effects, shows that this relation should really be:

$$\boldsymbol{\mu} = -\frac{q_e}{m} L \tag{4.131}$$

This is a relatively new result, and for a very long time physicists believed that this relation can be obtained from quantum mechanics only [88]. You will even find this statement in Feynman[35].

An even less naive model is that of a charged rotating black hole. Here we have to roll out the whole heavy machinery of General Relativity and analyze the so-called Kerr-Newman solution. For a specific case of the so called extremal black hole, which spins so fast that its horizon vanishes, the same "quantum" relation between the angular momentum and the magnetic moment of the black hole is obtained [76]. In view of the later result, which shows that this should be the case also in special relativity, this shouldn't be surprising. It shows a certain consistency of the classical relativistic model of an electron.

Yet, neither of these models, regardless of the degree of naivety, can account for the peculiar features of spin-1/2. Neither, in particular, can explain why it is so impossible to capture an electron or any other fermion in a state, which would be between $\pm\frac{1}{2}\hbar$. Classically, of course, all directions should be allowed.

The right way to attack this problem, again, is to look at the quantum mechanical equations that describe spin particles and attempt to restructure them into a Hamilton-Jacobi form, so that classical inferences can be made.

The starting point can be the non-relativistic Pauli equation, which looks as follows:

$$i\hbar\frac{\partial \boldsymbol{\Psi}}{\partial t} = -\frac{\hbar^2}{2m}\left(\boldsymbol{\nabla} - iq_e\boldsymbol{A}\right)^2\boldsymbol{\Psi} + V\boldsymbol{\Psi} + \mu\left(\boldsymbol{\sigma}\cdot\boldsymbol{B}\right)\boldsymbol{\Psi} \tag{4.132}$$

where $\boldsymbol{A}$ is the magnetic potential, V is the electric potential, and $\boldsymbol{\Psi}$ is a two component spinor function. Working with this equation leads to the Bohm, Schiller and Tiomno model, in which the particle is thought of as a rigid body rotating around an axis in the direction of the spin. This picture works quite well: we do get very rapid separation of particle beams in the Stern-Gerlach experiment, and a very rapid flip of the spin onto the direction of the magnetic field. But this model gets into serious difficulties when it is extended to a many-body system.

A better approach, which can be extended to a many-body system without difficulties, is to use the *real* spin equation, i.e., the Dirac equation, of which the Pauli equation is a non-relativistic approximation. The picture that emerges then is very different. In the Dirac equation based Hamilton-Jacobi theory the magnetic moment attributed to the spin derives from a circulating movement of a point particle and not from a rotation of an extended object. In other words, it is not an intrinsic localized property of the particle, but instead a property of the general motion of the particle through space as determined by the Dirac based Hamilton-Jacobi equation [14].

This harks back to the old idea of *zitterbewegung*, which was popular in the 1930s. The velocity eigenvalues in the Dirac equation are always $\pm c$. So people hypothesized that electrons must move along some zig-zagging trajectories with speed of light. The zig-zagging itself would be too fast to see, and the average trajectory would then be sub-luminal.

Derivation of classical relativistic equations of motion for spinning particles also has a *zitterbewegung* term, which can whack a spinning particle around its average trajectory. These derivations by Mathisson [74] and Lubański [72] go back to 1930s. Later in early fifties Papapetrou [83] derived similar equations of motion with *zitterbewegung* terms for spinning particles in General Relativity.

4.3.6 Polarization of the Photon

And what about the photon? What does Schrödinger equation for photons look like? The energy-momentum relation for photons is not $E = p^2/(2m)$ because photons are massless particles, which move with the speed of light. The relation is $E^2 - p^2c^2 = 0$ instead.

> Recall that photons don't have an electric charge, so they are not subject to any *potentials* other than gravitational potential. But to describe interaction of photons with gravity we would have to reach for the General Theory of Relativity, which is quite outside the scope of this lecture. Within the simple quantum mechanics of quantum computing, that is of concern to us here, photons don't interact with potentials.

We can try Schrödinger's trick on this relation, and this yields the following equation for the photon wavefunction $\boldsymbol{\Psi}$:

$$-\hbar^2 \frac{\partial^2}{\partial t^2} \boldsymbol{\Psi} = -\hbar^2 c^2 \nabla^2 \boldsymbol{\Psi}$$

Dividing this by $\hbar^2 c^2$ and moving the laplacian to the left hand side yields:

$$\nabla^2 \boldsymbol{\Psi} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \boldsymbol{\Psi} = 0 \tag{4.133}$$

But this is a normal wave equation and $\hbar$ has cancelled out!

We know that photons have polarization, therefore we would expect $\boldsymbol{\Psi}$ to be a composite object, which is why we used a bold font for it. Classical electrodynamics delivers two equations that look like the Schrödinger photon equation (4.133):

$$\nabla^2 \boldsymbol{E} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \boldsymbol{E} = 0 \tag{4.134}$$

$$\nabla^2 \boldsymbol{B} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \boldsymbol{B} = 0 \tag{4.135}$$

where $\boldsymbol{E}$ is an electric field vector and $\boldsymbol{B}$ is a magnetic field vector. These equations describe propagation of electromagnetic waves in vacuum. Vectors

$\boldsymbol{E}$ and $\boldsymbol{B}$ are always perpendicular to the direction of wave propagation and to each other and describe polarization of the wave. The energy density of the field is proportional to E^2. It is therefore quite natural to think of the electric field vector as the Schrödinger wave function of the photon, and its square as probability density.

But this Schrödinger picture still needs to be supplemented with Planck-Einstein-deBroglie relations which bind the wave frequency and length to the energy and momentum of photons:

$$E = \hbar\omega \tag{4.136}$$

$$\boldsymbol{p} = \hbar\boldsymbol{k} \tag{4.137}$$

where E is the energy (not the length of the $\boldsymbol{E}$ vector, mind you!) and $\boldsymbol{p}$ is the momentum of the photon. The frequency and the wave number of a photon combine into the speed of light:

$$c = \frac{\omega}{k} \tag{4.138}$$

Photon polarization can be linear, circular and elliptical. We can use symbols $\mid x\rangle$ and $\mid y\rangle$ to describe linearly polarized photons in the x or y directions. By combining states $\mid x\rangle$ and $\mid y\rangle$ it is easy to obtain circularly polarized photons:

$$\mid R\rangle = \frac{1}{\sqrt{2}}\left(\mid x\rangle + i \mid y\rangle\right) \tag{4.139}$$

$$\mid L\rangle = \frac{1}{\sqrt{2}}\left(\mid x\rangle - i \mid y\rangle\right) \tag{4.140}$$

These simple equations can be inverted so that linear polarization states can be expressed in terms of circular polarization states:

$$\mid x\rangle = \frac{1}{\sqrt{2}}\left(\mid R\rangle + \mid L\rangle\right) \tag{4.141}$$

$$\mid y\rangle = \frac{-i}{\sqrt{2}}\left(\mid R\rangle - \mid L\rangle\right) \tag{4.142}$$

It is often more convenient to reason in terms of circular polarization, because it does not favour any direction perpendicular to the direction of motion.

Using Schrödinger equation for photons, i.e., the Maxwell wave equation, we can easily find transition amplitudes between various polarization states. For example assume that we have two polarizers, and we pass photon beam through the first, which polarizes light in the x' direction and then through the second, which polarizes light in the x direction. What is the probability amplitude $\langle x \mid x'\rangle$ that describes this experiment?

Before light was filtered through the first polarizer the photon wave function was:

$$\boldsymbol{E} = E_{x'}\boldsymbol{e}_{x'} + E_{y'}\boldsymbol{e}_{y'}$$

From Maxwell theory we know that the first polarizer lets $E_{x'}\boldsymbol{e}_{x'}$ component of the beam only and filters off the other component. Now we need to express this component in terms of $\boldsymbol{e}_x$ and $\boldsymbol{e}_y$ of the second polarizer (see Figure 4.7).

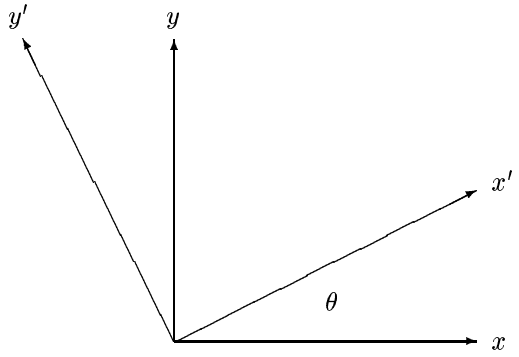


Figure 4.7: Rotating a system of reference

In the system or reference associated with the second polarizer:

$$\boldsymbol{e}_{x'} = e_{x'}{}^{x}\boldsymbol{e}_x + e_{x'}{}^{y}\boldsymbol{e}_y = \cos\theta\boldsymbol{e}_x + \sin\theta\boldsymbol{e}_y$$

consequently

$$E_{x'}\boldsymbol{e}_{x'} = E_{x'}\left(\cos\theta\boldsymbol{e}_x + \sin\theta\boldsymbol{e}_y\right)$$

The x component is the only one that will pass through the second polarizer: $E_{x'}\cos\theta\boldsymbol{e}_x$.

Because the intensity of the beam incident on the second polarizer is proportional to:

$$I_{\text{incident}} \propto E_{x'}^2$$

and the intensity of the transmitted beam is:

$$I_{\text{transmitted}} \propto E_{x'}^2\cos^2\theta$$

the attenuation is $\cos^2\theta$.

This relation translates into:

$$\mid x'\rangle = \cos\theta \mid x\rangle + \sin\theta \mid y\rangle$$

So that the probability amplitude for the transition is

$$\langle x \mid x'\rangle = \cos\theta\langle x \mid x\rangle + \sin\theta\langle x \mid y\rangle = \cos\theta$$

because $\langle x \mid y\rangle = 0$. The probability of the transition is then

$$\left|\langle x \mid x'\rangle\right|^2 = \cos^2\theta$$

4.4 The Berry Phase

In this section we're going to have a look at a very peculiar quantum mechanical effect, which has been discovered only in 1983, even though it is quite elementary and, as you will see, does not involve much heavy mathematics to describe it. The reason it took so long to discover is because people didn't look in the right direction.

Imagine the following situation. A quantum system is moved adiabatically along a closed trajectory. Our task is to describe the change to the system at the point where the trajectory closes. At first glance we would expect that system to evolve back onto its original state since the change has been adiabatic and we end up back at the point of origin. But it turns out that it is not so. In general the phase of the system is going to change due to the excursion in a special way.

This has certain implications for quantum computing. One can think, for example, about a quantum computing gate, which would be based on moving a qubit around in a special way, rather than irradiating the qubit with a pulse. In practice moving qubits is just too hard, so we end up moving the universe around them instead, and this translates back into irradiating the qubit with a pulse. But this time the pulse is going to be more complex than the pulse we've seen in sections that talked about the ammonia molecule.

For the time being though, let us assume that we're going to move the qubit, or some other quantum system, for our discussion is going to be fairly general, around a closed trajectory.

4.4.1 Moving a Qubit in a Circle

As we move the system, the Hamiltonian changes along the trajectory, and the state of the system changes in sympathy with the Hamiltonian. The equation of motion that describes this excursion is:

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t} \mid \Psi(t)\rangle = \boldsymbol{H}(\boldsymbol{r}(t)) \mid \Psi(t)\rangle \tag{4.143}$$

where $\boldsymbol{H}$ is the Hamiltonian and $\boldsymbol{r}$ is a variable that describes the motion. You can think of it for the time being as a guiding vector, but it can be any set of parameters that are used inside the Hamiltonian, so the excursion does not have to occur in the normal 3-dimensional geometric space. It can take place in some other parameter space.

If the motion is adiabatic, then we can also assume that eigenvectors of the Hamiltonian evolve smothly from one position to another one as follows:

$$\boldsymbol{H}(\boldsymbol{r}(t)) \mid n(\boldsymbol{r}(t))\rangle = E_n(\boldsymbol{r}(t)) \mid n(\boldsymbol{r}(t))\rangle \tag{4.144}$$

If a system starts in an eigenstate we can try the following solution

$$\mid \Psi(t)\rangle = e^{-(i/\hbar)\int_0^t E_n(\boldsymbol{r}(t'))\,\mathrm{d}t'} e^{i\gamma_n(t)} \mid n(\boldsymbol{r}(t))\rangle \tag{4.145}$$

In other words, the system is going to remain in the eigenstate, although the eigenstate itself will change, and on top of all that the system is going to accumulate a dynamic and geometrical phase change.

Substituting this solution into the Schrödinger equation yields an equation for $\gamma_n(t)$. But first, before we do this, a few simple evaluations:

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}e^{-(i/\hbar)\int_0^t E_n(\boldsymbol{r}(t'))\mathrm{d}t'} = -\frac{i}{\hbar}E_n(\boldsymbol{r}(t))e^{-(i/\hbar)\int_0^t E_n(\boldsymbol{r}(t'))\mathrm{d}t'} \\
&\frac{\mathrm{d}}{\mathrm{d}t}e^{i\gamma_n(t)} = i\frac{\mathrm{d}\gamma_n(t)}{\mathrm{d}t}e^{i\gamma_n(t)} \\
&\frac{\mathrm{d}}{\mathrm{d}t} \mid n(\boldsymbol{r}(t))\rangle = \mid \boldsymbol{\nabla} n(\boldsymbol{r}(t))\rangle \frac{\mathrm{d}\boldsymbol{r}(t)}{\mathrm{d}t}
\end{aligned}$$

And now let's plug all this into our Schrödinger equation

$$\begin{aligned}
i\hbar\frac{\mathrm{d}}{\mathrm{d}t} \mid \Psi(t)\rangle &= E_n(\boldsymbol{r}(t)) \mid \Psi(t)\rangle - \hbar\frac{\mathrm{d}\gamma_n(t)}{\mathrm{d}t} \mid \Psi(t)\rangle \\
&\quad + i\hbar e^{-(i/\hbar)\int_0^t E_n(\boldsymbol{r}(t'))\mathrm{d}t'} e^{i\gamma_n(t)} \mid \boldsymbol{\nabla} n(\boldsymbol{r}(t))\rangle \frac{\mathrm{d}\boldsymbol{r}(t)}{\mathrm{d}t} \\
&= E_n(\boldsymbol{r}(t)) \mid \Psi(t)\rangle
\end{aligned}$$

In order for this equation to be satisfied, we must make the following vanish:

$$-\frac{\mathrm{d}\gamma_n(t)}{\mathrm{d}t} \mid \Psi(t)\rangle + ie^{-(i/\hbar)\int_0^t E_n(\boldsymbol{r}(t'))\mathrm{d}t'} e^{i\gamma_n(t)} \mid \boldsymbol{\nabla} n(\boldsymbol{r}(t))\rangle \frac{\mathrm{d}\boldsymbol{r}(t)}{\mathrm{d}t} \tag{4.146}$$

Because the two exponenses reside also inside $\mid \Psi(t)\rangle$, we can cancel them out and obtain:

$$\frac{\mathrm{d}\gamma_n(t)}{\mathrm{d}t} \mid n(\boldsymbol{r}(t))\rangle = i \mid \boldsymbol{\nabla} n(\boldsymbol{r}(t))\rangle \frac{\mathrm{d}\boldsymbol{r}(t)}{\mathrm{d}t} \tag{4.147}$$

Now we multiply both sides from the left by $\langle n(\boldsymbol{r}(t)) \mid$ to get:

$$\frac{\mathrm{d}\gamma_n(t)}{\mathrm{d}t} = i\langle n(\boldsymbol{r}(t)) \mid \boldsymbol{\nabla} n(\boldsymbol{r}(t))\rangle \frac{\mathrm{d}\boldsymbol{r}(t)}{\mathrm{d}t} \tag{4.148}$$

and this is our equation for $\gamma_n(t)$.

The solution to this equation is:

$$\gamma_n(C) = i\int_C \langle n(\boldsymbol{r}) \mid \boldsymbol{\nabla} n(\boldsymbol{r})\rangle \,\mathrm{d}\boldsymbol{r} \tag{4.149}$$

where C is a curve traced by $\boldsymbol{R}(t)$. Observe that γ_n depends on the curve itself, not on how slowly or how quickly it is traced by $\boldsymbol{r}(t)$.

It is easy to see that the integral is purely imaginary, so that $\gamma_n(t)$ is real. This follows from the fact that $\langle n \mid n\rangle = 1$:

$$\begin{aligned}
0 &= \boldsymbol{\nabla}\langle n \mid n\rangle \\
&= \langle \boldsymbol{\nabla} n \mid n\rangle + \langle n \mid \boldsymbol{\nabla} n\rangle \\
&= \langle n \mid \boldsymbol{\nabla} n\rangle^* + \langle n \mid \boldsymbol{\nabla} n\rangle \\
&= 2\mathrm{Re}\langle n \mid \boldsymbol{\nabla} n\rangle
\end{aligned}$$

We are now going to rewrite our solution further. First let us replace the curve integral over C with a surface integral making use of the Stokes theorem, and making use of the fact that ours is a closed curve, which encloses a surface, i.e., $C = \partial S$:

$$\oint_{\partial S} \langle n \mid \boldsymbol{\nabla} n \rangle \, \mathrm{d}\boldsymbol{r} = \int_S \boldsymbol{\nabla} \times \langle n \mid \boldsymbol{\nabla} n \rangle \cdot \mathrm{d}^2 \boldsymbol{S} \tag{4.150}$$

Now notice that

$$\boldsymbol{\nabla} \times \langle n \mid \boldsymbol{\nabla} n \rangle = \langle \boldsymbol{\nabla} n \mid \times \mid \boldsymbol{\nabla} n \rangle$$

This is because $\boldsymbol{\nabla} \times \boldsymbol{\nabla} n = 0$. Furthermore we can decompose $\mid \boldsymbol{\nabla} n \rangle$ into the basis states, which yields:

$$\gamma_n(\partial S) = i \int_S \sum_{m \neq n} \langle \boldsymbol{\nabla} n \mid m \rangle \times \langle m \mid \boldsymbol{\nabla} n \rangle \cdot \mathrm{d}^2 \boldsymbol{S} \tag{4.151}$$

We can transform this solution even further by evaluating $\langle m \mid \boldsymbol{\nabla} n \rangle$. We begin by taking a gradient of our adiabatic equation:

$$\boldsymbol{\nabla} \left\{ \boldsymbol{H} \mid n \rangle = E_n \mid n \rangle \right\} \tag{4.152}$$

which yields

$$(\boldsymbol{\nabla} \boldsymbol{H}) \mid n \rangle + \boldsymbol{H} \mid \boldsymbol{\nabla} n \rangle = (\boldsymbol{\nabla} E_n) \mid n \rangle + E_n \mid \boldsymbol{\nabla} n \rangle \tag{4.153}$$

Now we multiply both sides from the left by $\langle m \mid$:

$$\langle m \mid \boldsymbol{\nabla} \boldsymbol{H} \mid n \rangle + \langle m \mid \boldsymbol{H} \mid \boldsymbol{\nabla} n \rangle = \boldsymbol{\nabla} E_n \langle m \mid n \rangle + E_n \langle m \mid \boldsymbol{\nabla} n \rangle \tag{4.154}$$

The term $\langle m \mid \boldsymbol{H} \mid \boldsymbol{\nabla} n \rangle$ translates into $E_m \langle m \mid \boldsymbol{\nabla} n \rangle$, and the term $\boldsymbol{\nabla} E_n \langle m \mid n \rangle$ vanishes, which leaves us with:

$$\langle m \mid \boldsymbol{\nabla} \boldsymbol{H} \mid n \rangle = (E_n - E_m) \langle m \mid \boldsymbol{\nabla} n \rangle \tag{4.155}$$

or

$$\langle m \mid \boldsymbol{\nabla} n \rangle = \frac{\langle m \mid \boldsymbol{\nabla} \boldsymbol{H} \mid n \rangle}{E_n - E_m} \tag{4.156}$$

Substituting this into equation (4.151) yields

$$\gamma_n(\partial S) = i \int_S \sum_{m \neq n} \frac{\langle n \mid \boldsymbol{\nabla} H \mid m \rangle \times \langle m \mid \boldsymbol{\nabla} H \mid n \rangle}{(E_n - E_m)^2} \cdot \mathrm{d}^2 \boldsymbol{S} \tag{4.157}$$

This is the celebrated Berry phase equation, named after M. V. Berry from the H. H. Wills Physics Laboratory of the University of Bristol, who derived it in 1983 [12].

Equation (4.157) represents a rather fundamental observation. It is therefore amazing that it took until 1983 to derive it, sic! This also shows that there may be always something interesting left to discover even in disciplines as well established as Quantum Mechanics.

4.4.2 Berry Phase in the Vicinity of a Degeneracy

A look at equation (4.157) shows that we are going to see problems if surface S crosses through a point in the Hamiltonian parameter space (parametrized by $\boldsymbol{r}$) where a degeneracy occurs, i.e., where two or more different states have the same eigen-energy. But the integral in (4.157) does not depend on the choice of the surface, so it may be possible to deform the surface so as to stay away from the degeneracy. A real problem will occur only if the edge of the surface crosses the degeneracy, or if the degeneracy forms along an infinite line piercing the surface so that no deformation of the surface can take us around it.

If the surface in question is close to the degeneracy point though, then the states which become degenerate at the point dominate the integral.

Let us assume that the degeneracy occurs at $\boldsymbol{r} = \boldsymbol{r}^*$, and let us assume, for simplicity, that just two states become degenerate at $\boldsymbol{r}^*$. A small distance away from $\boldsymbol{r}^*$ the energies of the two states are going to be different, so let us call the one with the higher energy $\mid +\rangle$ and the one with the lower energy $\mid -\rangle$.

The Berry phase for the $\mid +\rangle$ state is then

$$\gamma_+(\partial S) = i \int_S \boldsymbol{V}_+(\boldsymbol{r}) \cdot \mathrm{d}^2\boldsymbol{S} \tag{4.158}$$

where up to $\mathcal{O}(\boldsymbol{r}^2)$:

$$\boldsymbol{V}_+(\boldsymbol{r}) = \frac{\langle +(\boldsymbol{r}) \mid \boldsymbol{\nabla}\boldsymbol{H}(\boldsymbol{r}^*) \mid -(\boldsymbol{r})\rangle \times \langle -(\boldsymbol{r}) \mid \boldsymbol{\nabla}\boldsymbol{H}(\boldsymbol{r}^*) \mid +(\boldsymbol{r})\rangle}{\left(E_+(\boldsymbol{r}) - E_-(\boldsymbol{r})\right)^2} \tag{4.159}$$

It is easy to see that

$$\boldsymbol{V}_-(\boldsymbol{r}) = -\boldsymbol{V}_+(\boldsymbol{r}) \tag{4.160}$$

$$\gamma_-(\partial S) = -\gamma_+(\partial S) \tag{4.161}$$

We can always change the coordinates parametrizing the Hamiltonian so that $\boldsymbol{r}^* = \boldsymbol{0}$ and we can always scale the energy so that $E_\pm(\boldsymbol{r}^*) = 0$ too.

This being the case, and staying with $\boldsymbol{r}$ sufficiently close to $\boldsymbol{0}$ so that nonlinear terms in $\boldsymbol{r}$ can be neglected the two-state Hamiltonian can be parametrized as follows:

$$\boldsymbol{H}(\boldsymbol{r}) = \frac{1}{2}\left(x\boldsymbol{\sigma}_x + y\boldsymbol{\sigma}_y + z\boldsymbol{\sigma}_z\right) \tag{4.162}$$

where $\boldsymbol{\sigma}_x$, $\boldsymbol{\sigma}_y$, and $\boldsymbol{\sigma}_z$ are Pauli matrices. The two eigenvalues that correspond to this Hamiltonian are:

$$E_\pm(\boldsymbol{r}) = \pm\frac{1}{2}\sqrt{x^2 + y^2 + z^2} = \pm\frac{1}{2}r \tag{4.163}$$

$$E_+(\boldsymbol{r}) - E_-(\boldsymbol{r}) = r \tag{4.164}$$

The Hamiltonian eigenvectors $\mid +\rangle$ and $\mid -\rangle$ are *not* in general the eigenvectors of $\boldsymbol{\sigma}_z$, i.e., $\binom{1}{0}$ and $\binom{0}{1}$ in the basis in which Pauli matrices look the way they normally look. The latter are Hamiltonian eigenvectors *only* if $\boldsymbol{r} = r\boldsymbol{e}_z$, where

$\boldsymbol{e}_z$ is a unit vector pointing in the z direction. Otherwise the Hamiltonian eigenvectors will be rotated away from $\binom{1}{0}$ and $\binom{0}{1}$ by some unitary transformation $\boldsymbol{U}$:

$$\begin{aligned} |+\rangle &= \boldsymbol{U}\begin{pmatrix}1\\0\end{pmatrix} \\ |-\rangle &= \boldsymbol{U}\begin{pmatrix}0\\1\end{pmatrix} \end{aligned}$$

The gradient of Hamiltonian (4.162) is simply:

$$\boldsymbol{\nabla H} = \frac{1}{2}\begin{pmatrix}\boldsymbol{\sigma}_x\\ \boldsymbol{\sigma}_y\\ \boldsymbol{\sigma}_z\end{pmatrix} = \frac{1}{2}\vec{\boldsymbol{\sigma}}, \tag{4.165}$$

which yields

$$\begin{aligned} \boldsymbol{V}_+(\boldsymbol{r}) &= \frac{\langle +(\boldsymbol{r}) \mid \vec{\boldsymbol{\sigma}} \mid -(\boldsymbol{r})\rangle \times \langle -(\boldsymbol{r}) \mid \vec{\boldsymbol{\sigma}} \mid +(\boldsymbol{r})\rangle}{4r^2} \\ &= \frac{\left\langle \binom{1}{0}\boldsymbol{U}^\dagger(\boldsymbol{r}) \mid \vec{\boldsymbol{\sigma}} \mid \boldsymbol{U}(\boldsymbol{r})\binom{0}{1}\right\rangle \times \left\langle \binom{0}{1}\boldsymbol{U}^\dagger(\boldsymbol{r}) \mid \vec{\boldsymbol{\sigma}} \mid \boldsymbol{U}(\boldsymbol{r})\binom{1}{0}\right\rangle}{4r^2} \\ &= \frac{\left\langle \binom{1}{0} \mid \boldsymbol{U}^\dagger(\boldsymbol{r})\vec{\boldsymbol{\sigma}}\boldsymbol{U}(\boldsymbol{r}) \mid \binom{0}{1}\right\rangle \times \left\langle \binom{0}{1} \mid \boldsymbol{U}^\dagger(\boldsymbol{r})\vec{\boldsymbol{\sigma}}\boldsymbol{U}(\boldsymbol{r}) \mid \binom{1}{0}\right\rangle}{4r^2} \end{aligned}$$

The degeneracy itself is an isolated point $\boldsymbol{r} = \boldsymbol{0}$ at which x, y, and z vanish. Degeneracies such as this one are said to have *codimension* 3.

Let us assume, for simplicity, that $\boldsymbol{r} = r\boldsymbol{e}_z$. In this case $\boldsymbol{U} = \boldsymbol{1}$, $|+\rangle = \binom{1}{0}$, $|-\rangle = \binom{0}{1}$ and vector $\boldsymbol{V}_+$ simply becomes:

$$\begin{aligned} \boldsymbol{V}_+^x &= \frac{\langle + \mid \boldsymbol{\sigma}_y \mid -\rangle\langle - \mid \boldsymbol{\sigma}_z \mid +\rangle - \langle + \mid \boldsymbol{\sigma}_z \mid -\rangle\langle - \mid \boldsymbol{\sigma}_y \mid +\rangle}{4r^2} = 0 \\ \boldsymbol{V}_+^y &= \frac{\langle + \mid \boldsymbol{\sigma}_z \mid -\rangle\langle - \mid \boldsymbol{\sigma}_x \mid +\rangle - \langle + \mid \boldsymbol{\sigma}_x \mid -\rangle\langle - \mid \boldsymbol{\sigma}_z \mid +\rangle}{4r^2} = 0 \\ \boldsymbol{V}_+^z &= \frac{\langle + \mid \boldsymbol{\sigma}_x \mid -\rangle\langle - \mid \boldsymbol{\sigma}_y \mid +\rangle - \langle + \mid \boldsymbol{\sigma}_y \mid -\rangle\langle - \mid \boldsymbol{\sigma}_x \mid +\rangle}{4r^2} \\ &= \frac{i\langle + \mid +\rangle\langle - \mid -\rangle - (-i)\langle + \mid +\rangle\langle - \mid -\rangle}{4r^2} \\ &= \frac{i}{2r^2} \end{aligned}$$

This result is easy to generalize to any possible angle between $\boldsymbol{r}$ and $\boldsymbol{e}_z$ by expressing it in vector notation:

$$\boldsymbol{V}_+(\boldsymbol{r}) = i\frac{\boldsymbol{r}}{2r^3} \tag{4.166}$$

The corresponding Berry phase is now:

$$\gamma_\pm(\partial S) = \mp\frac{1}{2}\int_S \frac{\boldsymbol{r}}{r^3}\cdot \mathrm{d}^2\boldsymbol{S} \tag{4.167}$$

Because the integral in equation (4.167) simply represents the solid angle $\Omega(S)$ that the surface defined by ∂S subtends with respect to the degeneracy point, we can go this one step further and say that

$$\gamma_{\pm}(\partial S) = \mp\frac{1}{2}\Omega(S), \tag{4.168}$$

which is really a very elegant result.

4.4.3 Special Case: Spin 1/2 in a Magnetic Field

The spin 1/2 case corresponds exactly to the degeneracy case discussed in the previous section with $\boldsymbol{r}$ standing for magnetic field $\boldsymbol{B}$. Berry phase for spin 1/2 particles is therefore given by

$$\gamma_{\pm 1/2}(\partial S) = \mp\frac{1}{2}\int_S \frac{\boldsymbol{B}}{B^3}\cdot \mathrm{d}^2\boldsymbol{S} \tag{4.169}$$

We can use this equation in order to answer the question: what is Berry phase for rotating spin 1/2 immersed in a homogeneous magnetic field by 360°? It is actually easier to do it the other way round, i.e., to rotate the magnetic field around the spin. The contour ∂S can be imagined as a circle of radius B, the centre point of which is coincident with $\boldsymbol{B} = 0$, and with a centrifugal vector $\boldsymbol{B}$ attached to every point on the circle. The interior of the circle is not suitable for surface S, because it passes through $\boldsymbol{B} = 0$. But a half-hemisphere, which rests on the circle is just fine. We can then imagine that field $\boldsymbol{B}$ pierces the hemisphere radially and that its normal component has length B everywhere on this surface, including the circle at its base. The flux of $-\boldsymbol{B}/(2B^3)$ through this surface is then

$$-\frac{B}{2B^3}\frac{4\pi B^2}{2} = -\pi \tag{4.170}$$

And so the Berry phase is $-\pi$ and the result of rotating the field around the spin 1/2 particle is to multiply the state of the particle by

$$e^{i\gamma} = e^{-i\pi} = -1 \tag{4.171}$$

This is actually nothing new: rotating a spin 1/2 particle by 360° results in reversing the sign of its quantum state. One has to make *two* full rotations of a spin 1/2 particle in order to bring it back to the original state.

This fact can be derived in various ways. For example we could use the part of $\mathrm{SL}(2,\mathbb{C})$ that corresponds to rotations and rotate a spinor by 360°, or we could use the Hamilton quaternion formula, and these two approaches are basically the same. But here we have demonstrated this by using the Schrödinger equation and abstaining from the hocus pocus of group representation theory. We showed that this special feature of spin 1/2 follows from the dynamics of quantum mechanics.

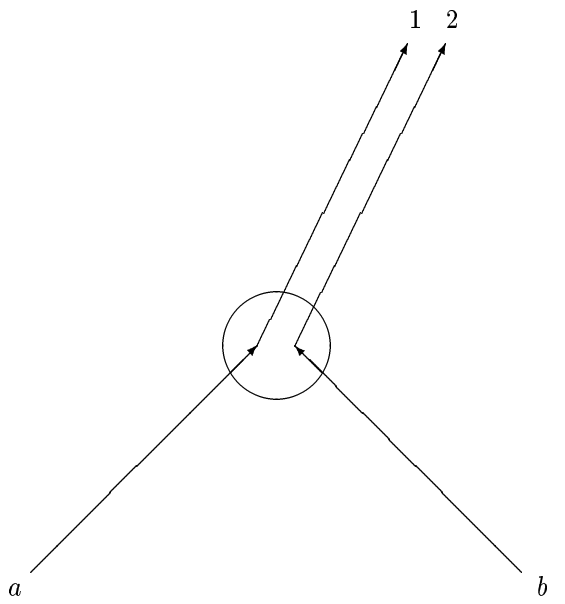


Figure 4.8: A double scattering into a nearby final state

4.5 Multiparticle Systems

Our introduction to Quantum Mechanics has been concerned with single qubits so far. Yet, for any non-trivial computation we need to have lots of them. A single qubit does not computer make. In this section we are going to look at multi-qubit systems. You will learn in the process some quite amazing facts about n-body Quantum Mechanics, about non-locality of Quantum Mechanics, about the kind of interactions that do not exist in classical mechanics at all, and about teleportation too. On the way we are also going to have a second look at the Nuclear Magnetic Resonance, and discuss spin-spin coupling, thermal equilibrium, magnetization readout, and decoherence.

4.5.1 Double Scattering Experiments

Consider the following scattering experiment. We have two particles, which originate at some points a and b. As long as we can *distinguish* these particles we are going to refer to them as "particle a" and "particle b" too. The particles move towards each other and then they interact within a circled region shown in Figure 4.8. From this point onwards they move towards their new destinations labeled 1 and 2 in Figure 4.8.

What is a probability amplitude that describes such a scattering event? Suppose probability amplitudes that particle a goes to 1 and that b goes to 2

are:

$$\langle 1 \mid a \rangle, \quad \text{and} \quad \langle 2 \mid b \rangle$$

The amplitude of both these events happening together is a product of amplitudes for separate processes:

$$\langle 1 \mid a \rangle \langle 2 \mid b \rangle \tag{4.172}$$

so that the resulting probability is a product of probabilities corresponding to the processes:

$$|\langle 1 \mid a \rangle \langle 2 \mid b \rangle|^2 = |\langle 1 \mid a \rangle|^2 \, |\langle 2 \mid b \rangle|^2 = |a_1|^2 \, |b_2|^2 \tag{4.173}$$

This is very much as you would expect probabilities to behave. The logical AND operation translates into a product of probabilities.

Similarly, a probability amplitude that particle a goes to 1 *and* that at the same time particle b goes to 2 is:

$$\langle 2 \mid a \rangle \langle 1 \mid b \rangle \tag{4.174}$$

and the resulting probability is:

$$|\langle 2 \mid a \rangle \langle 1 \mid b \rangle|^2 = |a_2|^2 \, |b_1|^2 \tag{4.175}$$

If there is a single large counter at 1 and 2, which does not resolve between the two locations, then the probability that the counter is going to register two particles is going to be a sum of both possible processes, i.e., that particle a goes to 1 and particle b goes to 2 – this is the first process – and then that particle a goes to 2 and particle b goes to 1 – this is the second process:

$$P_2 = |a_1|^2 \, |b_2|^2 + |a_2|^2 \, |b_1|^2 \tag{4.176}$$

If the two locations 1 and 2 are very close together, and if the interacting region is very small, then the amplitudes a_1 and a_2 are bound to be very close too. Similarly for the amplitudes b_1 and b_2:

$$a_1 \approx a_2 \approx a \qquad \text{and} \qquad b_1 \approx b_2 \approx b \tag{4.177}$$

Probability P_2 then becomes:

$$P_2 = 2 \, |a|^2 \, |b|^2 \tag{4.178}$$

But if both particles are identical Bose particles then this is not what happens. Instead the amplitude for the process turns out to be:

$$\langle 1 \mid a \rangle \langle 2 \mid b \rangle + \langle 2 \mid a \rangle \langle 1 \mid b \rangle \tag{4.179}$$

and the corresponding probability is:

$$P_2 = |a_1 b_2 + a_2 b_1|^2 \approx |2ab|^2 = 4 \, |a|^2 \, |b|^2 \tag{4.180}$$

The probability of finding two identical bosons in an identical final state is twice as large as would be the case for non-identical particles. Bose particles like each other. They tend to clump and behave in an identical manner. This carries to macroscopic systems, for example superfluids, superconductors, and lasers all derive their properties from this Bose particles' affinity for each other.

On the other hand, if the particles are Fermions then the amplitude is:

$$\langle 1 \mid a\rangle\langle 2 \mid b\rangle - \langle 2 \mid a\rangle\langle 1 \mid b\rangle \tag{4.181}$$

And the corresponding probability is:

$$P_2 = |a_1 b_2 - a_2 b_1|^2 \approx |ab - ab|^2 = 0 \tag{4.182}$$

The probability of finding two identical fermions in an identical final state is... zero! Fermi particles try to stay away from each other. No two Fermi particles can occupy the same state. This leads to the Pauli exclusion principle, and ultimately to the "granularity" of what we call "matter" (as opposed to "fields", which do not have this granular feel in thermodynamic limit).

The multiplication of amplitudes for a two-particle system, and their addition or subtraction (as shown above) leads to the following general description of multiplarticle states in quantum mechanics.

Individually each particle is described by a vector in Hilbert space $\mathcal{H}$. If we have two particles, we have two Hilbert spaces, which are formally placed next to each other. Such formal placement, if it obeys additionally rules of linearity for everyone of its components, is called a *tensor product* and is denoted by $\otimes$:

$$\begin{aligned} \mid a\rangle_1 &\in \mathcal{H}_1 \\ \mid b\rangle_2 &\in \mathcal{H}_2 \\ \mid a\rangle_1 \otimes \mid b\rangle_2 &\in \mathcal{H}_1 \otimes \mathcal{H}_2 \end{aligned}$$

If two particles are in such a state they can be easily separated, conceptually and physically. But only particles of different types can be in such states. As we have seen in the example with *identical* Bose and Fermi particles, we can have more *entangled bi-partite* states, for example:

$$\begin{aligned} \mid \Psi_{12}\rangle &= \mid a\rangle_1 \otimes \mid b\rangle_2 + \mid b\rangle_1 \otimes \mid a\rangle_2 \\ \mid \Phi_{12}\rangle &= \mid a\rangle_1 \otimes \mid b\rangle_2 - \mid b\rangle_1 \otimes \mid a\rangle_2 \end{aligned}$$

Particles in these states cannot be easily separated. If you do something to one, it affects the other one and vice versa. Even if the particles are separated by a large physical distance. This has recently been confirmed for distances stretching over some 20 km or more.

Then when you take the actual amplitudes for multi-particle states, you end up with products of amplitudes, which then turn into products of probabilities, unless interference effects affect the outcome. So here, again, we see how the

probability interpretation of quantum mechanics implies the use of tensor products for multi-particle states. If we could drop the probability interpretation, we could drop both the unitarity and the tensor product together with it.

If you have to do a lot of computations with multiparticle systems tensor notation becomes very tedious and so physicists often resort to the following shortcuts:

$$\begin{aligned} \mid \Psi_{12}\rangle &= \mid a\rangle_1 \mid b\rangle_2 + \mid b\rangle_1 \mid a\rangle_2 = \mid ab\rangle + \mid ba\rangle \\ \mid \Phi_{12}\rangle &= \mid a\rangle_1 \mid b\rangle_2 - \mid b\rangle_1 \mid a\rangle_2 = \mid ab\rangle - \mid ba\rangle \end{aligned}$$

In summary, they tend to drop the $\otimes$ sign (very often), and they tend to lump multiple *kets* into a single *ket* (often enough). But you should remember that what hides behind this frivolity is a good old-fashioned tensor product.

The realization that identical particles should be treated differently from non-identical particles goes back to Gibbs, who realized that the phase space volume for N identical particles is $N!$ smaller than what one would have thought it to be if each particle was distinguishable from others. This helped him to calculate entropy of a mixture of two gases of *the same kind* correctly.

In quantum mechanics particle identity leads to even more profound consequences. One can show in various ways that in three and more dimensions two distinct types of particles must exist, and these indeed correspond to bosons and fermions, for whom we must either symmetrize or anti-symmetrize their respective bi-partite wave functions. This, in turn, translates into a special type of repulsion for fermions or attraction for bosons. The fermionic repulsion is the force that supports degenerate matter against gravitational collapse in white dwarfs and in neutron stars. The bosonic attraction is what produces Bose condensates: a very special state of matter, in which light can be slowed down to a crawl.

The separation of quantum particles into bosons and fermions can be related to how multi-particle amplitudes are affected by performing permutations on identical particles. But in two dimensions something very strange happens and this division of all particles into fermions and bosons breaks down. It turns out that we cannot use permutations in this case. Instead we have to use the so called *braids*. This derives from the fact that in two dimensions *it is not enough to specify the initial and final configurations for a system of N particles to completely characterize the system. It is also necessary to specify how the different particle trajectories wind or braid around each other, as the system evolves* [58]. Such particles in the 2-dimensional world, which are neither fermions nor bosons, are called anyons. Anyons were considered a mere mathematical curiosity until Laughlin showed in 1983 that the Fractional Quantum Hall Effect, observed in quantum sheets immersed in very high magnetic fields and cooled to nearly absolute zero, implied that conductivity in these materials was due to anyons [65].

4.5.2 The Computational Basis

Suppose we have a molecule with some ^{1}H, ^{13}C, ^{19}F, and ^{15}N nuclei. Everyone of these nuclei is magnetically active and has spin 1/2. Everyone of these is also going to have some chemical shift associated with it so that its resonant frequency is going to be different from other atoms in the molecule. This way we can talk to each of them on a separate channel without having other nuclei listen to the messages. The spin of each nucleus can be *up* or *down* and we can associate $\mid 1\rangle$ with *up* and $\mid 0\rangle$ with *down*. A particular state of nuclear spins in the molecule can then be, for example,

$$\mid 0\rangle_{\rm H}\otimes \mid 1\rangle_{\rm C}\otimes \mid 1\rangle_{\rm F}\otimes \mid 0\rangle_{\rm N}$$

Because every nucleus here is different, we don't have to symmetrize or anti-symmetrize this state.

We can establish a convention for our particular quantum register associated with this molecule, which says that we always stick to the order shown above, so that we don't have to write the H, C, F, and N subscripts all the time:

$$\mid 0\rangle_{\rm H}\otimes \mid 1\rangle_{\rm C}\otimes \mid 1\rangle_{\rm F}\otimes \mid 0\rangle_{\rm N} =\mid 0\rangle\otimes \mid 1\rangle\otimes \mid 1\rangle\otimes \mid 0\rangle$$

Furthermore, we can switch to the flippant notation used by the physicists and drop the $\otimes$ sign:

$$\mid 0\rangle \mid 1\rangle \mid 1\rangle \mid 0\rangle =\mid 0110\rangle$$

The latter looks already quite like a content of a classical 4-bit register. And this is what it is, the difference being that here we are working with a quantum 4-qubit register.

It's quantumness becomes apparent when you consider the following manipulation. Recall that it took $\pi\hbar/(2\mu\mathcal{E}_0)$ seconds to flip the ammonia molecule from the $\mid +\rangle$ to the $\mid -\rangle$ state or from the $\mid -\rangle$ to the $\mid +\rangle$ state. We have also seen that an identical Hamiltonian:

$$\boldsymbol{H} = A\boldsymbol{\sigma}_z + \mu\mathcal{E}\boldsymbol{\sigma}_x$$

was responsible for flipping spins, where this time μ and $\mathcal{E}$ would relate to magnetic moment and magnetic field. Now, if we were to halve the duration of the pulse to $\pi\hbar/(4\mu\mathcal{E}_0)$, we would put the spin of the nucleus in a superposition of *up* and *down*:

$$|\uparrow\rangle \rightarrow \frac{1}{\sqrt{2}}\left(|\uparrow\rangle + |\downarrow\rangle\right) = \frac{1}{\sqrt{2}}\left(\mid 1\rangle + \mid 0\rangle\right)$$

Suppose that we put our quantum register in a state

$$\mid 0\rangle \mid 0\rangle \mid 0\rangle \mid 0\rangle$$

first, and then rotate each qubit half-way.

How to do the former? The way would be to leave the register immersed in a strong magnetic field, but otherwise isolated from the rest of the universe. The

register should be also cooled to as close to absolute zero as possible. Eventually through *spontanteous emmission* all qubits, which can still do this, will emmit a photon, and flip to a lower energy state.

After the *half way* rotation the resulting state of the register will be:

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

This can be rewritten in the following way:

$$\begin{aligned}
&\frac{1}{\sqrt{2}^4}\big(\mid 0\rangle \mid 0\rangle \mid 0\rangle \mid 0\rangle + \mid 0\rangle \mid 0\rangle \mid 0\rangle \mid 1\rangle \\
&\quad + \mid 0\rangle \mid 0\rangle \mid 1\rangle \mid 0\rangle + \mid 0\rangle \mid 0\rangle \mid 1\rangle \mid 1\rangle + \ldots \\
&\quad + \mid 1\rangle \mid 1\rangle \mid 1\rangle \mid 1\rangle\big) \\
&= \frac{1}{\sqrt{2}^4}(\mid 0000\rangle + \mid 0001\rangle + \mid 0010\rangle + \mid 0011\rangle + \ldots + \mid 1111\rangle) \\
&= \frac{1}{\sqrt{2}^4}(\mid \mathbf{0}\rangle + \mid \mathbf{1}\rangle + \mid \mathbf{2}\rangle + \ldots + \mid \mathbf{15}\rangle)
\end{aligned}$$

where we have used the bold decimal notation **1** through **15** to abbreviate binary qubit representations of the quantum states of the register. We have also *interpreted* qubit positions in the four-qubit sequence in such a way that the rightmost qubit corresponds to the 0^{th} binary position (i.e., $2^0 = 0001_2 = 1_{10}$) and the leftmost qubit corresponds to the 3^{rd} (i.e., $2^3 = 1000_2 = 8_{10}$) binary position.

As you see, by rotating all spins half-way we have filled the register with a superposition of all integer numbers from 0 through 15. This cannot be done in a classical computer, where only one number can reside in a register at a time. The power of quantum computing derives from this ability to cram all n-bit integer numbers into a single quantum register and then process all of them simultaneously.

Almost every quantum algorithms we are going to study in this course is going to begin with this trick.

We can now think of this object,

$$\mid \mathbf{\Psi}\rangle = \frac{1}{\sqrt{2}^4}(\mid \mathbf{0}\rangle + \mid \mathbf{1}\rangle + \mid \mathbf{2}\rangle + \ldots + \mid \mathbf{15}\rangle) \tag{4.183}$$

as a vector in a new 16-dimensional Hilbert space with $\mid \mathbf{0}\rangle$, $\mid \mathbf{1}\rangle$, ..., and $\mid \mathbf{15}\rangle$ the basis vectors of this new Hilbert space. This basis is called the *computational basis*, because it corresponds to integers. Coefficients $1/\sqrt{2}^4$ are probability amplitudes, $\langle \mathbf{0} \mid \mathbf{\Psi}\rangle$, $\langle \mathbf{1} \mid \mathbf{\Psi}\rangle$, ..., $\langle \mathbf{15} \mid \mathbf{\Psi}\rangle$. In other words, if our register is in state $\mid \mathbf{\Psi}\rangle$ then the probability of finding it in, e.g., $\mid \mathbf{7}\rangle$ upon a measurement is $\left(1/\sqrt{2}^4\right)^2 = 1/16.$ and the probability of finding it in, e.g., $\mid \mathbf{11}\rangle$ is the same.

It is generally the case that a tensor product of Hilbert spaces is a Hilbert space, with a scalar product extended to the new Hilbert space in the most

natural way. If $\langle a_1 \mid b_1 \rangle$ and $\langle a_2 \mid b_2 \rangle$ are scalar products in $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively then a scalar product in $\mathcal{H}_1 \otimes \mathcal{H}_2$ is

$$\langle a_1 a_2 \mid b_1 b_2 \rangle = \langle a_1 \mid b_1 \rangle \langle a_2 \mid b_2 \rangle \tag{4.184}$$

4.5.3 Dynamics of Multiparticle Systems

Because multiparticle states are still vectors in a Hilbert space, even though this Hilbert space is now composed of multiple smaller Hilbert spaces, the basic Schrödinger equation we derived in the section about Quantum Evolution, still applies to them. The difference is that when we write

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} \mid \boldsymbol{\Psi}(t) \rangle = \boldsymbol{H} \mid \boldsymbol{\Psi}(t) \rangle \tag{4.185}$$

where $\mid \boldsymbol{\Psi}(t) \rangle$ is defined as in, e.g., equation (4.183), the Hamiltonian matrix is 16×16 instead of 2×2, and the coefficients of $\mid \boldsymbol{\Psi}(t) \rangle$, let us call them $C_0(t)$, $C_1(t)$, ..., $C_{15}(t)$, correspond to projections of $\mid \boldsymbol{\Psi}(t) \rangle$ on multiparticle basis states, e.g., $\mid 0 \rangle \otimes \mid 1 \rangle \otimes \mid 0 \rangle \otimes \mid 1 \rangle = \mid \mathbf{5} \rangle$.

Suppose that the rightmost qubit in the 4-qubit register would evolve *in separation from other qubits* according to some 1-qubit Hamiltonian $\boldsymbol{H}_0$, then the next qubit would evolve in separation from other qubits according to another 1-qubit Hamiltonian $\boldsymbol{H}_1$, and for the remaining two qubits we would have Hamiltonians $\boldsymbol{H}_2$ and $\boldsymbol{H}_3$. Suppose that when you combine all 4 qubits into a 4-qubit register, they still evolve separately, i.e., there is no dynamic coupling between them. What is the Hamiltonian $\boldsymbol{H}$ for this 4-qubit register going to look like?

Recall that the physical meaning of $\boldsymbol{H}$ is energy. If there is no interaction between qubits, then the energy of a 4-qubit register is going to be the sum of energies of individual qubits, in other words:

$$\boldsymbol{H} = \boldsymbol{H}_0 + \boldsymbol{H}_1 + \boldsymbol{H}_2 + \boldsymbol{H}_3 \tag{4.186}$$

The problem with this equation though is that we have four 2×2 matrices on the right hand side and one 16×16 matrix on the left hand side. The reason for this disparity is an intuitive and notational shortcut. What we really mean is:

$$\begin{aligned} \boldsymbol{H} &= \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \otimes \boldsymbol{H}_0 + \mathbf{1} \otimes \mathbf{1} \otimes \boldsymbol{H}_1 \otimes \mathbf{1} \\ &+ \mathbf{1} \otimes \boldsymbol{H}_2 \otimes \mathbf{1} \otimes \mathbf{1} + \boldsymbol{H}_3 \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \end{aligned} \tag{4.187}$$

The tensor notation applied to operators has the following meaning. If $\boldsymbol{U} = \boldsymbol{U}_3 \otimes \boldsymbol{U}_2 \otimes \boldsymbol{U}_1 \otimes \boldsymbol{U}_0$, where $\boldsymbol{U}$ is a 16×16 operator and $\boldsymbol{U}_i$ are 2×2 operators, and if $\mid \boldsymbol{q} \rangle = \mid q_3 \rangle \otimes \mid q_2 \rangle \otimes \mid q_1 \rangle \otimes \mid q_0 \rangle$, where $\mid \boldsymbol{q} \rangle$ is a vector in a 16-dimensional Hilbert space, and $\mid q_i \rangle$ are vectors in four 2-dimensional Hilbert spaces, then

$$\begin{aligned} \boldsymbol{U} \mid \boldsymbol{q} \rangle &= \boldsymbol{U}_3 \otimes \boldsymbol{U}_2 \otimes \boldsymbol{U}_1 \otimes \boldsymbol{U}_0 \mid q_3 \rangle \otimes \mid q_2 \rangle \otimes \mid q_1 \rangle \otimes \mid q_0 \rangle \\ &= \boldsymbol{U}_3 \mid q_3 \rangle \otimes \boldsymbol{U}_2 \mid q_2 \rangle \otimes \boldsymbol{U}_1 \mid q_1 \rangle \otimes \boldsymbol{U}_0 \mid q_0 \rangle \end{aligned}$$

Every matrix element for $\boldsymbol{U}$ can be evaluated as follows:

$$\begin{aligned} U_{q'q} &= \langle \boldsymbol{q}' \mid \boldsymbol{U} \mid \boldsymbol{q} \rangle \\ &= \langle q'_3 q'_2 q'_1 q'_0 \mid \boldsymbol{U}_3 \otimes \boldsymbol{U}_2 \otimes \boldsymbol{U}_1 \otimes \boldsymbol{U}_0 \mid q_3 q_2 q_1 q_0 \rangle \\ &= \langle q'_3 \mid \boldsymbol{U}_3 \mid q_3 \rangle \langle q'_2 \mid \boldsymbol{U}_2 \mid q_2 \rangle \\ &\quad \langle q'_1 \mid \boldsymbol{U}_1 \mid q_1 \rangle \langle q'_0 \mid \boldsymbol{U}_0 \mid q_0 \rangle \end{aligned}$$

For practice let us evaluate $H_{7\,12}$ for Hamiltonian given by equation (4.187):

$$\begin{aligned} H_{7\,12} &= \langle \mathbf{7} \mid \boldsymbol{H} \mid \mathbf{12} \rangle \\ &= \langle 0111 \mid \boldsymbol{H}_3 + \boldsymbol{H}_2 + \boldsymbol{H}_1 + \boldsymbol{H}_0 \mid 1100 \rangle \\ &= \langle 0 \mid \boldsymbol{H}_3 \mid 1 \rangle \langle 1 \mid 1 \rangle \langle 1 \mid 0 \rangle \langle 1 \mid 0 \rangle + \langle 0 \mid 1 \rangle \langle 1 \mid \boldsymbol{H}_2 \mid 1 \rangle \langle 1 \mid 0 \rangle \langle 1 \mid 0 \rangle \\ &\quad + \langle 0 \mid 1 \rangle \langle 1 \mid 1 \rangle \langle 1 \mid \boldsymbol{H}_1 \mid 0 \rangle \langle 1 \mid 0 \rangle + \langle 0 \mid 1 \rangle \langle 1 \mid 1 \rangle \langle 1 \mid 0 \rangle \langle 1 \mid \boldsymbol{H}_0 \mid 0 \rangle \\ &= 0 \end{aligned}$$

From this we can already see that a lot of entries in this matrix are going to be zero. Only entries for which there are at least three digits identical in the corresponding slots in the *bra* and *ket* vectors are going to deliver any non-zero results. For example $\langle 0001 \mid$ and $\mid 0000 \rangle$ may yield a non-zero result when bracketing $\boldsymbol{H}_0$ as may also $\langle 1010 \mid$ and $\mid 1011 \rangle$. But the same pairs will yield zero when bracketing any other $\boldsymbol{H}_{i \neq 0}$, because of the orthogonality of the *bra* and *ket* states for the 0^{th} qubit.

Let us now assume for simplicity that Hamiltonians $\boldsymbol{H}_i$ are diagonal in their respective computational bases, so that

$$\boldsymbol{H}_i \mid 0_i \rangle = E_{i-} \mid 0_i \rangle \quad \text{and} \tag{4.188}$$

$$\boldsymbol{H}_i \mid 1_i \rangle = E_{i+} \mid 1_i \rangle \tag{4.189}$$

It is easy to see that in this case *all* non-diagonal terms in $\boldsymbol{H}$ will vanish. But for every diagonal term we are going to end up with some combination of $E_{3\pm} + E_{2\pm} + E_{1\pm} + E_{0\pm}$. For example:

$$\begin{aligned} H_{7\,7} &= \langle 0111 \mid \boldsymbol{H}_3 + \boldsymbol{H}_2 + \boldsymbol{H}_1 + \boldsymbol{H}_0 \mid 0111 \rangle \\ &= E_{3-} + E_{2+} + E_{1+} + E_{0+} \end{aligned}$$

and

$$\begin{aligned} H_{10\,10} &= \langle 1010 \mid \boldsymbol{H}_3 + \boldsymbol{H}_2 + \boldsymbol{H}_1 + \boldsymbol{H}_0 \mid 1010 \rangle \\ &= E_{3+} + E_{2-} + E_{1+} + E_{0-} \end{aligned}$$

Because every qubit here can have two different energy levels and we have four different qubits, we have altogether $2^4 = 16$ different combinations in which these energies can add, which corresponds to 16 diagonal terms of matrix $\boldsymbol{H}$.

Does this mean that by merely placing all four qubits in a single ensemble we are going to observe $\binom{16}{2} = 120$ different transitions lines, whereas previously

we could see only 4 transitions: $E_{i+} - E_{i-}$? The answer is that no, we are not going to see any more transitions now, and this is because not all levels in a Hamiltonian can be stridden by a transision. Whether a given transition is possible or not possible depends also on *selection rules*. In our case the selection rules simply state that transitions correspond to a flip of a qubit within the 4-qubit register. So there will be single photon transitions between adjacent even and odd states such as $|\ \mathbf{1}\rangle \rightarrow |\ \mathbf{0}\rangle$ and $|\ \mathbf{7}\rangle \rightarrow |\ \mathbf{6}\rangle$, which correspond to the flip of the zeroth qubit. There will be also single photon transitions between state pairs such as $|\ \mathbf{15}\rangle \rightarrow |\ \mathbf{13}\rangle$ and $|\ \mathbf{7}\rangle \rightarrow |\ \mathbf{5}\rangle$, which correspond to the flip of the first qubit. And then we're going to have two-photon transitions such as $|\ \mathbf{7}\rangle \rightarrow |\ \mathbf{4}\rangle$, which correspond to the simultaneous flip of the zeroth and first qubits.

As is the case with vector tensor products, physicists have the tendency to drop the $\otimes$ sign, when taking a tensor product of operators too. We shall follow this tradition reserving the $\otimes$ sign for situations when the tensor character of the product needs to be emphasized.

4.5.4 Spin-Spin Couplings in NMR

Consider again a molecule which contains some combination of magnetically active nuclei of ^{1}H, ^{1}3C, ^{1}9F and ^{1}5N. All these nuclei have spin 1/2. Apart from interacting with their environment, i.e., the background magnetic field, the pulse field, and the electron shell screening, the spins also interact with each other. The spin-spin coupling can be direct, through interaction of magnetic dipoles associated with the spins, or indirect, through the electron cloud. The direct dipolar coupling of two spins is described by the following Hamiltonian:

$$\boldsymbol{H}_{\text{direct}} = \frac{\gamma}{r^3}\left(\boldsymbol{\sigma}_1 \cdot \boldsymbol{\sigma}_2 - 3\,\boldsymbol{\sigma}_1 \cdot \boldsymbol{n}\,\boldsymbol{\sigma}_2 \cdot \boldsymbol{n}\right) \tag{4.190}$$

where γ is a coupling constant (which contains the Planck constant $\hbar$, contributions from magnetic moments of participating spins, and various other coefficients), r is the distance between the spins, $\boldsymbol{\sigma}_i$ is the "vector" of Pauli matrices acting on spin i, and $\boldsymbol{n}$ is the unit length vector in the direction of the line that joins the two nuclei.

The meaning of the notation used above is as follows. The symbol $\boldsymbol{\sigma}_1 \cdot \boldsymbol{\sigma}_2$ really means

$$\boldsymbol{\sigma}_x \otimes \boldsymbol{\sigma}_x + \boldsymbol{\sigma}_y \otimes \boldsymbol{\sigma}_y + \boldsymbol{\sigma}_z \otimes \boldsymbol{\sigma}_z$$

and the symbol $\boldsymbol{\sigma}_1 \cdot \boldsymbol{n}\,\boldsymbol{\sigma}_2 \cdot \boldsymbol{n}$ means

$$(\boldsymbol{\sigma}_x n_x + \boldsymbol{\sigma}_y n_y + \boldsymbol{\sigma}_z n_z) \otimes (\boldsymbol{\sigma}_x n_x + \boldsymbol{\sigma}_y n_y + \boldsymbol{\sigma}_z n_z)$$

In a low-viscosity solution the molecule keeps tumbling quite fast all the time. In the process the coupling term averages away. This can be seen by integrating $\boldsymbol{H}_{\text{direct}}$ for vector $\boldsymbol{n}$ running over the whole sphere.

The electron cloud mediated coupling is described by a Hamiltonian term, which looks as follows:

$$\boldsymbol{H}_{\text{mediated}} = \hbar\omega_{12}\boldsymbol{\sigma}_1 \cdot \boldsymbol{\sigma}_2 \tag{4.191}$$

This term is often rewritten using $\boldsymbol{\sigma}_+$ and $\boldsymbol{\sigma}_-$ matrices, which are defined as follows:

$$\boldsymbol{\sigma}_+ = \boldsymbol{\sigma}_x + i\boldsymbol{\sigma}_y \tag{4.192}$$
$$\boldsymbol{\sigma}_- = \boldsymbol{\sigma}_x - i\boldsymbol{\sigma}_y \tag{4.193}$$

Observe that:

$$\begin{aligned}
&(\boldsymbol{\sigma}_x + i\boldsymbol{\sigma}_y) \otimes (\boldsymbol{\sigma}_x - i\boldsymbol{\sigma}_y) \\
&\quad = \boldsymbol{\sigma}_x \otimes \boldsymbol{\sigma}_x - i\boldsymbol{\sigma}_x \otimes \boldsymbol{\sigma}_y + i\boldsymbol{\sigma}_y \otimes \boldsymbol{\sigma}_x + \boldsymbol{\sigma}_y \otimes \boldsymbol{\sigma}_y \\
&(\boldsymbol{\sigma}_x - i\boldsymbol{\sigma}_y) \otimes (\boldsymbol{\sigma}_x + i\boldsymbol{\sigma}_y) \\
&\quad = \boldsymbol{\sigma}_x \otimes \boldsymbol{\sigma}_x + i\boldsymbol{\sigma}_x \otimes \boldsymbol{\sigma}_y - i\boldsymbol{\sigma}_y \otimes \boldsymbol{\sigma}_x + \boldsymbol{\sigma}_y \otimes \boldsymbol{\sigma}_y
\end{aligned}$$

hence

$$\begin{aligned}
&\boldsymbol{\sigma}_+ \otimes \boldsymbol{\sigma}_- + \boldsymbol{\sigma}_- \otimes \boldsymbol{\sigma}_+ \\
&\quad = 2\left(\boldsymbol{\sigma}_x \otimes \boldsymbol{\sigma}_x + \boldsymbol{\sigma}_y \otimes \boldsymbol{\sigma}_y\right)
\end{aligned}$$

therefore

$$\begin{aligned}
\boldsymbol{H}_{\text{mediated}} &= \hbar\omega_{12}\left(\boldsymbol{\sigma}_x \otimes \boldsymbol{\sigma}_x + \boldsymbol{\sigma}_y \otimes \boldsymbol{\sigma}_y + \boldsymbol{\sigma}_z \otimes \boldsymbol{\sigma}_z\right) \\
&= \hbar\omega_{12}\boldsymbol{\sigma}_z \otimes \boldsymbol{\sigma}_z + \frac{\hbar\omega_{12}}{2}\left(\boldsymbol{\sigma}_+ \otimes \boldsymbol{\sigma}_- + \boldsymbol{\sigma}_- \otimes \boldsymbol{\sigma}_+\right)
\end{aligned}$$

When mediated couplings are weak and when the resonance frequencies of interacting nuclei are well separated then the term proportional to $\boldsymbol{\sigma}_+ \otimes \boldsymbol{\sigma}_- + \boldsymbol{\sigma}_- \otimes \boldsymbol{\sigma}_+$ averages away and we are left with

$$\boldsymbol{H}_{\text{mediated}} = \hbar\omega_{12}\boldsymbol{\sigma}_z \otimes \boldsymbol{\sigma}_z \tag{4.194}$$

which does not go away.

This $\boldsymbol{H}_{\text{mediated}}$ term makes quantum computing possible, because without it we wouldn't be able to couple qubits into two qubit gates. The part of the solution of the two spin Schrödinger equation that is due to this term:

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t} \mid \Psi\rangle_{12} = \ldots + \hbar\omega_{12}\boldsymbol{\sigma}_z \otimes \boldsymbol{\sigma}_z \mid \Psi\rangle_{12} + \ldots$$

looks as follows:

$$\mid \Psi(t)\rangle_{12} = e^{-i\omega_{12}t\boldsymbol{\sigma}_z \otimes \boldsymbol{\sigma}_z} \mid \Psi(0)\rangle_{12}$$

There is, in other words, a separate frequency associated with this coupling. Signals of this frequency sent to the molecule are received by the coupling part of the molecule, which then evolves according to the strength, polarization, and duration of the pulse.

Of course, if you don't send any signals to the molecule, the coupling part still contributes to its evolution in the way shown above.

We are going to see an example of this in the next section.

4.5.5 The Controlled NOT Gate

The controlled NOT gate is a gate that acts on two qubits. The first qubit is called a control qubit, and the second qubit is the data qubit. If the control qubit is $\mid 0\rangle$ then the data qubit is left alone. If the control qubit is $\mid 1\rangle$ the data qubit is flipped. The gate, for which we are going to use the $\oplus$ symbol, can be therefore described by the following table:

$$\begin{aligned}
\oplus :\mid 0\rangle \mid 0\rangle &\rightarrow \mid 0\rangle \mid 0\rangle \\
\oplus :\mid 0\rangle \mid 1\rangle &\rightarrow \mid 0\rangle \mid 1\rangle \\
\oplus :\mid 1\rangle \mid 0\rangle &\rightarrow \mid 1\rangle \mid 1\rangle \\
\oplus :\mid 1\rangle \mid 1\rangle &\rightarrow \mid 1\rangle \mid 0\rangle
\end{aligned}$$

or, using the computational basis:

$$\begin{aligned}
\oplus :\mid \mathbf{0}\rangle &\rightarrow \mid \mathbf{0}\rangle \\
\oplus :\mid \mathbf{1}\rangle &\rightarrow \mid \mathbf{1}\rangle \\
\oplus :\mid \mathbf{2}\rangle &\rightarrow \mid \mathbf{3}\rangle \\
\oplus :\mid \mathbf{3}\rangle &\rightarrow \mid \mathbf{2}\rangle
\end{aligned}$$

Its corresponding matrix $\oplus_{ij}$ in the computational basis is thus:

$$\begin{aligned}
\oplus_{00} &= \langle \mathbf{0} \mid \oplus \mid \mathbf{0}\rangle = \langle \mathbf{0} \mid \mathbf{0}\rangle = 1 \\
\oplus_{01} &= \langle \mathbf{0} \mid \oplus \mid \mathbf{1}\rangle = \langle \mathbf{0} \mid \mathbf{1}\rangle = 0 \\
\oplus_{02} &= \langle \mathbf{0} \mid \oplus \mid \mathbf{2}\rangle = \langle \mathbf{0} \mid \mathbf{3}\rangle = 0 \\
\oplus_{03} &= \langle \mathbf{0} \mid \oplus \mid \mathbf{3}\rangle = \langle \mathbf{0} \mid \mathbf{2}\rangle = 0 \\
\oplus_{10} &= \langle \mathbf{1} \mid \oplus \mid \mathbf{0}\rangle = \langle \mathbf{1} \mid \mathbf{0}\rangle = 0 \\
\oplus_{11} &= \langle \mathbf{1} \mid \oplus \mid \mathbf{1}\rangle = \langle \mathbf{1} \mid \mathbf{1}\rangle = 1 \\
\oplus_{12} &= \langle \mathbf{1} \mid \oplus \mid \mathbf{2}\rangle = \langle \mathbf{1} \mid \mathbf{3}\rangle = 0 \\
\oplus_{13} &= \langle \mathbf{1} \mid \oplus \mid \mathbf{3}\rangle = \langle \mathbf{1} \mid \mathbf{2}\rangle = 0 \\
\oplus_{20} &= \langle \mathbf{2} \mid \oplus \mid \mathbf{0}\rangle = \langle \mathbf{2} \mid \mathbf{0}\rangle = 0 \\
\oplus_{21} &= \langle \mathbf{2} \mid \oplus \mid \mathbf{1}\rangle = \langle \mathbf{2} \mid \mathbf{1}\rangle = 0 \\
\oplus_{22} &= \langle \mathbf{2} \mid \oplus \mid \mathbf{2}\rangle = \langle \mathbf{2} \mid \mathbf{3}\rangle = 0 \\
\oplus_{23} &= \langle \mathbf{2} \mid \oplus \mid \mathbf{3}\rangle = \langle \mathbf{2} \mid \mathbf{2}\rangle = 1 \\
\oplus_{30} &= \langle \mathbf{3} \mid \oplus \mid \mathbf{0}\rangle = \langle \mathbf{3} \mid \mathbf{0}\rangle = 0 \\
\oplus_{31} &= \langle \mathbf{3} \mid \oplus \mid \mathbf{1}\rangle = \langle \mathbf{3} \mid \mathbf{1}\rangle = 0 \\
\oplus_{32} &= \langle \mathbf{3} \mid \oplus \mid \mathbf{2}\rangle = \langle \mathbf{3} \mid \mathbf{3}\rangle = 1 \\
\oplus_{33} &= \langle \mathbf{3} \mid \oplus \mid \mathbf{3}\rangle = \langle \mathbf{3} \mid \mathbf{2}\rangle = 0
\end{aligned}$$

In summary:

$$\oplus = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{4.195}$$

Now we have to figure out how to implement this gate in, e.g., NMR. Recall the formula:

$$e^{i\boldsymbol{\sigma}_z\phi} = \mathbf{1}\cos\phi + i\boldsymbol{\sigma}_z\sin\phi \tag{4.196}$$

For $\phi = \pi/4$ we have $\sin\frac{\pi}{4} = \cos\frac{\pi}{4} = 1/\sqrt{2}$, so that

$$e^{i\boldsymbol{\sigma}_z\pi/4} = \frac{1}{\sqrt{2}}(\mathbf{1} + i\boldsymbol{\sigma}_z) \tag{4.197}$$

There is a similar formula for $e^{i\boldsymbol{\sigma}_z\otimes\boldsymbol{\sigma}_z\phi}$, namely

$$e^{i\boldsymbol{\sigma}_z\otimes\boldsymbol{\sigma}_z\phi} = \mathbf{1}\otimes\mathbf{1}\cos\phi + i\boldsymbol{\sigma}_z\otimes\boldsymbol{\sigma}_z\sin\phi \tag{4.198}$$

which for $\phi = \pi/4$ becomes

$$e^{i\boldsymbol{\sigma}_z\otimes\boldsymbol{\sigma}_z\pi/4} = \frac{1}{\sqrt{2}}(\mathbf{1}\otimes\mathbf{1} + i\boldsymbol{\sigma}_z\otimes\boldsymbol{\sigma}_z) \tag{4.199}$$

For $e^{-i\boldsymbol{\sigma}_z\pi/4}$ we just change the sign in front of i on the right hand side. Now we can easily see that:

$$\begin{aligned} e^{-i\boldsymbol{\sigma}_z\pi/4} \mid 0\rangle &= \frac{1}{\sqrt{2}}(1-i) \mid 0\rangle \\ e^{-i\boldsymbol{\sigma}_z\pi/4} \mid 1\rangle &= \frac{1}{\sqrt{2}}(1+i) \mid 1\rangle \end{aligned}$$

and

$$\begin{aligned} e^{-i\boldsymbol{\sigma}_z\pi/4}\otimes e^{-i\boldsymbol{\sigma}_z\pi/4} \mid 00\rangle &= \frac{1}{2}(1-i)(1-i) \mid 00\rangle = -i \mid 00\rangle \\ e^{-i\boldsymbol{\sigma}_z\pi/4}\otimes e^{-i\boldsymbol{\sigma}_z\pi/4} \mid 01\rangle &= \frac{1}{2}(1-i)(1+i) \mid 01\rangle = \mid 01\rangle \\ e^{-i\boldsymbol{\sigma}_z\pi/4}\otimes e^{-i\boldsymbol{\sigma}_z\pi/4} \mid 10\rangle &= \frac{1}{2}(1+i)(1-i) \mid 10\rangle = \mid 10\rangle \\ e^{-i\boldsymbol{\sigma}_z\pi/4}\otimes e^{-i\boldsymbol{\sigma}_z\pi/4} \mid 11\rangle &= \frac{1}{2}(1+i)(1+i) \mid 11\rangle = i \mid 11\rangle \end{aligned}$$

Similarly

$$\begin{aligned} e^{i\boldsymbol{\sigma}_z\otimes\boldsymbol{\sigma}_z\pi/4} \mid 00\rangle &= \frac{1}{\sqrt{2}}(1+i) \mid 00\rangle \\ e^{i\boldsymbol{\sigma}_z\otimes\boldsymbol{\sigma}_z\pi/4} \mid 01\rangle &= \frac{1}{\sqrt{2}}(1-i) \mid 01\rangle \\ e^{i\boldsymbol{\sigma}_z\otimes\boldsymbol{\sigma}_z\pi/4} \mid 10\rangle &= \frac{1}{\sqrt{2}}(1-i) \mid 10\rangle \\ e^{i\boldsymbol{\sigma}_z\otimes\boldsymbol{\sigma}_z\pi/4} \mid 01\rangle &= \frac{1}{\sqrt{2}}(1+i) \mid 11\rangle \end{aligned}$$

Now, let us combine $e^{i\boldsymbol{\sigma}_z\otimes\boldsymbol{\sigma}_z\pi/4}$ and $e^{-i\boldsymbol{\sigma}_z\pi/4}\otimes e^{-i\boldsymbol{\sigma}_z\pi/4}$:

$$\begin{aligned}
e^{i\boldsymbol{\sigma}_z\otimes\boldsymbol{\sigma}_z\pi/4}e^{-i\boldsymbol{\sigma}_z\pi/4}\otimes e^{-i\boldsymbol{\sigma}_z\pi/4}\mid 00\rangle &= \frac{1}{\sqrt{2}}(1+i)(-i)\mid 00\rangle = \frac{1}{\sqrt{2}}(1-i)\mid 00\rangle\\
e^{i\boldsymbol{\sigma}_z\otimes\boldsymbol{\sigma}_z\pi/4}e^{-i\boldsymbol{\sigma}_z\pi/4}\otimes e^{-i\boldsymbol{\sigma}_z\pi/4}\mid 01\rangle &= \frac{1}{\sqrt{2}}(1-i)(1)\mid 01\rangle = \frac{1}{\sqrt{2}}(1-i)\mid 01\rangle\\
e^{i\boldsymbol{\sigma}_z\otimes\boldsymbol{\sigma}_z\pi/4}e^{-i\boldsymbol{\sigma}_z\pi/4}\otimes e^{-i\boldsymbol{\sigma}_z\pi/4}\mid 10\rangle &= \frac{1}{\sqrt{2}}(1-i)(1)\mid 10\rangle = \frac{1}{\sqrt{2}}(1-i)\mid 10\rangle\\
e^{i\boldsymbol{\sigma}_z\otimes\boldsymbol{\sigma}_z\pi/4}e^{-i\boldsymbol{\sigma}_z\pi/4}\otimes e^{-i\boldsymbol{\sigma}_z\pi/4}\mid 11\rangle &= \frac{1}{\sqrt{2}}(1+i)(i)\mid 11\rangle = -\frac{1}{\sqrt{2}}(1-i)\mid 11\rangle
\end{aligned}$$

OK, let us now multiply all the four equations from the left by $1/\sqrt{2}\,(1+i)$ and let us switch to the computational basis to get:

$$\begin{aligned}
\frac{1}{\sqrt{2}}(1+i)e^{i\boldsymbol{\sigma}_z\otimes\boldsymbol{\sigma}_z\pi/4}e^{-i\boldsymbol{\sigma}_z\pi/4}\otimes e^{-i\boldsymbol{\sigma}_z\pi/4}\mid \mathbf{0}\rangle &= \mid \mathbf{0}\rangle\\
\frac{1}{\sqrt{2}}(1+i)e^{i\boldsymbol{\sigma}_z\otimes\boldsymbol{\sigma}_z\pi/4}e^{-i\boldsymbol{\sigma}_z\pi/4}\otimes e^{-i\boldsymbol{\sigma}_z\pi/4}\mid \mathbf{1}\rangle &= \mid \mathbf{1}\rangle\\
\frac{1}{\sqrt{2}}(1+i)e^{i\boldsymbol{\sigma}_z\otimes\boldsymbol{\sigma}_z\pi/4}e^{-i\boldsymbol{\sigma}_z\pi/4}\otimes e^{-i\boldsymbol{\sigma}_z\pi/4}\mid \mathbf{2}\rangle &= \mid \mathbf{2}\rangle\\
\frac{1}{\sqrt{2}}(1+i)e^{i\boldsymbol{\sigma}_z\otimes\boldsymbol{\sigma}_z\pi/4}e^{-i\boldsymbol{\sigma}_z\pi/4}\otimes e^{-i\boldsymbol{\sigma}_z\pi/4}\mid \mathbf{3}\rangle &= -\mid \mathbf{3}\rangle
\end{aligned}$$

The corresponding matrix is

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

This is the *controlled-$\boldsymbol{\sigma}_z$ gate*. If the first (control) qubit is $\mid 0\rangle$ the gate doesn't do anything to the second (data) qubit. But if the first qubit is $\mid 1\rangle$ then $\boldsymbol{\sigma}_z$ is applied to the data qubit.

There is a transformation called a Hadamard gate. This transformation is defined as follows:

$$\boldsymbol{H} = \frac{1}{\sqrt{2}}\,(\boldsymbol{\sigma}_x + \boldsymbol{\sigma}_z) \tag{4.200}$$

Do not mistake the Hadamard transformation for the Hamiltonian, even though we have used the same symbol here for both. It is unfortunate that there are only so many letters in the alphabet. In Quantum Computing we don't really make much use of a Hamiltonian at all, therefore this usually doesn't lead to confusion. Anyhow, it is easy to check that:

$$\begin{aligned}
\boldsymbol{H}\cdot\boldsymbol{H} &= \mathbf{1}\\
\boldsymbol{H}\cdot\boldsymbol{\sigma}_z\cdot\boldsymbol{H} &= \boldsymbol{\sigma}_x
\end{aligned}$$

This means that if we place a Hadamard gate on the data line both before and after the controlled-$\boldsymbol{\sigma}_z$ gate, then the combined effect will be a controlled-$\boldsymbol{\sigma}_x$ gate. And, when you look at the $\boldsymbol{\sigma}_x$ matrix, you realize quickly that it is a NOT matrix. And so with the help of two Hadamard gates bracketing our controlled-$\boldsymbol{\sigma}_z$ gate we have constructed a controlled-NOT gate.

Now, how to construct a Hadamard gate? This is a much easier gate to construct, because it is a single qubit gate, which does not involve any couplings. This gate is in fact equivalent to rotating a qubit half-way, and we already know how to do this. The Hadamard gate matrix looks as follows:

$$\boldsymbol{H} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{4.201}$$

This implies that it acts in the following way on $\mid 0\rangle$ and $\mid 1\rangle$:

$$\begin{aligned} \boldsymbol{H} \mid 0\rangle &= \frac{1}{\sqrt{2}}(\mid 0\rangle + \mid 1\rangle) \\ \boldsymbol{H} \mid 1\rangle &= \frac{1}{\sqrt{2}}(\mid 0\rangle - \mid 1\rangle) \end{aligned}$$

But this is not the only way to implement a controlled-NOT gate NMR. You will find the following alternative prescription in [67]:

$$\begin{aligned} \oplus^{1\leftrightarrow 2} = \; & e^{-i(\pi/4)\boldsymbol{\sigma}_y^{(1)}} e^{i(\pi/4)\boldsymbol{\sigma}_x^{(1)}} e^{i(\pi/4)\boldsymbol{\sigma}_y^{(1)}} \\ & e^{-i(\pi/4)\boldsymbol{\sigma}_x^{(2)}} e^{i(\pi/4)\boldsymbol{\sigma}_y^{(2)}} e^{-i(\pi/4)\boldsymbol{\sigma}_z^{(1)}\otimes\boldsymbol{\sigma}_z^{(2)}} e^{-i(\pi/4)\boldsymbol{\sigma}_y^{(2)}} \end{aligned} \tag{4.202}$$

where superscripts (1) and (2) mean that the operator acts on spin 1 or spin 2.

EXERCISE

Prove that this is indeed a controlled-NOT gate. Which is the control qubit, and which is the data qubit?

The crucial feature of this expression is again the coupling term $e^{-i(\pi/4)\boldsymbol{\sigma}_z^{(1)}\otimes\boldsymbol{\sigma}_z^{(2)}}$.

Observe that in both prescriptions for the controlled-NOT gate the coupling term, and, in case of the first prescription also the non-coupled terms, represent free evolution of the system according to a free NMR Hamiltonian. This is what the system does anyway when you don't treat it with RF pulses. The trick therefore will be how to *stop* this free evolution of the molecule when you don't want any couplings and when you don't want any controlled-NOT gates.

We are going to discuss procedures for doing just this in the next section.

The moral of this section is that the spin-spin coupling in the NMR Hamiltonian is crucial in delivering a controlled-NOT gate. You will learn later that this is one of the fundamental gates for quantum computation and, in particular, that *all* multi-input quantum gates can be built from this gate and from other single qubit gates. We can therefore make an even stronger statement: the spin-spin coupling in the NMR Hamiltonian makes quantum computation possible.

4.5.6 Halting and Reversing Time

A free Hamiltonian for a molecule suspended in a strong vertical magnetic field and in absence of radiation pulses comprises two terms:

$$\boldsymbol{H} = -\frac{1}{2}\sum_i \hbar\omega_i \boldsymbol{\sigma}_z^{(i)} + \sum_{i<j} \hbar\omega_{ij} \boldsymbol{\sigma}_z^{(i)} \otimes \boldsymbol{\sigma}_z^{(j)} \tag{4.203}$$

The solution to the corresponding Schrödinger equation is

$$|\ \boldsymbol{\Psi}(t)\rangle = \prod_i e^{i\omega_i t \boldsymbol{\sigma}_z^{(i)}/2} \prod_{i<j} e^{-i\omega_{ij} t \boldsymbol{\sigma}_z^{(i)} \otimes \boldsymbol{\sigma}_z^{(j)}}\ |\ \boldsymbol{\Psi}(\mathbf{0})\rangle \tag{4.204}$$

You can see that normally all spins in a molecule are coupled with each other. Moreover you can also see that the whole system is going to evolve in some way, even if it is left to itself. The coupling terms will make every spin pair evolve in a way that is used in the controlled-NOT gate. So if we tried to use this system without doing anything special to it, we would end up having controlled-NOT gates all over the place, with everything coupled and we would have individual spins evolving too. But no controlled *computation* could be carried out.

The way to resolve this problem is to develop a procedure that stops time for all couplings and for all individual spins in the molecule. Then if we need a selected coupling for a controlled-NOT gate, or if we want to push a given spin forward in time by just about that much, we can *release* just the coupling or just the spin, so that the required free evolution for them takes place for a precisely determined amount of time. Then we stop time for the coupling and for the spin again and perform whatever other operations are required to finish the controlled-NOT gate.

The key to the time reversal procedure is the observation that Pauli matrices anticommute with each other and that they all square to one. Hence, for example,

$$\begin{aligned} \boldsymbol{\sigma}_x\boldsymbol{\sigma}_z + \boldsymbol{\sigma}_z\boldsymbol{\sigma}_x &= 0 \quad |\cdot \boldsymbol{\sigma}_x \\ \boldsymbol{\sigma}_x\boldsymbol{\sigma}_z\boldsymbol{\sigma}_x + \boldsymbol{\sigma}_z &= 0 \quad |- \boldsymbol{\sigma}_z \\ \boldsymbol{\sigma}_x\boldsymbol{\sigma}_z\boldsymbol{\sigma}_x &= -\boldsymbol{\sigma}_z \end{aligned}$$

And the same also holds for a tensor product of two $\boldsymbol{\sigma}_z$ matrices:

$$\boldsymbol{\sigma}_x^{(2)}\boldsymbol{\sigma}_z^{(1)} \otimes \boldsymbol{\sigma}_z^{(2)}\boldsymbol{\sigma}_x^{(2)} = -\boldsymbol{\sigma}_z^{(1)} \otimes \boldsymbol{\sigma}_z^{(2)}$$

And since

$$e^{i\boldsymbol{\sigma}_z^{(1)} \otimes \boldsymbol{\sigma}_z^{(2)}\phi} = \mathbf{1} \otimes \mathbf{1} \cos\phi + i\boldsymbol{\sigma}_z^{(1)} \otimes \boldsymbol{\sigma}_z^{(2)} \sin\phi$$

it is easy to see that

$$\boldsymbol{\sigma}_x^{(2)} e^{i\boldsymbol{\sigma}_z^{(1)} \otimes \boldsymbol{\sigma}_z^{(2)}\phi} \boldsymbol{\sigma}_x^{(2)} = e^{-i\boldsymbol{\sigma}_z^{(1)} \otimes \boldsymbol{\sigma}_z^{(2)}\phi}$$

and similarly, though with slightly less fuss:

$$\boldsymbol{\sigma}_x e^{i\boldsymbol{\sigma}_z\phi} \boldsymbol{\sigma}_x = e^{-i\boldsymbol{\sigma}_z\phi}$$

So by bracketing the free evolution term with the NOT gate, because this is what $\boldsymbol{\sigma}_x$ is, we can reverse the time flow both for a single spin and for couplings.

And how do we go about halting time? It's easy. We let it go for a little while forward, then we make it go backward for the same amount of time. For a selected coupling between spins 1 and 2 this could look as follows:

$$\begin{aligned} & e^{i\omega_{12}\Delta t\boldsymbol{\sigma}_z^{(1)}\otimes\boldsymbol{\sigma}_z^{(2)}}e^{-i\omega_{12}\Delta t\boldsymbol{\sigma}_z^{(1)}\otimes\boldsymbol{\sigma}_z^{(2)}} \\ & \quad = \boldsymbol{\sigma}_x^{(2)}e^{-i\omega_{12}\Delta t\boldsymbol{\sigma}_z^{(1)}\otimes\boldsymbol{\sigma}_z^{(2)}}\boldsymbol{\sigma}_x^{(2)}e^{-i\omega_{12}\Delta t\boldsymbol{\sigma}_z^{(1)}\otimes\boldsymbol{\sigma}_z^{(2)}} \end{aligned}$$

Operators of the form $e^{i\boldsymbol{A}}$ and $e^{i\boldsymbol{B}}$ do not commute in general. This is easy to understand, when you recall that $e^{i\boldsymbol{A}}$ represents a unitary operator, which is a rotation in Hilbert space. And rotations do not commute in general, not even in $\mathbb{R}^3$. In general

$$\left[e^{a\boldsymbol{A}}, e^{b\boldsymbol{B}}\right] = ab\left[\boldsymbol{A}, \boldsymbol{B}\right] + \text{higher order terms}$$

But it is quite easy to see that if $\boldsymbol{A}$ is any $n \times n$ complex matrix and $a, b \in \mathbb{R}$ then [4]:

$$e^{(a+b)\boldsymbol{A}} = e^{a\boldsymbol{A}}e^{b\boldsymbol{A}} = e^{b\boldsymbol{A}}e^{a\boldsymbol{A}}$$

This implies that matrices of the form $e^{i\boldsymbol{\sigma}_z\phi}$ with various values and signs of ϕ commute, and therefore we can cancel a given coupling *globally*, rather than *locally*, i.e., all we need to make sure of is that we reverse time for the coupling for exactly half of the total time used by a computational procedure, to eliminate its effect. Elimination of a coupling between spins 1 and 2 occurs when we reverse time flow for just one of the spins. If we reverse time flow for both spins at the same time, then the coupling will keep evolving forward in time.

The procedure to stop all couplings in a molecule would therefore be to

1. divide the whole computation time into an even number of short intervals of length Δt,

2. come up with a scheme to reverse time for each spin in selected intervals in such a way that every coupling is going to move forward and backward for the same number of intervals

Consider a simple case of two spins, whose resonance frequencies are ω_1 and ω_2. The time halting scheme for the two spin coupling can be described in terms of the following table:

	Δt	Δt
ω_1	$\rightarrow$	$\rightarrow$
ω_2	$\rightarrow$	$\leftarrow$

where $\rightarrow$ signifies that time flows forward for this spin and in this time interval and $\leftarrow$ signifies that time flows backward for this spin and in this time interval. Here we have just two spins and one coupling. The first spin moves forward all the time, but the second spin moves forward in the first interval and backward

in the second interval and therefore the coupling between them moves forward in the first interval and backward in the second interval.

Now consider a case of 4 spins. The corresponding table that describes a decoupling for this system is:

	Δt	Δt	Δt	Δt
ω_1	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\rightarrow$
ω_2	$\rightarrow$	$\rightarrow$	$\leftarrow$	$\leftarrow$
ω_3	$\rightarrow$	$\leftarrow$	$\leftarrow$	$\rightarrow$
ω_4	$\rightarrow$	$\leftarrow$	$\rightarrow$	$\leftarrow$

The trick is to ensure that any two rows must disagree in exactly half of the entries.

This table can be converted into an appropriate sequence of pulses by sending a brief $\boldsymbol{\sigma}_x$ pulse to a given spin just before and after its corresponding $\leftarrow$ time segment. But observe that some economies are possible. For example if for a given spin two reversed time segments are adjacent, there is no need to send two $\boldsymbol{\sigma}_x$ pulses to the spin between the segments, because $\boldsymbol{\sigma}_x\boldsymbol{\sigma}_x = \mathbf{1}$.

> The resulting sequence of $\boldsymbol{\sigma}_x$ pulses sent to various spins at various frequencies constitutes a kind of music. You can associate a pitch with every frequency, and you can associate bars with time segments and bar lines with borders between the time segments. The duration of the $\boldsymbol{\sigma}_x$ notes should be very short, so we can think of them as semiquavers played at the beginning and end of each sequence of bars representing time reversal.
>
> The result of playing this music is that the evolution of all mediated spin-spin couplings in the system stops. But looking at the tables you can see that the first spin continues to evolve, although evolution is stopped for all other spins, because, like the couplings, they are all pushed forward for half the time, and then pushed backward for half the time too.
>
> You will see below that just a small change in our procedure is going to stop not only the evolution of the couplings but also the evolution of all individual spins, including the first one too.

The tables shown above can be converted to matrices by replacing $\rightarrow$ with $+1$ and $\leftarrow$ with -1. The first table for a two spin system would then look like this:

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

and the second table, which describes decoupling for a system of four spins would look like this:

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix}$$

The first matrix, the 2×2 one, looks like the Hadamard matrix we have encountered in the previous section. It turns out that $n \times n$ Hadamard matrices will do the job here for a system of n spins, and they'll do it very efficiently too.

The definition of a Hadamard matrix of order n is that it is an $n \times n$ matrix with entries ± 1, such that

$$\boldsymbol{H}\boldsymbol{H}^T = n\mathbf{1} \tag{4.205}$$

The rows of a Hadamard matrix are pair-wise orthogonal, therefore any two rows agree in exactly half of the entries, which is what is required to stop time for all couplings in a system of n spins. Columns of a Hadamard matrix are pair-wise orthogonal too.

Permutations or negations of rows or columns in Hadamard matrices leave the orthogonality condition unchanged. We can therefore transform Hadamard matrices into each other by these means. Hadamard matrices, which look like the ones above, i.e., with their first row and first column all comprising $+1$ are said to be *normalized*.

In order to stop the evolution of all couplings and all spins at the same time, we need to do the following. First let us construct an oversized normalized Hadamard matrix for a system of $n+2$ spins, where n is an even number. Then we are simply going to remove the top row, which contains $+1$ terms only, and the next row below it, so that what's left is a matrix $n \times (n+2)$, in which *all* rows sum up to zeros. This matrix stops not only all couplings, but also evolution of all individual spins in the system. The evolution table for a system of two spins would then look like the bottom half of the table we have originally used for a 4 spin system:

	Δt	Δt	Δt	Δt
ω_1	$\rightarrow$	$\leftarrow$	$\leftarrow$	$\rightarrow$
ω_1	$\rightarrow$	$\leftarrow$	$\rightarrow$	$\leftarrow$

Observe that this time every spin evolves forward for half the time, and then backward for half the time too, and... on top of this, the coupling itself also evolves forward for half the time and then backward for half the time.

To activate just a particular coupling without allowing other couplings or any individual spins to evolve, all we need to do is to make rows corresponding to spins, which need to be coupled, identical. For example, the following table implements a *re-coupling* between spins 3 and 4, out of 9, with all other couplings disabled and with all spins, including spins 3 and 4 halted in their evolution too:

	Δt	Δt	Δt	Δt	Δt	Δt	Δt	Δt	Δt	Δt	Δt	Δt
ω_1	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\leftarrow$	$\leftarrow$	$\leftarrow$	$\rightarrow$	$\leftarrow$	$\rightarrow$	$\leftarrow$	$\leftarrow$
ω_2	$\rightarrow$	$\leftarrow$	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\leftarrow$	$\leftarrow$	$\leftarrow$	$\rightarrow$	$\leftarrow$	$\rightarrow$	$\leftarrow$
ω_3	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\leftarrow$	$\leftarrow$	$\rightarrow$	$\leftarrow$	$\leftarrow$	$\rightarrow$	$\leftarrow$	$\leftarrow$	$\rightarrow$
ω_4	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\leftarrow$	$\leftarrow$	$\rightarrow$	$\leftarrow$	$\leftarrow$	$\rightarrow$	$\leftarrow$	$\leftarrow$	$\rightarrow$
ω_5	$\rightarrow$	$\leftarrow$	$\leftarrow$	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\leftarrow$	$\leftarrow$	$\leftarrow$	$\rightarrow$	$\leftarrow$	$\rightarrow$
ω_6	$\rightarrow$	$\rightarrow$	$\leftarrow$	$\leftarrow$	$\rightarrow$	$\rightarrow$	$\leftarrow$	$\rightarrow$	$\leftarrow$	$\leftarrow$	$\rightarrow$	$\leftarrow$
ω_7	$\rightarrow$	$\leftarrow$	$\leftarrow$	$\leftarrow$	$\leftarrow$	$\leftarrow$	$\leftarrow$	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\rightarrow$
ω_8	$\rightarrow$	$\leftarrow$	$\rightarrow$	$\leftarrow$	$\leftarrow$	$\rightarrow$	$\rightarrow$	$\leftarrow$	$\leftarrow$	$\rightarrow$	$\rightarrow$	$\leftarrow$
ω_9	$\rightarrow$	$\rightarrow$	$\leftarrow$	$\rightarrow$	$\leftarrow$	$\leftarrow$	$\rightarrow$	$\leftarrow$	$\leftarrow$	$\leftarrow$	$\rightarrow$	$\rightarrow$

4.5.7 The Feynman Quantum Computer

In this section we are going to look at a specific example of a multi-qubit system, so that you can go through the process of deriving a Hamiltonian for such a system, then solving the equations of motion and making some observations about the evolution of the computer.

Feynman computer discussed here uses a 3 qubit program counter register and a 1 qubit data register, on which we perform a simple computation. As the computation proceeds we keep looking at the counter every now and then to assess whether the computation has completed. But we don't look at the data register itself, since this would destroy the computation. We look at the data register only when the counter register tells us that it is now safe to do so.

The computation we are going to perform comprises two steps. a $\sqrt{\neg}$ gate is applied at each step, so that the final computation delivers $\sqrt{\neg}\sqrt{\neg} = \neg$, i.e., the logical NOT gate.

The $\sqrt{\neg}$ gate

The logical NOT gate can be described by the following simple matrix:

$$\neg = \boldsymbol{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{4.206}$$

The way it acts on the two basis states of a qubit are:

$$\neg \mid 0\rangle = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \mid 1\rangle \tag{4.207}$$

$$\neg \mid 1\rangle = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \mid 0\rangle \tag{4.208}$$

$$\tag{4.209}$$

The square root of NOT is a gate such that the square of its matrix is equal to the matrix of the NOT gate.

The following satisfies this requirement:

$$\sqrt{\neg} = \frac{1}{\sqrt{2}} \begin{pmatrix} e^{i\pi/4} & e^{-i\pi/4} \\ e^{-i\pi/4} & e^{i\pi/4} \end{pmatrix} \tag{4.210}$$

Because $e^{i\pi/4} = \frac{1}{\sqrt{2}}(1+i)$ and $e^{-i\pi/4} = \frac{1}{\sqrt{2}}(1-i)$ this can be also written as

$$\sqrt{\neg} = \frac{1}{2} \begin{pmatrix} 1+i & 1-i \\ 1-i & 1+i \end{pmatrix} \tag{4.211}$$

It is easy to show that $\sqrt{\neg}$ is unitary and that $\sqrt{\neg}\sqrt{\neg} = \neg$.

$$\sqrt{\neg}\sqrt{\neg} \;=\; \frac{1}{\sqrt{2}} \begin{pmatrix} e^{i\pi/4} & e^{-i\pi/4} \\ e^{-i\pi/4} & e^{i\pi/4} \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} e^{i\pi/4} & e^{-i\pi/4} \\ e^{-i\pi/4} & e^{i\pi/4} \end{pmatrix}$$

$$\begin{aligned} &= \frac{1}{2}\begin{pmatrix} e^{i\pi/2}+e^{-i\pi/2} & e^0+e^0 \\ e^0+e^0 & e^{-i\pi/2}+e^{i\pi/2} \end{pmatrix} \\ &= \frac{1}{2}\begin{pmatrix} 2\cos\frac{\pi}{2} & 2 \\ 2 & 2\cos\frac{\pi}{2} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{aligned}$$

Similarly:

$$\begin{aligned} \sqrt{\neg}\sqrt{\neg}^\dagger &= \frac{1}{\sqrt{2}}\begin{pmatrix} e^{i\pi/4} & e^{-i\pi/4} \\ e^{-i\pi/4} & e^{i\pi/4} \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} e^{-i\pi/4} & e^{i\pi/4} \\ e^{i\pi/4} & e^{-i\pi/4} \end{pmatrix} \\ &= \frac{1}{2}\begin{pmatrix} e^0+e^0 & e^{i\pi/2}+e^{-i\pi/2} \\ e^{-i\pi/2}+e^{i\pi/2} & e^0+e^0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

In summary $\sqrt{\neg}$ is unitary and it squares to $\neg$.

What is the result of this operator acting on $\mid 0\rangle$ and $\mid 1\rangle$?

$$\begin{aligned} \sqrt{\neg}\mid 0\rangle &= \frac{1}{\sqrt{2}}\begin{pmatrix} e^{i\pi/4} & e^{-i\pi/4} \\ e^{-i\pi/4} & e^{i\pi/4} \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \frac{1}{\sqrt{2}}\begin{pmatrix} e^{i\pi/4} \\ e^{-i\pi/4} \end{pmatrix} = \frac{1}{\sqrt{2}}\left(e^{i\pi/4}\mid 0\rangle + e^{-i\pi/4}\mid 1\rangle\right) \end{aligned}$$

and similarly:

$$\sqrt{\neg}\mid 1\rangle = \frac{1}{\sqrt{2}}\left(e^{-i\pi/4}\mid 0\rangle + e^{i\pi/4}\mid 1\rangle\right)$$

In a Feynman computer the gates are entangled with counter qubits, so that by looking at the counter qubit we can tell if a given operation has completed. In order to implement a $\neg$ circuit using two $\sqrt{\neg}$ gates we will need 3 counter qubits and one state qubit. When the $\sqrt{\neg}$ gate acts on the whole system it leaves counter qubits alone and rotates the state qubit only:

$$\sqrt{\neg}_4 = \mathbf{1}\otimes\mathbf{1}\otimes\mathbf{1}\otimes\sqrt{\neg} \tag{4.212}$$

The full circuit of Feynman computer implements

$$\sqrt{\neg}_4\sqrt{\neg}_4$$

The $\sqrt{\neg}_4$ can be represented in a matrix form. In order to do that we need to enumerate states of the 4-qubit system. For example:

$$\begin{aligned} \mid 0000\rangle &= \mid \mathbf{0}\rangle \\ \mid 0001\rangle &= \mid \mathbf{1}\rangle \\ \mid 0010\rangle &= \mid \mathbf{2}\rangle \\ \mid 0011\rangle &= \mid \mathbf{3}\rangle \\ \mid 0100\rangle &= \mid \mathbf{4}\rangle \\ \mid 0101\rangle &= \mid \mathbf{5}\rangle \\ \mid 0110\rangle &= \mid \mathbf{6}\rangle \end{aligned}$$

$$
\begin{aligned}
\mid 0111\rangle &= \mid \mathbf{7}\rangle \\
\mid 1000\rangle &= \mid \mathbf{8}\rangle \\
\mid 1001\rangle &= \mid \mathbf{9}\rangle \\
\mid 1010\rangle &= \mid \mathbf{10}\rangle \\
\mid 1011\rangle &= \mid \mathbf{11}\rangle \\
\mid 1100\rangle &= \mid \mathbf{12}\rangle \\
\mid 1101\rangle &= \mid \mathbf{13}\rangle \\
\mid 1110\rangle &= \mid \mathbf{14}\rangle \\
\mid 1111\rangle &= \mid \mathbf{15}\rangle
\end{aligned}
$$

It is easy to see what the $\sqrt{\neg}_4$ matrix is going to look like in this basis. The matrix will be 16×16. It will comprise diagonal blocks 2×2, which correspond to the normal 2×2 version of $\sqrt{\neg}$. For example, consider terms $\sqrt{\neg}_4[8,8]$ (where indexes run from 0 through 15), $\sqrt{\neg}_4[8,9]$, $\sqrt{\neg}_4[9,8]$, and $\sqrt{\neg}_4[9,9]$. Column 8 of $\sqrt{\neg}_4$ represents the result of acting with matrix $\sqrt{\neg}_4[n,m]$ on vector

$$
\begin{pmatrix}
0 \\
\vdots \\
0 \\
1 \leftarrow 8^{\text{th}} \text{ location} \\
0 \\
\vdots \\
0
\end{pmatrix}
$$

again the location is measured from 0 through 15. This is vector $\mid \mathbf{8}\rangle = \mid 1000\rangle$. Acting with $\sqrt{\neg}_4$ on this vector yields:

$$
\begin{aligned}
\sqrt{\neg}_4 \mid \mathbf{8}\rangle &= \sqrt{\neg}_4 \mid 1000\rangle = \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \otimes \sqrt{\neg} \mid 1000\rangle \\
&= \mid 100\rangle \otimes \sqrt{\neg} \mid 0\rangle \\
&= \mid 100\rangle \otimes \frac{1}{\sqrt{2}} \left(e^{i\pi/4} \mid 0\rangle + e^{-i\pi/4} \mid 1\rangle \right) \\
&= \frac{1}{\sqrt{2}} \left(e^{i\pi/4} \mid 1000\rangle + e^{-i\pi/4} \mid 1001\rangle \right) = \frac{1}{\sqrt{2}} \left(e^{i\pi/4} \mid \mathbf{8}\rangle + e^{-i\pi/4} \mid \mathbf{9}\rangle \right)
\end{aligned}
$$

and similarly

$$
\begin{aligned}
\sqrt{\neg}_4 \mid \mathbf{9}\rangle &= \sqrt{\neg}_4 \mid 1001\rangle = \mid 100\rangle \otimes \frac{1}{\sqrt{2}} \left(e^{-i\pi/4} \mid 0\rangle + e^{i\pi/4} \mid 1\rangle \right) \\
&= \frac{1}{\sqrt{2}} \left(e^{-i\pi/4} \mid \mathbf{8}\rangle + e^{i\pi/4} \mid \mathbf{9}\rangle \right)
\end{aligned}
$$

Because the 8^{th} column of matrix $\sqrt{\neg}_4$ is the result of this matrix acting on vector $\mid \mathbf{8}\rangle$ and the 9^{th} column of this matrix is the result of it acting on vector

$\mid \mathbf{9}\rangle$ we have found that in the vicinity of [8,8] the matrix has the form:

$$\frac{1}{\sqrt{2}}\begin{pmatrix} \ddots & \vdots & \vdots & \\ \cdots & e^{i\pi/4} & e^{-i\pi/4} & \cdots \\ \cdots & e^{-i\pi/4} & e^{i\pi/4} & \cdots \\ & \vdots & \vdots & \ddots \end{pmatrix}$$

But this is our original $\sqrt{\neg}$, and, since this computation can be applied in the same manner to all other vectors we arrive at the following form of matrix $\sqrt{\neg}_4$:

$$\sqrt{\neg}_4 = \begin{pmatrix} \sqrt{\neg} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sqrt{\neg} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sqrt{\neg} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sqrt{\neg} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sqrt{\neg} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sqrt{\neg} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sqrt{\neg} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sqrt{\neg} \end{pmatrix} \tag{4.213}$$

where $\mathbf{0}$ is a 2×2 zero matrix.

The Hamiltonian

The operator $\sqrt{\neg}_4$ defines the behaviour of the $\sqrt{\neg}$ gate as applied to the state qubit and to the counter qubits. But this is not the Hamiltonian matrix, which we could plug into the Schrödinger equation in order to simulate the behaviour of the whole system.

There are various ways in which gates such as this one can be implemented. Feynman's prescription is based on his model of spin waves propagating in a 1-D chain of molecules. The resulting Hamiltonian is given by the following formula:

$$\boldsymbol{H} = \sum_{i=0}^{k-1} \boldsymbol{c}_{i+1}\boldsymbol{a}_i\boldsymbol{M}_{i+1} + (\boldsymbol{c}_{i+1}\boldsymbol{a}_i\boldsymbol{M}_{i+1})^\dagger \tag{4.214}$$

Here $\boldsymbol{M}_i$ is the i^{th} gate, $\boldsymbol{c}_i$ is a "creation" operator that promotes a qubit from $\mid 0\rangle$ to $\mid 1\rangle$ in the i^{th} location, and $\boldsymbol{a}_i$ is the annihilation operator that demotes a qubit in the i^{th} location from $\mid 1\rangle$ to $\mid 0\rangle$

The creation and annihilation operators are defined operationally as follows:

$$\boldsymbol{c} \mid 0\rangle = \mid 1\rangle \tag{4.215}$$

$$\boldsymbol{c} \mid 1\rangle = \mathbf{0} \tag{4.216}$$

$$\boldsymbol{a} \mid 0\rangle = \mathbf{0} \tag{4.217}$$

$$\boldsymbol{a} \mid 1\rangle = \mid 0\rangle \tag{4.218}$$

$\mathbf{0}$ is the zero vector. It is obtained by taking any vector, e.g., $\mid 0\rangle$ and multiplying it by zero:

$$\mathbf{0} = 0 \mid 0\rangle \tag{4.219}$$

But $0 \mid 1\rangle = \mathbf{0}$ too. Using matrix notation:

$$\boldsymbol{c} \mid 0\rangle = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \mid 1\rangle \tag{4.220}$$

$$\boldsymbol{c} \mid 1\rangle = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \mathbf{0} \tag{4.221}$$

$$\boldsymbol{a} \mid 0\rangle = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \mathbf{0} \tag{4.222}$$

$$\boldsymbol{a} \mid 1\rangle = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \mid 0\rangle \tag{4.223}$$

The operators $\boldsymbol{c}_i$ and $\boldsymbol{a}_i$ are defined as follows:

$$\boldsymbol{a}_0 = \boldsymbol{a} \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \tag{4.224}$$

$$\boldsymbol{a}_1 = \mathbf{1} \otimes \boldsymbol{a} \otimes \mathbf{1} \otimes \mathbf{1} \tag{4.225}$$

$$\boldsymbol{a}_2 = \mathbf{1} \otimes \mathbf{1} \otimes \boldsymbol{a} \otimes \mathbf{1} \tag{4.226}$$

$$\boldsymbol{c}_0 = \boldsymbol{c} \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \tag{4.227}$$

$$\boldsymbol{c}_1 = \mathbf{1} \otimes \boldsymbol{c} \otimes \mathbf{1} \otimes \mathbf{1} \tag{4.228}$$

$$\boldsymbol{c}_2 = \mathbf{1} \otimes \mathbf{1} \otimes \boldsymbol{c} \otimes \mathbf{1} \tag{4.229}$$

The last qubit is reserved for the computation itself.

What does the Hamiltonian matrix in our basis $\mid \mathbf{0}\rangle, \ldots, \mid \mathbf{15}\rangle$ look like? For our 2-gate computer the Hamiltonian is given by:

$$\boldsymbol{H} = \boldsymbol{c}_1 \boldsymbol{a}_0 \sqrt{\neg}_4 + \left(\boldsymbol{c}_1 \boldsymbol{a}_0 \sqrt{\neg}_4\right)^\dagger + \boldsymbol{c}_2 \boldsymbol{a}_1 \sqrt{\neg}_4 + \left(\boldsymbol{c}_2 \boldsymbol{a}_1 \sqrt{\neg}_4\right)^\dagger \tag{4.230}$$

We could evaluate the Hamiltonian matrix by applying this Hamiltonian laboriously to all vectors in our basis, but it is easy to see that most of those applications will result in zeros. Consider the first term:

$$\boldsymbol{c}_1 \boldsymbol{a}_0 \sqrt{\neg}_4 \mid xyzs\rangle$$

This will generate a $\mathbf{0}$ vector whenever there is $\mid 0\rangle$ in the 0^{th} position or $\mid 1\rangle$ in the 1^{st} position. The only vectors that will survive the slaughter will therefore be:

$$\begin{aligned} \mid 1000\rangle &= \mid \mathbf{8}\rangle \\ \mid 1001\rangle &= \mid \mathbf{9}\rangle \\ \mid 1010\rangle &= \mid \mathbf{10}\rangle \\ \mid 1011\rangle &= \mid \mathbf{11}\rangle \end{aligned}$$

In turn

$$\boldsymbol{c}_2 \boldsymbol{a}_1 \sqrt{\neg}_4 \mid xyzs\rangle$$

will generate a $\mathbf{0}$ vector whenever there is $\mid 1\rangle$ in the 2^{nd} position or $\mid 0\rangle$ in the 1^{st} position. The only vectors that are going to survive this bloodbath will be:

$$\begin{aligned} \mid 0100\rangle &= \mid \mathbf{4}\rangle \\ \mid 0101\rangle &= \mid \mathbf{5}\rangle \\ \mid 1100\rangle &= \mid \mathbf{12}\rangle \\ \mid 1101\rangle &= \mid \mathbf{13}\rangle \end{aligned}$$

The effect of the other part of the Hamiltonian:

$$\left(\boldsymbol{c}_1\boldsymbol{a}_0\sqrt{\neg}_4\right)^\dagger + \left(\boldsymbol{c}_2\boldsymbol{a}_1\sqrt{\neg}_4\right)^\dagger$$

will be to add symmetric but complex-conjugated terms under the diagonal, so we don't have to worry about those. In summary, instead of evaluating the effect of $\boldsymbol{c}_1\boldsymbol{a}_0\sqrt{\neg}_4$ and of $\boldsymbol{c}_2\boldsymbol{a}_1\sqrt{\neg}_4$ on 16 vectors we only need to do that for 8 vectors. But we can save work even here. Consider the application of $\boldsymbol{c}_1\boldsymbol{a}_0\sqrt{\neg}_4$ to $\mid 100\rangle\otimes\mid s\rangle$, i.e., to either $\mid \mathbf{8}\rangle$ or to $\mid \mathbf{9}\rangle$:

$$\begin{aligned} \boldsymbol{c}_1\boldsymbol{a}_0\sqrt{\neg}_4 \mid 100\rangle\otimes\mid s\rangle &= \boldsymbol{c}_1\boldsymbol{a}_0 \mid 100\rangle \otimes \sqrt{\neg} \mid s\rangle \\ &= \boldsymbol{a} \mid 1\rangle \otimes \boldsymbol{c} \mid 0\rangle\otimes \mid 0\rangle \otimes \sqrt{\neg} \mid s\rangle = \mid 0\rangle \mid 1\rangle \mid 0\rangle\sqrt{\neg} \mid s\rangle \\ &= \mid 010\rangle \otimes \sqrt{\neg} \mid s\rangle \end{aligned}$$

Now if we substitute either $\mid 0\rangle$ or $\mid 1\rangle$ in place of $\mid s\rangle$ we will place a small 2×2 $\sqrt{\neg}$ matrix in the location given by

$$\begin{array}{rl} \text{columns} & \mid 1000\rangle = \mid \mathbf{8}\rangle \\ & \mid 1001\rangle = \mid \mathbf{9}\rangle \\ \text{rows} & \mid 0100\rangle = \mid \mathbf{4}\rangle \\ & \mid 0101\rangle = \mid \mathbf{5}\rangle \end{array}$$

The other two vectors, which the first term in the Hamiltonian acts on (without destroying them) are $\mid \mathbf{10}\rangle$ and $\mid \mathbf{11}\rangle$:

$$\boldsymbol{c}_1\boldsymbol{a}_0\sqrt{\neg}_4 \mid 101\rangle\otimes\mid s\rangle = \mid 011\rangle \otimes \sqrt{\neg} \mid s\rangle$$

This therefore will place the 2×2 $\sqrt{\neg}$ matrix in the following location:

$$\begin{array}{rl} \text{columns} & \mid 1010\rangle = \mid \mathbf{10}\rangle \\ & \mid 1011\rangle = \mid \mathbf{11}\rangle \\ \text{rows} & \mid 0110\rangle = \mid \mathbf{6}\rangle \\ & \mid 0111\rangle = \mid \mathbf{7}\rangle \end{array}$$

Repeating the same reasoning with the terms that survive $\boldsymbol{c}_2\boldsymbol{a}_1\sqrt{\neg}_4$ yields:

$$\begin{aligned} \boldsymbol{c}_2\boldsymbol{a}_1\sqrt{\neg}_4 \mid 010\rangle\otimes\mid s\rangle &= \mid 001\rangle \otimes \sqrt{\neg} \mid s\rangle \\ \boldsymbol{c}_2\boldsymbol{a}_1\sqrt{\neg}_4 \mid 110\rangle\otimes\mid s\rangle &= \mid 101\rangle \otimes \sqrt{\neg} \mid s\rangle \end{aligned}$$

The resulting 2×2 $\sqrt{\neg}$ matrices will therefore be placed in

$$\begin{array}{rl} \text{columns} & |\,0100\rangle = |\,\mathbf{4}\rangle \\ & |\,0101\rangle = |\,\mathbf{5}\rangle \\ \text{rows} & |\,0010\rangle = |\,\mathbf{2}\rangle \\ & |\,0011\rangle = |\,\mathbf{3}\rangle \end{array}$$

and

$$\begin{array}{rl} \text{columns} & |\,1100\rangle = |\,\mathbf{12}\rangle \\ & |\,1101\rangle = |\,\mathbf{13}\rangle \\ \text{rows} & |\,1010\rangle = |\,\mathbf{10}\rangle \\ & |\,1011\rangle = |\,\mathbf{11}\rangle \end{array}$$

Thus the upper part of the Hamiltonian matrix looks as follows:

$$\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & \mathbf{0} & \sqrt{\neg} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & & \mathbf{0} & \mathbf{0} & \sqrt{\neg} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & & & \mathbf{0} & \mathbf{0} & \sqrt{\neg} & \mathbf{0} & \mathbf{0} \\ & & & & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & & & & & \mathbf{0} & \sqrt{\neg} & \mathbf{0} \\ & & & & & & \mathbf{0} & \mathbf{0} \\ & & & & & & & \mathbf{0} \end{pmatrix} \tag{4.231}$$

All we need to do now in order to finish the Hamiltonian matrix is to fill the lower part of the matrix with the transposed and complex-conjucated mirror of the upper part:

$$\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sqrt{\neg} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sqrt{\neg}^{\dagger} & \mathbf{0} & \mathbf{0} & \sqrt{\neg} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sqrt{\neg} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sqrt{\neg}^{\dagger} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sqrt{\neg}^{\dagger} & \mathbf{0} & \mathbf{0} & \sqrt{\neg} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sqrt{\neg}^{\dagger} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \tag{4.232}$$

The unitary evolution operator

With the Hamiltonian (4.232) in place we can write the Schrödinger equation for the system:

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t} |\,\Psi(t)\rangle = \boldsymbol{H} |\,\Psi(t)\rangle \tag{4.233}$$

With the initial condition

$$|\,\Psi(0)\rangle = |\,1000\rangle$$

or

$$\mid \Psi(0)\rangle = \mid 1001\rangle$$

We should then let the computer evolve according to the Schrödinger equation until it reaches state

$$\mid 001\rangle \otimes \mid s\rangle$$

The computer is constructed in such a way that when the counter qubits become $\mid 001\rangle$ the $\mid s\rangle$ qubit is guaranteed to become the inverse of whatever it was set to at the beginning. In other words, the computation will yield a $\neg$ gate. This is accomplished by linking together two $\sqrt{\neg}$ gates and coupling them to counter qubits. Because the Hamiltonian does not depend on time. We can solve the Schrödinger equation analytically. The solution is simply given by:

$$\mid \Psi(t)\rangle = e^{-i\boldsymbol{H}t/\hbar} \mid \Psi(0)\rangle \tag{4.234}$$

How to evaluate $e^{-i\boldsymbol{H}t/\hbar}$ though? The most straightforward way is to use the expansion:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots \tag{4.235}$$

In our case this will become:

$$e^{-i\boldsymbol{H}t/\hbar} = \mathbf{1} - i\frac{t}{\hbar}\boldsymbol{H} - \frac{t^2}{2!\hbar^2}\boldsymbol{HH} + i\frac{t^3}{3!\hbar^3}\boldsymbol{HHH} + \frac{t^4}{4!\hbar^4}\boldsymbol{HHHH} + \dots \tag{4.236}$$

Because t can be anything, whereas $\boldsymbol{H}$ is fixed, it may be convenient to take $t/\hbar$ out of it and simply evaluate:

$$e^{-i\boldsymbol{H}} = \boldsymbol{U}_0 = \mathbf{1} - i\boldsymbol{H} - \frac{1}{2!}\boldsymbol{HH} + i\frac{1}{3!}\boldsymbol{HHH} + \frac{1}{4!}\boldsymbol{HHHH} + \dots \tag{4.237}$$

then

$$e^{-i\boldsymbol{H}t/\hbar} = \boldsymbol{U}_0^{t/\hbar} \tag{4.238}$$

Because $\boldsymbol{H}$ is a 16×16 matrix, but with a clear 2×2 blocking structure, we can keep those blocks and the final result for $\boldsymbol{U}_0$ is again a 16×16 matrix with a clear 2×2 blocking structure. Matrix $\boldsymbol{U}_0$ looks as follows:

$$\begin{pmatrix}
\mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \boldsymbol{B} & \boldsymbol{D} & \mathbf{0} & \boldsymbol{G} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & -\boldsymbol{D}^\dagger & \boldsymbol{A} & \mathbf{0} & \boldsymbol{D} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{B} & \mathbf{0} & \boldsymbol{D} & \boldsymbol{G} & \mathbf{0} \\
\mathbf{0} & \boldsymbol{G} & -\boldsymbol{D}^\dagger & \mathbf{0} & \boldsymbol{B} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & -\boldsymbol{D}^\dagger & \mathbf{0} & \boldsymbol{A} & \boldsymbol{D} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{G} & \mathbf{0} & -\boldsymbol{D}^\dagger & \boldsymbol{B} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1}
\end{pmatrix} \tag{4.239}$$

where

$$\begin{aligned} \boldsymbol{A} &= \cos\sqrt{2}\,\mathbf{1} \\ \boldsymbol{B} &= \frac{1}{2}\left(\cos\sqrt{2}+1\right)\mathbf{1} \\ \boldsymbol{G} &= \frac{1}{2}\left(\cos\sqrt{2}-1\right)\boldsymbol{\sigma}_x \\ \boldsymbol{D} &= \frac{-i\sin\sqrt{2}}{\sqrt{2}}\sqrt{\neg} \end{aligned}$$

And how does this come about? Well, in order to see this, we would have to multiply the Hamiltonian laboriously great many times, perform the additions in the exponential expansion, keep all computations on the analytical level and then observe what every term in the matrix converges to. This is a job for a symbolic manipulation system.

Having done all that how do we then raise $\boldsymbol{U}_0$ to the power of $t/\hbar$, where $t/\hbar$ can be any positive real number? At this stage we have to bite the bullet and diagonalize matrix $\boldsymbol{U}_0$, because an operation such as $\boldsymbol{U}_0^{t/\hbar}$ acquires its very meaning this way. This operation will yield 16 eigenvalues, λ_0 through λ_{15}, and 16 eigenvectors, $\mid v_0\rangle$ through $\mid v_{15}\rangle$. Our initial value vector $\mid \Psi(0)\rangle$ can be written in term of the eigenvectors as:

$$\mid \Psi(0)\rangle = \sum_{i=0}^{15} \mid v_i\rangle\langle v_i \mid \Psi(0)\rangle \tag{4.240}$$

When the operator $\boldsymbol{U}_0^{t/\hbar}$ acts on the eigenvectors it stretches them by $\lambda_i^{t/\hbar}$, so that

$$\mid \Psi(t)\rangle = \boldsymbol{U}_0^{t/\hbar}\sum_{i=0}^{15} \mid v_i\rangle\langle v_i \mid \Psi(0)\rangle = \sum_{i=0}^{15} \lambda_i^{t/\hbar} \mid v_i\rangle\langle v_i \mid \Psi(0)\rangle \tag{4.241}$$

In order to interpret the solution in physical terms, i.e., in terms of our vectors $\mid \mathbf{0}\rangle$ through $\mid \mathbf{15}\rangle$ we will have to write $\mid \Psi(t)\rangle$ in terms of these vectors. This can be done simply by writing the eigenvectors in terms of $\mid \mathbf{0}\rangle$ through $\mid \mathbf{15}\rangle$:

$$\begin{aligned} \mid \Psi(t)\rangle &= \sum_{i=0}^{15} \lambda_i^{t/\hbar} \mid v_i\rangle\langle v_i \mid \Psi(0)\rangle \\ &= \sum_{i=0}^{15} \lambda_i^{t/\hbar} \left(\sum_{k=0}^{15} \mid \boldsymbol{k}\rangle\langle \boldsymbol{k} \mid v_i\rangle\right) \langle v_i \mid \Psi(0)\rangle \end{aligned} \tag{4.242}$$

But for integer values of $t/\hbar$ you can simply apply $\boldsymbol{U}_0$ $t/\hbar$-times to vector $\mid \Psi(0)\rangle$ and... you'll be done!

The Evolution of the Feynman Computer

The file that contains the simulator lives in the directory

/afs/ovpit.indiana.edu/common/mathematica/quantum_computing/unix/nbooks3

file

feynman.nb

In order to gain access to this directory you must have a valid AFS token in our AFS cell.

Here is an example of how I go about running it. My machine is

`beige.ucs.indiana.edu`

and I don't have Mathematica installed on it. But Mathematica is installed on the Nations and on the Ships clusters, and both mount AFS. So I am going to run Mathematica on one of the Ships clusters machines, called `barruc`, and put a display window on my X11 server that runs on `beige`.

I use `slogin` to connect to `barruc` from `beige`. This sets up my `DISPLAY` variable on `barruc`, which points to a fake X11 display called `barruc:10.0` on that machine. That display in fact corresponds to a socket `ssh-201214-agent`, which lives in `/tmp/ssh-gustav`. Whatever falls into that socket is encrypted and sent to the `slogin` process that runs on `beige`. The process then unpacks the data and displays them on my X11 display.

If you don't use `slogin` then you may have to use `xauth` on `barruc` and on your own workstation to set up an appropriate entry in the `.Xauthority` file on `barruc`, and then point `DISPLAY` to your workstation.

There is one more thing that you have to do before you start Mathematica. You need to load Mathematica fonts into your X11 server that runs on your workstation:

```
beige $ xset +fp /afs/ovpit.indiana.edu/common/mathematica/Fonts/X
beige $ xset fp rehash
beige $
```

Now, on `barruc` I start Mathematica thusly:

```
barruc 29% /usr/afsws/bin/klog gustav
Password:
barruc 30% /usr/local/bin/mathematica &
```

and after a rather long wait the Mathematica windows pop up on my display. The wait is long if the Ships cluster NFS server is busy, if the network is busy (either in the Ships laboratory or between Franklin Hall and the Ships laboratory or both), if `barruc` itself is busy, or if `beige` is busy. All these things add up.

Now I load the notebook by pulling the `File` menu and selecting `Open`. A window pops up, which asks about the `File name`. Here I simply type the full path name of the file (with `/afs/ovpit...` at the front), press `OK`, and the simulator gets loaded.

The notebook contains a lot of stuff, which must be *evaluated* in order to put Mathematica's kernel in the right state for computations. To evaluate the whole book select

```
Kernel
   -> Evaluation
         -> Evaluate Notebook
```

This brings up a window, which asks you if you want to automatically evaluate all the initialization cells in this notebook, to which you should answer `Yes`. You should see Mathematica running through the lot at a fairly reasonable pace. Wait until the evaluation is finished. You will see Mathematica adding tags to all the statements that will have been evaluated and to all the answers as well. The window will assume a title `Running.../afs/ovpit.indiana.edu...` and may flicker and scroll many times. When the notebook gets processed to the end you will see various plots appear in it.

The first section of the notebook that is of interest to us is "Running the Quantum Computer for a Fixed Length of Time". In this section the computer, as defined by the unitary operator $\boldsymbol{U}_0$, is first run for $t/\hbar = 0.5$, and this is what comes out:

$$\begin{aligned} \boldsymbol{U}_0^{0.5} \mid 1000\rangle \;=\; & -0.119878 \mid 0011\rangle \\ & +0.229681(1-i) \mid 0100\rangle \\ & -0.229681(1+i) \mid 0101\rangle \\ & +0.880122 \mid 1000\rangle \end{aligned}$$

You can see that this is a very short time and the computer barely started evolving. We're still very close to the initial condition $\mid 1000\rangle$. But if we extend the running time to $t/\hbar = 2$ the probability of obtaining the right outcome of a measurement and thus completing the computation becomes very high:

$$\begin{aligned} \boldsymbol{U}_0^{2} \mid 1000\rangle \;=\; & -0.975682 \mid 0011\rangle \\ & +0.10892(1-i) \mid 0100\rangle \\ & -0.10892(1+i) \mid 0101\rangle \\ & +0.0243184 \mid 1000\rangle \end{aligned}$$

The simulator described in the notebook does more than just solve the Schrödinger equation. It can also simulate the measurement process. As the computer evolves, every now and then we are going to look at the cursor qubits in order to find out if the computation has completed. This process throws the whole state of the computer into a state compatible with the read-out of the cursor. The evolution then commences from that state. But what is the read-out state of the cursor? In order to select one of several possible outcomes we need to apply a probabilistic selection to the Schrödinger wave function, a selection weighted by the function itself.

In summary, this is what the simulator does:

1. It evolves an initial condition for one time step of length $t/\hbar = 1$
2. It looks at the counter qubits, thus performing a measurement. The result of this measurement is probabilistic, but the probabilities themselves are

derived from the quantum state of the computer reached as the result of the evolution. The measurement projects the state reached by the computer onto a state compatible with the result of the measurement.

3. The evolution now commences from the state reached and continues for another time step.

The computation keeps going until it is completed, i.e., until the cursor qubits are found in $\mid 001\rangle$.

Here is an example.

At time 0 the state of the system is

$$t = 0, \mid \Psi(0)\rangle = \mid 1000\rangle$$

We start the computer and look at the cursor qubits when $t/\hbar = 1$. Now the state of the system is:

$$\begin{aligned} t/\hbar = 1, \mid \Psi(1)\rangle \quad = \quad & -0.422028 \mid 0011\rangle \\ & +(0.349228 - 0.349228i) \mid 0100\rangle \\ & -(0.349228 + 0.349228i) \mid 0101\rangle \\ & +0.577972 \mid 1000\rangle \end{aligned}$$

When we perform the measurement the cursor is found in state

$$\mid 010\rangle$$

The measurements projects $\mid \Psi(1)\rangle$ onto

$$\begin{aligned} \mid \Psi'(1)\rangle \quad = \quad & (0.5 - 0.5i) \mid 0100\rangle \\ & -(0.5 + 0.5i) \mid 0101\rangle \end{aligned}$$

This state now becomes an initial condition for an unobserved evolution of the Feynman computer between $t/\hbar = 1$ and $t/\hbar = 2$. When we reach $t/\hbar = 2$ the computer finds itself in the following state:

$$\begin{aligned} t/\hbar = 2, \mid \Psi(2)\rangle \quad = \quad & -0.698456 \mid 0011\rangle \\ & +(0.0779718 - 0.0779718i) \mid 0100\rangle \\ & -(0.0779718 + 0.0779718i) \mid 0101\rangle \\ & -0.698456 \mid 1000\rangle \end{aligned}$$

Again we perform the measurement and find that the cursor is in state

$$\mid 100\rangle$$

which projects $\mid \Psi(2)\rangle$ onto

$$\mid \Psi'(2)\rangle \quad = \quad - \mid 1000\rangle$$

$t/\hbar$	$\mid \Psi(t)\rangle$	observation	$\mid \Psi'(t)\rangle$
1	$-0.422028 \mid 0011\rangle$ $+(0.349228 - 0.349228i) \mid 0100\rangle$ $-(0.349228 + 0.349228i) \mid 0101\rangle$ $+0.577972 \mid 1000\rangle$	$\mid 010\rangle$	$(0.5 - 0.5i) \mid 0100\rangle$ $-(0.5 + 0.5i) \mid 0101\rangle$
2	$-0.698456 \mid 0011\rangle$ $+(0.079718 - 0.079718i) \mid 0100\rangle$ $-(0.079718 + 0.079718i) \mid 0101\rangle$ $-0.698456 \mid 1000\rangle$	$\mid 100\rangle$	$- \mid 1000\rangle$
3	$-0.422028 \mid 0011\rangle$ $-(0.349228 - 0.349228i) \mid 0100\rangle$ $+(0.349228 + 0.349228i) \mid 0101\rangle$ $-0.577972 \mid 1000\rangle$	$\mid 010\rangle$	$(-0.5 + 0.5i) \mid 0100\rangle$ $+(0.5 + 0.5i) \mid 0101\rangle$
4	$0.698456 \mid 0011\rangle$ $-(0.0779718 - 0.0779718i) \mid 0100\rangle$ $+(0.0779718 + 0.0779718i) \mid 0101\rangle$ $+0.698456 \mid 1000\rangle$	$\mid 100\rangle$	$\mid 1000\rangle$
5	$-0.422028 \mid 0011\rangle$ $+(0.349228 - 0.349228i) \mid 0100\rangle$ $-(0.349228 + 0.349228i) \mid 0101\rangle$ $+0.577972 \mid 1000\rangle$	$\mid 010\rangle$	$(0.5 - 0.5i) \mid 0100\rangle$ $-(0.5 + 0.5i) \mid 0101\rangle$
6	$-0.698456 \mid 0011\rangle$ $+(0.0779718 - 0.0779718i) \mid 0100\rangle$ $-(0.0779718 + 0.0779718i) \mid 0101\rangle$ $-0.698456 \mid 1000\rangle$	$\mid 001\rangle$	$- \mid 0011\rangle$

Table 4.1: The evolution of the Feynman computer. Measurements are made every $\Delta t/\hbar = 1$.

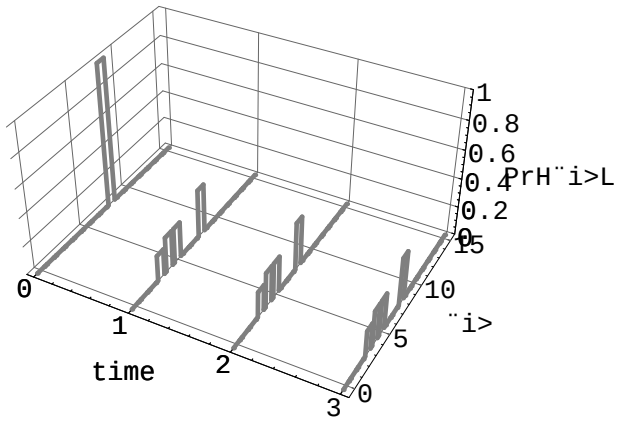


Figure 4.9: The evolution of the Feynman computer. Measurements are made every $\Delta t/\hbar = 1$.

Bummer! We shouldn't have looked, because this observation moved the computer back to the beginning.

Eventually, after 4 more steps we find the cursor in $|\ 001\rangle$ and the state qubit is now $|\ 1\rangle$, so the computation is completed. See table 4.1.

Figure 4.9 shows this evolution of the Feynman computer.

Figure 4.10 show the same evolution, but this time it also shows the results of the measurements.

4.5.8 Nonlocality and Teleportation

An astute observer should have noticed something unusual and profound about the measurements we made in the Feynman computer model: observing the counter register affected the content of the data register. Would this still happen if the data register and the counter register were separated by a very large distance? Would it happen instantaneously? Since the notion of simultaneity depends on a system of reference, in whose system of reference would this happen simultaneously? Would this not happen simultaneously in another system of reference?

The answers to this questions are:

1. Yes, the data register will get affected by a measurement made on the counter register regardless of how far away they are – as long as they remain *entangled*.

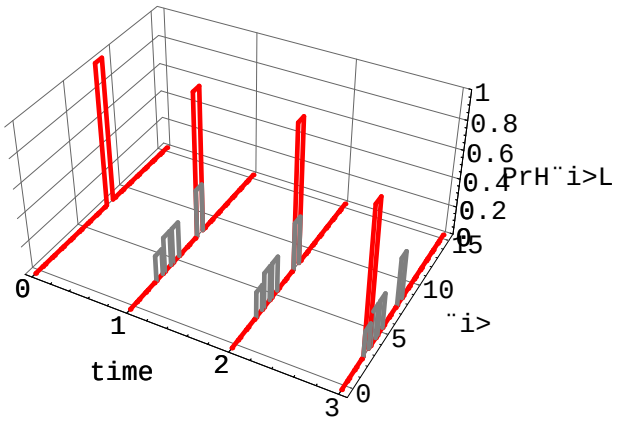


Figure 4.10: The evolution of the Feynman computer with the results of the measurements shown. As before the measurements are made every $\Delta t/\hbar = 1$.

2. As far as we know and as far as we were able to determine this experimentally, this will happen instantaneously – with the correction that the measurement process itself is not instantaneous – it has a *sub-Planck* structure [106].

3. We don't know what this process is going to look like when analyzed from various systems of reference. This would be a very difficult thing to test experimentally, because there is no transfer of retrievable information associated with this process. There is definitely a clash here between quantum mechanics and special relativity. One can show that special relativistic behaviour is restored in thermodynamic limit, but it seems to be violated in quantum regime [14].

   Should we bring ether back? I have met some distinguished physicists, who believe so, sic! And they even admitted this in public! But another possibility is that for *entangled* quantum systems a wormhole forms, which transmits *some* information between the components.

   At the end of the day, we don't really know what a quantum particle is. We only know what it looks like, when it interacts with a macroscopic system. And attempts to visualize a quantum particle on the basis of equations of quantum mechanics invariably yield a non-local object [14], whose interaction with a macroscopic observer is point-like.

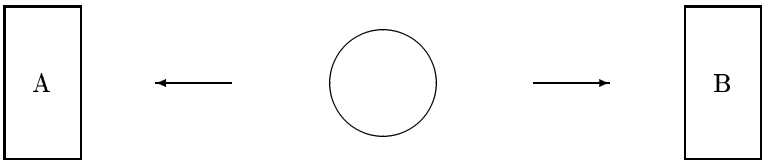


Figure 4.11: A decaying atom of Cesium emits two correlated photons, whose polarisation is then measured by Alice on the left and by Bob on the right.

Two and Three Photons: Bell's Inequality

Consider an experiment sketched in figure 4.11

We have an atom of Cesium in a laboratory somewhere in France (Earth), which decays and emits two *entangled*, or *correlated* photons, which rush in opposite directions. One photon speeds away towards Maxwell mountain on Venus, where Alice waits with a measuring apparatus. The other photon speeds away towards Olympus Mons on Mars, where Bob waits for it with his measuring device.

The two photons are correlated. If Alice and Bob measure linear polarizations of their respective photons then the basis on the Alice's side is:

$$\mid x\rangle_A, \mid y\rangle_A \tag{4.243}$$

and the basis on the Bob's side is

$$\mid x\rangle_B, \mid y\rangle_B \tag{4.244}$$

The combined basis for the bi-partite system is therefore:

$$\mid x_A x_B\rangle, \mid x_A y_B\rangle, \mid y_A x_B\rangle, \mid y_A y_B\rangle \tag{4.245}$$

and a general state of the bi-partite system can be described in the following way:

$$\mid \Psi\rangle = \Psi_{x_A x_B} \mid x_A x_B\rangle + \Psi_{x_A y_B} \mid x_A y_B\rangle + \Psi_{y_A x_B} \mid y_A x_B\rangle + \Psi_{y_A y_B} \mid y_A y_B\rangle \tag{4.246}$$

Now, if Alice measures her photon, and finds that it has passed through a vertical polarizer, then the state of the bipartite system filters into

$$\Psi_{y_A x_B} \mid y_A x_B\rangle + \Psi_{y_A y_B} \mid y_A y_B\rangle \tag{4.247}$$

multiplied by a scaling factor.

If now Bob makes a measurement the bi-partite system will go either into:

$$\mid y_A x_B\rangle \tag{4.248}$$

or into

$$\mid y_A y_B\rangle \tag{4.249}$$

In other words, Alice affects the result of Bob's measurement by her own measurement, and vice versa. The full state of the original bi-partite system can be ascertained only if repetitive measurements (so that probability distributions can be obtained) are made both by Alice and Bob.

This point can be made even more succinctly if we have a closer look at the photons that are emitted by the cesium atom. The photons are circularly polarized, one left and the other one right, and moving in the opposite directions, so that, in effect, the process conserves both the momentum and the angular momentum of the cesium atom. The combined state of both photons is (up to a scaling factor):

$$\begin{aligned} \mid \Psi\rangle &= (\mid x\rangle + i \mid y\rangle) \otimes (\mid x\rangle - i \mid y\rangle) \\ &= \mid x\rangle\otimes \mid x\rangle + i \mid y\rangle\otimes \mid x\rangle - i \mid x\rangle\otimes \mid y\rangle + \mid y\rangle\otimes \mid y\rangle \end{aligned}$$

Now, remember that photons are Bose particles, therefore a state describing two photons has to be described by a symmetric tensor. This means that we have to add to the above another bi-partite state for which the values of states in which both particles are found are swapped over:

$$\mid x\rangle\otimes \mid x\rangle + i \mid x\rangle\otimes \mid y\rangle - i \mid y\rangle\otimes \mid x\rangle + \mid y\rangle\otimes \mid y\rangle$$

If you add these two together what's left is:

$$\mid \Psi\rangle = \frac{1}{\sqrt{2}} (\mid x\rangle\otimes \mid x\rangle + \mid y\rangle\otimes \mid y\rangle), \tag{4.250}$$

where the proportionality coefficient has been set so that $|\langle\Psi \mid \Psi\rangle|^2 = 1$.

This state has a very interesting property. If Alice measures that her photon is polarized vertically, she projects it onto $\mid y\rangle\otimes \mid y\rangle$, so Bob is *guaranteed* to find that his photon is vertically polarized too. If Alice finds that her photon is polarized horizontally, she projects it onto $\mid x\rangle\otimes \mid x\rangle$ and Bob is now *guaranteed* to find his photon in horizontal polarization too.

If Bob and Alice could find that in this process their photons are differently polarized, e.g., Alice would find hers in $\mid x\rangle$ and Bob would find his in $\mid y\rangle$, then a certain amount of *twist* in the electromagnetic field would be associated with this pair of photons. That twist would imply that a certain amount of spin must have leaked from the cesium atom. But if the total angular momentum of the cesium atom did not change as both photons were emitted, this would imply that angular momentum is not conserved in this process.

But what if Bob's polarizer is set at a different angle? In that case his basis is $\mid x'\rangle$ and $\mid y'\rangle$.

Recall figure 4.7, page 88, in section 4.3.6. In that section we demonstrated that $\mid x\rangle$ and $\mid y\rangle$ for photons transform the same way $\boldsymbol{e}_x$ and $\boldsymbol{e}_y$ do, i.e.,

$$\begin{aligned} \mid x'\rangle &= \cos\theta \mid x\rangle + \sin\theta \mid y\rangle \\ \mid y'\rangle &= -\sin\theta \mid x\rangle + \cos\theta \mid y\rangle \end{aligned}$$

and inversely:

$$\begin{aligned} \mid x\rangle &= \cos\theta \mid x'\rangle - \sin\theta \mid y'\rangle \\ \mid y\rangle &= \sin\theta \mid x'\rangle + \cos\theta \mid y'\rangle \end{aligned}$$

Dropping those pesky $\otimes$ symbols and substituting the latter in equation (4.250) yields:

$$\begin{aligned} \mid \Psi\rangle &= \frac{1}{\sqrt{2}}\left(\mid x\rangle \mid x\rangle + \mid y\rangle \mid y\rangle\right) \\ &= \frac{1}{\sqrt{2}}\left(\mid x\rangle\left(\cos\theta \mid x'\rangle - \sin\theta \mid y'\rangle\right) + \mid y\rangle\left(\sin\theta \mid x'\rangle + \cos\theta \mid y'\rangle\right)\right) \\ &= \frac{1}{\sqrt{2}}\left(\cos\theta \mid x\rangle \mid x'\rangle - \sin\theta \mid x\rangle \mid y'\rangle + \sin\theta \mid y\rangle \mid x'\rangle + \cos\theta \mid y\rangle \mid y'\rangle\right) \end{aligned}$$

Even though this state does *not* look very symmetric, it is still the same symmetric state described by equation (4.250), but this time we use different basis vectors for Alice and for Bob, so the symmetry of the state is no longer apparent. In order to figure out if a given bi-partite state is indeed symmetric or not, you should write it using the same basis for both particles. This is only possible if these are identical particles, of course. But it is only for *identical* particles that we need to worry about our bi-partite states being symmetric or anti-symmetric.

Looking at this state we can now read that

$$\begin{aligned} P_{xx'} &= \frac{1}{2}\cos^2\theta \\ P_{xy'} &= \frac{1}{2}\sin^2\theta \\ P_{yx'} &= \frac{1}{2}\sin^2\theta \\ P_{yy'} &= \frac{1}{2}\cos^2\theta \end{aligned}$$

and $P_{xx'} + P_{xy'} + P_{yx'} + P_{yy'} = 1$.

Now assume that there are 3 similarly correlated photons and three observers with 3 polarizers. The state of the photons is given by:

$$\mid \Psi\rangle = \frac{1}{\sqrt{2}}\left(\mid x_A\rangle \mid x_B\rangle \mid x_C\rangle + \mid y_A\rangle \mid y_B\rangle \mid y_C\rangle\right), \tag{4.251}$$

where the $x_{A,B,C}$ and $y_{A,B,C}$ directions are the same for all three observers. To construct a state like this we would have to find a process that emits 3 photons and for which the total angular momentum of the emitter is preserved. But this wouldn't be quite enough, because for three photons you could have some twist carried away by photons A and B, which would then be balanced by photon C. So to realize a state such as this one physically isn't going to be easy. But this is a *Gedankenexperiment*.

Quantum mechanics tells us that Alice, Bob and Cecilia (we can place Cecilia conveniently on the Moon) will affect each other's measurements, even if the corresponding measurement events are space-like separated. On the other hand Special Theory of Relativity says that no interaction mediated by material particles can be transmitted faster than the speed of light[1]. So according to the Special Theory of Relativity if the measurement events are space-like separated, the measurements should be indpendent of each other. In this case the following probability argument would apply:

$$\begin{aligned} P_{x_A y_B} &= P_{x_A y_B x_C} + P_{x_A y_B y_C} \\ P_{x_B y_C} &= P_{x_A x_B y_C} + P_{y_A x_B y_C} \\ P_{x_A y_B} + P_{x_B y_C} &= P_{x_A y_B x_C} + (P_{x_A y_B y_C} + P_{x_A x_B y_C}) + P_{y_A x_B y_C} \\ &= P_{x_A y_B x_C} + P_{x_A y_C} + P_{y_A x_B y_C} \\ &\geq P_{x_A y_C} \end{aligned} \tag{4.252}$$

This is called a *Bell inequality*[7]. But for the tri-partite state given by equation (4.251) we have:

$$\begin{aligned} & P_{x_A y_B} + P_{x_B y_C} - P_{x_A y_C} \\ & = \frac{1}{2}\sin^2(\theta_B - \theta_A) + \frac{1}{2}\sin^2(\theta_C - \theta_B) - \frac{1}{2}\sin^2(\theta_C - \theta_A) \\ & \ngeq 0 \end{aligned} \tag{4.253}$$

If you choose $\theta_A = 0$ and then plot the left hand side in function of θ_B and θ_C you will find some points on the graph for which the left hand side of equation (4.253) is less than zero, sic! This is shown in figures 4.12 and 4.13.

If you don't trust figures, substitute $x = 15°$ and $y = 30°$, then:

$$\sin^2 15° + \sin^2 15° - \sin^2 30° = -0.116$$

This is indeed what the Aspect, Dalibard and Roger experiment demonstrated in 1982 using a system of two correlated photons. But for two photons the situation is more complex, because if you have three polarizers and two photons only, then you cannot write the inequality (4.252). This is because either Alice or Bob will have to measure the same photon twice, in which case two out of the three measurements cannot be independent, even according to classical physics.

But in February this year Pan, Bouwmeester, Daniell, Weinfurter, and Zeilinger reported an observation of quantum non-locality in a three-photon experiment [82] which is more similar to the one discussed here.[2]

> Tripartite entangled states such as the one given by equation (4.251) are called Greenberger-Horne-Zeilinger states, or GHZ (pronounced *gee-aich-zee* not *giga-hertz*) states for short. A state like this indicates that the

[1] This is not quite correct: Special Theory of Relativity allows for the existence of Tachyons – particles moving *always* with the speed greater than the speed of light.

[2] According to John Preskill[87] a good reference on Bell inequalities is a book by Asher Peres, "Quantum Theory: Concepts and Methods", chapter 6 [86].

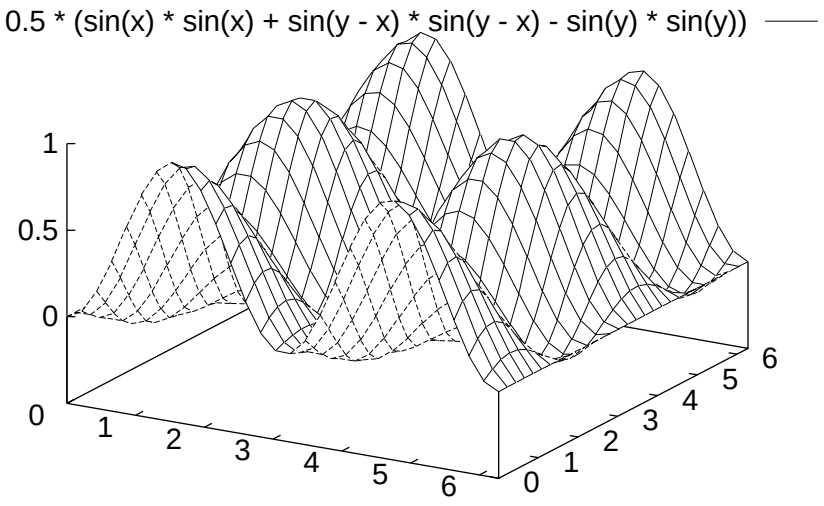


Figure 4.12: $P_{x_A y_B} + P_{x_B y_C} - P_{x_A y_C}$ for $\theta_A = 0$

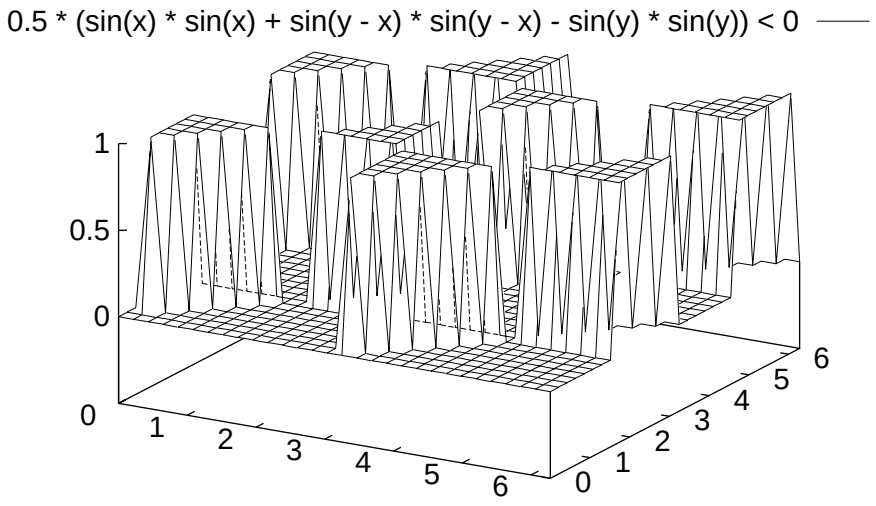


Figure 4.13: $P_{x_A y_B} + P_{x_B y_C} - P_{x_A y_C} < 0$ for $\theta_A = 0$

three photons in question are in a quantum superposition of the state $\mid x_A\rangle \mid x_B\rangle \mid x_C\rangle$ (all three are horizontally polarized) and $\mid y_A\rangle \mid y_B\rangle \mid y_C\rangle$ (all three are vertically polarized) with *none* of the photons having a well defined state on its own [82].

Consider first using circular polarizers to measure photons A and B, and a linear polarizer rotated by 45° with respect to the original polarizer at the source of the three photons. The representation of the GHZ state in this new basis is:

$$\begin{aligned} \mid \Psi\rangle \quad = \quad & \frac{1}{2}\Big(\mid R\rangle_A \mid L\rangle_B \mid x'\rangle_C + \mid L\rangle_A \mid R\rangle_B \mid x'\rangle_C + \\ & \mid R\rangle_A \mid R\rangle_B \mid y'\rangle_C + \mid L\rangle_A \mid L\rangle_B \mid y'\rangle_C \Big) \end{aligned} \tag{4.254}$$

An interesting feature of this expression is that the measurements described by it are highly randomized. Photon A exhibits left or right polarization with the same probability of 50%. On the other hand, given any two specific measurements on photons A and B the state of photon C can be predicted with certainty.

Similar expressions can be constructed by cyclic permutations, one in which photons A and C are polarized circularly and photon B is polarized linearly and another one in which photons B and C are polarized circularly and photon A is polarized linearly.

If the detected polarization is due to some internal local property that the photon carries with it, it should be possible to assign a simple label number to this property. Let these labels be called X for linear polarization and Y for circular polarization. Let us now assume that we can make the following assignments:

polarization	*label*
$\mid x'\rangle$	$X = 1$
$\mid y'\rangle$	$X = -1$
$\mid R\rangle$	$Y = 1$
$\mid L\rangle$	$Y = -1$

This seems to work quite well for predictions associated with state (4.254), where the outcome of each possible measurement yields

$$\begin{aligned} \mid R\rangle_A \mid L\rangle_B \mid x'\rangle_C \quad &\rightarrow \quad Y_A Y_B X_C = 1 \times -1 \times 1 = -1 \\ \mid L\rangle_A \mid R\rangle_B \mid x'\rangle_C \quad &\rightarrow \quad Y_A Y_B X_C = -1 \times 1 \times 1 = -1 \\ \mid R\rangle_A \mid R\rangle_B \mid y'\rangle_C \quad &\rightarrow \quad Y_A Y_B X_C = 1 \times 1 \times -1 = -1 \\ \mid L\rangle_A \mid L\rangle_B \mid y'\rangle_C \quad &\rightarrow \quad Y_A Y_B X_C = -1 \times -1 \times -1 = -1 \end{aligned}$$

By cyclic permutations this will hold also for $Y_A X_B Y_C$ and $X_A Y_B Y_C$. We thus arrive at the following rule:

$$Y_A Y_B X_C = Y_A X_B Y_C = X_A Y_B Y_C = -1 \tag{4.255}$$

Multiplying all three equations yields:

$$Y_A Y_B X_C Y_A X_B Y_C X_A Y_B Y_C = X_A X_B X_C = -1^3 = -1 \tag{4.256}$$

because

$$Y_A^2 = Y_B^2 = Y_C^2 = 1 \tag{4.257}$$

What this means is that if we were to replace circular polarizers for photons A and B with linear polarizers and thus measured all three photons versus linear polarizers, all at the same 45° angle to the original direction at the source, we should detect only the combinations for which the product of the X labels is -1, i.e.,

$$\begin{array}{l} \mid x'\rangle_A \mid x'\rangle_B \mid y'\rangle_C \\ \mid x'\rangle_A \mid y'\rangle_B \mid x'\rangle_C \\ \mid y'\rangle_A \mid x'\rangle_B \mid x'\rangle_C \\ \mid y'\rangle_A \mid y'\rangle_B \mid y'\rangle_C \end{array}$$

This is a prediction that derives from the assumption that the measured photon polarization is due to the photon carrying a labeled property with it before it arrives at the polarizer, which then affects the polarization detected.

The quantum prediction however is different. Rewriting the GHZ state in the $\mid x'\rangle, \mid y'\rangle$ basis yields:

$$\begin{aligned} \mid \Psi\rangle \;=\; & \frac{1}{2}\Big(\mid x'\rangle_A \mid x'\rangle_B \mid x'\rangle_C + \mid x'\rangle_A \mid y'\rangle_B \mid y'\rangle_C + \\ & \mid y'\rangle_A \mid x'\rangle_B \mid y'\rangle_C + \mid y'\rangle_A \mid y'\rangle_B \mid x'\rangle_C \Big) \end{aligned} \tag{4.258}$$

According to quantum mechanics the product of X labels for *all* possible outcomes of this experiment is going to be $+1$, *not* -1. Not a single possibility predicted by the "classical" reasoning outlined above is also predicted by quantum theory.

The beauty of this approach is that a measurement performed on a GHZ state clearly demonstrates the discrepancy between a theory based on a *local hidden variable* model, and quantum mechanics, without resorting to statistical inequalities, which may be quite difficult to detect experimentally. Instead the conflict arises for quite definite predictions and can be exhibited by a single measurement.

The measurement performed by Pan, Bouwmeester, Daniell, Weinfurter and Zeilinger [82] demonstrated the correctness of the quantum prediction using 200 fs pulses of ultraviolet light ($\lambda = 394\,\text{nm}$), which generated pairs of polarization entangled photons. These were then processed further so as to create a tri-partite GHZ state while dropping the forth photon. Two years earlier, in 1998, Laflamme, Knill, Zurek, Catasti and Mariappan performed an NMR experiment which demonstrated GHZ entanglement too [62].

Teleporting a Qubit

Consider a qubit in an $\mid x_1\rangle, \mid y_1\rangle$ basis

$$\mid \Psi\rangle = a \mid x_1\rangle + b \mid y_1\rangle = \begin{pmatrix} a \\ b \end{pmatrix} \tag{4.259}$$

where $\mid x_1\rangle$ and $\mid y_1\rangle$ don't have to correspond to polarization directions any more. They can be spin variables, for example. But $\mid x\rangle$ and $\mid y\rangle$ are easier to

write and also to read than $|\uparrow\rangle$ and $|\downarrow\rangle$ (why? because our eyes and brains are used to x and y more than to $\uparrow$ and $\downarrow$), so I'll just use this.

In order to transmit this qubit to another location we need to entangle it first with a bi-partite state, for example:

$$\mid \Phi\rangle = \frac{1}{\sqrt{2}} (\mid x_2\rangle \mid y_3\rangle - \mid y_2\rangle \mid x_3\rangle) \tag{4.260}$$

Although this state is antisymmetric, it is possible to realize it using bosons. For example, if we have two photons of different colour, they are distinguishable, and therefore they can exist in non-symmetric or anti-symmetric states as well as in symmetric ones.

How does one entangle a qubit? One puts it sufficiently close to the other qubits so that a tri-partite state forms:

$$\begin{aligned} \mid \Psi\rangle \otimes \mid \Phi\rangle &= \frac{1}{\sqrt{2}} (a \mid x_1\rangle + b \mid y_1\rangle) \otimes (\mid x_2\rangle \mid y_3\rangle - \mid y_2\rangle \mid x_3\rangle) \\ &= \frac{a}{\sqrt{2}} (\mid x_1\rangle \mid x_2\rangle \mid y_3\rangle - \mid x_1\rangle \mid y_2\rangle \mid x_3\rangle) \qquad (4.261) \\ &\quad + \frac{b}{\sqrt{2}} (\mid y_1\rangle \mid x_2\rangle \mid y_3\rangle - \mid y_1\rangle \mid y_2\rangle \mid x_3\rangle) \qquad (4.262) \end{aligned}$$

Once the tri-partite state has been formed Alice can stay home with particles 1, 2, their dog and their children, and Bob is sent away to Australia with particle 3. All three particles must be kept in a very, very deep freezer so that the entanglement is maintained. This is the really difficult bit in this whole business – the algebra itself is quite trivial.

In order to transmit the state of particle 1 to Bob Alice needs to measure her remaining two particles against the so called *Bell operator basis* introduced by Braunstein, Mann and Revzen [19]:

$$\begin{aligned} \mid \Psi^A\rangle &= \frac{1}{\sqrt{2}} (\mid x_1\rangle \mid y_2\rangle - \mid y_1\rangle \mid x_2\rangle) \qquad (4.263) \\ \mid \Psi^B\rangle &= \frac{1}{\sqrt{2}} (\mid x_1\rangle \mid y_2\rangle + \mid y_1\rangle \mid x_2\rangle) \qquad (4.264) \\ \mid \Psi^C\rangle &= \frac{1}{\sqrt{2}} (\mid x_1\rangle \mid x_2\rangle - \mid y_1\rangle \mid y_2\rangle) \qquad (4.265) \\ \mid \Psi^D\rangle &= \frac{1}{\sqrt{2}} (\mid x_1\rangle \mid x_2\rangle + \mid y_1\rangle \mid y_2\rangle) \qquad (4.266) \\ & \qquad (4.267) \end{aligned}$$

In order to see what this measurement will do to our tri-partite state (4.262) we need to evaluate the following products:

$$\begin{aligned} &\langle \Psi^A \mid \Psi \otimes \Phi\rangle \\ &\langle \Psi^B \mid \Psi \otimes \Phi\rangle \\ &\langle \Psi^C \mid \Psi \otimes \Phi\rangle \\ &\langle \Psi^D \mid \Psi \otimes \Phi\rangle \end{aligned}$$

This is actually less tedious than it seems at first glance, because once you write it all down on paper you can see very easily which terms become zeros and which survive:

$$\begin{aligned}
\langle \Psi^A \mid \Psi \otimes \Phi \rangle &= \frac{1}{\sqrt{2}} \left(\langle x_1 \mid \langle y_2 \mid - \langle y_1 \mid \langle x_2 \mid \right) \Bigg| \\
&\quad \left(\frac{a}{\sqrt{2}} \left(\mid x_1 \rangle \mid x_2 \rangle \mid y_3 \rangle - \mid x_1 \rangle \mid y_2 \rangle \mid x_3 \rangle \right) \right. \\
&\quad \left. + \frac{b}{\sqrt{2}} \left(\mid y_1 \rangle \mid x_2 \rangle \mid y_3 \rangle - \mid y_1 \rangle \mid y_2 \rangle \mid x_3 \rangle \right) \right) \\
&= -\frac{a}{2} \mid x_3 \rangle - \frac{b}{2} \mid y_3 \rangle \qquad (4.268)
\end{aligned}$$

$$\begin{aligned}
\langle \Psi^B \mid \Psi \otimes \Phi \rangle &= \frac{1}{\sqrt{2}} \left(\langle x_1 \mid \langle y_2 \mid + \langle y_1 \mid \langle x_2 \mid \right) \Bigg| \\
&\quad \left(\frac{a}{\sqrt{2}} \left(\mid x_1 \rangle \mid x_2 \rangle \mid y_3 \rangle - \mid x_1 \rangle \mid y_2 \rangle \mid x_3 \rangle \right) \right. \\
&\quad \left. + \frac{b}{\sqrt{2}} \left(\mid y_1 \rangle \mid x_2 \rangle \mid y_3 \rangle - \mid y_1 \rangle \mid y_2 \rangle \mid x_3 \rangle \right) \right) \\
&= -\frac{a}{2} \mid x_3 \rangle + \frac{b}{2} \mid y_3 \rangle \qquad (4.269)
\end{aligned}$$

$$\begin{aligned}
\langle \Psi^C \mid \Psi \otimes \Phi \rangle &= \frac{1}{\sqrt{2}} \left(\langle x_1 \mid \langle x_2 \mid - \langle y_1 \mid \langle y_2 \mid \right) \Bigg| \\
&\quad \left(\frac{a}{\sqrt{2}} \left(\mid x_1 \rangle \mid x_2 \rangle \mid y_3 \rangle - \mid x_1 \rangle \mid y_2 \rangle \mid x_3 \rangle \right) \right. \\
&\quad \left. + \frac{b}{\sqrt{2}} \left(\mid y_1 \rangle \mid x_2 \rangle \mid y_3 \rangle - \mid y_1 \rangle \mid y_2 \rangle \mid x_3 \rangle \right) \right) \\
&= \frac{a}{2} \mid y_3 \rangle + \frac{b}{2} \mid x_3 \rangle \qquad (4.270)
\end{aligned}$$

$$\begin{aligned}
\langle \Psi^D \mid \Psi \otimes \Phi \rangle &= \frac{1}{\sqrt{2}} \left(\langle x_1 \mid \langle x_2 \mid + \langle y_1 \mid \langle y_2 \mid \right) \Bigg| \\
&\quad \left(\frac{a}{\sqrt{2}} \left(\mid x_1 \rangle \mid x_2 \rangle \mid y_3 \rangle - \mid x_1 \rangle \mid y_2 \rangle \mid x_3 \rangle \right) \right. \\
&\quad \left. + \frac{b}{\sqrt{2}} \left(\mid y_1 \rangle \mid x_2 \rangle \mid y_3 \rangle - \mid y_1 \rangle \mid y_2 \rangle \mid x_3 \rangle \right) \right) \\
&= \frac{a}{2} \mid y_3 \rangle - \frac{b}{2} \mid x_3 \rangle \qquad (4.271)
\end{aligned}$$

In summary:

$$\mid \Psi \otimes \Phi \rangle \qquad (4.272)$$

$$= \frac{1}{2}\left(\mid \Psi^A\rangle \begin{pmatrix} -a \\ -b \end{pmatrix} + \mid \Psi^B\rangle \begin{pmatrix} -a \\ b \end{pmatrix} + \mid \Psi^C\rangle \begin{pmatrix} b \\ a \end{pmatrix} + \mid \Psi^D\rangle \begin{pmatrix} -b \\ a \end{pmatrix} \right)$$

$$= \frac{1}{2}\Bigg(\mid \Psi^A\rangle \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} + \mid \Psi^B\rangle \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix}$$

$$+ \mid \Psi^C\rangle \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} + \mid \Psi^D\rangle \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} \Bigg) \qquad (4.273)$$

The result we have obtained is as follows: when Alice performs her measurement, the tri-partite system ends up in one of the four states, projected onto $\mid \Psi^A\rangle$, or $\mid \Psi^B\rangle$, or $\mid \Psi^C\rangle$, or $\mid \Psi^D\rangle$. This puts Bob's particle correspondingly into one of the following states: $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix}$, or $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix}$, or $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix}$, or $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix}$. The matrices that multiply $\begin{pmatrix} a \\ b \end{pmatrix}$ all square to $\mathbf{1}$ with the exception of the last one, which squares to $\mathbf{-1}$. All that Bob needs to do in order to put his particle in the state $\begin{pmatrix} a \\ b \end{pmatrix}$ is to apply the corresponding matrix (or the corresponding unitary transformation) to his particle. But how is he to know which is the corresponding matrix? Well, Alice has to tell him, i.e., Alice must call him using some conventional information transfer mechanism and tell him which of the four Bell operator vectors she ended up projecting her two particles on. Alice does not know this in advance. She has to perform a measurement to find out.

What is the experimental status of quantum teleportation? Can this really be done?

It has been done. Perhaps the first ones to demonstrate quantum teleportation were Bouwmeester, Pan, Mattle, Eibl, Weinfurter and Zeilinger [16] in December 1997.

> In order to demonstrate quantum teleportation entangled states must be produced on demand. The group that performed the first teleportation experiment [16] is the same group that demonstrated quantum non-locality using the tri-partite Greenberger-Horne-Zeilinger state some 2 years later [82]. They are located at the University of Innsbruck in Austria.
>
> In order to teleport a qubit we need to generate a maximally entangled bi-partite state first. Then we must additionally entangle the bi-partite state with the qubit to be teleported. Finally, we must come up with a procedure that identifies clearly all four Bell states for the quantum system held by Alice. But pairs of entangled photons can be easily produced and they can be just as easily projected onto two of the four Bell states.
>
> Pairs of entangled photons are generated inside a nonlinear crystal, where an incoming pump photon can decay spontaneously into two photons. If these are to correspond to particles 2 and 3, the resulting state is:
>
> $$\mid \Psi\rangle = \frac{1}{\sqrt{2}}(\mid x\rangle_2 \mid y\rangle_3 - \mid y\rangle_2 \mid x\rangle_3) \qquad (4.274)$$

To project photons 1 and 2 onto a Bell state they have to be made indistinguishable. This is accomplished by superposing the two photons at a beam splitter. Then when they are both either reflected or transmitted the corresponding amplitudes for these two processes must be added. Unitarity implies that the reflection amplitude obtains an additional minus sign. If an antisymmetric state is fed into the system constructive interference results. The projection of the $(1,2)$ pair onto a Bell state is thus accomplished by placing detectors in each of the outputs of the beam splitter and registering simultaneous detections.

Photons 1 and the photon that ultimately splits into 2 and 3 are generated by using a pulsed pump beam. They are then sent through narrow-bandwith filters, which produce coherence time much longer than the pump pulse length. This is based on $\Delta E \Delta t \approx \hbar$.

The pulses used in the experiment had duration of 200 fs, and they occurred at a frequency of 76 MHz. Photon wavelength was 788 nm with 4 nm bandwidth, which resulted in a coherence time of 520 fs.

In their experiment Bowmeester, Pan, Mattle, Eibl, Weinfurter and Zeilinger teleported a polarization state of a photon, but they did not perform the final rotation of the other photon, which would really put it in the full teleported state. In fact their procedure destroyed the teleported state so that it could not emerge as a freely propagating state for further examination and exploitation [20]. In short, their teleportation wasn't complete.

Then in 1998 Nielsen, Knill and LaFlamme [81] demonstrated a complete quantum teleportation using nuclear magnetic resonance. In fact they implemented the complete Brassard Teleportation Circuit, which we are going to discuss in the next section, in NMR. Teleportation was accomplished over interatomic distances. In the process a quantum state was transferred from a carbon nucleus to a hydrogen nucleus in molecules of trichloroethylene.

A molecule of trichloroethylene, $C\,Cl\,H\,C\,Cl_2$, comprises two atoms of carbon mutually connected by a strong double bond. One carbon atom, let's call it carbon-1, is then saturated with hydrogen and chlorine, and the other carbon atom, let's call it carbon-2, is saturated with two chlorine atoms. Hydrogen, ^{1}H, and chlorine, ^{35}Cl, are both magnetically active. But carbon ^{12}C is not. For NMR experiments such as the one described by Nielsen, Knill, and LaFlamme, a ^{13}C isotope is used in place of ^{12}C. ^{13}C is magnetically active.

The NMR demonstration of teleportation transferred a spin state from the carbon-2 nucleus to the hydrogen nucleus. The carbon-1 nucleus was used as an intermediary (the *ancilla*). The state was teleported over a distance of only a few Å.

The unitary operations described in the next section (The Brassard Teleportation Circuit) were implemented using non-selective radio frequency pulses tuned to the Larmor frequencies of the nuclear spins, and delays allowing entanglement to form through the interaction of neighbouring nuclei.

The Bell basis measurement was based on a procedure originally suggested by Brassard. A direct Bell basis measurement cannot be performed

in NMR because here the measurement step extracts expectation values of σ_x and σ_y for each spin *averaged* over the ensemble of molecules. In other words, in NMR we measure a collective magnetization of the whole ensemble.

Brassard's suggestion was to rotate the state from the Bell basis into the computational basis, $|00\rangle, |01\rangle, |10\rangle, |11\rangle$ first, and then to perform a projective measurement in this basis.

Nielsen, Knill, and LaFlamme exploited the natural phase decoherence occurring on the carbon nuclei to achieve an effect similar to a projective measurement. The decoherence process is indistinguishable from measurement. In fact this is what *measurement* really is. So, by allowing carbon nuclei quantum states to decohere, they effectively used *the environment* to perform the projective measurement for them. Then they performed the final transformation in order to recreate the teleported qubit in its new location.

The NMR spectrometer used in the experiment was Bruker DRX-500. Larmor and coupling frequencies for the hydrogen and both carbon nuclei were determined experimentally:

$$\begin{aligned} \omega_{\mathrm{H}} &\approx 500.133491\,\mathrm{MHz} \\ \omega_{\mathrm{C}_1} &\approx 125.772580\,\mathrm{MHz} \\ \omega_{\mathrm{C}_2} &\approx \omega_{\mathrm{C}_1} - 911\,\mathrm{Hz} \\ J_{\mathrm{H\,C}_1} &\approx 201\,\mathrm{Hz} \\ J_{\mathrm{C}_1\,\mathrm{C}_2} &\approx 103\,\mathrm{Hz} \end{aligned}$$

There are also some other couplings present in this molecule. These can be effectively supressed by *multiple refocusings*.

An exhaustive discussion of how to implement selective coupling between NMR qubits can be found in [67].

Because paper is cheap, whereas lasers, NMR machines and optical benches are expensive, I have found only these two papers that presented genuine experimental results so far. But you will find great many theoretical papers outlining various teleportation schemes in Physical Review A.

One of such schemes, which as I have remarked, has actually been implemented by Nielsen, Knill, and LaFlamme in NMR, we are going to discuss in the next section.

The Brassard Teleportation Circuit

Figure 4.14 shows a quantum circuit. This is our first non-trivial quantum circuit, and we are going to analyze it in detail.

The circuit is relatively simple, comprising 3 qubit lines only. Each horizontal line represents a qubit. The square boxes with various letters in them represent single qubit gates, and the $\oplus$ symbol with a wire sticking out of it vertically is the controlled-NOT gate. The $\oplus$ symbol indicates the data line and the other end with a black dot on it indicates the control line. If the vertical line crosses a horizontal line without a fat black dot on it, it means that the coupling

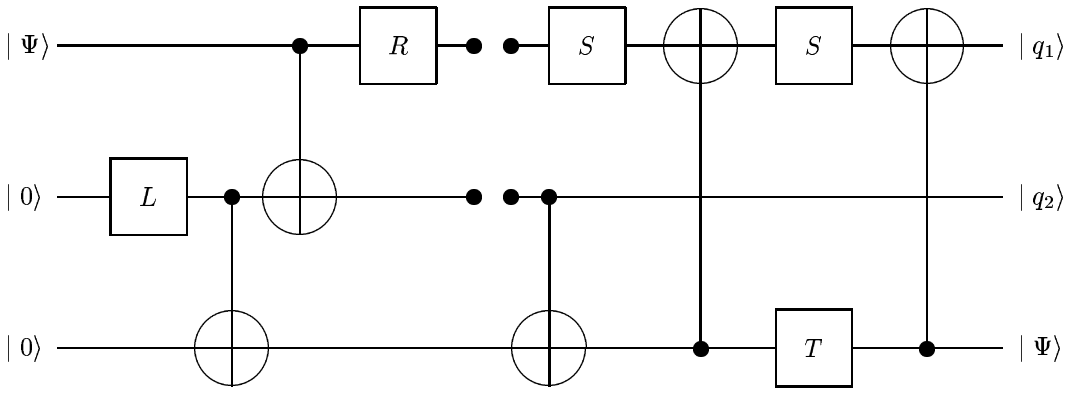


Figure 4.14: The Brassard teleportation circuit.

represented by the vertical line skips over the qubit. Time runs from left to right. But observe that time is not as strictly defined here as is the case in classical physics or in traditional quantum mechanics. Rather what flows from left to right is the order of events, each event being represented by an application of a gate.

There is an unusual feature in this circuit, which is seldom found in other circuits in this location, though most have it somewhere at the right end. The top two lines end in black dots about half-way through the circuit. This place indicates that a *measurement* is carried out here, or that the system is simply allowed to decohere naturally, which, as far as the qubits are concerned, is the same thing.

The circuit implements the process of teleportation. The top two lines represent Alice, and the bottom line represents Bob. You can see that we end up moving the state $|\Psi\rangle$ from the top line in the upper left corner of the circuit to the bottom line in the lower right corner of the circuit.

If we were to map this circuit on the molecule of trichloroethylene used by Nielsen, Knill and LaFlamme in their NMR teleportation experiment, then the top line would correspond to the Carbon-2 nucleus, the one that's connected to two Chlorine atoms, the middle line would correspond to the Carbon-1 nucleus, connected to a Chlorine and to a Hydrogen atom, and the bottom line would correpond to the Hydrogen atom. Chlorine atoms, although magnetically active, were not used in the computation, so they do not appear in the circuit.

All nuclei in the molecule of Trichloroethylene are coupled with each other. Yet in the circuit below we see couplings appear in 5 selected locations only. All the other couplings as well as these 5 couplings in places where they are not needed are supressed by *refocussing*, or *time halting*, which we have discussed in detail already.

The single qubit gates $\boldsymbol{L}$, $\boldsymbol{R}$, $\boldsymbol{S}$, and $\boldsymbol{T}$ used in the circuit, and the two qubit gate $\oplus$ have the following matrix definitions in the computational basis

$| 0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $| 1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

$$\boldsymbol{L} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \tag{4.275}$$

$$\boldsymbol{R} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \tag{4.276}$$

$$\boldsymbol{S} = \begin{pmatrix} i & 0 \\ 0 & 1 \end{pmatrix} \tag{4.277}$$

$$\boldsymbol{T} = \begin{pmatrix} -1 & 0 \\ 0 & -i \end{pmatrix} \tag{4.278}$$

$$\bigoplus = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{4.279}$$

We can think of the gate $\boldsymbol{L}$ as implementing a 90° qubit rotation (in the Bloch sphere) "to the left", then the gate $\boldsymbol{R}$ implements a 90° rotation "to the right". It is easy to see that $\boldsymbol{L} \cdot \boldsymbol{R} = \mathbf{1}$. Gates $\boldsymbol{S}$ and $\boldsymbol{T}$ represent a combination of rotations about the z axis with a multiplication by a fixed global phase-shift:

$$\boldsymbol{S} = \begin{pmatrix} i & 0 \\ 0 & 1 \end{pmatrix} = e^{i\pi/4} \begin{pmatrix} e^{i\pi/4} & 0 \\ 0 & e^{-i\pi/4} \end{pmatrix}$$

$$\boldsymbol{T} = \begin{pmatrix} -1 & 0 \\ 0 & -i \end{pmatrix} = e^{i\pi/4} \begin{pmatrix} e^{i3\pi/4} & 0 \\ 0 & e^{-i3\pi/4} \end{pmatrix}$$

It is computationally convenient to write definitions of these gates in the language of vectors rather than matrices:

$$L \mid 0\rangle = \frac{1}{\sqrt{2}} (\mid 0\rangle + \mid 1\rangle)$$

$$L \mid 1\rangle = \frac{1}{\sqrt{2}} (- \mid 0\rangle + \mid 1\rangle) \tag{4.280}$$

$$R \mid 0\rangle = \frac{1}{\sqrt{2}} (\mid 0\rangle - \mid 1\rangle)$$

$$R \mid 1\rangle = \frac{1}{\sqrt{2}} (\mid 0\rangle + \mid 1\rangle) \tag{4.281}$$

$$S \mid 0\rangle = i \mid 0\rangle$$

$$S \mid 1\rangle = \mid 1\rangle \tag{4.282}$$

$$T \mid 0\rangle = - \mid 0\rangle$$

$$T \mid 1\rangle = -i \mid 1\rangle \tag{4.283}$$

$$\oplus \mid 0\rangle \mid 0\rangle = \mid 0\rangle \mid 0\rangle$$

$$\oplus \mid 0\rangle \mid 1\rangle = \mid 0\rangle \mid 1\rangle$$

$$\begin{aligned} \oplus \mid 1\rangle \mid 0\rangle &= \mid 1\rangle \mid 1\rangle \\ \oplus \mid 1\rangle \mid 1\rangle &= \mid 1\rangle \mid 0\rangle \end{aligned} \tag{4.284}$$

Inputs applied to the circuit are $\mid 0\rangle$ to the two bottom lines and an arbitrary state $\mid \Psi\rangle$ applied to the top line. This is the state which is going to be teleported to the bottom line.

The first gate, $\boldsymbol{L}$, converts the input $\mid 0\rangle$ to $\frac{1}{\sqrt{2}}(\mid 0\rangle + \mid 1\rangle)$. After the gate has been applied the states on the three lines of the circuit are:

$$\begin{array}{l} \mid \Psi\rangle \\ \frac{1}{\sqrt{2}}(\mid 0\rangle + \mid 1\rangle) \\ \mid 0\rangle \end{array}$$

The next step entangles the two bottom lines using the controlled-NOT gate. Here we use a subscript c to point to the control line and subscript d to point to the data line.

$$\begin{aligned} &\oplus \frac{1}{\sqrt{2}}(\mid 0\rangle_c + \mid 1\rangle_c) \mid 0\rangle_d \\ &= \frac{1}{\sqrt{2}}(\oplus \mid 0\rangle_c \mid 0\rangle_d + \oplus \mid 1\rangle_c \mid 0\rangle_d) \\ &= \frac{1}{\sqrt{2}}(\mid 0\rangle_c \mid 0\rangle_d + \mid 1\rangle_c \mid 1\rangle_d) \end{aligned} \tag{4.285}$$

This is a new bi-partite state which now binds the two bottom lines together. Their state is *maximally entangled*. The state of the circuit at this stage is:

$$\begin{array}{l} \mid \Psi\rangle \\ \frac{1}{\sqrt{2}}(\mid 0\rangle \mid 0\rangle + \mid 1\rangle \mid 1\rangle) \end{array}$$

The next operation entangles $\mid \Psi\rangle$ with the middle qubit, which is already entangled with the bottom qubit. We continue with our convention, where the control qubit is marked with subscript c and the data qubit is marked with subscript d:

$$\begin{aligned} &\oplus \mid \Psi\rangle_c \frac{1}{\sqrt{2}}(\mid 0\rangle_d \mid 0\rangle + \mid 1\rangle_d \mid 1\rangle) \\ &= \frac{1}{\sqrt{2}} \oplus (a \mid 0\rangle_c + b \mid 1\rangle_c)(\mid 0\rangle_d \mid 0\rangle + \mid 1\rangle_d \mid 1\rangle) \\ &= \frac{1}{\sqrt{2}} \Big(a(\oplus(\mid 0\rangle_c \mid 0\rangle_d) \mid 0\rangle + \oplus(\mid 0\rangle_c \mid 1\rangle_d) \mid 1\rangle) \\ &\quad + b(\oplus(\mid 1\rangle_c \mid 0\rangle_d) \mid 0\rangle + \oplus(\mid 1\rangle_c \mid 1\rangle_d) \mid 1\rangle) \Big) \\ &= \frac{1}{\sqrt{2}} \Big(a(\mid 0\rangle_c \mid 0\rangle_d \mid 0\rangle + \mid 0\rangle_c \mid 1\rangle_d \mid 1\rangle) \\ &\quad + b(\mid 1\rangle_c \mid 1\rangle_d \mid 0\rangle + \mid 1\rangle_c \mid 0\rangle_d \mid 1\rangle) \Big) \end{aligned} \tag{4.286}$$

Now all three lines are entangled and the computer is no longer in a state that would let us isolate any of the lines.

The last operation before the measurement rotates the upper qubit "to the right". For ease of reading we mark the object qubit of this operation with subscript R:

$$\begin{aligned}
&\boldsymbol{R}\frac{1}{\sqrt{2}}\Big(a\,(\mid 0\rangle_R \mid 0\rangle \mid 0\rangle + \mid 0\rangle_R \mid 1\rangle \mid 1\rangle)\\
&\quad +b\,(\mid 1\rangle_R \mid 1\rangle \mid 0\rangle + \mid 1\rangle_R \mid 0\rangle \mid 1\rangle)\Big)\\
&= \frac{1}{\sqrt{2}}\Big(a\,(\boldsymbol{R} \mid 0\rangle_R \mid 0\rangle \mid 0\rangle + \boldsymbol{R} \mid 0\rangle_R \mid 1\rangle \mid 1\rangle)\\
&\quad +b\,(\boldsymbol{R} \mid 1\rangle_R \mid 1\rangle \mid 0\rangle + \boldsymbol{R} \mid 1\rangle_R \mid 0\rangle \mid 1\rangle)\Big)\\
&= \frac{1}{\sqrt{2}}\left(a\left(\frac{1}{\sqrt{2}}(\mid 0\rangle_R - \mid 1\rangle_R) \mid 0\rangle \mid 0\rangle + \frac{1}{\sqrt{2}}(\mid 0\rangle_R - \mid 1\rangle_R) \mid 1\rangle \mid 1\rangle\right)\right.\\
&\quad \left. +b\left(\frac{1}{\sqrt{2}}(\mid 0\rangle_R + \mid 1\rangle_R) \mid 1\rangle \mid 0\rangle + \frac{1}{\sqrt{2}}(\mid 0\rangle_R + \mid 1\rangle_R) \mid 0\rangle \mid 1\rangle\right)\right)\\
&= \frac{1}{2}\Big(a\,(\mid 0_R 00\rangle - \mid 1_R 00\rangle + \mid 0_R 11\rangle - \mid 1_R 11\rangle)\\
&\quad +b\,(\mid 0_R 10\rangle + \mid 1_R 10\rangle + \mid 0_R 01\rangle + \mid 1_R 01\rangle)\Big) \qquad (4.287)
\end{aligned}$$

At this stage we reach the measurement point. At this point the upper two qubits are either measured or they are allowed to decohere naturally, as was the case in the experiment with Trichloroethylene, with the result that they collapse jointly onto one of the following states: $\mid 00\rangle$, $\mid 01\rangle$, $\mid 10\rangle$, or $\mid 11\rangle$.

This process forces the bottom qubit into a state that is commensurate with whatever the upper qubits become and with the original quantum state of all three qubits.

In order to see what happens next we need to carry out our analysis for all possible outcomes of the measurement. This is going to be rather tedious, so we shall pick up just one possible outcome and go ahead with anlysing what happens in this case leaving the analysis of the remaining three channels to the reader as an EXERCISE.

Suppose the upper two qubits decohere to $\mid 01\rangle$. This filters the state of the system into a state proportional to (up to a new normalization constant):

$$a \mid 011\rangle + b \mid 010\rangle \qquad (4.288)$$

Observe that every other outcome of the measurement on the upper two wires will produce a similar result, i.e.,

$$a \mid \text{something}\rangle + b \mid \text{something different}\rangle$$

No possible measurement outcome results in information loss about either a or b.

The first operation on the right hand side applies $\oplus$ to the second and third qubit. We continue with our convention of marking control and data qubits with c and d:

$$\begin{aligned} & a \mid 0\rangle \oplus (\mid 1\rangle_c \mid 1\rangle_d) + b \mid 0\rangle \oplus (\mid 1\rangle_c \mid 0\rangle_d) \\ & \quad = a \mid 0\rangle \mid 1\rangle_c \mid 0\rangle_d + b \mid 0\rangle \mid 1\rangle_c \mid 1\rangle_d \end{aligned} \tag{4.289}$$

Next we pass the upper line through the $\boldsymbol{S}$ gate, which in this case simply multiplies $\mid 0\rangle$ by i, so that the state of the system becomes:

$$ia \mid 0\rangle \mid 1\rangle \mid 0\rangle + ib \mid 0\rangle \mid 1\rangle \mid 1\rangle \tag{4.290}$$

The next operation is quite difficult to write down symbolically because it couples the top and the bottom line of the circuit and it is the bottom line that controls the gate. Again our subscript convention helps:

$$\begin{aligned} & \oplus (ia \mid 0\rangle_d \mid 1\rangle \mid 0\rangle_c + ib \mid 0\rangle_d \mid 1\rangle \mid 1\rangle_c) \\ & \quad = ia \mid 0\rangle_d \mid 1\rangle \mid 0\rangle_c + ib \mid 1\rangle_d \mid 1\rangle \mid 1\rangle_c \end{aligned}$$

Now we apply the $\boldsymbol{S}$ gate to the top qubit and the $\boldsymbol{T}$ gate to the bottom one:

$$\begin{aligned} & ia\boldsymbol{S} \mid 0\rangle_S \mid 1\rangle\boldsymbol{T} \mid 0\rangle_T + ib\boldsymbol{S} \mid 1\rangle_S \mid 1\rangle\boldsymbol{T} \mid 1\rangle_T \\ & \quad = iai \mid 0\rangle_S \mid 1\rangle(-1) \mid 0\rangle_T + ib(1) \mid 1\rangle_S \mid 1\rangle(-i) \mid 1\rangle_T \\ & \quad = a \mid 0\rangle \mid 1\rangle \mid 0\rangle + b \mid 1\rangle \mid 1\rangle \mid 1\rangle \end{aligned}$$

And finally we apply a yet another upside controlled-NOT gate:

$$\begin{aligned} & \oplus (a \mid 0\rangle_d \mid 1\rangle \mid 0\rangle_c + b \mid 1\rangle_d \mid 1\rangle \mid 1\rangle_c) \\ & \quad = a \mid 0\rangle_d \mid 1\rangle \mid 0\rangle_c + b \mid 0\rangle_d \mid 1\rangle \mid 1\rangle_c \\ & \quad = \mid 0\rangle \mid 1\rangle \begin{pmatrix} a \\ b \end{pmatrix} \end{aligned}$$

You may say that in this case all these operations merely brought us back to where we were after we applied the controlled-NOT gate to the bottom two lines after the mid-circuit measurement. But this is just fine in this case. You should check yourself what is going to happen if states of the upper two qubits measured in the mid-circuit end up in the three other combinations, i.e., $\mid 00\rangle$, $\mid 10\rangle$ and $\mid 11\rangle$. The beauty of the Brassard circuit is that the same operations will always untangle the state of the bottom qubit putting it eventually back in $\mid \Psi\rangle$ every time.

Also, observe that the state of the middle qubit does not change after the measurement in the mid-circuit at all, so the qubit just stays as is, and the state of the upper qubit ultimately does not change either. Since these two qubits have already decohered, the bottom qubit ends up being forced into $\mid \Psi\rangle$ at the end of the computation.

This is why the Nielsen, Knill and LaFlamme experiment worked so nicely. After the four initial gates the experimenters simply waited a little allowing the

quantum spin state on the two carbon nuclei to collapse, or decohere. They simply used the environment to carry out "the measurement". Then they completed the computation by sending the remaining 6 gates to the molecule and, pronto, the spin of the Hydrogen nucleus was forced into the $|\ \Psi\rangle$ state.

What did Bohm think about all this?

People drilling Quantum Mechanics for meaning noticed fairly early that whereas Quantum Mechanics was strange enough for a single quantum particle, it was definitely way too strange for multiparticle systems. The famous paper by Einstein, Podolsky and Rosen about the "EPR Paradox" [33] goes all the way back to 1935. It took 29 years for Einstein and his colleague's objections to be coined into a mathematical form of Bell inequalities [7], and it took further 18 years for Bell inequalities to be tested experimentally by Aspect, Dalibard and Roger [1]. When it comes to things that are really profound progress tends to be slow. Recall that the Berry's paper about geometric phase shift was published in 1983 only.

The Schrödinger wave equation for two particles of identical mass looks as follows:

$$i\hbar\frac{\partial}{\partial t}\Psi\left(\boldsymbol{r}_1,\boldsymbol{r}_2,t\right)=\left[-\frac{\hbar^2}{2m}\left(\nabla_1^2+\nabla_2^2\right)+V\right]\Psi\left(\boldsymbol{r}_1,\boldsymbol{r}_2,t\right) \tag{4.291}$$

The requirement that the wave function should be symmetric or anti-symmetric if the particles are indistinguishable, is external to the Schrödinger equation.

Function Ψ is a complex-valued function on the configuration space of two particles. We can *always* rewrite this function in the polar representation:

$$\Psi\left(\boldsymbol{r}_1,\boldsymbol{r}_2,t\right)=R\left(\boldsymbol{r}_1,\boldsymbol{r}_2,t\right)e^{iS(\boldsymbol{r}_1,\boldsymbol{r}_2,t)/\hbar} \tag{4.292}$$

where R and S are two real functions on the configuration space. Upon such substitution equation (4.291) splits naturally into two equations for S and $P=R^2=\Psi^*\Psi$:

$$\frac{\partial S}{\partial t}+\frac{1}{2m}\left[\left(\boldsymbol{\nabla}_1 S\right)^2+\left(\boldsymbol{\nabla}_2 S\right)^2\right]+V+Q=0 \tag{4.293}$$

$$\frac{\partial P}{\partial t}+\boldsymbol{\nabla}_1\cdot\left(P\,\boldsymbol{\nabla}_1 S/m\right)+\boldsymbol{\nabla}_2\cdot\left(P\,\boldsymbol{\nabla}_2 S/m\right)=0 \tag{4.294}$$

where

$$Q=-\frac{\hbar^2}{2m}\frac{\left(\nabla_1^2+\nabla_2^2\right)R}{R} \tag{4.295}$$

Equation for S can be still interpreted as the Hamilton-Jacobi equation with the momenta of the two particles given by:

$$\boldsymbol{p}_1=\boldsymbol{\nabla}_1 S\quad\text{and}\quad\boldsymbol{p}_2=\boldsymbol{\nabla}_2 S \tag{4.296}$$

The above implies that the two particles are "guided" by the quantum potential Q in a correlated way. Quantum potential Q does not necessarily fall with the

distance between the two particles. The particles may therefore be strongly coupled even at long distances. As was the case in the single particle example, where the particle interacted *non-locally* with various distant obstacles, e.g., with the *other* slit in the double slit experiment, so here the particle interacts *non-locally* with the other particle.

There are no retardation terms in the quantum potential Q, so this non-local interaction is instananeous. This should not be so surprising perhaps, because, after all the Schrödinger equation is non-relativistic.

But one can carry out similar reasoning in context of the Dirac equation and the non-locality will still be present. Yet, one can also demonstrate that the form of the Quantum Potential that arises in the relativistic theory will not allow signals to be transmitted faster than light.

Another important observation that can be made here is that the quantum potential Q depends on the quantum state of the *whole* system in a way that cannot be described as *interaction* between the two particles. The system is indivisible.

This indivisible entanglement and non-locality, in principle, applies to the whole universe, and, in principle, it is impossible to disentangle one part of the universe from another one. But it turns out that it is possible to obtain an approximate separation of an entangled quantum system into multiple portions, which can be then considered in isolation from each other. This is what happens in typical laboratory situations, and in the thermodynamic limit, where quantum potential becomes negligible [14].

4.6 The Measurement

So far we have not talked much about the measurement itself, although in various places we have smuggled some ideas already, based on, should we say, *common knowledge* about things quantum mechanical.

There exists a very elaborate theory of quantum measurement, which arose in response to somewhat dogmatic role that the measurement played in early quantum mechanics. Physicists are people, who, by and large, don't like dogmas and axioms. Physics theories always arise from observations of nature in the first place, observations, which are then coded into mathematics. They are not *mathematical* theories. There is always a place in physics for asking questions and for trying to subvert even the most established theories. In fact most Nobel prizes result from such subversive activities.

This is very different from mathematics, where all theories rest on the foundations of axioms and primary notions, from which theorems and further definitions are then developed by logical reasoning.

In physics, and especially in quantum physics, logical reasoning may often lead astray.

In the following few sections we will not present the whole theory of measurement or even its small part. As has been the case in this course so far, we will only discuss what is relevant to quantum computing.

4.6.1 The Density Operator

In context of real life physics we cannot measure the wavefunction directly. The only thing we can measure are the so called *observables* such as energy, spin, momentum, charge, and these correspond to eigenvalues of various Hermitian operators, and probability distributions. The shape and form of the wave function must then be inferred from the combination of theory and experimental data.

In order to measure probability distributions we must work with a statistical ensemble of quantum systems. We cannot collect probability distributions from a single quantum system, because every time we probe it, the system collapses to an eigenstate of whatever operation we use to do the probing.

An example of a statistical ensemble of quantum states is the collection of molecules in a magnetically inactive suspension in an NMR experiment. When the measurement is finally made, and we'll explain down the road what is actually measured and how, we end up finding that a certain percentage of molecules was in state $\mid \Psi_1\rangle$, then some other percentage was in state $\mid \Psi_2\rangle$, and yet some other percentage may have been in state $\mid \Psi_3\rangle$ and so on. The percentages measured are classical probabilities of finding a molecule in some such state. The probabilities can be associated for example with intensity of spectral lines that correspond to various energy levels, or with intensity of beams in a Stern-Gerlach experiment, or with strength of magnetization of an NMR sample in various directions.

Once the probabilities have been collected, we can assemble them in an object called the *density operator*, the definition of which is as follows:

$$\boldsymbol{\rho} = \sum_i p_i \mid \Psi_i\rangle\langle\Psi_i \mid \tag{4.297}$$

where p_i is a probability of finding the quantum system in state $\mid \Psi_i\rangle$.

As the ensemble evolves unitarily in time thusly

$$\mid \Psi(t_2)\rangle = \boldsymbol{U}(t_2, t_1) \mid \Psi(t_1)\rangle \tag{4.298}$$

its corresponding density operator evolves too:

$$\begin{aligned} \boldsymbol{\rho}(t_2) &= \sum_i p_i \mid \Psi(t_2)\rangle\langle\Psi(t_2) \mid \\ &= \sum_i p_i \boldsymbol{U}(t_2, t_1) \mid \Psi(t_1)\rangle\langle\Psi(t_1) \mid \boldsymbol{U}^\dagger(t_2, t_1) \\ &= \boldsymbol{U}(t_2, t_1)\boldsymbol{\rho}(t_1)\boldsymbol{U}^\dagger(t_2, t_1) \end{aligned} \tag{4.299}$$

A density operator does not define the wave function uniquely. In fact it is possible, as you will see later, to write down and even measure a density operator to which no single wave function corresponds. Such density operators are said to described *mixed* or *impure* states.

Consider the following two vectors

$$\begin{aligned} \mid a\rangle &= \sqrt{\frac{3}{4}} \mid 0\rangle + \sqrt{\frac{1}{4}} \mid 1\rangle \\ \mid b\rangle &= \sqrt{\frac{3}{4}} \mid 0\rangle - \sqrt{\frac{1}{4}} \mid 1\rangle \end{aligned}$$

Now consider a density operator that corresponds to these two vectors with equal probability of $1/2$ assigned to each state

$$\begin{aligned} \boldsymbol{\rho} &= \frac{1}{2} \mid a\rangle\langle a \mid + \frac{1}{2} \mid b\rangle\langle b \mid \\ &= \frac{3}{4} \mid 0\rangle\langle 0 \mid + \frac{1}{4} \mid 1\rangle\langle 1 \mid \end{aligned}$$

As you see the same density operator describes another statistical ensemble, in which state $\mid 0\rangle$ appears with probability $3/4$ and state $\mid 1\rangle$ appears with probability $1/4$.

EXERCISE Demonstrate this equality.

4.6.2 Projective Measurement

A quantum mechanical measurement can be thought of as a filtration process. In the introductory section, in which we talked about probability amplitudes you saw this filtration in action. We filtered light beams to produce beams with specified polarization.

In the language of Hilbert space such a filtration process is described by a projection operator $\boldsymbol{P}_m$ associated with some observable $\boldsymbol{M}$ and its eigenvalue M_m and eigenstate $\mid \Psi_m\rangle$. The projection extracts the component of a quantum state that points in the direction of $\mid \Psi_m\rangle$:

$$\boldsymbol{P}_m \mid \Psi\rangle = \boldsymbol{P}_m \sum_i \alpha_i \mid \Psi_i\rangle = \alpha_m \mid \Psi_m\rangle \tag{4.300}$$

After the measurement has been completed the system remains in state $\mid \Psi_m\rangle$. The probability of finding it there is 1 until the system evolves away from this state by natural evolution, or by some other means.

The probability of finding $\mid \Psi\rangle$ in $\mid \Psi_m\rangle$ is

$$\begin{aligned} \langle \Psi \mid \boldsymbol{P}_m \mid \Psi\rangle &= \langle \Psi \mid \alpha_m \mid \Psi_m\rangle \\ &= \sum_i \alpha_i^* \langle \Psi_i \mid \alpha_m \mid \Psi_m\rangle \\ &= \alpha_m^* \alpha_m \langle \Psi_m \mid \Psi_m\rangle \\ &= \alpha_m^* \alpha_m = p_m \end{aligned}$$

But also observe that

$$\langle \Psi \mid \boldsymbol{P}_m^\dagger \boldsymbol{P}_m \mid \Psi\rangle = \langle \Psi_m \mid \alpha_m^* \alpha_m \mid \Psi_m\rangle = p_m$$

Hence

$$\boldsymbol{P}_m^\dagger = \boldsymbol{P}_m = \boldsymbol{P}_m \boldsymbol{P}_m \tag{4.301}$$

Our projections, associated with eigenvectors of $\boldsymbol{M}$ are *idempotent* and Hermitian.

Projections associated with eigenvectors of $\boldsymbol{M}$ are orthogonal and they add up to $\mathbf{1}$:

$$\boldsymbol{P}_i \boldsymbol{P}_j = \delta_{ij} \boldsymbol{P}_i \quad \text{and} \quad \sum_i \boldsymbol{P}_i = \mathbf{1} \tag{4.302}$$

Wrapping all the above we can described the state immediately after the measurement as:

$$\begin{aligned} \mid \Psi_m \rangle &= \frac{1}{\alpha_m} \boldsymbol{P}_m \mid \Psi \rangle \\ &= \frac{\boldsymbol{P}_m \mid \Psi \rangle}{\sqrt{\langle \Psi \mid \boldsymbol{P}_m \mid \Psi \rangle}} \end{aligned}$$

If $\mid \Psi_i \rangle$ are eigenstates of some operator $\boldsymbol{M}$ then an eigenvalue M_i is associated with every eigenvector. The *average* value of $\boldsymbol{M}$ on $\mid \Psi \rangle$ is

$$\begin{aligned} \langle \Psi \mid \boldsymbol{M} \mid \Psi \rangle &= \langle \Psi \mid \sum_i M_i \alpha_i \mid \Psi_i \rangle \\ &= \sum_j \langle \Psi_j \mid \alpha_j^* \sum_i M_i \alpha_i \mid \Psi_i \rangle \\ &= \sum_m \alpha_m^* \alpha_m M_m \langle \Psi_m \mid \Psi_m \rangle \\ &= \sum_m p_m M_m \end{aligned}$$

This should be understood as follows: imagine that you have a statistical ensemble of quantum systems, all in the same state $\mid \Psi \rangle$. The density matrix of this ensemble would be simply $\boldsymbol{\rho} = \mid \Psi \rangle \langle \Psi \mid$. Then if you perform a measurement $\boldsymbol{M}$ on this ensemble, you'll get $\langle M \rangle$ as an answer.

4.6.3 Projective Measurements and the Density Operator

The relationships describing projective measurements can be translated into the formalism of density operator – extending the usefulness of the description, because this time we may apply it to a mixture of various quantum states within our statistical ensemble. The complication is that in the density operator formalism classical ensemble probabilities combine with quantum probabilities, which may easily lead to confusion. Let us use greek indexes for the classical probabilities then and latin indexes for the quantum ones, so that

$$\boldsymbol{\rho} = \sum_\mu p_\mu \mid \Psi_\mu \rangle$$

and

$$\mid \Psi_\mu \rangle = \sum_i \alpha_{\mu i} \mid \Psi_i \rangle$$

where $\mid \Psi_i \rangle$ are the basis vectors of some observable Hermitian operator $\boldsymbol{M}$. How does measurement affect $\boldsymbol{\rho}$?

$$\begin{aligned} \boldsymbol{P}_m \boldsymbol{\rho} \boldsymbol{P}_m^\dagger &= \sum_\mu p_\mu \boldsymbol{P}_m \mid \Psi_\mu \rangle\langle \Psi_\mu \mid \boldsymbol{P}_m^\dagger \\ &= \sum_\mu p_\mu \alpha_{\mu m} \mid \Psi_m \rangle\langle \Psi_m \mid \alpha_{\mu m}^* \\ &= \sum_\mu p_\mu p_{\mu m} \mid \Psi_m \rangle\langle \Psi_m \mid \\ &= p_m \mid \Psi_m \rangle\langle \Psi_m \mid \end{aligned}$$

The result is *not* a dinkum density operator, because its trace isn't 1, and, as you will see in the next section, the trace of the density operator is always 1. However, $\boldsymbol{P}_m \boldsymbol{\rho} \boldsymbol{P}_m^\dagger / p_m$ is a fine density operator of the state that the measurement has produced. You will see a nicer formula for this state further down.

Probabilities p_m can be extracted from $\boldsymbol{\rho}$ using the so called *trace* relationships. These are beloved by quantum physicists, but they can look quite obscure to the unaided eye. Yet, if you rewrite them in terms of *index notation*, which is my preferred way of doing things, because I happen to have a General Relativity background, they turn out to be rather trivial. This is one such relationship that comes useful:

$$\langle \Psi \mid \boldsymbol{P}^\dagger \boldsymbol{P} \mid \Psi \rangle = \sum_{jkl} \Psi_j P_k{}^j P^k{}_l \Psi^l = \mathrm{tr}_{jm} \sum_{kl} P_k{}^j P^k{}_l \Psi^l \Psi_m = \mathrm{tr} \left(\boldsymbol{P}^\dagger \boldsymbol{P} \mid \Psi \rangle\langle \Psi \mid \right)$$

From this we get

$$p_{\mu m} = \langle \Psi_\mu \mid \boldsymbol{P}_m^\dagger \boldsymbol{P}_m \mid \Psi_\mu \rangle = \mathrm{tr} \left(\boldsymbol{P}_m^\dagger \boldsymbol{P}_m \mid \Psi_\mu \rangle\langle \Psi_\mu \mid \right) \tag{4.303}$$

and now applying it to p_m:

$$\begin{aligned} p_m &= \sum_\mu p_\mu p_{\mu m} = \sum_\mu p_\mu \, \mathrm{tr} \left(\boldsymbol{P}_m^\dagger \boldsymbol{P}_m \mid \Psi_\mu \rangle\langle \Psi_\mu \mid \right) \\ &= \mathrm{tr} \left(\boldsymbol{P}_m^\dagger \boldsymbol{P}_m \sum_\mu p_\mu \mid \Psi_\mu \rangle\langle \Psi_\mu \mid \right) \\ &= \mathrm{tr} \left(\boldsymbol{P}_m^\dagger \boldsymbol{P}_m \boldsymbol{\rho} \right) \end{aligned}$$

We can now write the formula for the density operator of a state *resulting* from a measurement as:

$$\mid \Psi_m \rangle\langle \Psi_m \mid = \boldsymbol{P}_m \boldsymbol{\rho} \boldsymbol{P}_m / p_m = \frac{\boldsymbol{P}_m \boldsymbol{\rho} \boldsymbol{P}_m}{\mathrm{tr} \left(\boldsymbol{P}_m^\dagger \boldsymbol{P}_m \boldsymbol{\rho} \right)} \tag{4.304}$$

Recall that for projections associated with eigenvectors of a Hermitian operator $\boldsymbol{M}$ we have $\boldsymbol{P}_m^\dagger \boldsymbol{P}_m = \boldsymbol{P}_m$, so the relation above is also written as

$$p_m = \mathrm{tr}\,(\boldsymbol{P}_m \boldsymbol{\rho}) \tag{4.305}$$

From this it follows that an expectation value of $\boldsymbol{M}$ on state $\boldsymbol{\rho}$ is given by:

$$\langle \boldsymbol{M} \rangle = \mathrm{tr}\,(\boldsymbol{M}\boldsymbol{\rho}) \tag{4.306}$$

EXERCISE Prove this relationship attending to details.

4.6.4 Other Properties of the Density Operator

In this section we are going to summarize briefly the properties of the density operator.

1. The density operator is Hermitian. This is easy to see from the basic definition. For a pure state:

   $$\boldsymbol{\rho} = \mid \Psi\rangle\langle\Psi \mid$$

   and this is clearly Hermitian. For a mixed state $\boldsymbol{\rho}$ is a sum of Hermitian operators and so it is Hermitian.

2. Trace of $\boldsymbol{\rho}$ is 1. This is also easy to see from the basic definition:

   $$\mathrm{tr}\boldsymbol{\rho} = \sum_i p_i = 1$$

3. Density operator is positive. This means that for any state $\mid \phi\rangle$ in $\mathcal{H}$ the following holds: $\langle \phi \mid \boldsymbol{\rho} \mid \phi \rangle \geq 0$. This is easy to see by direct evaluation:

   $$\begin{aligned} \langle \phi \mid \boldsymbol{\rho} \mid \phi \rangle &= \sum_i p_i \langle \phi \mid \Psi_i \rangle \langle \Psi_i \mid \phi \rangle \\ &= \sum_i p_i \left| \langle \phi \mid \Psi_i \rangle \right|^2 \\ &\geq 0 \end{aligned}$$

The inverse is also true, i.e., any operator that is Hermitian, has trace 1 and is positive, is a density operator for some statistical ensemble of quantum states.

4.6.5 Density Operator of a Single Qubit: The Bloch Sphere

The most general Hermitian 2×2 matrix can be formed by multiplying matrices from $\{\mathbf{1}, \boldsymbol{\sigma}_x, \boldsymbol{\sigma}_y, \boldsymbol{\sigma}_z\}$ by real coefficients. Since every Pauli matrix $\boldsymbol{\sigma}_i$ is traceless, all trace that a $\boldsymbol{\rho}$ has must come from $\mathbf{1}$, which we must multiply by $1/2$ in order

to get $\mathrm{tr}\boldsymbol{\rho} = 1$. Consequently, the most general form of a density matrix for a qubit is:

$$\boldsymbol{\rho} = \frac{1}{2}\left(\mathbf{1} + r_x\boldsymbol{\sigma}_x + r_y\boldsymbol{\sigma}_y + r_z\boldsymbol{\sigma}_z\right) \tag{4.307}$$

$$= \frac{1}{2}\begin{pmatrix} 1+r_z & r_x - ir_y \\ r_x + ir_y & 1-r_z \end{pmatrix} \tag{4.308}$$

A positive definite operator must be Hermitian, which we have already ensured, it must have a positive trace, which we have ensured too and it must have a positive determinant: $\det\boldsymbol{\rho} \geq 0$, which, on evaluating the determinant from equation (4.308) yields

$$1 - r_z^2 - r_x^2 - r_y^2 \geq 0$$

This implies that

$$r_x^2 + r_y^2 + r_z^2 \leq 1$$

There is a one-to-one correspondence here between the possible density matrices of a single qubit and the points *in* the ball $0 \leq |\boldsymbol{r}| \leq 1$. This ball is called a *Bloch Sphere* – though it is a ball, really, and not a sphere.

For points on the surface of the sphere the determinant of $\boldsymbol{\rho}$ vanishes. We can always diagonalize $\boldsymbol{\rho}$ and then the determinant is $\rho_{11}\rho_{22}$. For this to be zero, one or the other must be zero. For the trace to be 1, one or the other must be 1. In short

$$\boldsymbol{\rho} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{or} \quad \boldsymbol{\rho} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

In other words $\boldsymbol{\rho} = \mid 0'\rangle\langle 0' \mid$ or $\boldsymbol{\rho} = \mid 1'\rangle\langle 1' \mid$. Remember that we may have rotated our basis in order to diagonalize $\boldsymbol{\rho}$ so these states $\mid 0'\rangle$ and $\mid 1'\rangle$ may not necessarily point up and down within the Bloch Sphere. But they all represent the so called *pure* states, so that

$$\boldsymbol{\rho}(\boldsymbol{n}) = \mid \Psi(\boldsymbol{n})\rangle\langle\Psi(\boldsymbol{n}) \mid$$

where $\boldsymbol{n}$ is a unit length vector pointing in some direction within the Bloch Sphere.

EXERCISE Using equation (4.308) show that

$$\mathrm{tr}\boldsymbol{\rho}^2 = \left(1 + r^2\right)/2$$

The result of this exercise gives us an easy criterion for distinguishing between *pure* and *impure* states. For pure states we have

$$r^2 = 1 \quad \rightarrow \quad \mathrm{tr}\boldsymbol{\rho}^2 = 1$$

For impure states

$$r^2 < 1 \quad \rightarrow \quad \mathrm{tr}\boldsymbol{\rho}^2 < 1$$

EXERCISE Show that state

$$\boldsymbol{\rho} = \frac{3}{4} \mid 0\rangle\langle 0 \mid + \frac{1}{4} \mid 1\rangle\langle 1 \mid$$

is impure.

We can now support our previous definition of pure and impure (or *mixed* states with mathematical rigour. A given state described by $\boldsymbol{\rho}$ is a mixture if there is no such single state $\mid \psi\rangle$ that $\boldsymbol{\rho} = \mid \psi\rangle\langle\psi \mid$. This is equivalent to $\mathrm{tr}\boldsymbol{\rho}^2 < 1$ and to $r < 1$.

Vector $\boldsymbol{r}$ can be given a nice physical meaning. Evaluate the expectation value for the $\boldsymbol{n} \cdot \boldsymbol{\sigma}$ operator, where $\boldsymbol{n}$ is a unit length vector pointing anywhere, on $\boldsymbol{\rho}(\boldsymbol{r})$:

$$\begin{aligned}
\mathrm{tr}\left(\boldsymbol{n} \cdot \boldsymbol{\sigma} \frac{1}{2}\left(\mathbf{1} + \boldsymbol{r} \cdot \boldsymbol{\sigma}\right)\right) &= \frac{1}{2}\mathrm{tr}\left(\boldsymbol{n} \cdot \boldsymbol{\sigma} + \boldsymbol{n} \cdot \boldsymbol{\sigma}\boldsymbol{r} \cdot \boldsymbol{\sigma}\right) \\
&= \frac{1}{2}\mathrm{tr}\left(\boldsymbol{n} \cdot \boldsymbol{\sigma}\boldsymbol{r} \cdot \boldsymbol{\sigma}\right) \\
&= \frac{1}{2}\mathrm{tr}\left(\sum_{ij} n_i r_j \boldsymbol{\sigma}_i \boldsymbol{\sigma}_j\right) \\
&= \frac{1}{2}\mathrm{tr}\left(\sum_{i} n_i r_i \mathbf{1}\right) \\
&= \sum_{i} n_i r_i = \boldsymbol{n} \cdot \boldsymbol{r}
\end{aligned}$$

The vector $\boldsymbol{r}$ can therefore be thought of as an expectation value of spin polarization, and it can be obtained by measuring $\boldsymbol{n} \cdot \boldsymbol{\sigma}$ along each direction $\boldsymbol{e}_x$, $\boldsymbol{e}_y$ and $\boldsymbol{e}_z$.

4.6.6 Partial Trace

Consider a system that is a combination of two quantum systems mutually entangled with each other. Let its density matrix be $\boldsymbol{\rho}^{AB}$, where AB is just a mark and not an exponent. The density matrix for the subsystem A can be obtained from the density matrix for the combined system by evaluating a *partial trace* over the B component:

$$\boldsymbol{\rho}^A = \mathrm{tr}_B\left(\boldsymbol{\rho}^{AB}\right)$$

The partial trace is defined as follows:

$$\mathrm{tr}_B\left(\mid a_1\rangle\langle a_2 \mid \otimes \mid b_1\rangle\langle b_2 \mid\right) = \mid a_1\rangle\langle a_2 \mid \mathrm{tr}\left(\mid b_1\rangle\langle b_2 \mid\right)$$

This definition stems from the following relationship, which describes two quantum systems $\boldsymbol{\rho}$ and $\boldsymbol{\sigma}$. The systems combine into $\boldsymbol{\rho} \otimes \boldsymbol{\sigma}$. Evaluating a partial trace over the $\boldsymbol{\sigma}$ component for this combined system yields pure $\boldsymbol{\rho}$:

$$\boldsymbol{\rho}^A = \mathrm{tr}_B\left(\boldsymbol{\rho} \otimes \boldsymbol{\sigma}\right) = \boldsymbol{\rho}\mathrm{tr}\,\boldsymbol{\sigma} = \boldsymbol{\rho}$$

It is possible to present more involved arguments in favour of this constructions even for systems that are maximally entangled. The argument demonstrates that partial trace is a unique operation that gives rise to the correct description of observable quantities for subsystems of a composite system.

The partial trace operation yields some very interesting results for maximally entangled quantum states. Consider the following state:

$$\mid \Psi\rangle = \frac{1}{\sqrt{2}} (\mid 00\rangle + \mid 11\rangle)$$

The density matrix for this state is

$$\boldsymbol{\rho} = \frac{1}{2} (\mid 00\rangle\langle 00 \mid + \mid 11\rangle\langle 00 \mid + \mid 00\rangle\langle 11 \mid + \mid 11\rangle\langle 11 \mid)$$

A partial trace over the second qubit is evaluated by replacing $\mid x\rangle\langle y \mid$ for this qubit with $\langle x \mid y\rangle$, and so

$$\begin{aligned} \mathrm{tr}_2\, \boldsymbol{\rho} &= \frac{1}{2} (\mid 0\rangle\langle 0 \mid \langle 0 \mid 0\rangle + \mid 1\rangle\langle 0 \mid \langle 1 \mid 0\rangle + \mid 0\rangle\langle 1 \mid \langle 0 \mid 1\rangle + \mid 1\rangle\langle 1 \mid \langle 1 \mid 1\rangle) \\ &= \frac{1}{2} (\mid 0\rangle\langle 0 \mid + \mid 1\rangle\langle 1 \mid) = \frac{1}{2}\mathbf{1} \end{aligned}$$

Observe that $\mathrm{tr}\boldsymbol{\rho}^2 = 1/2 < 1$, which implies that this state is a mixture.

Even though the original two qubit state was pure, i.e., we knew everything about it, after tracing away the second qubit we end up with a state about which we do not have maximal knowledge.

The opposite is generally true too.

> If you have a mixture, it is always possible to construct a higher dimensional state comprising more qubits (or particles), which is pure, and the partial trace of which returns the original mixture. This procedure is called *purification*.

Coming back to our example, the property that a partial trace of a pure maximally entangled state returns a mixture is one of the characteristics of entanglement.

> Consider the teleportation state of the Brassard circuit just before the decoherence takes place. The state of the circuit is then given by
>
> $$\begin{aligned} \frac{1}{2}\Big(&\mid 00\rangle(a \mid 0\rangle + b \mid 1\rangle) + \mid 01\rangle(b \mid 0\rangle + a \mid 1\rangle) \\ &+ \mid 10\rangle(-a \mid 0\rangle + b \mid 1\rangle) + \mid 11\rangle(b \mid 0\rangle - a \mid 1\rangle)\Big) \end{aligned}$$
>
> Everyone of the terms in the sum has $1/4^{\mathrm{th}}$ probability of occuring, therefore the density matrix for the system after the decoherence is
>
> $$\begin{aligned} \boldsymbol{\rho} = \frac{1}{4}\Big(&\mid 00\rangle\langle 00 \mid (a \mid 0\rangle + b \mid 1\rangle)(a^*\langle 0 \mid + b^*\langle 1 \mid) \\ &+ \mid 01\rangle\langle 01 \mid (b \mid 0\rangle + a \mid 1\rangle)(b^*\langle 0 \mid + a^*\langle 1 \mid) \\ &+ \mid 10\rangle\langle 10 \mid (a \mid 0\rangle - b \mid 1\rangle)(a^*\langle 0 \mid - b^*\langle 1 \mid) \\ &+ \mid 11\rangle\langle 11 \mid (b \mid 0\rangle - a \mid 1\rangle)(b^*\langle 0 \mid - a^*\langle 1 \mid)\Big) \end{aligned}$$

Tracing this over the first two qubits yields

$$\begin{aligned}\boldsymbol{\rho}^B &= \frac{1}{4}\Big((a\mid 0\rangle + b\mid 1\rangle)(a^*\langle 0\mid + b^*\langle 1\mid) \\ &\qquad +(b\mid 0\rangle + a\mid 1\rangle)(b^*\langle 0\mid + a^*\langle 1\mid) \\ &\qquad +(a\mid 0\rangle - b\mid 1\rangle)(a^*\langle 0\mid - b^*\langle 1\mid) \\ &\qquad +(b\mid 0\rangle - a\mid 1\rangle)(b^*\langle 0\mid - a^*\langle 1\mid)\Big) \\ &= \frac{1}{2}\Big((|a|^2+|b|^2)\mid 0\rangle\langle 0\mid + (|a|^2+|b|^2)\mid 1\rangle\langle 1\mid\Big) \\ &= \mathbf{1}/2\end{aligned}$$

We can see immediately that the state is not pure, but by now we should expect it. However, we also see that the state has no dependence on the state being teleported. It is blank. The information must be supplied by conventional means in order to complete teleportation. There is no way that teleportation can be used to transfer data faster than speed of light.

4.6.7 The NMR Measurement

The measurement in NMR is carried out by switching Helmholtz coils from the emitter to the receiver mode and detecting the magnetization of the sample in the $\boldsymbol{e}_x \times \boldsymbol{e}_y$ plane. The single qubit Hermitian operator that corresponds to this magnetization is proportional to

$$\boldsymbol{\sigma}_x + \boldsymbol{\sigma}_y$$

On averaging this outcome over the whole statistical ensemble described by the density operator $\boldsymbol{\rho}$ we get the expectation value:

$$\langle i\boldsymbol{\sigma}_x + \boldsymbol{\sigma}_y\rangle = \mathrm{tr}\ ((\boldsymbol{\sigma}_x + \boldsymbol{\sigma}_y)\,\boldsymbol{\rho})$$

The measurement is not instantaneous. As the measurement goes on the density operator evolves in time, so that what is really observed is

$$\mathrm{tr}\ \left((\boldsymbol{\sigma}_x + \boldsymbol{\sigma}_y)\, e^{-i\boldsymbol{H}t/\hbar}\boldsymbol{\rho}e^{i\boldsymbol{H}t/\hbar}\right)$$

In a multi-qubit molecule we can observe this separately for each qubit, as long as these are sufficiently frequency shifted from each other. The Helmoltz coils and electronics used in the measurement convert this observable into a voltage signal, which, for a given kth spin, becomes

$$V(t) = V_0\ \mathrm{tr}\ \left(\left(\boldsymbol{\sigma}_x^{(k)} + \boldsymbol{\sigma}_y^{(k)}\right) e^{-i\boldsymbol{H}t/\hbar}\boldsymbol{\rho}e^{i\boldsymbol{H}t/\hbar}\right)$$

where V_0 is a constant dependent on the geometry and other electronic properties of the coils. This signal will oscillate with the resonance frequency of the kth spin, $\omega_0^{(k)}$. The apparatus will mix this signal with an oscillator locked at

this frequency, and the result will be Fourier transformed to reveal peaks in the vicinity of $\omega_0^{(k)}$, but not the $\omega_0^{(k)}$ itself.

This is a very subtle, non-invasive measurement, which is why NMR practitioners were perhaps the first to observe decoherence. The magnetization signal decays exponentially due to various factors, such as decoherence, inhomogeneity of the magnetic field, thermalization of spins, and presence of spin-spin coupling, which were supressed during the computation. The resulting density matrix for a single qubit system in this context depends on time and this dependence can be described phenomenologically as follows:

$$\boldsymbol{\rho} = \begin{pmatrix} (a-a_0)e^{-t/\tau_1} + a_0 & be^{-t/2\tau_2} \\ b^* e^{-t/2\tau_2} & (a_0-a)e^{-t/\tau_1} + 1 - a_0 \end{pmatrix}$$

where τ_1 is called a longitudinal and τ_2 a transverse relaxation rate.

Observe that for $t = 0$ we have

$$\boldsymbol{\rho} = \begin{pmatrix} a & b \\ b^* & 1-a \end{pmatrix}$$

which is a general form of the density matrix. For $t \to \infty$ the density matrix becomes

$$\boldsymbol{\rho} = \begin{pmatrix} a_0 & 0 \\ 0 & 1-a_0 \end{pmatrix}$$

This vanishing of the off-diagonal terms implies the vanishing of superposition.

It is possible to calculate τ_1 and τ_2 theoretically, and it is also easy to measure these two rates. These activities provided NMR practitioners with the first glimpse of quantum decoherence. Because NMR is a technique used primarily by chemists, and only more recently also by physicians and biologists, for a long time physicists were not quite aware that their colleagues in the chemistry departments actually observed the collapse of the wave function as it unfolded in time. Some may have a problem with this even today.

Temporal and Spatial Averaging

When the computation is to begin the NMR sample is allowed to thermalize. This results in its density operator becoming

$$\boldsymbol{\rho}(T) = \frac{e^{-\boldsymbol{H}/kT}}{\mathrm{tr}e^{-\boldsymbol{H}/kT}} \tag{4.309}$$

where k is the Boltzmann constant and T is temperature. This means that it is going to be very difficult to force the whole sample into a pure state, so that a well defined initial state can be provided for the computation, for example all qubits in the whole ensemble being in state $\mid 0\rangle$.

As was the case with reversing and stopping time, there is a trick, which lets us do this indirectly. The trick is in averaging results of runs with permuted initial states.

Consider an ensemble of two-qubit systems described by:

$$\boldsymbol{\rho}_1 = \begin{pmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & d \end{pmatrix} \tag{4.310}$$

where $a+b+c+d = 1$. By sending various combinations of controlled-NOT gates to the ensemble, it is possible to permute the populations so that the following two states can be also generated for successive computation runs:

$$\boldsymbol{\rho}_2 = \begin{pmatrix} a & 0 & 0 & 0 \\ 0 & c & 0 & 0 \\ 0 & 0 & d & 0 \\ 0 & 0 & 0 & b \end{pmatrix} \tag{4.311}$$

$$\boldsymbol{\rho}_2 = \begin{pmatrix} a & 0 & 0 & 0 \\ 0 & d & 0 & 0 \\ 0 & 0 & b & 0 \\ 0 & 0 & 0 & c \end{pmatrix} \tag{4.312}$$

$$\tag{4.313}$$

Now let us perform an identical computation on each of these ensembles. Every computation, however long, can be described in terms of a single unitary operation $\boldsymbol{U}$. The results of the computation on the ensembles will therefore be:

$$\begin{array}{l} \boldsymbol{U}\boldsymbol{\rho}_1\boldsymbol{U}^\dagger \\ \boldsymbol{U}\boldsymbol{\rho}_2\boldsymbol{U}^\dagger \\ \boldsymbol{U}\boldsymbol{\rho}_3\boldsymbol{U}^\dagger \end{array}$$

Let us now add all three results and see what comes out:

$$\begin{aligned} &\boldsymbol{U}\boldsymbol{\rho}_1\boldsymbol{U}^\dagger + \boldsymbol{U}\boldsymbol{\rho}_2\boldsymbol{U}^\dagger + \boldsymbol{U}\boldsymbol{\rho}_3\boldsymbol{U}^\dagger \\ &= \boldsymbol{U}\left(\boldsymbol{\rho}_1 + \boldsymbol{\rho}_2 + \boldsymbol{\rho}_3\right)\boldsymbol{U}^\dagger \\ &= \boldsymbol{U}\begin{pmatrix} 3a & 0 & 0 & 0 \\ 0 & b+c+d & 0 & 0 \\ 0 & 0 & b+c+d & 0 \\ 0 & 0 & 0 & b+c+d \end{pmatrix}\boldsymbol{U}^\dagger \\ &= \boldsymbol{U}\begin{pmatrix} 3a & 0 & 0 & 0 \\ 0 & 1-a & 0 & 0 \\ 0 & 0 & 1-a & 0 \\ 0 & 0 & 0 & 1-a \end{pmatrix}\boldsymbol{U}^\dagger \\ &= \boldsymbol{U}\begin{pmatrix} 1-a+4a-1 & 0 & 0 & 0 \\ 0 & 1-a & 0 & 0 \\ 0 & 0 & 1-a & 0 \\ 0 & 0 & 0 & 1-a \end{pmatrix}\boldsymbol{U}^\dagger \end{aligned}$$

$$= \boldsymbol{U}\left((4a-1)\begin{pmatrix}1&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix}+(1-a)\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}\right)\boldsymbol{U}^\dagger$$
$$= \boldsymbol{U}\left((4a-1)\mid 00\rangle\langle 00\mid +(1-a)\mathbf{1}\right)\boldsymbol{U}^\dagger$$

Let us now evaluate an expectation value of, say, $\boldsymbol{\sigma}_x+\boldsymbol{\sigma}_y$ on $\boldsymbol{U}\left(\boldsymbol{\rho}_1+\boldsymbol{\rho}_2+\boldsymbol{\rho}_3\right)\boldsymbol{U}^\dagger$:

$$\begin{aligned}\langle\boldsymbol{\sigma}_x+\boldsymbol{\sigma}_y\rangle &= \mathrm{tr}\left(\left(\boldsymbol{\sigma}_x+\boldsymbol{\sigma}_y\right)\boldsymbol{U}\left((4a-1)\mid 00\rangle\langle 00\mid +(1-a)\mathbf{1}\right)\boldsymbol{U}^\dagger\right)\\ &= (4a-1)\mathrm{tr}\left(\left(\boldsymbol{\sigma}_x+\boldsymbol{\sigma}_y\right)\boldsymbol{U}\mid 00\rangle\langle 00\mid \boldsymbol{U}^\dagger\right)+(1-a)\mathrm{tr}\left(\left(\boldsymbol{\sigma}_x+\boldsymbol{\sigma}_y\right)\mathbf{1}\right)\\ &= (4a-1)\mathrm{tr}\left(\left(\boldsymbol{\sigma}_x+\boldsymbol{\sigma}_y\right)\boldsymbol{U}\mid 00\rangle\langle 00\mid \boldsymbol{U}^\dagger\right)\end{aligned}$$

The second term in the sum above vanished, because $\boldsymbol{\sigma}_x$ and $\boldsymbol{\sigma}_y$ are traceless.

The result of our summation of results obtained for computations on three different ensembles is as if the computation was performed on a pure state $\mid 00\rangle\langle 00\mid$.

The individual computations $\boldsymbol{U}\boldsymbol{\rho}_1\boldsymbol{U}^\dagger$, $\boldsymbol{U}\boldsymbol{\rho}_2\boldsymbol{U}^\dagger$ and $\boldsymbol{U}\boldsymbol{\rho}_3\boldsymbol{U}^\dagger$ can be performed one *after* another using the same apparatus, in which case the procedure will be called *temporal averaging* or *temporal labeling*.

Alternatively the computations can be carried out at the same time on three different systems, using three different, but identically prepared, samples.

Sometimes all three computations can be carried out on the same sample, if it is immersed in a machine that can apply different static fields and different signals to various parts of the same sample. This is really equivalent to performing three computations using three different machines, even though we have only one vial. The technique is then called *spatial averaging* or *spatial labeling*.

Averaging, temporal or spatial, is right at the core of NMR quantum computation, together with reversing and halting time (refocusing), because this is the only way that a *pseudo*-pure state can be delivered as a start-up value.

This procedure is sensitive to temperature. It turns out that the total signal that can be obtained this way decreases exponentially with the number of qubits. This problem can be alleviated by cooling the sample. Yet, there is a limit on the total number of qubits that even a super-cooled NMR computer can work with effectively.

State Tomography

As the computation unfolds it is always possible to stop it briefly and measure magnetization thus obtaining an insight into the state of the register during computation. This will not ruin the register, because NMR measurement is so non-invasive, and decoherence time is quite long.

We have already demonstrated that

$$\langle\boldsymbol{n}\cdot\boldsymbol{\sigma}\rangle = \boldsymbol{n}\cdot\boldsymbol{r}$$

Substituting $\boldsymbol{e}_x$, $\boldsymbol{e}_y$ and $\boldsymbol{e}_z$ in place of $\boldsymbol{n}$ yields

$$\langle \boldsymbol{\sigma}_x \rangle = \mathrm{tr}\,(\boldsymbol{\sigma}_x \boldsymbol{\rho}) = r_x \tag{4.314}$$
$$\langle \boldsymbol{\sigma}_y \rangle = \mathrm{tr}\,(\boldsymbol{\sigma}_y \boldsymbol{\rho}) = r_y \tag{4.315}$$
$$\langle \boldsymbol{\sigma}_z \rangle = \mathrm{tr}\,(\boldsymbol{\sigma}_z \boldsymbol{\rho}) = r_z \tag{4.316}$$

Since $\boldsymbol{\rho} = \frac{1}{2}(\mathbf{1} + \boldsymbol{r} \cdot \boldsymbol{\sigma})$, measuring r_x, r_y, and r_z, gives us $\boldsymbol{\rho}$.

This procedure is called *state tomography*. State tomography provides us with means to *debug* NMR programs.

4.7 Interaction with the Environment

In this section we are going to build on the partial trace formalism we introduced in section 4.6.6, in order to describe behaviour of quantum systems, which interact with the environment. We are in for two unpleasant surprises on this occasion. The surprises are that

- Measurements performed on systems that interact with the environment are not orthogonal projections.
- Evolution of systems that interact with the environment is not unitary.

Because every physical system interacts with the environment, one way or another, the above means that the basic canons of quantum mechanics, i.e., unitary evolution and orthogonal projective measurement, are, in a way, fiction. The interaction that derails these idealizations doesn't even have to result in exchange of energy, momentum, or other observables with the environment. As you will see, a quantum system, which is initially isolated and prepared in some nice condition, very quickly develops quantum entanglements with the environment and these sponge away quantum information from it. This results in the decay of superposition into a mixture of basis states.

The observations we are going to make are profound, especially for quantum computing and any experimental procedure that calls for very precise control of a quantum system.

We will then make use of what we are going to learn about in this section, when discussing error correction codes, whose purpose is to stabilize quantum computers.

4.7.1 The Measurement

Suppose we have a quantum system described by $\boldsymbol{\rho}_A$ which interacts with the environment, described by $\boldsymbol{\rho}_E$. Suppose that initially the system and the environment are disentangled. The resulting density matrix for the combined system is

$$\boldsymbol{\rho}_{AE} = \boldsymbol{\rho}_A \otimes \boldsymbol{\rho}_E$$

Measurements performed on the *combined* ensemble, that comprises our system *and* the environment, are orthogonal projections corresponding to some Hermitian operator, $\boldsymbol{H}$, that represents an observable, e.g., energy, and can be described by the family of projectors $\{\boldsymbol{P}_i\}$ on $\mathcal{H}_A \otimes \mathcal{H}_E$ such that

$$\sum_i \boldsymbol{P}_i = \mathbf{1} \quad \text{and} \quad \boldsymbol{P}_i \boldsymbol{P}_j = \delta_{ij} \boldsymbol{P}_i$$

The probability of obtaining a result E_i, the i^{th} eigenvalue of $\boldsymbol{H}$, when performing the measurement on the combined ensemble is

$$p_i = \mathrm{tr}\left(\boldsymbol{P}_i \boldsymbol{\rho}_{AE}\right)$$

Immediately after the measurement the density operator of the combined system is

$$\boldsymbol{\rho}'_{AE} = \frac{\boldsymbol{P}_i \boldsymbol{\rho}_{AE} \boldsymbol{P}_i^\dagger}{\mathrm{tr}\left(\boldsymbol{P}_i \boldsymbol{\rho}_{AE}\right)}$$

and it is going to evolve from this point on, according to whatever dynamics the combined system is subjected to after the measurement.

Suppose that our observer has access to subsystem A only. What the observer sees after the measurement is given by the partial trace over the environment variables:

$$\mathrm{tr}_E \boldsymbol{\rho}'_{AE} = \frac{\mathrm{tr}_E \left(\boldsymbol{P}_i \left(\boldsymbol{\rho}_A \otimes \boldsymbol{\rho}_E\right) \boldsymbol{P}_i^\dagger\right)}{\mathrm{tr}_{AE}\left(\boldsymbol{P}_i \left(\boldsymbol{\rho}_A \otimes \boldsymbol{\rho}_E\right)\right)}$$

In order to see what exactly happens here let us evaluate

$$p_i = \mathrm{tr}\left(\boldsymbol{P}_i \boldsymbol{\rho}_{AE}\right) = \mathrm{tr}_A \mathrm{tr}_E \left(\boldsymbol{P}_i \left(\boldsymbol{\rho}_A \otimes \boldsymbol{\rho}_E\right)\right)$$

The operator $\boldsymbol{P}_i$ acts on $\mathcal{H}_A \otimes \mathcal{H}_E$, which is $\dim \mathcal{H}_A \times \dim \mathcal{H}_E$ dimensional. The operator itself can be represented by a $(\dim \mathcal{H}_A \times \dim \mathcal{H}_E) \times (\dim \mathcal{H}_A \times \dim \mathcal{H}_E)$ matrix. But the matrix can be constructed in such a way that indexes pertaining to system A and to the environment remain separated. Reserving Latin indexes for system A and Greek indexes for the environment we can represent $\boldsymbol{P}_i$ by the following *vierbein*[3]:

$$P^{(i)}_{k\alpha\, l\beta}$$

Then $\boldsymbol{P}_i \left(\boldsymbol{\rho}_A \otimes \boldsymbol{\rho}_E\right)$ would be represented by:

$$\sum_l P^{(i)}_{k\alpha\, l\beta} \rho^{(A)}_{lm} \rho^{(E)}_{\gamma\delta}$$

Taking the double trace results in *saturating* all indexes (this means summing over them in the index notation parlance) in this expression so as to deliver a

[3] *Vierbein* means four-legs in German. This word has been used traditionally to describe repers in space-time, so its use here is not traditional. All I mean by it here is an object with four indexes.

scalar value:

$$\begin{aligned}
&\mathrm{tr}_A \mathrm{tr}_E \left(\boldsymbol{P}_i \left(\boldsymbol{\rho}_A \otimes \boldsymbol{\rho}_E \right) \right) \\
&= \sum_{kl} \sum_{\alpha\beta} P^{(i)}_{k\alpha\, l\beta} \rho^{(A)}_{lk} \rho^{(E)}_{\beta\alpha} \\
&= \sum_{k} \sum_{l} \left(\sum_{\alpha\beta} P^{(i)}_{k\alpha\, l\beta} \rho^{(E)}_{\beta\alpha} \right) \rho^{(A)}_{lk} \\
&= \mathrm{tr} \left(\sum_{l} \left(\sum_{\alpha\beta} P^{(i)}_{m\alpha\, l\beta} \rho^{(E)}_{\beta\alpha} \right) \rho^{(A)}_{ln} \right) \\
&= \mathrm{tr} \sum_{l} F^{(i)}_{ml} \rho^{(A)}_{ln}
\end{aligned}$$

where

$$F^{(i)}_{ml} = \sum_{\alpha\beta} P^{(i)}_{m\alpha\, l\beta} \rho^{(E)}_{\beta\alpha} \tag{4.317}$$

Matrix $F^{(i)}_{ml}$ represents an operator $\boldsymbol{F}_i = \mathrm{tr}_E \left(\boldsymbol{P}_i \boldsymbol{\rho}_E \right)$ acting on $\mathcal{H}_A$.

Operators $\boldsymbol{F}_i$ describe measurements on the subsystem A coupled to some environment E. They are numbered by index i, which runs through the number of dimensions of $\mathcal{H}_A \otimes \mathcal{H}_E$. There are way too many of them to be mutually orthogonal within $\mathcal{H}_A$. Consequently measurements on A are NOT orthogonal projections, sic! Because *every* quantum system is a subsystem of a larger system, which is in turn a subsystem or a yet larger system ... this chain of inclusions eventually ending with the Universe, no real-life measurement can really be an orthogonal projection. One can at best strive to get as close to this concept as possible experimentally by trying to isolate the system under observation and/or by trying to perform operations on that system as quickly as possible, so that the rest of the Universe doesn't have enough time to affect the outcome.

To derive a non-index expression for $\boldsymbol{F}_i$, let us assume that we have the density operator for the environment in its diagonal form:

$$\boldsymbol{\rho}_E = \sum_{|\mu\rangle \in \mathcal{H}_E} p_\mu \mid \mu \rangle \langle \mu \mid$$

then acting with $\boldsymbol{P}_i$ on it yields

$$\boldsymbol{P}_i \boldsymbol{\rho}_E = \sum_{|\mu\rangle \in \mathcal{H}_E} p_\mu \boldsymbol{P}_i \mid \mu \rangle \langle \mu \mid$$

Taking trace over the environment replaces $\mid \mu \rangle \ldots \langle \mu \mid$ with $\langle \mu \mid \mu \rangle \ldots$:

$$\mathrm{tr}_E \left(\boldsymbol{P}_i \boldsymbol{\rho}_E \right) = \sum_{|\mu\rangle \in \mathcal{H}_E} p_\mu \langle \mu \mid \boldsymbol{P}_i \mid \mu \rangle$$

This still leaves all the indexes that pertain to subsystem A, i.e., the Latin indexes, unsaturated.

Although operators $\boldsymbol{F}_i$ are not orthogonal to each other within $\mathcal{H}_A$, their other properties are quite the same as the properties of orthogonal projections. And so

1. $\boldsymbol{F}_i$ are Hermitian. This follows from the fact that both $\boldsymbol{P}_i$ and $\boldsymbol{\rho}_E$ are Hermitian:

$$F^{(i)}_{lm} = \sum_{\alpha\beta} P_{l\alpha\, m\beta} \rho^{(E)}_{\alpha\beta} = \sum_{\alpha\beta} P^*_{m\beta\, l\alpha} \rho^{(E)*}_{\beta\alpha} = \left(F^{(i)}_{ml}\right)^*$$

2. $\boldsymbol{F}_i$ are positive. This follows from the fact that $\boldsymbol{P}_i$ is positive. For an arbitrary vector $\mid \Psi\rangle_A \in \mathcal{H}_A$ we have:

$$_A\langle\Psi \mid \boldsymbol{F}_i \mid \Psi\rangle_A = \sum_{\mid\mu\rangle\in\mathcal{H}_E} p_\mu \, {}_A\langle\Psi \mid \langle\mu \mid \boldsymbol{P}_i \mid \Psi\rangle_A \mid \mu\rangle \geq 0$$

3. All operators $\boldsymbol{F}_i$ sum up to $\mathbf{1}_A$. This follows from the fact that $\boldsymbol{P}_i$ sum up to $\mathbf{1}_{AE} = \mathbf{1}_A \otimes \mathbf{1}_E$:

$$\begin{aligned}
&\sum_{\mid i\rangle\in\mathcal{H}_A\otimes\mathcal{H}_E} \boldsymbol{F}_i \\
&\quad = \sum_{\mid\mu\rangle\in\mathcal{H}_E} p_\mu \langle\mu \mid \sum_{\mid i\rangle\in\mathcal{H}_A\otimes\mathcal{H}_E} \boldsymbol{P}_i \mid \mu\rangle \\
&\quad = \sum_{\mid\mu\rangle\in\mathcal{H}_E} p_\mu \langle\mu \mid \mathbf{1}_A \otimes \mathbf{1}_E \mid \mu\rangle \\
&\quad = \mathbf{1}_A \sum_{\mid\mu\rangle\in\mathcal{H}_E} p_\mu \langle\mu \mid \mu\rangle \\
&\quad = \mathbf{1}_A \sum_{\mid\mu\rangle\in\mathcal{H}_E} p_\mu = \mathbf{1}_A 1 = \mathbf{1}_A
\end{aligned}$$

Unfortunately there is no simple expression for $\boldsymbol{\rho}'_A$ in terms of $\boldsymbol{\rho}_A$ and $\boldsymbol{F}_i$.

The procedure displayed above can be reversed, i.e., given $\boldsymbol{\rho}_A$ and a family of $\boldsymbol{F}_i$ it is possible to coopt such $\boldsymbol{\rho}_E$ and define such $\boldsymbol{P}_i$ that

$$\mathrm{tr}_{AE}\left(\boldsymbol{P}_i\left(\boldsymbol{\rho}_A \otimes \boldsymbol{\rho}_E\right)\right) = \mathrm{tr}_A\left(\boldsymbol{F}_i \boldsymbol{\rho}_A\right)$$

but neither $\boldsymbol{\rho}_E$ nor $\boldsymbol{P}_i$ are unique.

4.7.2 The Evolution

Let us assume for simplicity that at the beginning of the evolution the environment is in state $\mid \mathbf{0}\rangle_E \, {}_E\langle\mathbf{0} \mid$. This implies, for example, a liquid helium bath, or something similar.

$$\boldsymbol{\rho}_{AE} = \boldsymbol{\rho}_A \otimes \mid \mathbf{0}\rangle_E \, {}_E\langle\mathbf{0} \mid$$

The evolution of the combined system is unitary:

$$\boldsymbol{U}_{AE}\,(\boldsymbol{\rho}_A \otimes \mid \mathbf{0}\rangle_{E\ E}\langle \mathbf{0} \mid)\,\boldsymbol{U}_{AE}^{\dagger}$$

What is this evolution going to look like if we ignore the environment. The answer is given by *tracing the environment away*.

$$\mathrm{tr}_E\left(\boldsymbol{U}_{AE}\,(\boldsymbol{\rho}_A \otimes \mid \mathbf{0}\rangle_{E\ E}\langle \mathbf{0} \mid)\,\boldsymbol{U}_{AE}^{\dagger}\right)$$

We can have a closer look at this expression by expanding the part of $\boldsymbol{U}_{AE}$ that acts on the environment in the environment's basis $\mid \mu\rangle$:

$$\boldsymbol{U}_{AE} \mid \mathbf{0}\rangle_E = \sum_{\mid \mu\rangle \in \mathcal{H}_E} \mid \mu\rangle\langle \mu \mid \boldsymbol{U}_{AE} \mid \mathbf{0}\rangle_E$$

We can now make use of this expression in rewriting $\boldsymbol{U}_{AE} \mid \mathbf{0}\rangle_{E\ E}\langle \mathbf{0} \mid \boldsymbol{U}_{AE}^{\dagger}$:

$$\boldsymbol{U}_{AE} \mid \mathbf{0}\rangle_{E\ E}\langle \mathbf{0} \mid \boldsymbol{U}_{AE}^{\dagger} = \sum_{\mid \mu\rangle, \mid \nu\rangle \in \mathcal{H}_E} \mid \mu\rangle\langle \mu \mid \boldsymbol{U}_{AE} \mid \mathbf{0}\rangle_{E\ E}\langle \mathbf{0} \mid \boldsymbol{U}_{AE}^{\dagger} \mid \nu\rangle\langle \nu \mid$$

Taking a trace over the environment, means replacing

$$\mid \mu\rangle \ldots \langle \nu \mid$$

with

$$\langle \nu \mid \mu\rangle \ldots = \delta_{\mu\nu} \ldots$$

which yields:

$$\sum_{\mid \mu\rangle \in \mathcal{H}_E} \langle \mu \mid \boldsymbol{U}_{AE} \mid \mathbf{0}\rangle_E \ldots {}_E\langle \mathbf{0} \mid \boldsymbol{U}_{AE}^{\dagger} \mid \mu\rangle$$

The $\ldots$ in this formula shows a slot for $\boldsymbol{\rho}_A$. Let us insert $\boldsymbol{\rho}_A$ in this slot then:

$$\begin{aligned}
&\mathrm{tr}\,\boldsymbol{U}_{AE}\,(\boldsymbol{\rho}_A \otimes \mid \mathbf{0}\rangle_{E\ E}\langle \mathbf{0} \mid)\,\boldsymbol{U}_{AE}^{\dagger} \\
&\quad = \sum_{\mid \mu\rangle \in \mathcal{H}_E} \langle \mu \mid \boldsymbol{U}_{AE} \mid \mathbf{0}\rangle_E\, \boldsymbol{\rho}_A\, {}_E\langle \mathbf{0} \mid \boldsymbol{U}_{AE}^{\dagger} \mid \mu\rangle \\
&\quad = \sum_{\mid \mu\rangle \in \mathcal{H}_E} \boldsymbol{M}_\mu \boldsymbol{\rho}_A \boldsymbol{M}_\mu^{\dagger}
\end{aligned}$$

where

$$\boldsymbol{M}_\mu = \langle \mu \mid \boldsymbol{U}_{AE} \mid \mathbf{0}\rangle_E$$

This expression defines what is called a *superoperator* or a *quantum operation*:

$$\mathfrak{A} : \boldsymbol{\rho}_A \to \boldsymbol{\rho}'_A = \sum_{\mid \mu\rangle \in \mathcal{H}_E} \boldsymbol{M}_\mu \boldsymbol{\rho}_A \boldsymbol{M}_\mu^{\dagger} \tag{4.318}$$

$\mathfrak{A}$ maps an operator on $\mathcal{H}_A$ into another operator on $\mathcal{H}_A$. The resulting evolution of $\boldsymbol{\rho}_A$ is *not* unitary. This type of evolution is unitary *only* if there is only one

$\boldsymbol{M}_\mu$ in the expansion (4.318). The expansion in terms of the family of operators $\boldsymbol{M}_i$ is called *the operator sum representation of the superoperator* $\mathfrak{A}$.

Operators $\boldsymbol{M}_i$ have the following nice property:

$$\begin{aligned}
&\sum_{|\mu\rangle\in\mathcal{H}_E} \boldsymbol{M}_\mu^\dagger \boldsymbol{M}_\mu \\
&= \sum_{|\mu\rangle\in\mathcal{H}_E} {}_E\langle \mathbf{0} \mid \boldsymbol{U}_{AE}^\dagger \mid \mu\rangle\langle\mu \mid \boldsymbol{U}_{AE} \mid \mathbf{0}\rangle_E \\
&= {}_E\langle \mathbf{0} \mid \boldsymbol{U}_{AE}^\dagger \boldsymbol{U}_{AE} \mid \mathbf{0}\rangle_E \\
&= {}_E\langle \mathbf{0} \mid \mathbf{1}_A \otimes \mathbf{1}_E \mid \mathbf{0}\rangle_E \\
&= \mathbf{1}_A \, {}_E\langle \mathbf{0} \mid \mathbf{1}_E \mid \mathbf{0}\rangle_E = \mathbf{1}_A
\end{aligned}$$

The superoperator $\mathfrak{A}$ as defined in terms of the operator sum representation has the following properties:

1. It preserves Hermiticity, i.e., if $\boldsymbol{\rho}_A$ is Hermitian then so is $\mathfrak{A}(\boldsymbol{\rho}_A)$:

$$\begin{aligned}
\left(\boldsymbol{\rho}_A^\dagger\right)' &= \sum_{|\mu\rangle\in\mathcal{H}_E} \boldsymbol{M}_\mu \boldsymbol{\rho}_A^\dagger \boldsymbol{M}_\mu^\dagger \\
&= \sum_{|\mu\rangle\in\mathcal{H}_E} \boldsymbol{M}_\mu \boldsymbol{\rho}_A \boldsymbol{M}_\mu^\dagger = \boldsymbol{\rho}_A'
\end{aligned}$$

2. $\mathfrak{A}$ preserves trace 1:

$$\begin{aligned}
\mathrm{tr}_A \mathfrak{A}(\boldsymbol{\rho}_A) &= \mathrm{tr}_A \sum_{|\mu\rangle\in\mathcal{H}_E} \boldsymbol{M}_\mu \boldsymbol{\rho}_A \boldsymbol{M}_\mu^\dagger = \mathrm{tr}_A \left(\boldsymbol{\rho}_A \sum_{|\mu\rangle\in\mathcal{H}_E} \boldsymbol{M}_\mu^\dagger \boldsymbol{M}_\mu \right) \\
&= \mathrm{tr}_A (\boldsymbol{\rho}_A \mathbf{1}_A) = \mathrm{tr}_A \boldsymbol{\rho}_A = 1
\end{aligned}$$

3. $\mathfrak{A}$ preserves the positivity

$$\begin{aligned}
{}_A\langle \Psi \mid \mathfrak{A}(\boldsymbol{\rho}_A) \mid \Psi\rangle_A &= {}_A\langle \Psi \mid \sum_{|\mu\rangle\in\mathcal{H}_E} \boldsymbol{M}_\mu \boldsymbol{\rho}_A \boldsymbol{M}_\mu^\dagger \mid \Psi\rangle_A \\
&= \sum_{|\mu\rangle\in\mathcal{H}_E} ({}_A\langle \Psi \mid \boldsymbol{M}_\mu) \, \boldsymbol{\rho}_A \, (\boldsymbol{M}_\mu \mid \Psi\rangle_A) \\
&= \sum_{|\mu\rangle\in\mathcal{H}_E} {}_A\langle \phi_\mu \mid \boldsymbol{\rho}_A \mid \phi_\mu\rangle_A \geq 0
\end{aligned}$$

where we have used $\mid \phi_\mu\rangle_A = \boldsymbol{M}_\mu \mid \Psi\rangle_A$.

Given a family of operators $\boldsymbol{M}_\mu$ it is always possible to find such $\boldsymbol{U}_{AE}$ that

$${}_E\langle \mu \mid \boldsymbol{U}_{AE} \mid \mathbf{0}\rangle_E = \boldsymbol{M}_\mu$$

But the operator sum representation of a superoperator is *not* unique.

Superoperators can convert a pure state into a mixture. They let us analyze decoherence and other problems that are going to affect quantum computation or any other operation carried out on a quantum system. Under the action of superoperators states from $\mathcal{H}_A$ can become irreversibly entangled with states from $\boldsymbol{H}_E$, which leads to a loss of information.

Superoperators can be combined, i.e., $\mathfrak{A} \circ \mathfrak{B}$. This leads to the formation of a semigroup. But, unlike unitary operators, superoperators do not form a group, because there is no inverse. Phenomena described by superoperators are irreversible, e.g., decoherence is irreversible. And so wee see that superoperators define the arrow of time. This is the only place in Quantum Mechanics where this happens. Of course, it is clear by now that evolution dictated by superoperators is NOT unitary.

One can define a superoperator abstractly and then ask if an operator sum representation can be found. To this effect we can define $\mathfrak{A}$ as an operator that

1. preserves hermiticity of $\boldsymbol{\rho}_A$
2. preserves trace=1 of $\boldsymbol{\rho}_A$
3. preserves non-negativity of $\boldsymbol{\rho}_A$

In general there is no good reason to restrict $\mathfrak{A}$ to linearity. One can conceive of non-linear quantum dynamics. Some of these are even consistent with probabilistic interpretation. But we have to assume linearity if we want $\mathfrak{A}$ to have an operator sum representation. To this effect we also have to impose a stricter condition on the positivity of $\mathfrak{A}$: we have to assume that $\mathfrak{A}$ is *completely positive*. With all these assumption in place one can show that $\mathfrak{A}$ admits an operator sum representation. This is the subject of the Kraus Representation Theorem.

4.7.3 Three Quantum Channels

The three quantum channels discussed in this section provide a quite insightful model of perils that await quantum computers, and, for this matter, any other quantum system. Also they provide us with fine examples of the use of superoperators. Every channel is described by a superoperator. The channels are non-unitary effects that arise from interaction of quantum systems with the environment. The three we are going to study here are traditionally called:

depolarization channel this channel describes spin and phase flips that may occur in quantum registers;

phase damping channel this channel describes decoherence, i.e., the decay of a superposition into a mixture;

amplitude damping channel this channel describes spontaneous emmision.

Depolarization

We begin discussing this channel by providing a unitary description of a combined system comprising a single qubit and a 4-qubit "environment". As we have done in the section about superoperators, we assume that the environment is initially in state $\mid \mathbf{0}\rangle_E$. The single qubit, which is a subsystem A in this model, can spin flip with probability 1/3. The spin flip is described by $\boldsymbol{\sigma}_x$:

$$\mid \Psi\rangle_A \rightarrow \boldsymbol{\sigma}_x \mid \Psi\rangle_A$$

It can also phase-flip with probability 1/3. The phase-flip is described by $\boldsymbol{\sigma}_z$:

$$\mid \Psi\rangle_A \rightarrow \boldsymbol{\sigma}_z \mid \Psi\rangle_A$$

Finally, the state can both spin and phase flip at the same time. This is described by $\boldsymbol{\sigma}_y$:

$$\mid \Psi\rangle_A \rightarrow \boldsymbol{\sigma}_y \mid \Psi\rangle_A$$

There is a $1-p$ probability that the state will not suffer any of the flips described above. The $\boldsymbol{U}_{AE}$ operator describing this evolution on the combined system can now be defined as follows:

$$\begin{aligned}&\boldsymbol{U}_{AE}:\\ &\quad \mid \Psi\rangle_A \rightarrow \sqrt{1-p}\mid \Psi\rangle_A \otimes \mid \mathbf{0}\rangle_E\\ &\qquad +\sqrt{\frac{p}{3}}\left(\boldsymbol{\sigma}_x \mid \Psi\rangle_A \otimes \mid \mathbf{1}\rangle_E + \boldsymbol{\sigma}_y \mid \Psi\rangle_A \otimes \mid \mathbf{2}\rangle_E + \boldsymbol{\sigma}_z \mid \Psi\rangle_A \otimes \mid \mathbf{3}\rangle_E\right)\end{aligned}$$

Observe that the environment preserves the memory of what happens. And so, if the qubit spin flips, then $\mid \mathbf{0}\rangle_E$ switches to $\mid \mathbf{1}\rangle_E$. If the qubit phase-flips then $\mid \mathbf{0}\rangle_E$ is upgraded to $\mid \mathbf{3}\rangle_E$. And if the qubit phase-and-spin-flips the environment is upgraded to $\mid \mathbf{2}\rangle_E$.

In order to find the corresponding superoperator $\mathfrak{A}$ we need to use the definition of $\boldsymbol{U}_{AE}$ to find the family of operators $\boldsymbol{M}_\mu$, where $\mid \mu\rangle \in \mathcal{H}_E$, i.e., where $\mu = 0, 1, 2, 3$.

The general formula is:

$$\boldsymbol{M}_\mu = {}_E\langle \mu \mid \boldsymbol{U}_{AE} \mid \mathbf{0}\rangle_E$$

Specifying this for $\mid \mu\rangle = \mid \mathbf{0}\rangle_E, \mid \mathbf{1}\rangle_E, \mid \mathbf{2}\rangle_E, \mid \mathbf{3}\rangle_E$ yields:

$$\begin{aligned}\boldsymbol{M}_0 &= {}_E\langle \mathbf{0} \mid \boldsymbol{U}_{AE} \mid \mathbf{0}\rangle_E = \sqrt{1-p}\,\mathbf{1}_A\\ \boldsymbol{M}_1 &= {}_E\langle \mathbf{1} \mid \boldsymbol{U}_{AE} \mid \mathbf{0}\rangle_E = \sqrt{\frac{p}{3}}\,\boldsymbol{\sigma}_x\\ \boldsymbol{M}_2 &= {}_E\langle \mathbf{2} \mid \boldsymbol{U}_{AE} \mid \mathbf{0}\rangle_E = \sqrt{\frac{p}{3}}\,\boldsymbol{\sigma}_y\\ \boldsymbol{M}_3 &= {}_E\langle \mathbf{3} \mid \boldsymbol{U}_{AE} \mid \mathbf{0}\rangle_E = \sqrt{\frac{p}{3}}\,\boldsymbol{\sigma}_z\end{aligned}$$

Now we can make use of the operator sum representation of $\mathfrak{A}$ to evaluate its action of $\boldsymbol{\rho}_A$:

$$\begin{aligned}\mathfrak{A}(\boldsymbol{\rho}_A) &= \boldsymbol{M}_0\boldsymbol{\rho}_A\boldsymbol{M}_0^\dagger + \boldsymbol{M}_1\boldsymbol{\rho}_A\boldsymbol{M}_1^\dagger + \boldsymbol{M}_2\boldsymbol{\rho}_A\boldsymbol{M}_2^\dagger + \boldsymbol{M}_3\boldsymbol{\rho}_A\boldsymbol{M}_3^\dagger \\ &= (1-p)\boldsymbol{\rho} + \frac{p}{3}(\boldsymbol{\sigma}_x\boldsymbol{\rho}\boldsymbol{\sigma}_x + \boldsymbol{\sigma}_y\boldsymbol{\rho}\boldsymbol{\sigma}_y + \boldsymbol{\sigma}_z\boldsymbol{\rho}\boldsymbol{\sigma}_z)\end{aligned}$$

Depolarization can have specially devastating effect on entangled states: a single error of this type can completely randomize the state. To see this consider a state from the the Bell Operator Basis, which we have introduced in the section about teleportation:

$$\begin{aligned}\mid \Psi^A\rangle &= \frac{1}{\sqrt{2}}(\mid 0\rangle \mid 1\rangle - \mid 1\rangle \mid 0\rangle) \\ \mid \Psi^B\rangle &= \frac{1}{\sqrt{2}}(\mid 0\rangle \mid 1\rangle + \mid 1\rangle \mid 0\rangle) \\ \mid \Psi^C\rangle &= \frac{1}{\sqrt{2}}(\mid 0\rangle \mid 0\rangle - \mid 1\rangle \mid 1\rangle) \\ \mid \Psi^D\rangle &= \frac{1}{\sqrt{2}}(\mid 0\rangle \mid 0\rangle + \mid 1\rangle \mid 1\rangle)\end{aligned}$$

Suppose our start-up state is $\mid \Psi^D\rangle$. Pushing this state through the depolarization channel on the first qubit converts the initial density matrix $\mid \Psi^D\rangle\langle\Psi^D \mid$ to:

$$\boldsymbol{\rho}' = \left(1 - \frac{4}{3}p\right) \mid \Psi^D\rangle\langle\Psi^D \mid + \frac{p}{3}\mathbf{1} \tag{4.319}$$

For $p = 3/4$ we get

$$\boldsymbol{\rho}' = \mathbf{1}/4 \tag{4.320}$$

Remember that the $\mathbf{1}$ in this equation is $\mid \Psi^A\rangle\langle\Psi^A \mid + \mid \Psi^B\rangle\langle\Psi^B \mid + \mid \Psi^C\rangle\langle\Psi^C \mid + \mid \Psi^D\rangle\langle\Psi^D \mid$. The result is a thoroughly chaotic state. All information carried by $\mid \Psi^D\rangle\langle\Psi^D \mid$ has been lost, sic!

In order to see the outcome of $\mathfrak{A}$ acting on a general 2×2 density matrix, consider the Bloch Sphere representation:

$$\boldsymbol{\rho} = \frac{1}{2}(\mathbf{1} + \boldsymbol{r}\cdot\boldsymbol{\sigma})$$

For simplicity assume that $\boldsymbol{r} = r\boldsymbol{e}_z$:

$$\boldsymbol{\rho} = \frac{1}{2}(\mathbf{1} + r\boldsymbol{\sigma}_z)$$

We will make use of the following properties of Pauli matrices:

$$\begin{aligned}\boldsymbol{\sigma}_x\boldsymbol{\sigma}_z\boldsymbol{\sigma}_x &= -\boldsymbol{\sigma}_z \\ \boldsymbol{\sigma}_y\boldsymbol{\sigma}_z\boldsymbol{\sigma}_y &= -\boldsymbol{\sigma}_z \\ \boldsymbol{\sigma}_z\boldsymbol{\sigma}_z\boldsymbol{\sigma}_z &= \boldsymbol{\sigma}_z\end{aligned}$$

Substituting the Bloch Sphere representation and making use of the above in our expression for the superoperator of the depolarization channel yields:

$$\begin{aligned}
&(1-p)\frac{1}{2}\left(\mathbf{1}+r\boldsymbol{\sigma}_z\right)+\frac{p}{3}\left(\frac{1}{2}3\mathbf{1}+\frac{1}{2}r\left(-\boldsymbol{\sigma}_z-\boldsymbol{\sigma}_z+\boldsymbol{\sigma}_z\right)\right)\\
&=(1-p)\frac{1}{2}\left(\mathbf{1}+r\boldsymbol{\sigma}_z\right)+\frac{p}{2}\mathbf{1}-\frac{p}{6}r\boldsymbol{\sigma}_z\\
&=\frac{1-p}{2}\mathbf{1}+\frac{p}{2}\mathbf{1}+\frac{1-p}{2}r\boldsymbol{\sigma}_z-\frac{p}{6}r\boldsymbol{\sigma}_z\\
&=\frac{1}{2}\mathbf{1}+\frac{3-3p-p}{6}r\boldsymbol{\sigma}_z\\
&=\frac{1}{2}\left(\mathbf{1}+\left(1-\frac{4}{3}p\right)r\boldsymbol{\sigma}_z\right)
\end{aligned}$$

This shows the shrinking of the Bloch sphere. If we start in a pure state we get a mixture right away.

Phase Damping

Phase damping occurs when the initial system becomes gradually entangled with the environment. The following unitary description of this process assumes that the environment is spanned by $\mid \mathbf{0}\rangle_E$, $\mid \mathbf{1}\rangle_E$, and $\mid \mathbf{2}\rangle_E$. The initial state $\mid 0\rangle_A \mid \mathbf{0}\rangle_E$ has probability p of becoming entangled with $\mid 0\rangle_A \mid \mathbf{1}\rangle_E$ and a probability $1-p$ of not becoming entangled. The initial state $\mid 1\rangle_A \mid \mathbf{0}\rangle_E$ has probability p of becoming entangled with $\mid 1\rangle_A \mid \mathbf{2}\rangle_E$ and probability $1-p$ of not becoming entangled. Observe that the state $\mid \Psi\rangle_A$ itself does not change in the process at all. There is no exchange of energy or momentum, or any other physical observable with the environment:

$$\begin{aligned}
\mid 0\rangle_A \mid \mathbf{0}\rangle_E &\rightarrow \sqrt{1-p}\mid 0\rangle_A \mid \mathbf{0}\rangle_E+\sqrt{p}\mid 0\rangle_A \mid \mathbf{1}\rangle_E\\
\mid 1\rangle_A \mid \mathbf{0}\rangle_E &\rightarrow \sqrt{1-p}\mid 1\rangle_A \mid \mathbf{0}\rangle_E+\sqrt{p}\mid 1\rangle_A \mid \mathbf{2}\rangle_E
\end{aligned}$$

The first operator $\boldsymbol{M}_0$ is easy to find:

$$\boldsymbol{M}_0 = {}_E\langle \mathbf{0} \mid \boldsymbol{U}_{AE} \mid \mathbf{0}\rangle_E = \sqrt{1-p}\,\mathbf{1}_A$$

For the operator $\boldsymbol{M}_1$ we need to perform separate evaluation on $\mid 0\rangle_A$ and $\mid 1\rangle_A$ in order to find the corresponding matrix:

$$\begin{aligned}
\boldsymbol{M}_1 \mid 0\rangle_A &= {}_E\langle \mathbf{1} \mid \boldsymbol{U}_{AE} \mid 0\rangle_A \mid \mathbf{0}\rangle_E\\
&= {}_E\langle \mathbf{1} \mid \left(\sqrt{1-p}\mid 0\rangle_A \mid \mathbf{0}\rangle_E+\sqrt{p}\mid 0\rangle_A \mid \mathbf{1}\rangle_E\right)\\
&= \sqrt{p}\mid 0\rangle_A\\
\boldsymbol{M}_1 \mid 1\rangle_A &= {}_E\langle \mathbf{1} \mid \boldsymbol{U}_{AE} \mid 1\rangle_A \mid \mathbf{0}\rangle_E = 0
\end{aligned}$$

Hence the matrix representation of $\boldsymbol{M}_1$ is

$$\boldsymbol{M}_1 = \sqrt{p}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

Similarly for $\boldsymbol{M}_2$:

$$\begin{aligned}
\boldsymbol{M}_2 \mid 0\rangle_A &= {}_E\langle \mathbf{2} \mid \boldsymbol{U}_{AE} \mid 0\rangle_A \mid \mathbf{0}\rangle_E = 0 \\
\boldsymbol{M}_2 \mid 1\rangle_A &= {}_E\langle \mathbf{2} \mid \boldsymbol{U}_{AE} \mid 1\rangle_A \mid \mathbf{0}\rangle_E \\
&= {}_E\langle \mathbf{2} \mid \left(\sqrt{1-p} \mid 1\rangle_A \mid \mathbf{0}\rangle_E + \sqrt{p} \mid 1\rangle_A \mid \mathbf{2}\rangle_E\right) \\
&= \sqrt{p} \mid 1\rangle_A
\end{aligned}$$

Hence the matrix representation of $\boldsymbol{M}_2$ is

$$\boldsymbol{M}_2 = \sqrt{p}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Now we can assemble the superoperator $\mathfrak{A}$:

$$\begin{aligned}
\mathfrak{A}(\boldsymbol{\rho}_A) &= \boldsymbol{M}_0\boldsymbol{\rho}_A\boldsymbol{M}_0^\dagger + \boldsymbol{M}_1\boldsymbol{\rho}_A\boldsymbol{M}_1^\dagger + \boldsymbol{M}_2\boldsymbol{\rho}_A\boldsymbol{M}_2^\dagger \\
&= (1-p)\boldsymbol{\rho}_A + p\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\boldsymbol{\rho}_A\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + p\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\boldsymbol{\rho}_A\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}
\end{aligned}$$

It is easy to see that

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}$$

and

$$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & d \end{pmatrix}$$

So that the final form of $\mathfrak{A}(\boldsymbol{\rho}_A)$ becomes:

$$\begin{aligned}
\mathfrak{A}(\boldsymbol{\rho}_A) &= (1-p)\boldsymbol{\rho}_A + p\begin{pmatrix} \rho_{00} & 0 \\ 0 & \rho_{11} \end{pmatrix} \\
&= \begin{pmatrix} \rho_{00} & (1-p)\rho_{01} \\ (1-p)\rho_{10} & \rho_{11} \end{pmatrix}
\end{aligned}$$

Applying this operator n times results in:

$$\mathfrak{A}^n\boldsymbol{\rho}_A = \begin{pmatrix} \rho_{00} & (1-p)^n\rho_{01} \\ (1-p)^n\rho_{10} & \rho_{11} \end{pmatrix}$$

Suppose that the probability of the entanglement per unit time is Γ, so that $p = \Gamma\Delta t$. then $n = t/\Delta t$ and

$$(1-p)^n = (1-\Gamma\Delta t)^{t/\Delta t}$$

In the limit of $\Delta t \to 0$ this becomes

$$\lim_{\Delta t\to 0}(1-\Gamma\Delta t)^{t/\Delta t} = e^{-\Gamma t}$$

This follows from:

$$\lim_{m\to\infty}\left(1+\frac{z}{m}\right)^m = e^z$$

Which, in turn, can be proven easily by using

$$(x+y)^m = \sum_{k=0}^{m}\binom{m}{k}x^{m-k}y^k$$

and then taking the limit $m \to \infty$

EXERCISE Prove it.

And so our density operator becomes:

$$\mathfrak{A}^{t/\Delta t}(\boldsymbol{\rho}_A) = \begin{pmatrix} \rho_{00} & e^{-\Gamma t}\rho_{01} \\ e^{-\Gamma t}\rho_{10} & \rho_{11} \end{pmatrix}$$

Observe the exponential vanishing of the off-diagonal terms. We have already shown in the discussion about the NMR measurement that this implies vanishing of superposition, and its exponential replacement with a mixture.

So now you see how this happens.

What does phase damping look like in the Bloch Sphere representation? Let

$$\boldsymbol{\rho}_A = \frac{1}{2}(\mathbf{1}+\boldsymbol{r}\cdot\boldsymbol{\sigma}) = \frac{1}{2}\begin{pmatrix} 1+r_z & r_x - ir_y \\ r_x + ir_y & 1-r_z \end{pmatrix}$$

Then $\mathfrak{A}(\boldsymbol{\rho}_A)$ becomes

$$\mathfrak{A}(\boldsymbol{\rho}_A) = \frac{1}{2}\begin{pmatrix} 1+r_z & (1-p)(r_x - ir_y) \\ (1-p)(r_x + ir_y) & 1-r_z \end{pmatrix}$$

You see that spin becomes forced in the $\boldsymbol{e}_z$ direction, with spin components pointing in the $\boldsymbol{e}_x$ and $\boldsymbol{e}_y$ directions vanishing exponentially.

What physical process corresponds to the phase damping? There was no energy or momentum exchange in the unitary description of the combined system. The usual interpretation is that of a heavy atom, bombarded by a stream of low energy photons. This process does not visibly affect the energy and momentum of the atom. Also, the photons, being of very low energy cannot affect electronic transitions in the atom. Yet if the atom has been originally placed in a superposition, this interaction with the low energy photon bath will sponge the superposition away. All information contained in the superposition will be lost irretrievably to the photon bath.

Amplitude Damping

This channel describes the process of spontaneous emission. If qubit A is in the $|\,0\rangle_A$ state, then nothing happens. This is the ground state. But if qubit A is in the $|\,1\rangle_A$ state, then there is a probability p that it is going to decay to state $|\,0\rangle_A$ while emitting photon, which puts the environment, initially in state $|\,0\rangle_E$

in state $\mid 1\rangle_E$. But then there is also a probability $1-p$ that none of this will happen:

$$\begin{aligned}
\boldsymbol{U}_{AE} \mid 0\rangle_A \mid 0\rangle_E &= \mid 0\rangle_A \mid 0\rangle_E \\
\boldsymbol{U}_{AE} \mid 1\rangle_A \mid 0\rangle_E &= \sqrt{1-p} \mid 1\rangle_A \mid 0\rangle_E + \sqrt{p} \mid 0\rangle_A \mid 1\rangle_E
\end{aligned}$$

There are only two operators $\boldsymbol{M}_\mu$:

$$\begin{aligned}
\boldsymbol{M}_0 &= {}_E\langle 0 \mid \boldsymbol{U}_{AE} \mid 0\rangle_E \\
\boldsymbol{M}_1 &= {}_E\langle 1 \mid \boldsymbol{U}_{AE} \mid 0\rangle_E
\end{aligned}$$

We proceed as in the phase damping section, evaluating the action of $\boldsymbol{M}_0$ and $\boldsymbol{M}_1$ on both $\mid 0\rangle_A$ and $\mid 0\rangle_E$ in order to find the corresponding matrix representations of these operators. And so

$$\begin{aligned}
\boldsymbol{M}_0 \mid 0\rangle_A &= {}_E\langle 0 \mid \boldsymbol{U}_{AE} \mid 0\rangle_A \mid 0\rangle_E \\
&= {}_E\langle 0 \mid 0\rangle_A \mid 0\rangle_E \\
&= \mid 0\rangle_A \\
\boldsymbol{M}_0 \mid 1\rangle_A &= {}_E\langle 0 \mid \boldsymbol{U}_{AE} \mid 1\rangle_A \mid 0\rangle_E \\
&= {}_E\langle 0 \mid \left(\sqrt{1-p} \mid 1\rangle_A \mid 0\rangle_E + \sqrt{p} \mid 0\rangle_A \mid 1\rangle_E\right) \\
&= \sqrt{1-p} \mid 1\rangle_A
\end{aligned}$$

Hence

$$\boldsymbol{M}_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix}$$

Similarly:

$$\begin{aligned}
\boldsymbol{M}_1 \mid 0\rangle_A &= {}_E\langle 1 \mid \boldsymbol{U}_{AE} \mid 0\rangle_A \mid 0\rangle_E \\
&= {}_E\langle 1 \mid 0\rangle_A \mid 0\rangle_E \\
&= 0 \\
\boldsymbol{M}_1 \mid 1\rangle_A &= {}_E\langle 1 \mid \boldsymbol{U}_{AE} \mid 1\rangle \mid 0\rangle_E \\
&= {}_E\langle 1 \mid \left(\sqrt{1-p} \mid 1\rangle_A \mid 0\rangle_E + \sqrt{p} \mid 0\rangle_A \mid 1\rangle_E\right) \\
&= \sqrt{p} \mid 0\rangle_A
\end{aligned}$$

Hence

$$\boldsymbol{M}_1 = \begin{pmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{pmatrix}$$

The action of the superoperator $\mathfrak{A}$ on $\boldsymbol{\rho}_A$ can now be assembled thusly:

$$\begin{aligned}
\mathfrak{A}(\boldsymbol{\rho}_A) &= \boldsymbol{M}_0 \boldsymbol{\rho}_A \boldsymbol{M}_0^\dagger + \boldsymbol{M}_1 \boldsymbol{\rho}_A \boldsymbol{M}_1^\dagger \\
&= \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix} \boldsymbol{\rho}_A \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{pmatrix} + \begin{pmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{pmatrix} \boldsymbol{\rho}_A \begin{pmatrix} 0 & 0 \\ \sqrt{p} & 0 \end{pmatrix} \\
&= \begin{pmatrix} \rho_{00} & \sqrt{1-p}\rho_{01} \\ \sqrt{1-p}\rho_{10} & (1-p)\rho_{11} \end{pmatrix} + \begin{pmatrix} p\rho_{11} & 0 \\ 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} \rho_{00} + p\rho_{11} & \sqrt{1-p}\rho_{01} \\ \sqrt{1-p}\rho_{10} & (1-p)\rho_{11} \end{pmatrix}
\end{aligned}$$

Consider again $\mathfrak{A}^n(\boldsymbol{\rho}_A)$. The $(1-p)^n$ term eventually becomes $e^{-\Gamma t}$. As $t \to \infty$ $p \to 0$, so that

$$\mathfrak{A}^n(\boldsymbol{\rho}_A) \to \begin{pmatrix} \rho_{00}+\rho_{11} & 0 \\ 0 & 0 \end{pmatrix}$$

The system ends up in a ground state. Here it is possible for $\mathfrak{A}$ to start with a mixture and deliver a pure state at the end. All you need to do is to cool system A and wait long enough for it to drop to the ground state.

Using the Bloch Sphere representation for $\boldsymbol{\rho}_A$:

$$\boldsymbol{\rho}_A = \frac{1}{2}(\mathbf{1} + \boldsymbol{r}\cdot\boldsymbol{\sigma}) = \frac{1}{2}\begin{pmatrix} 1+r_z & r_x - ir_y \\ r_x + ir_y & 1-r_z \end{pmatrix}$$

yields the following expression for $\mathfrak{A}(\boldsymbol{\rho}_A)$

$$\mathfrak{A}(\boldsymbol{\rho}_A) = \frac{1}{2}\begin{pmatrix} 1+p+(1-p)r_z & \sqrt{1-p}\,(r_x - ir_y) \\ \sqrt{1-p}\,(r_x+ir_y) & 1-p-(1-p)r_z \end{pmatrix}$$

At $p=1$ this becomes

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

The convergence process squeezes the ball and pushes it towards the upper pole, so that eventually it becomes the point at the pole itself.

4.8 Midterm Assignment

1. In section about spin dynamics in NMR we have made use of the following three identities:

$$\begin{aligned} e^{i\omega\boldsymbol{\sigma}_z t/2}\boldsymbol{\sigma}_z e^{-i\omega\boldsymbol{\sigma}_z t/2} &= \boldsymbol{\sigma}_z \\ e^{i\omega\boldsymbol{\sigma}_z t/2}\boldsymbol{\sigma}_x e^{-i\omega\boldsymbol{\sigma}_z t/2} &= \boldsymbol{\sigma}_x\cos\omega t - \boldsymbol{\sigma}_y\sin\omega t \\ e^{i\omega\boldsymbol{\sigma}_z t/2}\boldsymbol{\sigma}_y e^{-i\omega\boldsymbol{\sigma}_z t/2} &= \boldsymbol{\sigma}_y\cos\omega t + \boldsymbol{\sigma}_x\sin\omega t \end{aligned}$$

   Prove these identities.

2. Prove that

$$\begin{aligned} \oplus^{1\leftrightarrow 2} = \; & e^{-i(\pi/4)\boldsymbol{\sigma}_y^{(1)}} e^{i(\pi/4)\boldsymbol{\sigma}_x^{(1)}} e^{i(\pi/4)\boldsymbol{\sigma}_y^{(1)}} \\ & e^{-i(\pi/4)\boldsymbol{\sigma}_x^{(2)}} e^{i(\pi/4)\boldsymbol{\sigma}_y^{(2)}} e^{-i(\pi/4)\boldsymbol{\sigma}_z^{(1)}\otimes\boldsymbol{\sigma}_z^{(2)}} e^{-i(\pi/4)\boldsymbol{\sigma}_y^{(2)}} \end{aligned}$$

   implements a controlled-NOT gate.

3. Analyze the behaviour of the Brassard Teleportation Circuit for the case of the top two lines decohering to $|\,00\rangle$, $|\,10\rangle$, and $|\,11\rangle$.

4. Show that for a Hermitian operator $\boldsymbol{M}: \mathcal{H} \to \mathcal{H}$ and for a density operator $\boldsymbol{\rho}: \mathcal{H} \to \mathcal{H}$ the expectation value for $\boldsymbol{M}$ on the ensemble described by $\boldsymbol{\rho}$ is $\langle \boldsymbol{M} \rangle = \mathrm{tr}\,(\boldsymbol{M}\boldsymbol{\rho})$

5. Show that for an ensemble of single qubit systems

$$\mathrm{tr}\boldsymbol{\rho}^2 = \left(1 + r^2\right)/2$$

where $\boldsymbol{\rho}$ is the density matrix for the ensemble and r is the length of vector $\boldsymbol{r}$ used in the Bloch Sphere representation of the ensemble.

Chapter 5

Gates and Circuits

5.1 Gates

5.1.1 The Toffoli Gate

Although by now you probably remember it well enough it is always worth emphasizing the difference between a bit and a qubit. A bit b is a scalar that belongs to $\{0,1\}$:

$$b \in \{0,1\}$$

whereas a qubit $\boldsymbol{q}$ is a vector that belongs to a two dimensional complex Hilbert space $\mathcal{H}$:

$$\boldsymbol{q} \in \mathcal{H}$$

A bit can assume values 0 or 1. A qubit can assume values $\mid 0\rangle = \binom{1}{0}$ or $\mid 1\rangle = \binom{0}{1}$ *and* all values in between. A bit cannot have a value between 0 and 1.

A classical n-bit register can contain one value from

$$\{0,1\} \times \{0,1\} \times \cdots \quad n \text{ times} \quad \cdots \times \{0,1\} = \{0,1\}^n$$

at a time. We can think of this value as a string of 1s and 0s. We could name these n-bit strings by interpreting them as binary representations of some numbers and then it would become clear that there are going to be 2^n such strings that can be placed in an n-bit register.

A reversible function that operates on an n bit state of a computer converts a string from $\{0,1\}^n$ into another string from $\{0,1\}^n$ in a one-to-one operation. If we were to think of these strings as binary representations of numbers then we would say that the reversible function converts a number from $\{0 \ldots 2^n - 1\}$ into another number from $\{0 \ldots 2^n - 1\}$ and that conversion is 1-to-1. In other words a reversible function must be a *permutation* on $\{0 \ldots 2^n - 1\}$.

From the theory of permutations we know that a set of m elements can be permuted in $m!$ ways. Hence the set of 2^n elements can be permuted in $2^n!$ ways. This is the total number of reversible functions on an n-bit register.

We can think of each of these permutations as a reversible gate. There are therefore $2^n!$ *reversible* gates acting on an n-bit register.

But classical gates do not have to be reversible, and, as a matter of fact, most of them aren't. If we were to allow non-reversible gates to act on our n-bit register, the total number of such gates, reversible and not, would be much, much larger.

Consider a gate that converts a string from $\{0,1\}^n$ into a single bit:

$$f : \{0,1\}^n \rightarrow \{0,1\}$$

there are 2^n possible numbers (from 0 to $2^n - 1$) in the *domain* of this function. The gate can be characterised by assigning either 0 or 1 to everyone of its 2^n inputs, for example

$$\begin{aligned} 0 &\rightarrow 1 \\ 1 &\rightarrow 0 \\ 2 &\rightarrow 0 \\ 3 &\rightarrow 1 \\ &\cdots \\ 2^n - 2 &\rightarrow 0 \\ 2^n - 1 &\rightarrow 0 \end{aligned}$$

We can think of all the values on the right hand side as a string that, in turn, can be thought of as a binary representation of a number. This string is 2^n elements long. Hence the numbers can range from 0 to $2^{2^n} - 1$. There are, in other words, 2^{2^n} possible right hand side strings, and each of them defines a different gate f.

Now consider a gate with n inputs and m outputs. Every one of its m outputs can be thought of as an independent single bit gate on an n-bit domain. So every one of the single bit outputs can be characterized by one of the 2^{2^n} values that characterize a gate such as f. Think of this output line as a digit (but not a digital digit and not a binary one either) that can hold a number from 0 to $2^{2^n} - 1$. Think of all m output lines as a register comprising m such digits, each of which can take values from 0 to $2^{2^n} - 1$. The numbers that can be addressed by this register now range from 0 to $\left(2^{2^n}\right)^m - 1$. Each of these numbers characterizes one $n \times m$ gate. There are, therefore, $\left(2^{2^n}\right)^m$ such gates.

Which is a larger number, $2^n!$ or $\left(2^{2^n}\right)^m$?

The total number of gates with 2 inputs and 2 outputs is

$$\left(2^{2^2}\right)^2 = 16^2 = 256$$

But the total number of *reversible* 2×2 gates is only

$$2^2! = 24$$

When you think of it, it is obvious that the number of reversible gates must be smaller, because reversible gates form a subset of an all-gates set. So the real question is how much smaller.

All reversible 2×2 gates (since they are *reversible* they have to have the same number of inputs and outputs, so we can simply call them 2-bit gates) can be represented by linear transformations, for example a *controlled-*$\neg$ gate is:

$$\{0,1\}^2 \ni (x,y) \to (x, y +_2 x) \in \{0,1\}^2$$

which in matrix notation can be represented as

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ x +_2 y \end{pmatrix}$$

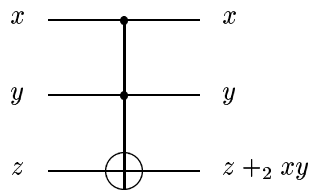
where $+_2$ is modulo-2 addition.[1]

But for $n \geq 3$ there are *nonlinear* gates that are reversible, for example the *Toffoli* gate about which more below. So those nonlinear 3- and more-bit reversible gates cannot be made by a combination of linear reversible 2-bit gates.

This means that

there are no universal reversible 2-bit gates in classical computing.

There are, however, universal reversible 3-bit gates in classical computing and one of them is the aforementioned Toffoli gate. The Toffoli gate is a *controlled-controlled-NOT* gate and its diagramatic representation is as follows:



This gate flips z if x and y are both 1 and leaves z alone if they aren't. x and y themselves remain unchanged. The nonlinearity of the gate is evident in the formula that describes the bottom output: $z +_2 xy$.

The Toffoli gate can be used as a *universal* gate for Boolean logic if

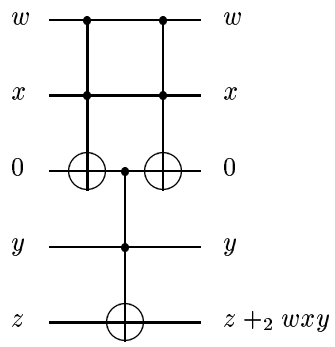
1. fixed input bits can be provided on some inputs,

2. some output bits can be ignored.

This can be shown as follows:

1. First demonstrate that an n-bit Toffoli gate can be constructed from 3-bit Toffoli gates allowing for scratch space.

[1]Note that this matrix representation has nothing to do with unitary matrices acting on qubits. In particular a 2-qubit unitary matrix would be 4×4, not 2×2.



This figure illustrates how you can implement a 4-input Toffoli gate by combining 3 3-input Toffoli gates. The point between the two gates on the 0 line in the middle of the diagram becomes 1 only if both w and x are 1. In order for z to toggle y must be 1 too. And so z toggles if all three, w, x, and y are 1, and doesn't otherwise.

2. Next demonstrate that by combining an n-bit Toffoli gate with $\neg$ gates we can alter the control string that triggers the action of the Toffoli gate. Well, this really should be obvious. We can insert $\neg$ gates in front of and behind any control point for the n-bit Toffoli gate and thus effectively change the sequence that triggers the toggle.

   The act of triggering the toggle effectively transposes two selected strings of 0s and 1s. Recall that the toggled line doesn't really have to be always at the bottom of the diagram. You can place $\oplus$ on any line and use other lines as controls.

   When a specific control sequence is applied to inputs only the two corresponding strings of 0s and 1s are transposed. All other strings remain unchanged.

3. Now notice that by combining multiple n-bit Toffoli gates with appropriately adjusted controls a circuit can be built that transposes *any* two specific n-bit strings, since such transpositions can be built from the ones that transpose a single bit only. This operation will leave all other strings unchanged.

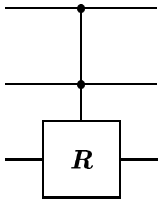
4. Any permutation (and hence any reversible gate) can then be implemented as an even or odd sequence of transpositions.

5. Observe that a $\neg$ gate can be implemented by a Toffoli gate with its two control lines tied to 1.

5.1.2 The Deutsch Gate

The Deutsch gate is a quantum gate, which is based on the idea of a Toffoli gate. It is a 3 input gate where the two top inputs control the action of the bottom

line. But this time the action is not a toggle. Instead it is a spin rotation by angle θ about the x axis. The figure below shows a diagrammatic representation of the Deutch gate.



The operation $\boldsymbol{R}$ is given by:

$$\boldsymbol{R} = -ie^{i(\theta/2)\boldsymbol{\sigma}_x} = -i\left(\cos\frac{\theta}{2} + i\boldsymbol{\sigma}_x \sin\frac{\theta}{2}\right)$$

We assume additionally that angle θ is *incommensurate* with π, i.e., that it is not a rational fraction of π.

The Deutsch gate has the following properties:

1. Because θ is not a rational fraction of π, consecutive applications of $\boldsymbol{R}$ to a qubit $\mid s\rangle$ will eventually reach any point on the $e^{i\lambda\boldsymbol{\sigma}_x} \mid s\rangle$ trajectory, or, in other words, for any fixed value of λ we can get arbitrarily close to $e^{i\lambda\boldsymbol{\sigma}_x} \mid s\rangle$ by applying $\boldsymbol{R}$ to $\mid s\rangle$ a finite number of times. Another way of saying the same thing is to state that

   powers of $\boldsymbol{R}$ *are* dense *in torus* $e^{i\lambda\boldsymbol{\sigma}_x}$.

2. The n^{th} power of the Deutsch gate is a *controlled-controlled-*$\boldsymbol{R}^n$ gate. Since θ is incommensurate with π we can make $n\theta/2$ arbitrarily close to a multiple of $\pi/2$ at which stage the Deutsch gate becomes a Toffoli gate. Since Toffoli gate is universal for classical computation then so is Deutch gate.

3. In the 8 dimensional 3-qubit space ($2^3 = 8$): $\mid 000\rangle, \mid 001\rangle, \ldots, \mid 111\rangle$ the Toffoli gate can be represented by the following 8×8 matrix

$$\boldsymbol{T} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

   The gate is an identity for all vectors of the form $\mid 00x\rangle$, $\mid 01x\rangle$, and $\mid 10x\rangle$, and it fires up for $\mid 11x\rangle$ where it converts $\mid 110\rangle$ into $\mid 111\rangle$ and vice versa, or, in other words, it converts $\mid \mathbf{6}\rangle$ into $\mid \mathbf{7}\rangle$.

4. What is a generator of this gate in the 8-dimensional computational space? We need to find such 8×8 matrix $\boldsymbol{A}$ that $\boldsymbol{T} = e^{i\boldsymbol{A}}$. Consider the following simple matrix:

$$\boldsymbol{\sigma}_x(|\,\mathbf{6}\rangle, |\,\mathbf{7}\rangle) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Now we need to take $e^{ia\boldsymbol{\sigma}_x(|\mathbf{6}\rangle,|\mathbf{7}\rangle)}$. The easiest way to go about it is to simply type it into Mathematica and then take its exponens. Here's what the conversation with Mathematica looks like:

```
In[35] A = {{0, 0, 0, 0, 0, 0, 0, 0},
            {0, 0, 0, 0, 0, 0, 0, 0},
            {0, 0, 0, 0, 0, 0, 0, 0},
            {0, 0, 0, 0, 0, 0, 0, 0},
            {0, 0, 0, 0, 0, 0, 0, 0},
            {0, 0, 0, 0, 0, 0, 0, 0},
            {0, 0, 0, 0, 0, 0, 0, 1},
            {0, 0, 0, 0, 0, 0, 1, 0}}
Out[35] {{0, 0, 0, 0, 0, 0, 0, 0},
         {0, 0, 0, 0, 0, 0, 0, 0},
         {0, 0, 0, 0, 0, 0, 0, 0},
         {0, 0, 0, 0, 0, 0, 0, 0},
         {0, 0, 0, 0, 0, 0, 0, 0},
         {0, 0, 0, 0, 0, 0, 0, 0},
         {0, 0, 0, 0, 0, 0, 0, 1},
         {0, 0, 0, 0, 0, 0, 1, 0}}
In[36] T = FullSimplify [ MatrixExp [  I   a   A] ]
Out[36] {{1, 0, 0, 0, 0, 0, 0, 0},
         {0, 1, 0, 0, 0, 0, 0, 0},
         {0, 0, 1, 0, 0, 0, 0, 0},
         {0, 0, 0, 1, 0, 0, 0, 0},
         {0, 0, 0, 0, 1, 0, 0, 0},
         {0, 0, 0, 0, 0, 1, 0, 0},
         {0, 0, 0, 0, 0, 0,   Cos [a], i Sin [a]},
         {0, 0, 0, 0, 0, 0, i Sin [a],   Cos [a]}}
```

It is now easy to see that when $a = \pi/2$ we get

$$e^{i(\pi/2)\boldsymbol{\sigma}_x(|\mathbf{6}\rangle,|\mathbf{7}\rangle)}$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & i \\ 0 & 0 & 0 & 0 & 0 & 0 & i & 0 \end{pmatrix}$$

This is not *exactly* a Toffoli gate, but if we add a $\times(-i) = \times e^{-i\pi/2}$ gate on the third line then we'll get exactly the Toffoli gate. So this 8×8 matrix, $\boldsymbol{\sigma}_x(|\ \mathbf{6}\rangle, |\ \mathbf{7}\rangle)$ is as close to a generator of the Toffoli gate as we are going to get in this lecture.

5. This generator, as I have already remarked above, swaps $|\ \mathbf{6}\rangle \leftrightarrow |\ \mathbf{7}\rangle$. By applying various combinations of Toffoli and $\neg$ gates (but the latter can be made of Toffoli gates too, remember) to $|\ 000\rangle \ldots |\ 111\rangle$ we can generate any transposition and ultimately any permutation of these 8 basis vectors, so that we can generate transformations such as $|\ \mathbf{4}\rangle \leftrightarrow |\ \mathbf{6}\rangle$ and others, or, more generally $|\ \boldsymbol{j}\rangle \leftrightarrow |\ \boldsymbol{k}\rangle$.

6. In particular we can generate $|\ \mathbf{5}\rangle \leftrightarrow |\ \mathbf{6}\rangle$ and take the corresponding generator of this transformation, which (up to an i) is going to look as follows

$$\boldsymbol{\sigma}_x(|\ \mathbf{5}\rangle, |\ \mathbf{6}\rangle) = \begin{pmatrix} \ddots & \vdots & \vdots & \vdots \\ \cdots & 0 & 1 & 0 \\ \cdots & 1 & 0 & 0 \\ \cdots & 0 & 0 & 0 \end{pmatrix}$$

The commutator of these two operators is

$$\begin{pmatrix} \ddots & \vdots & \vdots & \vdots \\ \cdots & 0 & 1 & 0 \\ \cdots & 1 & 0 & 0 \\ \cdots & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ddots & \vdots & \vdots & \vdots \\ \cdots & 0 & 0 & 0 \\ \cdots & 0 & 0 & 1 \\ \cdots & 0 & 1 & 0 \end{pmatrix}$$
$$- \begin{pmatrix} \ddots & \vdots & \vdots & \vdots \\ \cdots & 0 & 0 & 0 \\ \cdots & 0 & 0 & 1 \\ \cdots & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \ddots & \vdots & \vdots & \vdots \\ \cdots & 0 & 1 & 0 \\ \cdots & 1 & 0 & 0 \\ \cdots & 0 & 0 & 0 \end{pmatrix}$$
$$= \begin{pmatrix} \ddots & \vdots & \vdots & \vdots \\ \cdots & 0 & 0 & 1 \\ \cdots & 0 & 0 & 0 \\ \cdots & 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} \ddots & \vdots & \vdots & \vdots \\ \cdots & 0 & 0 & 0 \\ \cdots & 0 & 0 & 0 \\ \cdots & 1 & 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} \ddots & \vdots & \vdots & \vdots \\ \cdots & 0 & 0 & 1 \\ \cdots & 0 & 0 & 0 \\ \cdots & -1 & 0 & 0 \end{pmatrix}$$

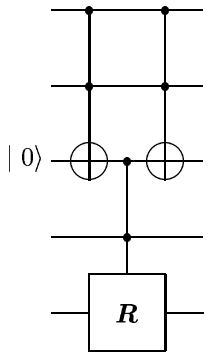
and this is $i\boldsymbol{\sigma}_y$ applied to $\mid \mathbf{5}\rangle$ and $\mid \mathbf{7}\rangle$.

7. Similarly we can generate $\boldsymbol{\sigma}_z$ between any $\mid \boldsymbol{j}\rangle$ and $\mid \boldsymbol{k}\rangle$ vectors, for example:
$$\left[\boldsymbol{\sigma}_x \left(\mid \boldsymbol{j}\right\rangle \mid \boldsymbol{k}\rangle\right), \boldsymbol{\sigma}_y \left(\mid \boldsymbol{j}\right\rangle \mid \boldsymbol{k}\rangle)\right] = i\boldsymbol{\sigma}_z \left(\mid \boldsymbol{j}\right\rangle \mid \boldsymbol{k}\rangle)$$

8. We can therefore reach any transformation generated by a linear combination of $\boldsymbol{\sigma}_x \left(\mid \boldsymbol{j}\right\rangle \mid \boldsymbol{k}\rangle)$, $\boldsymbol{\sigma}_y \left(\mid \boldsymbol{j}\right\rangle \mid \boldsymbol{k}\rangle)$, and $\boldsymbol{\sigma}_z \left(\mid \boldsymbol{j}\right\rangle \mid \boldsymbol{k}\rangle)$, and these span the whole SU(8) Lie Algebra.

9. Since using the generators of the Deutsch gate we can generate the whole Lie Algebra for SU(8), by taking exponens of the generated matrices we can deliver any element of SU(8). Consequently, the Deutsch gate is a universal gate for 3-qubit computation.

10. Proceeding similarly to how we did with the Toffoli gate we can construct the n-bit Deutsch gate. For example:

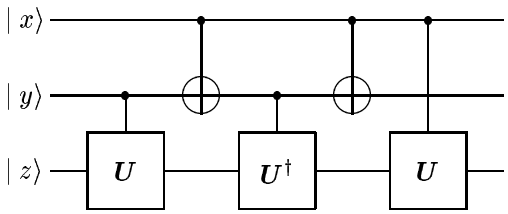


11. Proceeding as above we can then demonstrate that the n-bit Deutsch gate generates the whole SU (2^n). Consequently the Deutsch gate is a universal gate for quantum computation.

5.1.3 Universal 2-qubit Gates

So, a 3-qubit Deutch gate is a universal gate for quantum computation. But in quantum computation we can do even better. We can, as it turns out, build a 3-qubit Deutch gate from 2-qubit gates.

Consider the following quantum circuit



where $\boldsymbol{U}$ is a controlled unitary operator gate, and $\oplus$ is a controlled $\neg$ gate.

Let us analyze this circuit.

1. Assume that $|\, x\rangle = |\, 0\rangle$:

   If $|\, x\rangle = |\, 0\rangle$ then the first $\oplus$ gate leaves $|\, y\rangle$ alone and so does the second one. Also the last $\boldsymbol{U}$ gate on the $|\, z\rangle$ line is inactive. In effect we only have the $\boldsymbol{U}$ and $\boldsymbol{U}^\dagger$ gates on the $|\, z\rangle$ line.

   If $|\, y\rangle = |\, 0\rangle$ then, first, it stays that way, and, second, it inactivates the controlled $\boldsymbol{U}$ and $\boldsymbol{U}^\dagger$ gates on the $|\, z\rangle$ line, so $|\, z\rangle$ remains unchanged.

   If $|\, y\rangle = |\, 1\rangle$ then, first, it stays that way, and second, it *activates* $\boldsymbol{U}$ and $\boldsymbol{U}^\dagger$ on the $|\, z\rangle$ line. But because $\boldsymbol{U}$ is unitary $\boldsymbol{U}^\dagger = \boldsymbol{U}^{-1}$, so these two gates cancel each other in this context.

   In summary, $|\, x\rangle = |\, 0\rangle$ results in $|\, z\rangle$ unchanged.

2. Assume that $|\, y\rangle = |\, 0\rangle$ and that $|\, x\rangle = |\, 1\rangle$:

   In this case the two $\oplus$ gates on the $|\, y\rangle$ line toggle, and the last $\boldsymbol{U}$ gate on the $|\, z\rangle$ line is active too. But the first $\boldsymbol{U}$ gate is inactive. Because $|\, y\rangle$ is initially $|\, 0\rangle$ it is toggled to $|\, 1\rangle$ between the $\oplus$ gates and then back to $|\, 0\rangle$, and this activates the $\boldsymbol{U}^\dagger$ gate on the $|\, z\rangle$ line.

   In final effect we find that $|\, x\rangle$ remains unchanged, as does $|\, y\rangle$. The $|\, z\rangle$ state is passed through the $\boldsymbol{U}^\dagger$ gate first and then through the $\boldsymbol{U}$ gate, but $\boldsymbol{U}^\dagger\boldsymbol{U} = \boldsymbol{1}$, so $|\, z\rangle$ remains unchanged.

   In summary, having either $|\, x\rangle = |\, 0\rangle$ or $|\, y\rangle = |\, 0\rangle$ or both leaves $|\, z\rangle$ unchanged.

   In all situations considered so far $|\, x\rangle$ and $|\, y\rangle$ remain unchanged too.

3. Assume that $|\, x\rangle = |\, 1\rangle$ and that $|\, y\rangle = |\, 1\rangle$:

   In this situation

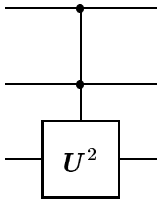
   (a) Both $\oplus$ gates on the $|\, y\rangle$ line toggle, so $|\, y\rangle$ emerges unchanged at the end, but...

   (b) the $\boldsymbol{U}^\dagger$ gate on the $|\, z\rangle$ line is inactivated by $|\, y\rangle$ becoming temporarily $|\, 0\rangle$ between the two $\oplus$ gates.

   (c) Both $\boldsymbol{U}$ gates on the $|\, z\rangle$ line are active, so initial $|\, z\rangle$ becomes transformed into $\boldsymbol{U}\boldsymbol{U}\,|\, z\rangle$.

   (d) $|\, x\rangle$ remains unchanged.

The whole circuit therefore behaves like a *controlled-controlled* $\boldsymbol{U}^2$ gate:



We can now choose $\boldsymbol{U}$ so that $\boldsymbol{U}^2 = \boldsymbol{R}$, e.g.,

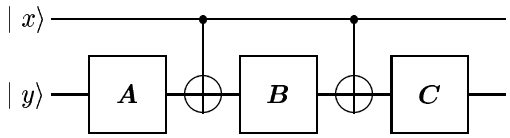
$$\begin{aligned} \boldsymbol{U}^2 &= -i\boldsymbol{R}_x(\theta) \\ \boldsymbol{U} &= e^{-i\pi/4}\boldsymbol{R}_x(\theta/2) \end{aligned}$$

and this yields the Deutsch gate, which is universal. Hence the combination of the 2-qubit $\oplus$ and $\boldsymbol{U}$ gates used in this circuit is universal too.

Why doesn't the same work for classical computing circuits? The reason for this is that in order to implement a Toffoli gate this way we would have to find such an operator $\boldsymbol{U}$ that $\boldsymbol{U}^2 = \neg$, in other words $\boldsymbol{U} = \sqrt{\neg}$. But $\sqrt{\neg}$ is a quantum gate. It cannot be implemented classically.

Now we are going to show that the controlled-$\boldsymbol{U}$ gates can be implemented as a combination of single qubit gates and the controlled-$\neg$ gate.

Consider the following circuit:



Let us analyze this circuit.

1. If $|\ x\rangle = |\ 0\rangle$ then
   (a) The $\oplus$ gates are inactive and $|\ y\rangle$ is subjected to the action of $\boldsymbol{A}$ followed by $\boldsymbol{B}$ followed by $\boldsymbol{C}$, or, in other words, $|\ y\rangle \rightarrow \boldsymbol{CBA}\ |\ y\rangle$.
   (b) Now, if $\boldsymbol{CBA} = \boldsymbol{1}$ then there is no change to $|\ y\rangle$. Such $\boldsymbol{C}$, $\boldsymbol{B}$, and $\boldsymbol{A}$ are easy to find. For example, let $\boldsymbol{A}$ be unitary. Then we can choose $\boldsymbol{C}$ and $\boldsymbol{B}$ be both $\sqrt{\boldsymbol{A}^\dagger}$. This way:
   $$\boldsymbol{CBA} = \sqrt{\boldsymbol{A}^\dagger}\sqrt{\boldsymbol{A}^\dagger}\boldsymbol{A} = \boldsymbol{A}^\dagger\boldsymbol{A} = \boldsymbol{1}$$
2. If $|\ x\rangle = |\ 1\rangle$ then
   (a) The $\oplus$ gates are active, i.e., they turn into $\neg = \boldsymbol{\sigma}_x$
   (b) Vector $|\ y\rangle$ now turns into
   $$|\ y\rangle \rightarrow \sqrt{\boldsymbol{A}^\dagger}\boldsymbol{\sigma}_x\sqrt{\boldsymbol{A}^\dagger}\boldsymbol{\sigma}_x\boldsymbol{A}\ |\ y\rangle$$
   where the operation $\sqrt{\boldsymbol{A}^\dagger}\boldsymbol{\sigma}_x\sqrt{\boldsymbol{A}^\dagger}\boldsymbol{\sigma}_x\boldsymbol{A} = \boldsymbol{U}$ is a unitary transformation, if $\boldsymbol{A}$ is unitary.

So this is, in effect, a controlled unitary gate, which, in combination with the controlled $\neg$ gate proved to be universal. In other words we have pushed the universality even one notch lower. We can state that

> *the combination of 2-qubit controlled $\neg$ and 1-qubit unitary gates is adequate for universal quantum computation.*

5.2 Simple Quantum Oracles

Oracles are devices, which are used to answer questions with a simple *yes* or *no*. The questions may be as elaborate as you can make them, the procedure that answers the questions may be lengthy and a lot of auxiliary data may get generated while the question is being answered. Yet all that comes out is just *yes* or *no*.

The oracle architecture is very suitable for quantum computers. The reason for this is that the read-out of a quantum system is probabilistic. Therefore if you pose a question the answer to which is given in the form of a wave function, you will have to carry out the computation on an ensemble of quantum computers to get anywhere. On the other hand if the computation can be designed in such a way that you do get your *yes* or *no* at the end, and some data reduction may be required to accomplish this, then a single quantum computer and a single quantum computation run may suffice.

In this section we are going to look at 4 increasingly complex oracles. The questions these oracles answer range from very silly to silly. Make no mistake: these are toy devices. Yet they demonstrate various techniques, some of which we are going to use later to answer more involved questions.

The oracles we are going to study are:

The Deutsch Oracle This oracle answers the following question. Suppose we have a function $f : \{0,1\} \to \{0,1\}$, which can be either *constant* or *balanced*. In this case the function is constant if $f(0) = f(1)$ and it is balanced if $f(0) \neq f(1)$. Classically it would take two evaluations of the function to tell whether it is one or the other. Quantumly, we can answer this question in one evaluation. The reason for this is that quantumly we can pack 0 and 1 into x at the same time, of course.

The Deutsch-Jozsa Oracle This oracle generalizes the Deutsch oracle to a function $f : \{0,1\}^n \to \{0,1\}$. We ask the same question: is the function constant or balanced. Here *balanced* means that the function is 0 on half of its arguments and 1 on the other half. Of course in this case the function may be *neither* constant nor balanced. In this case the oracle doesn't work: it may say *yes* or *no* and the answer will be meaningless.

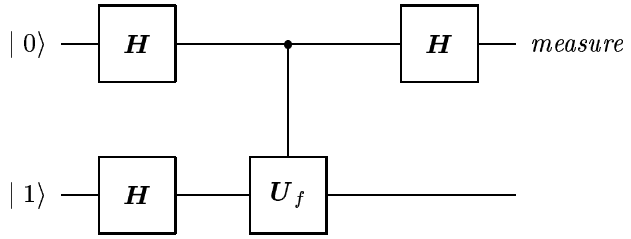
The Bernstein-Vazirani Oracle Suppose you have a function $f : \{0,1\}^n \to \{0,1\}$ of the form $f(\boldsymbol{x}) = \boldsymbol{a} \cdot \boldsymbol{x}$, where $\boldsymbol{a}$ is a constant vector of 0s and 1s and $\cdot$ is a scalar product. How many measurements are required to find $\boldsymbol{a}$? Classically you'd have to perform measurements for all possible

arguments and then solve a system of linear equations for $\boldsymbol{a}$. Quantumly $\boldsymbol{a}$ is delivered in one computational step on output lines of the oracle.

The Simon Oracle Suppose you have a function $f : \{0,1\}^n \rightarrow \{0,1\}^n$. The function is supposed to be 2-to-1, i.e., for every value of f there are always two such $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ that $f(\boldsymbol{x}_1) = f(\boldsymbol{x}_2)$. The function is also supposed to be periodic, meaning that there is such a binary vector $\boldsymbol{a}$ that $f(\boldsymbol{x} +_2 \boldsymbol{a}) = f(\boldsymbol{x}$, where $+_2$ designates addition modulo 2, i.e., $1 +_2 1 = 0$. The oracle returns period $\boldsymbol{a}$ in $\mathcal{O}(n)$ measurements. Of course, if you have a sufficiently large ensemble of quantum computers then a single computation will return the answer in the density operator.

5.2.1 The Deutsch Oracle

The circuit that implements the Deutsch Oracle is shown below:



Here $\boldsymbol{H}$ is the Hadamard gate, which we have already encountered before, when discussing the Brassard teleportation circuit and which is defined as follows:

$$\begin{aligned} \boldsymbol{H} \mid 0\rangle &= \frac{1}{\sqrt{2}}(\mid 0\rangle + \mid 1\rangle) \\ \boldsymbol{H} \mid 1\rangle &= \frac{1}{\sqrt{2}}(\mid 0\rangle - \mid 1\rangle) \end{aligned}$$

or in matrix notation:

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

and $\boldsymbol{U}_f$ is a controlled gate with the following definition:

$$\boldsymbol{U}_f \mid x\rangle \mid y\rangle = \mid x\rangle \mid y +_2 f(x)\rangle$$

Function f maps $\{0,1\}$ on $\{0,1\}$ and as such it can be either constant or balanced, and, as we have remarked in the preamble to this section, on a classical computer two measurements are required to figure out which.

Observe that the Hadamard operator on the upper line converts $\mid 0\rangle$ into $\frac{1}{\sqrt{2}}(\mid 0\rangle + \mid 1\rangle)$, this way we can feed simultaneously both $\mid 0\rangle$ and $\mid 1\rangle$ into x.

Let us analyze the circuit in detail:

1. The first pair of Hadamard transforms converts $\mid 0\rangle \mid 1\rangle$ into

$$\frac{1}{2}(\mid 0\rangle + \mid 1\rangle) \otimes (\mid 0\rangle - \mid 1\rangle)$$

2. Now we need to apply our controlled-$\boldsymbol{U}_f$ gate to this. There is a useful formula that we are going to derive now:

$$\boldsymbol{U}_f \mid x\rangle \otimes (\mid 0\rangle - \mid 1\rangle) = \mid x\rangle \otimes ((\mid 0\rangle - \mid 1\rangle) +_2 f(x))$$

If $f(x) = 0$ then

$$(\mid 0\rangle - \mid 1\rangle) +_2 f(x) = \mid 0\rangle - \mid 1\rangle = (-1)^0 (\mid 0\rangle - \mid 1\rangle) = (-1)^{f(x)} (\mid 0\rangle - \mid 1\rangle)$$

If $f(x) = 1$ then

$$(\mid 0\rangle - \mid 1\rangle) +_2 f(x) = \mid 1\rangle - \mid 0\rangle = (-1)^1 (\mid 0\rangle - \mid 1\rangle) = (-1)^{f(x)} (\mid 0\rangle - \mid 1\rangle)$$

This means that the same formula holds for all values of $f(x)$, therefore:

$$\boldsymbol{U}_f \mid x\rangle \otimes (\mid 0\rangle - \mid 1\rangle) = (-1)^{f(x)} \mid x\rangle \otimes (\mid 0\rangle - \mid 1\rangle)$$

Applying this to our state yields:

$$\begin{aligned} &\boldsymbol{U}_f \frac{1}{2}(\mid 0\rangle + \mid 1\rangle) \otimes (\mid 0\rangle - \mid 1\rangle) \\ &\quad = \frac{1}{2}\left((-1)^{f(0)} \mid 0\rangle + (-1)^{f(1)} \mid 1\rangle\right) \otimes (\mid 0\rangle - \mid 1\rangle) \end{aligned}$$

3. The last step in the circuit analysis applies the Hadamard gate to the first vector, i.e., to $(-1)^{f(0)} \mid 0\rangle + (-1)^{f(1)} \mid 1\rangle$:

$$\begin{aligned} &\frac{1}{2}\left((-1)^{f(0)}\boldsymbol{H} \mid 0\rangle + (-1)^{f(1)}\boldsymbol{H} \mid 1\rangle\right) \otimes (\mid 0\rangle - \mid 1\rangle) \\ &\quad = \frac{1}{2}\left((-1)^{f(0)}\frac{1}{\sqrt{2}}(\mid 0\rangle + \mid 1\rangle) + (-1)^{f(1)}\frac{1}{\sqrt{2}}(\mid 0\rangle - \mid 1\rangle)\right) \otimes (\mid 0\rangle - \mid 1\rangle) \\ &\quad = \frac{1}{2}\left(\mid 0\rangle \left((-1)^{f(0)} + (-1)^{f(1)}\right) + \mid 1\rangle \left((-1)^{f(0)} - (-1)^{f(1)}\right)\right) \otimes \frac{1}{\sqrt{2}}(\mid 0\rangle - \mid 1\rangle) \end{aligned}$$

Now observe that

if $f(x)$ is constant then

$$(-1)^{f(0)} - (-1)^{f(1)} = 0$$

and in this case the upper line vector evaluates to:

$$\frac{1}{2}\left(\mid 0\rangle \left((-1)^{f(0)} + (-1)^{f(1)}\right)\right) = \pm \mid 0\rangle$$

if $f(x)$ is balanced then

$$(-1)^{f(0)} + (-1)^{f(1)} = 0$$

and in this case the upper line vector evaluates to:

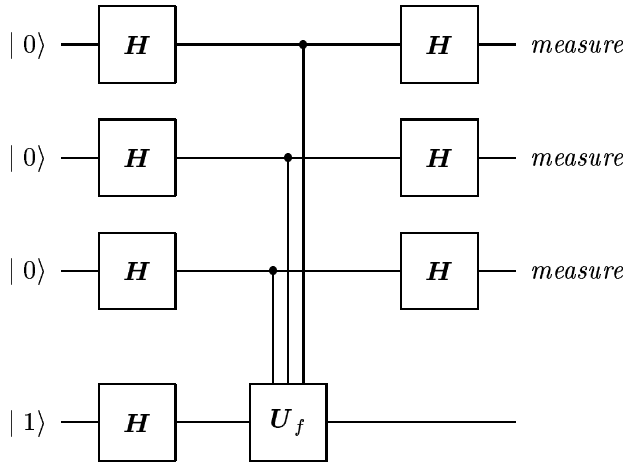
$$\frac{1}{2}\left(\mid 1\rangle \left((-1)^{f(0)} - (-1)^{f(1)}\right)\right) = \pm \mid 1\rangle$$

Consequently in order to find whether $f(x)$ is constant or balanced all that's required is to measure the upper line vector. If it's $\mid 0\rangle$ then $f(x)$ is constant, if it's $\mid 1\rangle$ then $f(x)$ is balanced.

You may ask what happens to the bottom line and why the values such as $(-1)^{f(x)}$ mysteriously shifted to the upper line and have not stayed with the bottom line. The answer is that the bottom line is allowed to decohere. As it does so, it collapses onto $\mid 0\rangle$ or $\mid 1\rangle$, thus forcing the parameters that describe the bipartite state onto the upper line. The Deutsch oracle is a very nice and simple demonstration of the essentials of quantum computing: first it shows the power of quantum parallelism, then it shows the importance of entanglement and non-locality in quantum computing. Every quantum computer is a little demonstration of the Einstein-Podolsky-Rosen paradox.

5.2.2 The Deutsch-Jozsa Oracle

This oracle looks much like the Deutsch oracle, only more so:



The Hadamard gates work as before, and the $\boldsymbol{U}_f$ gate is now controlled not by one but by n lines. Function f maps from $\{0,1\}^n$ to $\{0,1\}$, and, as before, it can be either *constant* or *balanced*. Our task is to determine which of the two by performing just one measurement. A classical oracle would require 2^n measurements, one for each value of the argument, to ascertain that f is constant.

So let us analyze this circuit now.

1. First we need to apply $\boldsymbol{H}$ to n vectors. We have already seen this done when we introduced the concept of a computational basis. The reasoning below repeats the calculation:

$$\begin{aligned}
&\boldsymbol{H} \mid 0\rangle \boldsymbol{H} \mid 0\rangle \cdots \boldsymbol{H} \mid 0\rangle \\
&= \frac{1}{\sqrt{2}}(\mid 0\rangle + \mid 1\rangle)\frac{1}{\sqrt{2}}(\mid 0\rangle + \mid 1\rangle)\cdots\frac{1}{\sqrt{2}}(\mid 0\rangle + \mid 1\rangle) \\
&= \frac{1}{2^{n/2}}(\mid 00\ldots0\rangle + \mid 00\ldots1\rangle + \ldots + \mid 11\ldots1\rangle) \\
&= \frac{1}{2^{n/2}}(\mid \mathbf{0}\rangle + \mid \mathbf{1}\rangle + \mid \mathbf{2}\rangle + \ldots + \mid 2^n - 1\rangle) \\
&= \frac{1}{2^{n/2}}\sum_{\boldsymbol{x}=0}^{2^n-1} \mid \boldsymbol{x}\rangle
\end{aligned}$$

   Applying $\boldsymbol{H}$ to the bottom line yields

$$\frac{1}{\sqrt{2}}(\mid 0\rangle - \mid 1\rangle)$$

   So the state of the whole computer becomes:

$$\frac{1}{2^{n/2}}\left(\sum_{\boldsymbol{x}=0}^{2^n-1} \mid \boldsymbol{x}\rangle\right) \otimes \frac{1}{\sqrt{2}}(\mid 0\rangle - \mid 1\rangle)$$

2. Now the n-line controlled $\boldsymbol{U}_f$ is applied, and, extending our result from the previous section, we get:

$$\frac{1}{2^{n/2}}\left(\sum_{\boldsymbol{x}=0}^{2^n-1}(-1)^{f(\boldsymbol{x})} \mid \boldsymbol{x}\rangle\right) \otimes \frac{1}{\sqrt{2}}(\mid 0\rangle - \mid 1\rangle)$$

3. Finally we have to apply the Hadamard transform to the top n lines again. But the top lines are no longer just $\mid 0\rangle$, so here we have to do some more thinking.

   Observe that the basic definition for the Hadamard transform can be rewritten as follows:

$$\begin{aligned}
\boldsymbol{H} \mid 0\rangle &= \frac{1}{\sqrt{2}}(\mid 0\rangle + \mid 1\rangle) = \frac{1}{\sqrt{2}}\left((-1)^{0\cdot0} \mid 0\rangle + (-1)^{0\cdot1} \mid 1\rangle\right) \\
\boldsymbol{H} \mid 1\rangle &= \frac{1}{\sqrt{2}}(\mid 0\rangle - \mid 1\rangle) = \frac{1}{\sqrt{2}}\left((-1)^{1\cdot0} \mid 0\rangle + (-1)^{1\cdot1} \mid 1\rangle\right)
\end{aligned}$$

   in summary:

$$\boldsymbol{H} \mid x\rangle = \frac{1}{\sqrt{2}}\sum_{y=0}^{1}(-1)^{x\cdot y} \mid y\rangle$$

In order to use this formula we have to figure out how to apply it not to an individual qubit, but to a tensor product of n qubits:

$$\begin{aligned}
&\boldsymbol{H} \mid x_1\rangle \otimes \boldsymbol{H} \mid x_2\rangle \otimes \ldots \otimes \boldsymbol{H} \mid x_n\rangle \\
&= \left(\frac{1}{\sqrt{2}} \sum_{y_1=0}^{1} (-1)^{x_1 \cdot y_1} \mid y_1\rangle\right) \otimes \left(\frac{1}{\sqrt{2}} \sum_{y_2=0}^{1} (-1)^{x_2 \cdot y_2} \mid y_2\rangle\right) \otimes \ldots \\
&\quad \otimes \left(\frac{1}{\sqrt{2}} \sum_{y_n=0}^{1} (-1)^{x_n \cdot y_n} \mid y_n\rangle\right) \\
&= \frac{1}{2^{n/2}} \sum_{y_1 y_2 \ldots y_n} (-1)^{x_1 \cdot y_1} (-1)^{x_2 \cdot y_2} \cdots (-1)^{x_n \cdot y_n} \mid y_1 y_2 \ldots y_n\rangle \\
&= \frac{1}{2^{n/2}} \sum_{\boldsymbol{y}=\boldsymbol{0}}^{2^n-1} (-1)^{\boldsymbol{x} \cdot \boldsymbol{y}} \mid \boldsymbol{y}\rangle
\end{aligned}$$

where

$$\boldsymbol{x} \cdot \boldsymbol{y} = x_1 \cdot y_1 +_2 x_2 \cdot y_2 +_2 \ldots +_2 x_n \cdot y_n$$

Now we can plug this expression into our formula to get:

$$\begin{aligned}
&\frac{1}{2^{n/2}} \left(\sum_{\boldsymbol{x}=\boldsymbol{0}}^{2^n-1} (-1)^{f(\boldsymbol{x})} \bigotimes^{n} \boldsymbol{H} \mid \boldsymbol{x}\rangle\right) \otimes \frac{1}{\sqrt{2}} (\mid 0\rangle - \mid 1\rangle) \\
&= \frac{1}{2^{n/2}} \left(\sum_{\boldsymbol{x}=\boldsymbol{0}}^{2^n-1} (-1)^{f(\boldsymbol{x})} \frac{1}{2^{n/2}} \sum_{\boldsymbol{y}=\boldsymbol{0}}^{2^n-1} (-1)^{\boldsymbol{x} \cdot \boldsymbol{y}} \mid \boldsymbol{y}\rangle\right) \otimes \frac{1}{\sqrt{2}} (\mid 0\rangle - \mid 1\rangle) \\
&= \frac{1}{2^n} \left(\sum_{\boldsymbol{x}=\boldsymbol{0}}^{2^n-1} \sum_{\boldsymbol{y}=\boldsymbol{0}}^{2^n-1} (-1)^{f(\boldsymbol{x})} (-1)^{\boldsymbol{x} \cdot \boldsymbol{y}} \mid \boldsymbol{y}\rangle\right) \frac{1}{\sqrt{2}} (\mid 0\rangle - \mid 1\rangle)
\end{aligned}$$

Well, that's it! Wasn't too painful, was it?

What can we deduct from our final formula? First observe that if $f(\boldsymbol{x})$ is constant then we can take it in front of the sum, and then the sum becomes:

$$\sum_{\boldsymbol{x}=\boldsymbol{0}}^{2^n-1} \sum_{\boldsymbol{y}=\boldsymbol{0}}^{2^n-1} (-1)^{\boldsymbol{x} \cdot \boldsymbol{y}} \mid \boldsymbol{y}\rangle$$

Now, let us fix $\mid \boldsymbol{y}\rangle$ and consider what

$$\sum_{\boldsymbol{x}=\boldsymbol{0}}^{2^n-1} (-1)^{\boldsymbol{x} \cdot \boldsymbol{y}} \mid \boldsymbol{y}\rangle$$

is going to be. If $\boldsymbol{y} \neq \boldsymbol{0}$ then

$$\sum_{\boldsymbol{x}=\boldsymbol{0}}^{2^n-1} (-1)^{\boldsymbol{x} \cdot \boldsymbol{y}}$$

must be zero, because $\boldsymbol{x} \cdot \boldsymbol{y}$ will "push as often to the right as to the left". So the only term that is going to survive in this case is for $\boldsymbol{y} = \boldsymbol{0}$. Consequently in this case the final state of the oracle is going to be:

$$\frac{1}{2^n}(-1)^{f(\boldsymbol{x})}\left(\sum_{\boldsymbol{x}=\boldsymbol{0}}^{2^n-1}(-1)^{\boldsymbol{x}\cdot\boldsymbol{0}} \mid \boldsymbol{0}\rangle\right)\otimes\frac{1}{\sqrt{2}}(\mid 0\rangle - \mid 1\rangle) = (-1)^{f(\boldsymbol{x})} \mid \boldsymbol{0}\rangle\otimes\frac{1}{\sqrt{2}}(\mid 0\rangle - \mid 1\rangle)$$

On the other hand if $f(\boldsymbol{x})$ is balanced then for $\mid \boldsymbol{y}\rangle = \mid \boldsymbol{0}\rangle$ we get

$$\sum_{\boldsymbol{x}=\boldsymbol{0}}^{2^n-1}(-1)^{f(\boldsymbol{x})}(-1)^{\boldsymbol{x}\cdot\boldsymbol{0}} \mid \boldsymbol{0}\rangle = \sum_{\boldsymbol{x}=\boldsymbol{0}}^{2^n-1}(-1)^{f(\boldsymbol{x})} \mid \boldsymbol{0}\rangle = 0$$

because $f(\boldsymbol{x})$ pushes as often to the right as it pushes to the left, on account of being balanced, so here we get that the probability amplitude of finding $\mid \boldsymbol{y}\rangle$ in $\mid \boldsymbol{0}\rangle$ is zero.

In summary, if $f(\boldsymbol{x})$ is constant then measuring control lines on exit *must* return $\mid 0\rangle$ on every line. If this is not the case then $f(\boldsymbol{x})$ must be balanced.

5.2.3 The Bernstein-Vazirani Oracle

The Bernstein Vazirani Oracle is the Deutsch Jozsa oracle with

$$f(\boldsymbol{x}) = \boldsymbol{a} \cdot \boldsymbol{x}$$

The final state of the oracle is:

$$\frac{1}{2^n}\left(\sum_{\boldsymbol{x}=\boldsymbol{0}}^{2^n-1}\sum_{\boldsymbol{y}=\boldsymbol{0}}^{2^n-1}(-1)^{\boldsymbol{a}\cdot\boldsymbol{x}}(-1)^{\boldsymbol{x}\cdot\boldsymbol{y}} \mid \boldsymbol{y}\rangle\right)\frac{1}{\sqrt{2}}(\mid 0\rangle - \mid 1\rangle)$$

As before, consider the sum over $\boldsymbol{x}$:

$$\sum_{\boldsymbol{x}=\boldsymbol{0}}^{2^n-1}(-1)^{\boldsymbol{a}\cdot\boldsymbol{x}}(-1)^{\boldsymbol{x}\cdot\boldsymbol{y}} \mid \boldsymbol{y}\rangle$$

If $\boldsymbol{a} \neq \boldsymbol{y}$ we are going to get zero, because the components of the sum will push as much to the right as they will push to the left. But if $\boldsymbol{a} = \boldsymbol{y}$ then we simply get

$$(-1)^{\boldsymbol{a}\cdot\boldsymbol{x}}(-1)^{\boldsymbol{a}\cdot\boldsymbol{x}} = 1$$

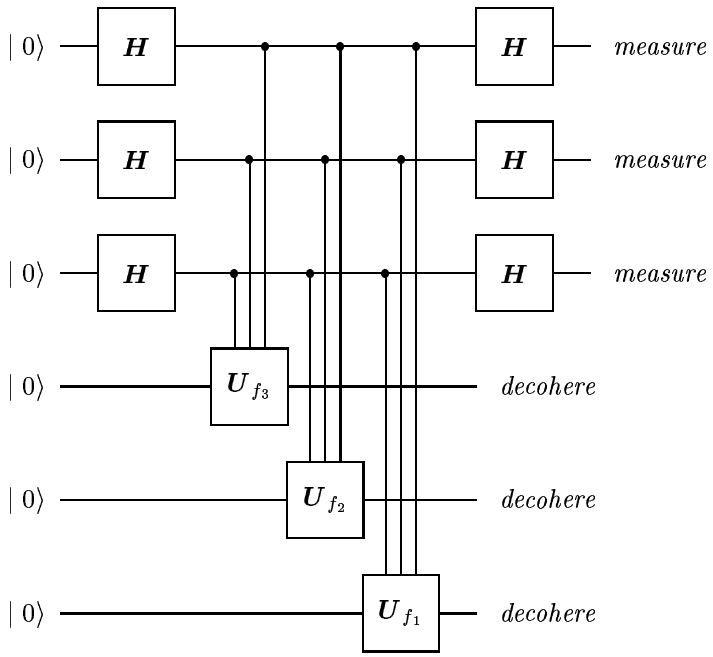
so that our final state becomes:

$$\left(\sum_{\boldsymbol{y}}\delta_{\boldsymbol{a},\boldsymbol{y}} \mid \boldsymbol{y}\rangle\right)\otimes\frac{1}{\sqrt{2}}(\mid 0\rangle - \mid 1\rangle) = \mid \boldsymbol{a}\rangle\otimes\frac{1}{\sqrt{2}}(\mid 0\rangle - \mid 1\rangle)$$

Thus, measuring the control lines, in this case, returns $\boldsymbol{a}$.

5.2.4 The Simon Oracle

An example of a Simon Oracle for a function $f : \{0,1\}^3 \rightarrow \{0,1\}^3$ is shown below.



In general the oracle comprises n lines at the top, which look the same as the top lines in oracles we have analyzed in previous sections, and then n lines at the bottom. Each of these n lines corresponds to a sub-function $f_i : \{0,1\}^n \rightarrow \{0,1\}$, n of which make up function f.

The boxes labelled $\boldsymbol{U}_{f_k}$ are controlled $\neg$ gates, where the control is provided by $f_k(\boldsymbol{x})$.

In summary, the Simon oracle tests a function:

$$f : \{0,1\}^n \rightarrow \{0,1\}^n$$

which in the drawing above was split into n scalar-valued functions

$$f_k : \{0,1\}^n \rightarrow \{0,1\}$$

The Simon oracle function must satisfy the following conditions:

1. f is 2-to-1, i.e., for every value of f there are always two different vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ such that $f(\boldsymbol{x}_1) = f(\boldsymbol{x}_2)$

2. f is periodic, i.e., there exists such vector $\boldsymbol{a}$ that $f(\boldsymbol{x} +_2 \boldsymbol{a}) = f(\boldsymbol{x})$

Of course, you may ask: if f is periodic then it should be more than just 2-to-1, because

$$f(\boldsymbol{x} +_2 \boldsymbol{a} +_2 \boldsymbol{a}) = f(\boldsymbol{x} +_2 \boldsymbol{a}) = f(\boldsymbol{x})$$

But remember that here we work within binary arithmetic and $+_2$ is a modulo-2 addition (or XOR), hence for every vector $\boldsymbol{a}$ $\boldsymbol{a}+_2\boldsymbol{a} = \boldsymbol{0}$ and therefore $\boldsymbol{x}+_2\boldsymbol{a}+_2\boldsymbol{a}$ takes us back to $\boldsymbol{x}$.

Assuming that function $f(\boldsymbol{x})$ satisfies these conditions, the oracle yields its period $\boldsymbol{a}$ in $\mathcal{O}(n)$ measurements.

This is a considerable improvement on a classical system designed to do the same, because the latter would have to be queried an exponential number of times (in n) in order to find $\boldsymbol{a}$.

Let us analyze the circuit and see how the oracle works:

1. Applying the Hadamard transform to the top n lines works the same way as we have already seen in the Deutsch-Jozsa oracle, so we can simply reuse the result obtained there (see step 1 for the analysis of the Deutsch-Jozsa circuit):

$$\frac{1}{2^{n/2}}\left(\sum_{\boldsymbol{x}=\boldsymbol{0}}^{2^n-1} \mid \boldsymbol{x}\rangle\right) \otimes \mid 0\rangle \mid 0\rangle \cdots \mid 0\rangle = \frac{1}{2^{n/2}}\left(\sum_{\boldsymbol{x}=\boldsymbol{0}}^{2^n-1} \mid \boldsymbol{x}\rangle\right) \otimes \mid \boldsymbol{0}\rangle$$

2. The application of the $\boldsymbol{U}_{f_k}$ gates at this stage converts the n bottom lines that carry $\mid \boldsymbol{0}\rangle$ into $\mid f_k(\boldsymbol{x})\rangle$. This is easy to see: for every individual $\mid 0\rangle$ line, if its corresponding $f_k(\boldsymbol{x})$ evaluates to 1 the line is flipped to $\mid 1\rangle$, if $f_k(\boldsymbol{x})$ evaluates to 0, the line stays $\mid 0\rangle$, consequently the line simply becomes $\mid f_k(\boldsymbol{x})\rangle$.

   In effect we get:

$$\frac{1}{2^{n/2}}\left(\sum_{\boldsymbol{x}=\boldsymbol{0}}^{2^n-1} \mid \boldsymbol{x}\rangle\right) \otimes \mid f(\boldsymbol{x})\rangle$$

3. Now allow the bottom n lines to decohere and this yields some value, which corresponds either to $f(\boldsymbol{x}_0)$ or to $f(\boldsymbol{x}_0 +_2 \boldsymbol{a})$. So this puts the top n lines into a superposition of these two vectors (it's Einstein-Podolsky-Rosen paradox again!), and the state of the computer becomes

$$\frac{1}{\sqrt{2}}\left(\mid \boldsymbol{x}_0\rangle + \mid \boldsymbol{x}_0 +_2 \boldsymbol{a}\rangle\right) \otimes \mid f(\boldsymbol{x}_0)\rangle$$

4. Applying the Hadamard transform to the top n lines now results in:

$$\frac{1}{\sqrt{2}}\frac{1}{2^{n/2}}\left(\sum_{\boldsymbol{y}=\boldsymbol{0}}^{2^n-1}\left((-1)^{\boldsymbol{x}_0\cdot\boldsymbol{y}} + (-1)^{(\boldsymbol{x}_0+_2\boldsymbol{a})\cdot\boldsymbol{y}}\right) \mid \boldsymbol{y}\rangle\right) \otimes \mid f(\boldsymbol{x}_0)\rangle$$

   Now we can divide all vectors $\boldsymbol{y}$ into two classes. For one $\boldsymbol{y}\cdot\boldsymbol{a} = 1$, and for the other one $\boldsymbol{y}\cdot\boldsymbol{a} = 0$. For the first class we'll end up generating:

$$(-1)^{\boldsymbol{x}_0\cdot\boldsymbol{y}} - (-1)^{\boldsymbol{x}_0\cdot\boldsymbol{y}} = 0$$

for every coefficient, so the only vectors $\boldsymbol{y}$ that are going to survive this are vectors perpendicular to $\boldsymbol{a}$, hence this sum evaluates to:

$$\frac{2}{\sqrt{2}}\frac{1}{2^{n/2}}\left(\sum_{\boldsymbol{y}\cdot\boldsymbol{a}=0}(-1)^{\boldsymbol{x}_0\cdot\boldsymbol{y}}\mid \boldsymbol{y}\rangle\right)\otimes\mid f(\boldsymbol{x}_0)\rangle$$

Measuring the top n lines now returns always a vector $\boldsymbol{y}$ which is perpendicular to $\boldsymbol{a}$. But it can be any vector from the superposition generated by the corresponding measurement on the bottom n lines. However, if we perform the measurement a sufficient number of times to obtain n different vectors $\boldsymbol{y}_k$, then we get n independent equations:

$$\begin{aligned}\boldsymbol{y}_1\cdot\boldsymbol{a} &= 0\\ \boldsymbol{y}_2\cdot\boldsymbol{a} &= 0\\ &\cdots\\ \boldsymbol{y}_n\cdot\boldsymbol{a} &= 0\end{aligned}$$

and these can be solved classically for $\boldsymbol{a}$.

This is a good place to stop and ask: what makes the Simon oracle so much more powerful than its classical equivalent. Our analysis of the circuit has already answered this question, but it's enlightening to summarize it here. *Repetitio mater studiorum est.*

1. Applying Hadamard transforms to $\mid 0\rangle$ inputs on the top n lines creates a superposition of all possible numbers from 0 to 2^n-1 in a single register. This superposition is then fed by the means of couplings to the second register, which evaluates function f. Function f is evaluated for all possible values of its argument at the same time. Classically we would need 2^n-1 separate evaluations to replace this single quantum step.

2. The use of controlled $\boldsymbol{U}_f$ gates entangles the whole computer.

3. When the bottom n lines are subjected to decoherence, the top n lines, which remain entangled with the bottom lines, are forced through the EPR paradox mechanism into superposition of some $\mid \boldsymbol{x}_0\rangle$ and $\mid \boldsymbol{x}_0 +_2 \boldsymbol{a}\rangle$. So we already have $\boldsymbol{a}$ waiting for us in the top n-qubit register although it is still mixed with $\boldsymbol{x}_0$. All that is required is to clear that $\boldsymbol{x}_0$ away. And this is done by performing further rotation of each qubit, which results in a superposition of vectors perpendicular to $\boldsymbol{a}$.

The three elements, the parallelism of (1), the entanglement of (2), and the EPR of (3), underlie just about every quantum computational algorithm. We can therefore end this section by stating that

> *The power of quantum computing derives from superposition and non-locality.*

5.3 Quantum Fourier Transform and its Applications

Quantum Fourier Transform is a magic box that replaces the last column of Hadamard Transforms in the Simon Oracle. The box is markedly more complex than Hadamard Transforms, but it lets us perform more effective extraction of data from the upper lines of various oracles than a simple Hadamard rotation.

Fourier transform is a tool for decomposing a signal, and it can be any signal, for example a computerised tomography measurement, or an acoustic signal, or a planetary motion[2], which is a signal of a kind too, into its constituent harmonics. But the inverse Fourier transform, which combines constituent harmonics, sometimes with changed amplitudes, into a new signal is a Fourier transform too: the only difference is in the change of the sign in front of the imaginary unit i. So whenever you have several harmonics that add up and interfere with various amplitudes forming beats and troughs, you see inverse Fourier transform in action.

Most often when a signal is received it is filtered before it is delivered to an analyzer. The receiver itself is a filtering device, because it is unlikely to be equally sensitive at all frequencies. If the receiver filter is described by $B(k)$ and the incoming signal $p(t)$ has a spectrum given by $P(k) = \boldsymbol{F}_k(p)$, then the spectrum of the received signal is going to be given by $B(k)P(k)$, and the received signal itself is going to be given by $p'(t) = \boldsymbol{F}_t^{-1}(B \cdot P)$. Now, if $b(t)$ is such a function that $B(k) = \boldsymbol{F}_k(b)$ then

$$p'(t) = \boldsymbol{F}_t^{-1}(B \cdot P) = \boldsymbol{F}_t^{-1}\left(\boldsymbol{F}_k(b) \cdot \boldsymbol{F}_k(p)\right) = \boldsymbol{F}_t^{-1}\left(\boldsymbol{F}_k(b \star p(t))\right) = b \star p(t),$$

where $\star$ is the convolution operator.

This works also in spacial domain – in any domain, in fact, because t is just a letter, and it can stand for anything. For example, the double slit experiment and the resulting interference pattern can be thought of as a convolution of the beam (the signal) with the slits (the filter).

In short, whenever we have interference, we have Fourier transform hiding behind. Whenever we have filters, we have convolution of a signal with a filter.

In quantum computing Quantum Fourier Transform is a device, which is used to force interference between qubits. The interference then enhances constructively certain aspects of the signal and destroys some other aspects: the constructive interference pattern is what we are after. This is the answer to our

[2]Decomposing a planetary motion into harmonics yields the old fashioned epicycles and deferents. There was nothing wrong with these. In fact their invention and then application to describe planetary motions was one of the greatest intellectual accomplishments of antiquity. Contrary to a popular belief epicycles and deferents were not invented by Ptolemy, who worked between 127 and 151AD in Alexandria, Egypt. Their application to describe planetary motions goes a long way back to Hipparchus, who lived in Nicaea, Bithynia, and died in Rhodes some time after 127BC. But their invention goes even further back to Apollonius of Perga, who was born about 262BC in Anatolia and who died about 190BC in Alexandria. From our perspective it often seems like Apollonius, Hipparchus and Ptolemy lived at the same time. But there is about 350 years difference between them. This is like difference between, say, us and Descartes. Descartes is hardly a contemporary of Stephen Hawking!

computation. The transform itself can be combined with a variety of filters, so as to produce a variety of computations.

5.3.1 Quantum Fourier Transform

Quantum Fourier Transform is a unitary operation on n qubits defined as follows:

$$\boldsymbol{F} : \mid \boldsymbol{x}\rangle \mapsto \frac{1}{\sqrt{2^n}} \sum_{\boldsymbol{y}=\boldsymbol{0}}^{2^n-1} e^{2\pi i x y/2^n} \mid \boldsymbol{y}\rangle, \tag{5.1}$$

where xy is a normal "decimal" multiplication of numbers x and y, which are represented by the quantum registers

$$\begin{aligned} \mid \boldsymbol{x}\rangle &= \mid x_{n-1}\rangle \otimes \mid x_{n-2}\rangle \otimes \cdots \otimes \mid x_0\rangle \\ \mid \boldsymbol{y}\rangle &= \mid y_{n-1}\rangle \otimes \mid y_{n-2}\rangle \otimes \cdots \otimes \mid y_0\rangle, \end{aligned}$$

where $\mid x_k\rangle$ and $\mid y_k\rangle$ are individual qubits.

Compare this with the notation in the section about Simon Oracle, where $\boldsymbol{x} \cdot \boldsymbol{y}$ meant

$$\boldsymbol{x} \cdot \boldsymbol{y} = x_0 \cdot y_0 +_2 x_1 \cdot y_1 +_2 x_2 \cdot y_2 +_2 \cdots +_2 x_{n-1} \cdot y_{n-1}$$

There we treated $\boldsymbol{x}$ and $\boldsymbol{y}$ as arrays of bits rather than integer numbers. Of course in computing a single integer number is implemented as an array of bits, but the point is how you interpret this array, and so xy in the Fourier Transform formula is not the same as $\boldsymbol{x} \cdot \boldsymbol{y}$ in the Simon Oracle formula. The former is an integer operation on two scalar numbers and the latter is a binary operation on two binary vectors. The former can be expressed in terms of a binary operation too, but it will not be $\boldsymbol{x} \cdot \boldsymbol{y}$.

Observe that once you know what $\boldsymbol{F}$ does to the basis vectors $\mid \boldsymbol{x}\rangle$, you can figure out what $\boldsymbol{F}$ does to any other vector. This other vector can be $\sum_{\boldsymbol{x}} f(\boldsymbol{x}) \mid \boldsymbol{x}\rangle$, which yields the following formula for Quantum Fourier Transform of function f:

$$\begin{aligned} \boldsymbol{F}\left(\sum_{\boldsymbol{x}=\boldsymbol{0}}^{N-1} f(\boldsymbol{x}) \mid \boldsymbol{x}\rangle\right) &= \sum_{\boldsymbol{x}=\boldsymbol{0}}^{N-1} f(\boldsymbol{x}) \boldsymbol{F}\left(\mid \boldsymbol{x}\rangle\right) \\ &= \frac{1}{\sqrt{N}} \sum_{\boldsymbol{x}=\boldsymbol{0}}^{N-1} f(\boldsymbol{x}) \sum_{\boldsymbol{y}=\boldsymbol{0}}^{N-1} e^{2\pi i x y/N} \mid \boldsymbol{y}\rangle \\ &= \frac{1}{\sqrt{N}} \sum_{\boldsymbol{y}=\boldsymbol{0}}^{N-1} \sum_{\boldsymbol{x}=\boldsymbol{0}}^{N-1} f(\boldsymbol{x}) e^{2\pi i x y/N} \mid \boldsymbol{y}\rangle \end{aligned}$$

From this formula the y^{th} component of $\boldsymbol{F}$ is

$$\boldsymbol{F}_{\boldsymbol{y}}(f) = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{x}=\boldsymbol{0}}^{N-1} f(\boldsymbol{x}) e^{2\pi i x y/N}$$

which is beginning to look quite like a normal Discrete Fourier Transform.

5.3.2 The QFT Circuit

In order to implement the circuit that calculates:

$$\frac{1}{\sqrt{2^n}} \sum_{\boldsymbol{y}=\boldsymbol{0}}^{2^n-1} e^{2\pi i x y/2^n} \mid \boldsymbol{y}\rangle$$

we shall deploy the trickery of the Fast Fourier Transform. Let us have a look at:

$$e^{2\pi i x y/2^n}$$

This expression is periodic in xy and the period is 2^n. The trick about the Fast Fourier Transform is that it only uses the terms of $e^{2\pi i x y/2^n}$ that correspond to the "first circle", i.e., the terms for which $xy/2^n < 1$. Let us evalute then $xy/2^n$ while truncating everything that would go onto the second and third circle:

$$\begin{aligned}\frac{xy}{2^n} &\equiv \frac{1}{2^n}\left(x_0 + x_1 2 + x_2 2^2 + x_3 2^3 + \cdots + x_{n-1}2^{n-1}\right) \\ &\quad \times \left(y_0 + y_1 2 + y_2 2^2 + y_3 2^3 + \cdots + y_{n-1}2^{n-1}\right) = \ldots\end{aligned}$$

Here we have decomposed x and y into their binary components, so that each of the x_k and y_k terms is either 0 or 1.

$$\begin{aligned}&= \frac{1}{2^n}\Big(y_0\left(x_0 + x_1 2 + x_2 2^2 + \cdots + x_{n-1}2^{n-1}\right) \\ &\quad + y_1 2^1\left(x_0 + x_1 2 + x_2 2^2 + \cdots + x_{n-1}2^{n-1}\right) \\ &\quad + \ldots \\ &\quad + y_{n-1}2^{n-1}\left(x_0 + x_1 2 + x_2 2^2 + \cdots + x_{n-1}2^{n-1}\right)\Big) \\ &= \frac{1}{2^n}\Big(y_0\left(x_0 + x_1 2 + x_2 2^2 + \cdots + x_{n-1}2^{n-1}\right) \\ &\quad + y_1\left(x_0 2 + x_1 2^2 + x_2 2^3 + \cdots + x_{n-2}2^{n-1}\right) \\ &\quad + y_2\left(x_0 2^2 + x_1 2^3 + x_2 2^4 + \cdots + x_{n-3}2^{n-1}\right) \\ &\quad + \cdots \\ &\quad + y_{n-1}x_0 2^{n-1}\Big) \\ &= y_0\left(\frac{x_0}{2^n} + \frac{x_1}{2^{n-1}} + \frac{x_2}{2^{n-2}} + \cdots + \frac{x_{n-1}}{2}\right) \\ &\quad + y_1\left(\frac{x_0}{2^{n-1}} + \frac{x_1}{2^{n-2}} + \frac{x_2}{2^{n-3}} + \cdots + \frac{x_{n-2}}{2}\right) \\ &\quad + y_2\left(\frac{x_0}{2^{n-2}} + \frac{x_1}{2^{n-3}} + \frac{x_2}{2^{n-4}} + \cdots + \frac{x_{n-3}}{2}\right) \\ &\quad + \cdots \\ &\quad + y_{n-1}\frac{x_0}{2} \\ &= \ldots\end{aligned}$$

There is a special notation, which covers the sums in the brackets:

$$\begin{aligned} \frac{x_0}{2} &\rightarrow (.x_0) \\ \frac{x_0}{2^2}+\frac{x_1}{2} &\rightarrow (.x_0x_1) \\ \frac{x_0}{2^3}+\frac{x_1}{2^2}+\frac{x_2}{2} &\rightarrow (.x_0x_1x_2) \\ &\cdots \end{aligned}$$

Using this notation:

$$\begin{aligned} \frac{xy}{2^n} \equiv\ & y_0(.x_0x_1\ldots x_{n-1})+y_1(.x_0x_1\ldots x_{n-2})+y_2(.x_0x_1\ldots x_{n-3})+\cdots \\ & +y_{n-1}(.x_0) \end{aligned}$$

So now we can write our Quantum Fourier Transform thusly:

$$\begin{aligned} \boldsymbol{F}\mid \boldsymbol{x}\rangle &= \frac{1}{\sqrt{2^n}}\sum_{\boldsymbol{y}=\boldsymbol{0}}^{2^n-1} e^{2\pi ixy/2^n}\mid \boldsymbol{y}\rangle \\ &= \frac{1}{\sqrt{2^n}}\sum_{\boldsymbol{y}=\boldsymbol{0}}^{2^n-1} e^{2\pi i(y_0(.x_0x_1\ldots x_{n-1})+y_1(.x_0x_1\ldots x_{n-2})+y_2(.x_0x_1\ldots x_{n-3})+\cdots+y_{n-1}(.x_0))}\mid \boldsymbol{y}\rangle \\ &= \frac{1}{\sqrt{2^n}}\sum_{\boldsymbol{y}=\boldsymbol{0}}^{2^n-1} e^{2\pi iy_0(.x_0x_1\ldots x_{n-1})}\mid y_0\rangle\otimes e^{2\pi iy_1(.x_0x_1\ldots x_{n-2})}\mid y_1\rangle \\ &\quad \otimes e^{2\pi iy_2(.x_0x_1\ldots x_{n-3})}\mid y_2\rangle\otimes\cdots\otimes e^{2\pi iy_{n-1}(.x_0)}\mid y_{n-1}\rangle \end{aligned}$$

Now observe that y_k is either 0 or 1. If it is 0 then the corresponding term is, for example,

$$e^{2\pi i0(.x_0x_1\ldots x_{n-3})}\mid 0\rangle =\mid 0\rangle$$

If it is 1 then the corresponding term is:

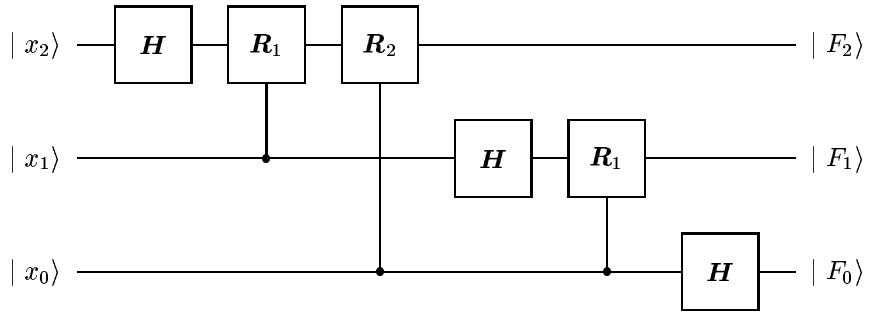
$$e^{2\pi i1(.x_0x_1\ldots x_{n-3})}\mid 1\rangle = e^{2\pi i(.x_0x_1\ldots x_{n-3})}\mid 1\rangle$$

The sum over all possible values of $\boldsymbol{y}$ will eventually assign both 0 and 1 to every y_k, therefore the following superposition is equivalent to the above:

$$\begin{aligned} \boldsymbol{F}\mid \boldsymbol{x}\rangle &= \frac{1}{\sqrt{2}}\left(\mid 0\rangle+e^{2\pi i(.x_0x_1\ldots x_{n-1})}\mid 1\rangle\right)\otimes\frac{1}{\sqrt{2}}\left(\mid 0\rangle+e^{2\pi i(.x_0x_1\ldots x_{n-2})}\mid 1\rangle\right)\otimes\cdots \\ &\quad \cdots\otimes\frac{1}{\sqrt{2}}\left(\mid 0\rangle+e^{2\pi i(.x_0)}\mid 1\rangle\right) \end{aligned} \tag{5.2}$$

And this already points to the way we can implement a QFT circuit.

Consider the following circuit:



Here, as before, $\boldsymbol{H}$ is the Hadamard operator and $\boldsymbol{R}_d$ is a controlled gate defined by:

$$\boldsymbol{R}_d = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/2^d} \end{pmatrix}$$

where d is the *distance* between the lines.

Let us analyze this circuit step by step:

1. After the first Hadamard gate the top line becomes

$$\frac{1}{\sqrt{2}}\sum_{y=0}^{1}(-1)^{x_2 y} \mid y\rangle = \frac{1}{\sqrt{2}}\sum_{y=0}^{1} e^{2\pi i x_2 y/2} \mid y\rangle = \frac{1}{\sqrt{2}}\left(\mid 0\rangle + e^{2\pi i(.x_2)} \mid 1\rangle\right)$$

2. The second step applies $\boldsymbol{R}_1$ to the top line under the control of the middle line. Observe that $\boldsymbol{R}_1$ does nothing to $\mid 0\rangle$ and phase shifts $\mid 1\rangle$. The phase shift factor is $e^{i\pi/2}$ if the control line $\mid x_1\rangle$ is $\mid 1\rangle$ and there is no phase shift if $\mid x_1\rangle = \mid 0\rangle$. We can therefore write that the phase shift inflicted by $\boldsymbol{R}_1$ on $\mid 1\rangle$ is *always* $e^{i\pi x/2}$, where x is the control signal.

   Applying this to states on the top and on the middle line yields

$$\begin{aligned}
&\boldsymbol{R_1} \mid x_1\rangle \otimes \frac{1}{\sqrt{2}}\left(\mid 0\rangle + e^{2\pi i(.x_2)} \mid 1\rangle\right) \\
&= \mid x_1\rangle \otimes \frac{1}{\sqrt{2}}\left(\mid 0\rangle + e^{2\pi i(.x_2)} e^{i\pi x_1/2} \mid 1\rangle\right) \\
&= \mid x_1\rangle \otimes \frac{1}{\sqrt{2}}\left(\mid 0\rangle + e^{2\pi i(x_2/2 + x_1/4)} \mid 1\rangle\right) \\
&= \mid x_1\rangle \otimes \frac{1}{\sqrt{2}}\left(\mid 0\rangle + e^{2\pi i(.x_1 x_2)} \mid 1\rangle\right)
\end{aligned}$$

3. The third gate applies $\boldsymbol{R}_2$ to the top line, but this time under the control of the bottom line. This operator, again, will do nothing to $\mid 0\rangle$, but will phase shift $\mid 1\rangle$ by additional $e^{i\pi x_0/4}$ so the state of the whole system now becomes:

$$\boldsymbol{R_2} \mid x_0\rangle \otimes \mid x_1\rangle \otimes \frac{1}{\sqrt{2}}\left(\mid 0\rangle + e^{2\pi i(.x_1 x_2)} \mid 1\rangle\right)$$

$$= \mid x_0\rangle \otimes \mid x_1\rangle \otimes \frac{1}{\sqrt{2}}\Big(\mid 0\rangle + e^{2\pi i(.x_1x_2)}e^{\pi i x_0/4} \mid 1\rangle\Big)$$
$$= \mid x_0\rangle \otimes \mid x_1\rangle \otimes \frac{1}{\sqrt{2}}\Big(\mid 0\rangle + e^{2\pi i(x_0/8+x_1/4+x_2/2)} \mid 1\rangle\Big)$$
$$= \mid x_0\rangle \otimes \mid x_1\rangle \otimes \frac{1}{\sqrt{2}}\Big(\mid 0\rangle + e^{2\pi i(.x_0x_1x_2)} \mid 1\rangle\Big)$$

4. Reasoning as above we can see immediately that the next two gates applied to $\mid x_1\rangle$ will convert it into

$$\frac{1}{\sqrt{2}}\Big(\mid 0\rangle + e^{2\pi i(.x_0x_1)} \mid 1\rangle\Big)$$

So that now the state of the computer is:

$$\mid x_0\rangle \otimes \frac{1}{\sqrt{2}}\Big(\mid 0\rangle + e^{2\pi i(.x_0x_1)} \mid 1\rangle\Big) \otimes \frac{1}{\sqrt{2}}\Big(\mid 0\rangle + e^{2\pi i(.x_0x_1x_2)} \mid 1\rangle\Big)$$

5. And finally the single Hadamard transform on the bottom line converts $\mid x_0\rangle$ to $(\mid 0\rangle + e^{2\pi i(.x_0)} \mid 1\rangle)/\sqrt{2}$, so that in effect the final state of the computer is:

$$\frac{1}{\sqrt{2}}\Big(\mid 0\rangle + e^{2\pi i(.x_0)} \mid 1\rangle\Big) \otimes \frac{1}{\sqrt{2}}\Big(\mid 0\rangle + e^{2\pi i(.x_0x_1)} \mid 1\rangle\Big) \otimes \frac{1}{\sqrt{2}}\Big(\mid 0\rangle + e^{2\pi i(.x_0x_1x_2)} \mid 1\rangle\Big)$$

But this is a 3-point Quantum Fourier Transform, so the circuit shown above is a QFT circuit. Furthermore the analysis of the circuit shows also how to make a 4-point QFT circuit and an n-point QFT circuit in general.

5.3.3 Finding the Period, The Shor Oracle

Let us now replace the last column of Hadamards in the Simon Oracle with the Quantum Fourier Transform box and see what we can do with this new circuit. This question was first posed by Peter Shor, who demonstrated, as you will see below, that the resulting oracle can be used to find period of an integer function

$$f : \mathbb{N} \ni x \mapsto f(x) \in \mathbb{N}$$

This time both the function and its period are defined in a normal way without any modulo-2 hocus-pocus. We are in the world of normal integer arithmetic now (though still restricted to what we can possibly pack into an n-bit register). If the period of the function is r then

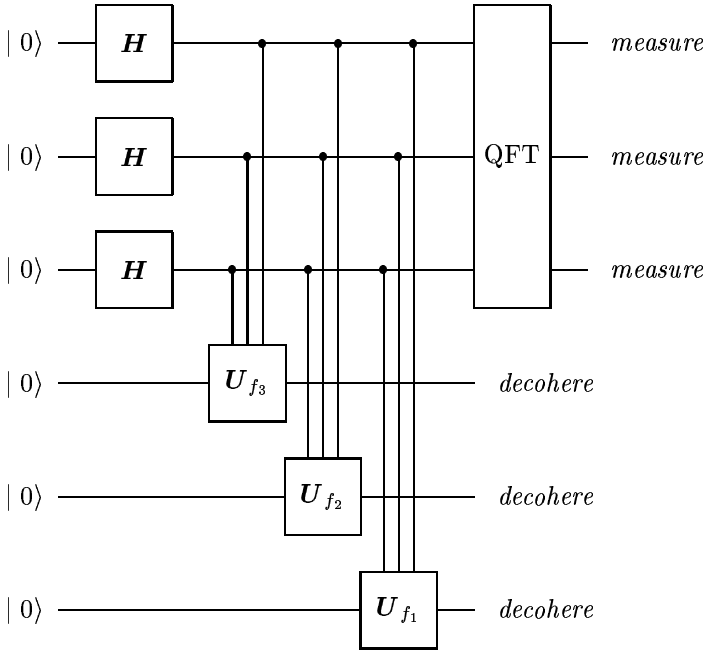
$$f(x) = f(x + kr),$$

where k is an integer.

In order to throw this problem on a computer (classical or quantum), we have to restrict our input and output values to a finite number of bits (or qubits), and we have to rewrite function f as a binary function thusly:

$$f : \{0,1\}^n \rightarrow \{0,1\}^n, \quad r \in [1, 2^n]$$

The figure below shows the new circuit, the Shor Oracle, which is the Simon Oracle with a QFT box replacing Hadamards in the upper right corner of the circuit.



Let us analyze this circuit then. The beginning of the analysis goes exactly the same way it did for the Simon circuit. After we cross the first array of the Hadamard gates we generate the superposition of all integer numbers between 0 and $2^n - 1$ on the top lines:

$$\frac{1}{2^{n/2}} \sum_{\boldsymbol{x}=\boldsymbol{0}}^{2^n-1} \mid \boldsymbol{x}\rangle$$

and then when we pass the array of $\boldsymbol{U}_{f_k}$ gates on the bottom lines we generate $f(\boldsymbol{x})$ for every $\mid \boldsymbol{x}\rangle$, so that the state of the computer is:

$$\frac{1}{2^{n/2}} \sum_{\boldsymbol{x}=\boldsymbol{0}}^{2^n-1} \mid \boldsymbol{x}\rangle \otimes \mid f(\boldsymbol{x})\rangle,$$

where

$$\mid f(\boldsymbol{x})\rangle = \mid f_{n-1}(\boldsymbol{x})\rangle \otimes \mid f_{n-2}(\boldsymbol{x})\rangle \otimes \cdots \otimes \mid f_0(\boldsymbol{x})\rangle$$

Now we allow the bottom lines to decohere, as we did in oracles discussed in the previous section, and we end up with some value of f that corresponds to some x_0. But the same value also corresponds to x_0+r and x_0+2r and so on. Because we haven't touched any of the top lines yet, the decoherence of the bottom lines forces the upper lines, via the EPR mechanism into the corresponding state, so that the state of the whole computer becomes:

$$\frac{1}{\sqrt{A}}\left(\sum_{j=0}^{A-1} \mid x_0 + jr\rangle\right) \otimes \mid f(x_0)\rangle,$$

where A is such an integer that we don't run outside the $[0, 2^n]$ segment as we jump from x_0 to $x_0 + r$, then to $x_0 + 2r$, and so on. Of course we can also jump from x_0 to $x_0 - r$, so it is good to think of x_0 as the lowest $x_0 \in [0, 2^n]$.

At this stage we can forget about the bottom lines and concentrate on what happens on the top n lines. The state of those lines is now passed through the QFT box, which yields:

$$\begin{aligned}
\boldsymbol{F}\left(\frac{1}{\sqrt{A}}\sum_{j=0}^{A-1} \mid x_0 + jr\rangle\right) &= \frac{1}{\sqrt{A2^n}}\sum_{\boldsymbol{y}=\boldsymbol{0}}^{2^n-1}\sum_{j=0}^{A} e^{2\pi i(x_0+jr)y/2^n} \mid \boldsymbol{y}\rangle \\
&= \frac{1}{\sqrt{A2^n}}\sum_{\boldsymbol{y}=\boldsymbol{0}}^{2^n-1} e^{2\pi i x_0 y/2^n}\sum_{j=0}^{A} e^{2\pi i j r y/2^n} \mid \boldsymbol{y}\rangle
\end{aligned}$$

When we perform a measurement at this stage the probability of finding a given $\mid \boldsymbol{y}\rangle$ is given by the square of an appropriate amplitude, i.e.,

$$\frac{A}{2^n}\left|\frac{1}{A}\sum_{j=0}^{A} e^{2\pi i j r y/2^n}\right|^2$$

Now, assume that period r divides 2^n exactly, i.e., without a remainder. Then $A = 2^n/r$ and $A/2^n = 1/r$. The probability amplitude then becomes:

$$\frac{1}{r}\left|\frac{1}{A}\sum_{j=0}^{A} e^{2\pi i j y/A}\right|^2$$

For $y = A$ this becomes:

$$\begin{aligned}
&\frac{1}{r}\left|\frac{1}{A}\left(e^{2\pi i 0} + e^{2\pi i 1} + e^{2\pi i 2} + \ldots\right)\right|^2 \\
&= \frac{1}{r}\left|\frac{1}{A}\left(1 + 1 + 1 + \ldots\right)\right|^2 \\
&= \frac{1}{r}\left|\frac{1}{A}A\right|^2 = \frac{1}{r}
\end{aligned}$$

For $y = 2A$ we get

$$\begin{aligned}
&\frac{1}{r}\left|\frac{1}{A}\left(e^{2\pi i 0} + e^{2\pi i 2} + e^{2\pi i 4} + \ldots\right)\right|^2 \\
&\quad = \frac{1}{r}\left|\frac{1}{A}(1 + 1 + 1 + \ldots \quad \text{as many as before!})\right|^2 \\
&\quad = \frac{1}{r}\left|\frac{1}{A}A\right|^2 = \frac{1}{r}
\end{aligned}$$

But if y is *incommensurate* with A then we'll hit a range of points around the circle, eventually filling the whole circle, so that the resulting interference will be totally destructive. In effect the measurement will return

$$y \in \{A, 2A, 3A, \ldots, rA\}$$

From this we can easily find A and knowing the range 2^n we can easily find the period $r = 2^n/A$.

What if r does not divide 2^n so that there is a bit less than r left? In that case y will get scattered around integer multiples of $2^n/r$ and the width of the peaks is going to be about 1. This still gives us a sufficiently precise idea about r, to find its exact value by trial and error without much additional computational effort.

We are going to discuss this case in more detail in section about Phase Estimation further down.

5.3.4 Breaking the RSA Encryption

The period finding procedure is the only quantum computation required to break the RSA Public Encryption Key. In this section we are going to discuss how this is done. This is not very enlightening from the Quantum Computing point of view, because whatever is of interest there we have covered already, but since this is what made Quantum Computing such a popular subject all of a sudden, and what also resulted in the injection of a significant amount of money into this area, we feel that we should cover this topic for completeness.

However we are not going to talk about it in the lecture, because we're running out of time, so this section is for you to read at your leasure.

The RSA Public Key Cryptography

The RSA Public Key Cryptography was invented by Ronald Rivest, Adi Shamir, and Leonard Adelman in 1978. The trick here is that encrypted messages can be passed from the sender to the receiver, and they can be decrypted by the receiver, without having to pass a secret decryption key between them.

Instead a public/private key pair is used. The public key, which can be safely published for all to know, is used to encrypt the message. The private

key, which is held by the owner, and which is never shown to anybody, is used to decrypt the message.

We can describe this process symbolically as follows:

$$\begin{aligned} E\,(m, k_{\text{public}}) &= m' \\ D\,(m', k_{\text{private}}) &= m \end{aligned}$$

where E is the encrypting function, D is the decryption function, m is the message, m' is the encrypted message, k_{public} is the public key and k_{private} is the private key.

In order for the public key encryption scheme to work, the private and the public keys must be linked in some way, but this way must be such that knowing the public key makes us none the wiser as to what the private key might be.

The RSA algorithm links both keys in the following way:

- pick two large prime numbers p and q and let $n = p \cdot q$;
- find a random integer d that is co-prime with $(p-1) \cdot (q-1)$. *Coprime* means that the Greatest Common Divisor of d and $(p-1) \cdot (q-1)$ is 1;
- let e be a modular inverse of d, i.e., $e \cdot d|_{\text{mod}\,(p-1)(q-1)} = 1$;
- the pair (e, n) becomes a public key now, and
- the pair (d, n) becomes a private key.
- the encryption rule is: $m'_i = m_i^e|_{\text{mod}\,n}$
- the decryption rule is: $m_i = m_i'^d|_{\text{mod}\,n}$

How difficult is breaking the RSA crypt going to be? Since the public key (e, n) is known, n is known too. Recall that $n = p \cdot q$. If we could find p and q, we could find $p-1$ and $q-1$ as well. Then since $e \cdot d|_{\text{mod}\,(p-1)(q-1)} = 1$, and since we know e we can find d and this is all that's needed to decrypt the message.

But assuming that n is really, very, very long, how hard can it be to find p and q such that $p \cdot q = n$? This problem is called *integer factoring*.

Although integer factoring seems to be a pretty trivial problem, it turns out that it is extremely hard. There is an algorithm for doing this called a Number Field Sieve algorithm. Let x be the number of bits needed to encode n. Then the time to factor n using the Number Field Sieve algorithm is proportional to

$$e^{\alpha x^{1/3} (\ln x)^{2/3}}$$

The problem, as you see, scales sub-exponentially. This makes it hard, because the scaling is faster than polynomial.

To see how hard this can be consider the challenge, which Rivest, Shamir and Adleman issued to the computer community in 1977. They challenged them to factor a 129-digit integer number. Eventually it took computer scientists 17

years to rise to the challenge. The number was eventually factored on a cluster of 1,600 computers.

What if the number is longer? According to Vazirani,

> *if every particle in the Universe was a classical computer running at full speed for the entire life of the Universe so far (about 12 billion years) that would be still insufficient to factor a 2,000 digits number.*

Breaking the RSA Crypt Quantumly

But Shor's oracle makes it possible to break the RSA encryption key quantumly in polynomial time. The trick that makes it possible is as follows. Let x is coprime with n. Define a function

$$f_n(a) = x^a|_{\mathrm{mod}\, n}$$

This function is periodic with the period being a function of x. If for a given x the period is r then

$$f_n(a+r) = x^{(a+r)}|_{\mathrm{mod}\, n} = x^a x^r|_{\mathrm{mod}\, n}$$

which implies that

$$x^r|_{\mathrm{mod}\, n} = 1$$

Now assume that r is even. Then

$$\begin{aligned} \left(x^{r/2}\right)^2 &= 1|_{\mathrm{mod}\, n} \quad \text{hence} \\ \left(x^{r/2}\right)^2 - 1^2 &= 0|_{\mathrm{mod}\, n} \quad \text{hence} \\ \left(x^{r/2}-1\right)\left(x^{r/2}+1\right) &= 0|_{\mathrm{mod}\, n} \end{aligned}$$

This implies that unless $x^{r/2} = \pm 1|_{\mathrm{mod}\, n}$ either $\left(x^{r/2}-1\right)$ or $\left(x^{r/2}+1\right)$ must have a nontrivial factor in common with n. The factor of n can then be found by calculating the Greatest Common Divisor of either $\left(x^{r/2}-1, n\right)$ or $\left(x^{r/2}+1, n\right)$.

In summary,

> *we have reduced the problem of factoring n to the problem of finding a period of $f_n(a) = x^a|_{\mathrm{mod}\, n}$, where x is co-prime with n. And to find the period we can use the Shor oracle.*

The only quantum question that remains is how to implement function $f_n(a)$ using quantum computing elements.

If $a < 2^n$ then we can decompose a into powers of 2:

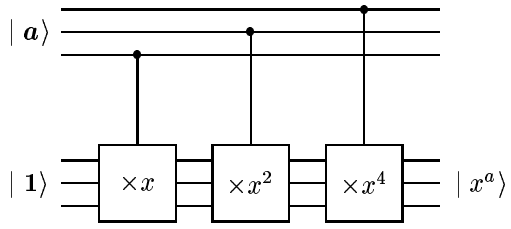
$$a = a_0 + a_1 2 + a_2 2^2 + \ldots + a_{n-1} 2^{n-1}$$

Making use of this decomposition we can now express x^a as follows:

$$\begin{aligned} x^a &= x^{a_0} x^{a_1 2} x^{a_2 2^2} \cdots x^{a_{n-1} 2^{n-1}} \\ &= x^{a_0} \left(x^2\right)^{a_1} \left(x^{2^2}\right)^{a_2} \cdots \left(x^{2^{n-1}}\right)^{a_{n-1}} \end{aligned}$$

where $a_k \in \{0,1\}$ – so that they can be thought of as control lines: if $a_k = 0$ then the corresponding $\left(x^{2^k}\right)^{a_k} = 1$. The powers of x^{2^k} can be evaluated classically by repeated squaring.

The resulting quantum circuit, which evaluates $\mid x^a\rangle$ looks as follows:



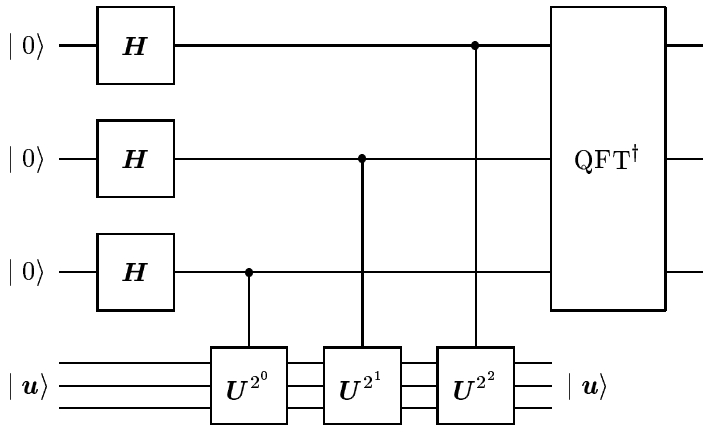
5.3.5 Phase Estimation

The period finding circuit is a special example of a broader category of circuits, all of which fall under the label of *Phase Estimation* circuits.

Let $\boldsymbol{U}$ be a unitary operator and let $\mid \boldsymbol{u}\rangle$ be its eigenvector such that

$$\boldsymbol{U} \mid \boldsymbol{u}\rangle = e^{2\pi i\phi} \mid \boldsymbol{u}\rangle$$

Consider the following circuit



where gates $\boldsymbol{U}^{2^k}$ with $k = 0, 1, 2$ are all controlled by the corresponding lines. Observe the similarity to the circuit we have drawn in the previous section.

Because $\boldsymbol{u}$ is the eigenvector of $\boldsymbol{U}$ once we have fed $\boldsymbol{u}$ into the 3 bottom lines, it stays there unchanged. This, courtesy of the EPR paradox, forces all the action into the top 3-qubit register. The Hadamards rotate $\mid 0\rangle$, applied to the input of each top register line, to $(\mid 0\rangle + \mid 1\rangle)/\sqrt{2}$. And so after the first $\boldsymbol{U}^{2^0}$ gate has been traversed, the combined state of line 0 of the top register and of the bottom register is

$$\frac{1}{\sqrt{2}}\left(\mid 0\rangle \mid \boldsymbol{u}\rangle + \mid 1\rangle e^{2\pi i\phi 2^0} \mid \boldsymbol{u}\rangle\right) = \frac{1}{\sqrt{2}}\left(\mid 0\rangle + e^{2\pi i\phi 2^0} \mid 1\rangle\right) \mid \boldsymbol{u}\rangle$$

Immediately before the application of the second $\boldsymbol{U}^{2^1}$ gate the state of the zeroth and first lines of the top register and of the bottom register is

$$\frac{1}{\sqrt{2}}(\mid 0\rangle + \mid 1\rangle)\,\frac{1}{\sqrt{2}}\left(\mid 0\rangle + e^{2\pi i\phi 2^0} \mid 1\rangle\right) \mid \boldsymbol{u}\rangle$$

When the second gate is traversed, this state changes to:

$$\frac{1}{\sqrt{2}}\left(\mid 0\rangle + e^{2\pi i\phi 2^1} \mid 1\rangle\right)\frac{1}{\sqrt{2}}\left(\mid 0\rangle + e^{2\pi i\phi 2^0} \mid 1\rangle\right) \mid \boldsymbol{u}\rangle$$

In a similar manner we can already see that after the third gate, $\boldsymbol{U}^{2^2}$ is traversed the state of the computer becomes:

$$\frac{1}{\sqrt{2}}\left(\mid 0\rangle + e^{2\pi i\phi 2^2} \mid 1\rangle\right)\frac{1}{\sqrt{2}}\left(\mid 0\rangle + e^{2\pi i\phi 2^1} \mid 1\rangle\right)\frac{1}{\sqrt{2}}\left(\mid 0\rangle + e^{2\pi i\phi 2^0} \mid 1\rangle\right) \mid \boldsymbol{u}\rangle$$

Remember that the products between vectors in this formula are all *tensor* products.

Suppose that ϕ in the formula above can be represented by the following expression *exactly*:

$$\phi = \frac{\phi_0}{2^3} + \frac{\phi_1}{2^2} + \frac{\phi_2}{2^1} = (.\phi_0\phi_1\phi_2)$$

where $\phi_0, \phi_1, \phi_2 \in \{0, 1\}$. Then

$$e^{2\pi i\phi 2^0} = e^{2\pi i\left(\phi_0/2^3 + \phi_1/2^2 + \phi_2/2^3\right)} = e^{2\pi i(.\phi_0\phi_1\phi_2)}$$

and

$$e^{2\pi i\phi 2^1} = e^{2\pi i\left(\phi_0/2^2 + \phi_1/2^1 + \phi_2\right)}$$

But because ϕ_2 is either 0 or 1, $e^{2\pi i\phi_2} = 1$, so we can drop it from the formula, which therefore yields:

$$e^{2\pi i\phi 2^1} = e^{2\pi i\left(\phi_0/2^2 + \phi_1/2^1\right)} = e^{2\pi i(.\phi_0\phi_1)}$$

By similar reasoning:

$$e^{2\pi i\phi 2^2} = e^{2\pi i(.\phi_0)}$$

Substituting this to our expression that describes the circuit after the application of the three $\boldsymbol{U}$ gates yields

$$\frac{1}{\sqrt{2}}\left(\mid 0\rangle + e^{2\pi i(.\phi_0)} \mid 1\rangle\right)\frac{1}{\sqrt{2}}\left(\mid 0\rangle + e^{2\pi i(.\phi_0\phi_1)} \mid 1\rangle\right)\frac{1}{\sqrt{2}}\left(\mid 0\rangle + e^{2\pi i(.\phi_0\phi_1\phi_2)} \mid 1\rangle\right) \mid \boldsymbol{u}\rangle$$

Ah, but what we have here is exactly the formula we used to implement Quantum Fourier Transform as a circuit, given by equation (5.2) on page 198. The only difference is that this is the formula for $\boldsymbol{F} \mid \boldsymbol{\phi}'\rangle$, where $\phi' = \phi_0 2^0 + \phi_1 2^1 + \phi_2 2^2$. In summary, the state of the circuit just before its upper register passes through $\mathrm{QFT}^\dagger$ is:

$$\left(\boldsymbol{F} \mid \boldsymbol{\phi}'\rangle\right) \otimes \mid \boldsymbol{u}\rangle$$

It is now clear that applying $\mathrm{QFT}^\dagger$, i.e., $\boldsymbol{F}^{-1}$ to the upper register is going to yield $\{\phi_0, \phi_1, \phi_2\}$ in the upper register when the measurement is made. It is now up to us whether we want to read these as $\phi' = \phi_0 2^0 + \phi_1 2^1 + \phi_2 2^2$ or as $\phi = \phi_0/2^3 + \phi_1/2^2 + \phi_2/2^1$.

Now, I want to torture you a little with the analysis of a far more difficult case. What if $\phi \neq (.\phi_0\phi_1\phi_2)$? This case is similar to another case we skipped above: the case of a period of function f not fitting exactly into $\{1, \ldots, 2^n\}$. The analysis is difficult, because this time we have to abandon the purely algebraic approach that is really very easy to follow, although at times somewhat tedious. The approach instead is going to be characteristic of mathematical analysis. We're going to plunge into the subtle world of inequalities.

Suppose b is a t-bit integer described by a string of zeros and ones $\{b_0, b_1, \ldots, b_{t-1}\}$. Dividing it by 2^t yields

$$\frac{b}{2^t} = (.b_0 b_1 \ldots b_{t-1})$$

Now assume that $(.b_0 b_1 \ldots b_{t-1})$ is the best t-bit approximation to ϕ that's still less than ϕ, i.e.,

$$(.b_0 b_1 \ldots b_{t-1}) < \phi$$

Let the difference between ϕ and $(.b_0 b_1 \ldots b_{t-1})$ be δ. Because $(.b_0 b_1 \ldots b_{t-1})$ is the *best* approximation of ϕ:

$$\delta = \phi - (.b_0 b_1 \ldots b_{t-1}) = \phi - \frac{b}{2^t} < \frac{1}{2^t}$$

From the analysis of the Phase Estimation Oracle we know that just before the application of the inverse Fourier Transform the state of the upper register is going to be:

$$\frac{1}{\sqrt{2^t}} \sum_{k=0}^{2^t-1} e^{2\pi i \phi k} \mid \boldsymbol{k}\rangle$$

There is no division by 2^t in the exponens, because this is implicit in the definition of $\phi = \phi'/2^t$.

The inverse Fourier Transform is given by the formula:

$$\mid \boldsymbol{k}\rangle \mapsto \frac{1}{\sqrt{2^t}} \sum_{l=0}^{2^t-1} e^{-2\pi i k l/2^t} \mid \boldsymbol{l}\rangle$$

Acting with the latter on the former yields

$$\frac{1}{2^t} \sum_{k=0}^{2^t-1} \sum_{l=0}^{2^t-1} e^{2\pi i \phi k} e^{-2\pi i k l/2^t} \mid \boldsymbol{l}\rangle$$

Now the probability amplitude of measuring $\mid \boldsymbol{l}\rangle$ is

$$\frac{1}{2^t} \sum_{k=0}^{2^t-1} e^{-2\pi i k l/2^t} e^{2\pi i \phi k} = \frac{1}{2^t} \sum_{k=0}^{2^t-1} \left(e^{2\pi i (\phi - l/2^t)}\right)^k = \frac{1}{2^t} \sum_{k=0}^{2^t-1} q^k$$

where

$$q = e^{2\pi i(\phi - l/2^t)}$$

The above is a sum of a geometric series, for which we have the following expression:

$$S_n = \frac{1 - q^{n+1}}{1 - q}$$

where $n = \{0, 1, \ldots\}$, e.g., $S_0 = 1$ and $S_1 = 1 + q$. And so the probability amplitude of measuring $|\, l\rangle$ becomes

$$\frac{1}{2^t} S_{2^t - 1} = \frac{1}{2^t} \frac{1 - q^{2^t}}{1 - q}$$

Substituting our expression for q into q^{2^t} yields:

$$q^{2^t} = e^{2\pi i(2^t \phi - l)}$$

so that the probability amplitude is

$$\frac{1}{2^t} \frac{1 - e^{2\pi i(2^t \phi - l)}}{1 - e^{2\pi i(\phi - l/2^t)}}$$

If $b/2^t$ is the closest to ϕ then what is of special interest to us is the vicinity of $l = b$. Let us thus replace l in the above formula with $b + l$ and let us define:

$$\alpha_l = \frac{1}{2^t} \frac{1 - e^{2\pi i(2^t \phi - (b+l))}}{1 - e^{2\pi i(\phi - (b+l)/2^t)}}$$

with $l = 0, \pm 1, \pm 2, \ldots$. Our hope is that α_l should peak at $l = 0$ and that the peak should be sharp. We can rewrite this expression even further making use of our definition of $\delta = \phi - b/2^t$ to get:

$$\alpha_l = \frac{1}{2^t} \frac{1 - e^{2\pi i(2^t \delta - l)}}{1 - e^{2\pi i(\delta - l/2^t)}}$$

In order to assess if there is a peak there at all and how sharp it is if it exists let us try to estimate the

probability of finding upon a measurement that l is more than e steps to the left or to the right of 0*:*

$$P = \sum_{l \in]-2^{t-1}, -(e+1)]} |\alpha_l|^2 + \sum_{l \in [e+1, 2^{t-1}]} |\alpha_l|^2$$

The exact expression for α_l is too complicated to help us here, so we are going to replace it with estimates instead. We want to demonstrate that P is *less than* something. To do this we need to show that $|\alpha_l|$ is *less than* something else. Because α_l is a fraction, its absolute value is a fraction of absolute values of its numerator and denominator:

$$|\alpha_l| = \frac{1}{2^t} \frac{\left|1 - e^{2\pi i(2^t \delta - l)}\right|}{\left|1 - e^{2\pi i(\delta - l/2^t)}\right|}$$

To show that $|\alpha_l|$ is *less than* something we need to show that the numerator in the above formula is *less than* something. But at the same

time we need to show that the denominator in the formula is *greater than* something else. And these estimates need to be tight, but not any tighter than necessary to show what we want to show, i.e., they should result in an expression that's easy to use for our purposes.

The numerator is of the form

$$\left|1 - e^{i\theta}\right|$$

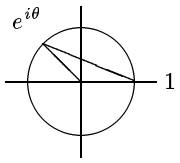
It assumes the largest value when $\theta = \pi$ and then $e^{i\theta} = -1$. So we can clearly see that:

$$\left|1 - e^{2\pi i(2^t\delta - l)}\right| \leq 2$$

Now let us have a look at the denominator. It is also of the form

$$\left|1 - e^{i\theta}\right|$$

but now we must find what it is *greater than* and here we can't be too blunt (it is obviously greater than 0, but this doesn't help). First observe that $1 - e^{i\theta}$ is a bowstring (or a chord) connecting the ends of the arc that corresponds to θ, see figure below:



$\left|1 - e^{i\theta}\right|$ is the length of the chord. The maximum length of the chord is 2 and it corresponds to $\theta = \pi$. The chord is always shorter than the arc, but it is always longer than $2 \times \text{arc}/\pi$. Because the radius of the circle here is 1, the length of the arc is simply equal to the angle θ, so we have the estimate:

$$\left|1 - e^{i\theta}\right| \geq \frac{2|\theta|}{\pi}$$

with $\theta \in [0, \pi]$.

For $\theta = \pi$ we get the equality. Then as the angle diminishes the left hand side becomes increasingly larger than the right hand side. For example for $\theta = \pi/2$ the length of the chord is $\sqrt{2} = 1.4142\ldots$, whereas the right hand side evaluates to $2\pi/(2\pi) = 1$. For very small angles, the length of the chord is θ and the right hand side is $2\theta/\pi = 0.6366\theta$, so the length of the chord is 1.57 times larger than the right hand side of our estimate.

The particular angle we're really interested in is

$$\theta = 2\pi\left(\delta - l/2^t\right)$$

hence

$$\left|1 - e^{i\theta}\right| \geq \frac{2}{\pi} 2\pi \left|\delta - l/2^t\right| = 4\left|\delta - l/2^t\right|$$

Making use of our estimates for the numerator and for the denominator we can now state that:

$$|\alpha_l| \leq \frac{1}{2^t}\frac{2}{4|\delta - l/2^t|} = \frac{1}{2^t}\frac{1}{2|\delta - l/2^t|} = \frac{1}{2^{t+1}}\frac{1}{|\delta - l/2^t|} = \frac{1}{2}\frac{1}{|2^t\delta - l|}$$

The probability of getting further to the right or to the left of b than e steps is therefore:

$$P \leq \frac{1}{4} \sum_{l \in]-2^{t-1}, -(e+1)]} \frac{1}{(2^t\delta - l)^2} + \frac{1}{4} \sum_{l \in [e+1, 2^{t-1}]} \frac{1}{(2^t\delta - l)^2}$$

Now, recall the definition of δ:

$$\delta = \phi - (.b_0 b_1 \ldots b_{t-1}) = \phi - \frac{b}{2^t} < \frac{1}{2^t}$$

therefore

$$0 < 2^t\delta < 1$$

and for positive values of l

$$\left(2^t\delta - l\right)^2 = \left(l - 2^t\delta\right)^2 > (l-1)^2$$

hence

$$\frac{1}{(2^t\delta - l)^2} < \frac{1}{(l-1)^2}$$

For negative values of l

$$\left(2^t\delta - l\right)^2 > l^2$$

hence for negative values of l

$$\frac{1}{(2^t\delta - l)^2} < \frac{1}{l^2}$$

Now our estimate for P simplifies further to:

$$P < \frac{1}{4} \sum_{l \in]-2^{t-1}, -(e+1)]} \frac{1}{l^2} + \frac{1}{4} \sum_{l \in [e+1, 2^{t-1}]} \frac{1}{(l-1)^2}$$

We can replace the $l-1$ term in the second sum simply with l by renaming the index of the sum: instead of running from $e+1$ to 2^{t-1} we're going to run it from e to $2^{t-1}-1$, and so:

$$P < \frac{1}{4} \sum_{l=-2^{t-1}+1}^{-(e+1)} \frac{1}{l^2} + \frac{1}{4} \sum_{l=e}^{2^{t-1}-1} \frac{1}{l^2}$$

Let us now extend the upper range of the first sum to e. This does not change the inequality, because by doing this we increment the value on the right hand side even further. But this will help us wrap this expression into a single sum, because $1/l^2 = 1/(-l)^2$, and so we get:

$$P < \frac{1}{2} \sum_{l=e}^{2^{t-1}-1} \frac{1}{l^2}$$

The sum on the right hand side is less than the integral

$$\int_e^\infty \frac{1}{(l-1)^2}\,\mathrm{d}l = \int_{e-1}^\infty \frac{1}{l'^2}\,\mathrm{d}l' = -\frac{1}{l'}\Big|_{e-1}^\infty = \frac{1}{e-1}$$

Here we have replaced $1/l^2$ in the sum with $1/(l-1)^2$ in the integral, in order to obtain the estimate from above. The replacement has the effect of shifting the integrated function to the right by one unit on the x axis, so as to be *above* the rectangles represented by the discrete sum. We wouldn't have to do it if the function was a rising one.

In summary:

$$P < \frac{1}{2(e-1)}$$

The probability of straying further away from b than e steps drops *faster than* $1/(2(e-1))$. There is clearly an interference peak in the vicinity of b.

This result brings home the importance of running quantum computations on statistical ensembles of quantum computers. If you have, say, 10 million registers in your sample, all performing the same computation, you will see a very well defined distribution upon the measurement, which will be easy to read and analyze. If instead you have just one quantum register, you may get almost anything on the final measurement, because $1/(2(e-1))$ isn't really a very sharp peak (for $e = 2, 3, 4, 5$ the probability is less than $1/2, 1/4, 1/6, 1/8$). You will have to repeat the computation and the measurement over and over, before you get a well defined distribution.

5.3.6 Discrete Logarithms

Consider the function

$$f(x_1, x_2) = a^{sx_1+x_2} \bmod N$$

where $x_1, x_2, s, a, N \in \mathbb{N}$ and r is the smallest positive integer such that

$$a^r \bmod N = 1$$

Observe that

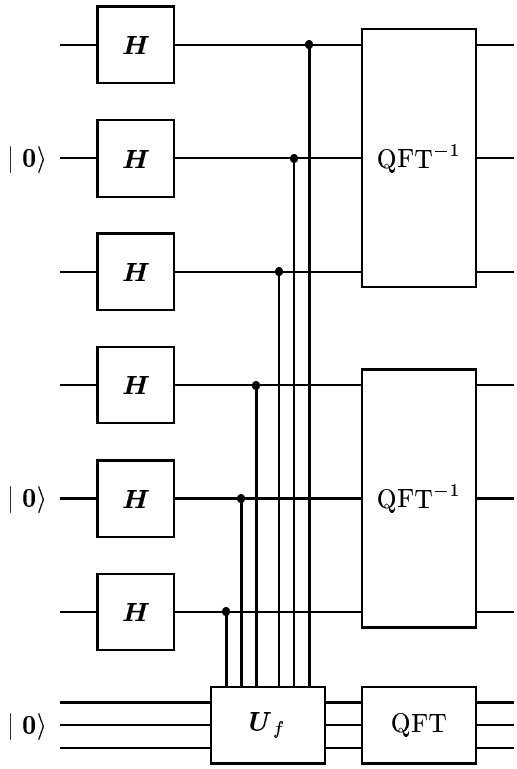
$$f(x_1 + l, x_2 - sl) = a^{s(x_1+l)+x_2-sl} = a^{sx_1+sl+x_2-sl} = a^{sx_1+x_2} = f(x_1, x_2)$$

So a pair $(l, -sl)$ constitutes a period of this two-argument function.

The question we would like to ask now is

given a and $b = a^s \bmod N$ what is s?

The oracle to answer this question is drawn below:



The top six lines correspond here to x_1 and x_2. The bottom three lines and the box on them implement

$$f(x_1, x_2) = b^{x_1} a^{x_2} = (a^s)^{x_1} a^{x_2} = a^{sx_1 + x_2}$$

After the column of Hadamard gates is traversed the state of the computer becomes:

$$\left(\frac{1}{\sqrt{2^t}} \sum_{x_1=0}^{2^t-1} \frac{1}{\sqrt{2^t}} \sum_{x_2=0}^{2^t-1} \mid \boldsymbol{x}_1 \rangle \mid \boldsymbol{x}_2 \rangle \right) \mid \mathbf{0} \rangle$$

Then we apply the $\boldsymbol{U}_f$ gate to the bottom register and the state of the computer becomes:

$$\frac{1}{\sqrt{2^t}} \sum_{x_1=0}^{2^t-1} \frac{1}{\sqrt{2^t}} \sum_{x_2=0}^{2^t-1} \mid \boldsymbol{x}_1 \rangle \mid \boldsymbol{x}_2 \rangle \mid a^{sx_1 + x_2} \rangle$$

Now let us rewrite $f(x_1, x_2)$ in terms of its own Fourier Transform:

$$\mid f(x_1, x_2) \rangle = \frac{1}{\sqrt{r}} \sum_{l=0}^{r-1} e^{2\pi i (sx_1 + x_2) l / r} \mid \hat{f}(sl, l) \rangle$$

We can therefore rewrite the state of the computer as

$$\frac{1}{2^t}\sum_{x_1=0}^{2^t-1}\sum_{x_2=0}^{2^t-1} \mid \boldsymbol{x}_1\rangle \mid \boldsymbol{x}_2\rangle \frac{1}{\sqrt{r}}\sum_{l=0}^{r-1} e^{2\pi i(sx_1+x_2)l/r} \mid \hat{f}(sl,l)\rangle$$
$$= \frac{1}{2^t}\frac{1}{\sqrt{r}}\sum_{l=0}^{r-1}\left(\sum_{x_1=0}^{2^t-1} e^{2\pi i s x_1 l/r} \mid \boldsymbol{x}_1\rangle\right)\left(\sum_{x_2=0}^{2^t-1} e^{2\pi i x_2 l/r} \mid \boldsymbol{x}_2\rangle\right) \mid \hat{f}(sl,l)\rangle$$

Applying QFT^{-1} to the top two register now yields:

$$\frac{1}{2^t}\frac{1}{\sqrt{r}}\sum_{l=0}^{r-1} \mid \frac{sl}{r}\rangle \mid \frac{l}{r}\rangle \mid \hat{f}\rangle$$

Measurement of the two upper registers is going to return a pair

$$\left\{\frac{sl}{r}, \frac{l}{r}\right\}$$

from which s can be readily obtained.

5.3.7 The Hidden Subgroup Problem

An astute observer would have noticed by now that all quantum algorithms presented so far, with the notable exception of the Brassard Teleportation Circuit, are very similar to each other. There is always a bottom register, with some function being evaluated on the output of that register. A superposition of all possible arguments is fed into the function box. The output of the function is allowed to decohere, which shifts some result back onto the upper register or a group of registers. Filters are then applied to the upper registers in order to extract information about some properties of the function. We have studied only two filters so far: the Hadamard operator and the Quantum Fourier Transform.

It is possible to rephrase the above in mathematical terms. We can think of the function as characterizing cosets of some subgroup K of some group G. The function varies from a coset to a coset, but is constant on any given coset. Thus the decoherence of the bottom register places the whole coset in the upper register. Through filtration and repetitive measurements we can then generate the subgroup K on the output of the upper register.

The algorithms described so far were therefore examples of procedures designed to crack a specific *hidden subgroup problem*. This is what made them all so similar.

5.4 Quantum Database Search

Imagine that you have a telephone number and a telephone book and that you're trying to find the person in the book the number belongs to. The problem is hard

because listings in the telephone book are ordered alphabetically, and not by number. The numbers themselves appear quite randomized, with the exception perhaps of the leading 3 digits or so which may be the same for a large number of people in a given locality.

This problem is not uncommon. In more general terms, if you have a very large unsorted data base with $N \gg 1$ items and if you need to locate an item with specific narrowly defined characteristics, then this is the kind of problem we're going to consider in this section.

In simpler terms, suitable for a quantum computation, we can define a characteristic function $f(x) : x \in \{0, 1, \ldots, N-1\} \mapsto \{0, 1\}$ such that for just one x_a $f(x_a) = 1$ and for all other x it is 0. The task of our quantum algorithm will be to find x_a. This is like solving an equation, but function f may not be given by any specific formula such as $x^2 + 2x - 1$, rather f may be defined arbitrarily, e.g., as a table. In this case one would normally have to search the whole table to find the corresponding x_a.

5.4.1 The State Marker

This time we are going to construct the oracle step by step rather than draw the whole circuit from the beginning and then analyze it.

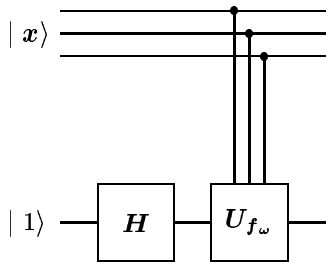
This problem is not quite like the hidden subgroup problem discussed in the previous section, but it has a sufficient number of common elements with it. And so our first step is to construct a controlled unitary gate that would implement function

$$f_\omega(x) = \begin{cases} 0 & \text{if} \quad x \neq \omega \\ 1 & \text{if} \quad x = \omega \end{cases}$$

The gate itself can be described by the formula:

$$\boldsymbol{U}_{f_\omega} : |\, \boldsymbol{x}\rangle \,|\, y\rangle \mapsto |\, \boldsymbol{x}\rangle \,|\, y +_2 f_\omega(x)\rangle$$

Now consider the following circuit. This is not the whole oracle. It is just one of its elements, but an important one.



Let us analyze it:

1. The Hadamard operator on the bottom line converts $|\, 1\rangle$ to $(|\, 0\rangle - |\, 1\rangle)/\sqrt{2}$

2. As we have seen before in the section about the Deutsch-Jozsa oracle, when a $+_2$ operator controlled by a function f_ω of $\mid \boldsymbol{x}\rangle$ acts on $(\mid 0\rangle - \mid 1\rangle)/\sqrt{2}$ the following results:

$$\boldsymbol{U}_{f_\omega} \mid \boldsymbol{x}\rangle \frac{1}{\sqrt{2}} (\mid 0\rangle - \mid 1\rangle) = (-1)^{f_\omega(x)} \mid \boldsymbol{x}\rangle \frac{1}{\sqrt{2}} (\mid 0\rangle - \mid 1\rangle)$$

3. Now consider $(-1)^{f_\omega(x)} \mid \boldsymbol{x}\rangle$. If $x = \omega$ then $f_\omega(x) = 1$, otherwise it's zero. Consequently for all $x \neq \omega$

$$(-1)^{f_\omega(x)} \mid \boldsymbol{x}\rangle = (-1)^0 \mid \boldsymbol{x}\rangle = \mid \boldsymbol{x}\rangle$$

but for the single $x = \omega$

$$(-1)^{f_\omega(\omega)} \mid \boldsymbol{\omega}\rangle = (-1)^1 \mid \boldsymbol{\omega}\rangle = - \mid \boldsymbol{\omega}\rangle$$

Given that $\mid \boldsymbol{\omega}\rangle$ is one of the basis states and that it's length is 1 we can sum up the above as follows:

$$(-1)^{f_\omega(x)} \mid \boldsymbol{x}\rangle = (\mathbf{1} - 2 \mid \boldsymbol{\omega}\rangle\langle\boldsymbol{\omega} \mid) \mid \boldsymbol{x}\rangle$$

4. The resulting state of the computer can therefore be written as:

$$(\mathbf{1} - 2 \mid \boldsymbol{\omega}\rangle\langle\boldsymbol{\omega} \mid) \mid \boldsymbol{x}\rangle \otimes \frac{1}{\sqrt{2}} (\mid 0\rangle - \mid 1\rangle)$$

Now suppose that we place a column of Hadamard operators in the upper register in front of this gate arrangement and generate a superposition of all values x can assume from 0 through $N - 1$. With this simple device we will have accomplished an important result: we will have labelled $\mid \boldsymbol{\omega}\rangle$ with a phase that's different from all other vectors $\mid \boldsymbol{x}\rangle$.

The operator

$$\mathbf{1} - 2 \mid \boldsymbol{\omega}\rangle\langle\boldsymbol{\omega} \mid$$

is a reflection that flips a component of $\mid \boldsymbol{x}\rangle$ that is parallel to $\mid \boldsymbol{\omega}\rangle$ and leaves all other components unchanged. A reflection in a mirror is an example of such a transformation. The vector $\boldsymbol{\omega}$ in this case is a line that is perpendicular to the surface of the mirror.

Reflections and rotations are closely related. Both preserve a scalar product of two vectors and therefore also a length of the vector. Both are described by orthogonal matrices, i.e., matrices $\boldsymbol{P}$ such that $\boldsymbol{P}^T \cdot \boldsymbol{P} = \mathbf{1}$. The difference is that for rotations $\det \boldsymbol{P} = +1$ whereas for reflections $\det \boldsymbol{P} = -1$.

Reflections and rotations can be combined into a group, which has a disconnected structure. It comprises two simply connected layers: the first one is a subgroup of pure rotations, and the second one is a set obtained by multiplying rotations by a single reflection. This second layer is *not* a subgroup, because it does not contain the identity.

Whereas you can get arbitrarily close to the identity within the first layer, the rotations, you cannot get arbitrarily close to the identity within the second layer. The reflected rotations stand apart. Which is why we say that the combined group of reflections and rotations is *disconnected*.

But if you combine two reflections or two rotated reflections you get back a pure rotation because $\det \boldsymbol{P} \cdot \det \boldsymbol{P} = (-1) \cdot (-1) = 1$.

Observe that if $\boldsymbol{R}$ is a rotation then in an odd-dimensional space $\boldsymbol{P} = -\boldsymbol{R}$ is a reflection, and vice versa. For example in a 3-dimensional space $\mathbf{1} - 2\boldsymbol{w} \otimes \boldsymbol{w}$ is a reflection *along* the direction of $\boldsymbol{w}$ but $2\boldsymbol{w} \otimes \boldsymbol{w} - \mathbf{1}$ is a rotation by 180° *about* the direction of $\boldsymbol{w}$. But this rotation can be also viewed as a reflection *about* the axis defined by $\boldsymbol{w}$. In odd-dimensional spaces reflections like that happen to be rotations too, but in even-dimensional spaces they are just reflections.

If you combine two reflections of the second type, i.e., reflections *about* two axes $\boldsymbol{P}_w = 2\boldsymbol{w} \otimes \boldsymbol{w} - \mathbf{1}$ and $\boldsymbol{P}_s = 2\boldsymbol{s} \otimes \boldsymbol{s} - \mathbf{1}$, where the angle between $\boldsymbol{w}$ and $\boldsymbol{s}$ is θ, then $\boldsymbol{P}_s \cdot \boldsymbol{P}_w$ is a *rotation* in the plane defined by $\boldsymbol{w}$ and $\boldsymbol{s}$ and the rotation angle is 2θ.

Show this!

If you combine $\boldsymbol{P}_w = \mathbf{1} - 2\boldsymbol{w} \otimes \boldsymbol{w}$ with $\boldsymbol{P}_s = 2\boldsymbol{s} \otimes \boldsymbol{s} - \mathbf{1}$ then this can be still thought of as a rotation by 2θ in a plane defined by $\boldsymbol{w}$ and $\boldsymbol{s}$ combined with a total reflection (the $-\mathbf{1}$ bit). But a total reflection does not change the direction of the vector, it only changes where it points.

5.4.2 The Grover Iteration

Define

$$| \boldsymbol{s}\rangle = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{x}=\mathbf{0}}^{N-1} | \boldsymbol{x}\rangle = \bigotimes^{N} \boldsymbol{H} \, | \, 0\rangle \, | \, 0\rangle \cdots | \, 0\rangle$$

Since $| \boldsymbol{\omega}\rangle$ is one of the basis states that $| \boldsymbol{x}\rangle$ runs through, and they are all perpendicular to each other, we know that

$$\langle \boldsymbol{\omega} \mid \boldsymbol{s}\rangle = \frac{1}{\sqrt{N}} = \cos\theta$$

The probability of finding $| \boldsymbol{\omega}\rangle$ when measuring $| \boldsymbol{s}\rangle$ is

$$|\langle \boldsymbol{\omega} \mid \boldsymbol{s}\rangle|^2 = \frac{1}{N}$$

Define a reflection:

$$\boldsymbol{P}_s = 2 \mid \boldsymbol{s}\rangle\langle \boldsymbol{s} \mid -\mathbf{1}$$

The superposition of $\boldsymbol{P}_s$ and $\boldsymbol{P}_\omega$ defined above

$$\boldsymbol{R}_G = \boldsymbol{P}_s \cdot \boldsymbol{P}_\omega = (2 \mid \boldsymbol{s}\rangle\langle \boldsymbol{s} \mid -\mathbf{1}) \cdot (\mathbf{1} - 2 \mid \boldsymbol{\omega}\rangle\langle \boldsymbol{\omega} \mid)$$

is a rotation by 2θ in the plane spanned by $| \boldsymbol{s}\rangle$ and $| \boldsymbol{\omega}\rangle$ combined with a total reflection. This rotation is called *Grover iteration*.

By applying this rotation repetitively to the input state $\mid \boldsymbol{s}\rangle$ we can zoom on the solution $\mid \boldsymbol{\omega}\rangle$.

Consider first a simple situation where $N = 4$. Then

$$\langle \boldsymbol{s} \mid \boldsymbol{\omega}\rangle = \frac{1}{\sqrt{4}} = \frac{1}{2} = \cos\theta, \quad \text{hence} \quad \theta = 60^\circ$$

A single Grover iteration rotates state $\mid \boldsymbol{s}\rangle$ by 120° and then it reflects it in the opposite direction. Now $60^\circ + 120^\circ = 180^\circ$ and when you do the reflection you end up with a vector that is parallel to $\mid \boldsymbol{\omega}\rangle$ and points in the same direction too, sic! In other words, a single Grover iteration converts $\mid \boldsymbol{s}\rangle$ into $\mid \boldsymbol{\omega}\rangle$.

Now consider a situation where N is large, i.e., $\gg 4$. A very large N implies that $\cos\theta$ is very small and therefore close to 90°.

Now we need to apply the Grover iteration $\boldsymbol{R}_G$ a sufficient number of times so that the state which is initially $\mid \boldsymbol{s}\rangle$ and nearly perpendicular to $\mid \boldsymbol{\omega}\rangle$ gets rotated onto a direction that is parallel to $\mid \boldsymbol{\omega}\rangle$. Let $\frac{\pi}{2} - \theta = \vartheta$. Then $\cos\theta = \sin\vartheta$. On every iteration we move by 2ϑ away from $\pi/2$. Eventually we are going to align with $\boldsymbol{\omega}$ after

$$\frac{\pi/2}{2\vartheta} = \frac{\pi/2}{2\cos\theta} = \frac{\pi/2}{2/\sqrt{N}} = \frac{\pi\sqrt{N}}{4}$$

iterations.

A classical computer would have to perform $\mathcal{O}(N)$ queries to find the answer. Grover algorithm is therefore quadratically faster than a classical algorithm. Although the improvement is not exponential, for very large N the saving is dramatic: for example if $N = 250 \times 10^6$, roughly speaking the number of people who live in the USA, $\sqrt{N} = 15,811$. There is a tremendous difference between making 250 million queries versus making only 15,811 queries.

5.4.3 Implementing the Iteration

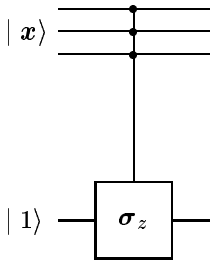
We have already shown at the beginning how to implement $\mathbf{1} - 2 \mid \boldsymbol{\omega}\rangle\langle\boldsymbol{\omega} \mid$. Here we're going to show how to implement $\boldsymbol{P}_s$. The combination of the two circuits yields one Grover iteration.

Because $\boldsymbol{P}_s = 2 \mid \boldsymbol{s}\rangle\langle\boldsymbol{s} \mid -\mathbf{1}$ and $\mid \boldsymbol{s}\rangle = \bigotimes^N \boldsymbol{H} \mid \mathbf{0}\rangle$, and because $\boldsymbol{H}^2 = \mathbf{1}$ we can rewrite $\boldsymbol{P}_s$ as follows:

$$\boldsymbol{P}_s = \bigotimes^N \boldsymbol{H}\,(2 \mid \mathbf{0}\rangle\langle\mathbf{0} \mid -\mathbf{1})\bigotimes^N \boldsymbol{H}$$

What stands between the Hadamards is an operator that flips states about the $\mid \mathbf{0}\rangle$ axis, i.e., it reverses every basis state with the exception of $\mid \mathbf{0}\rangle$.

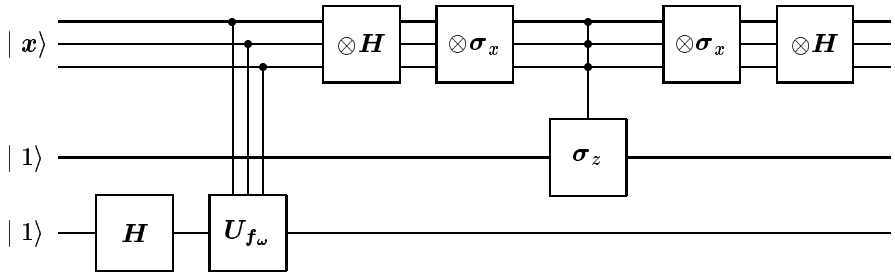
Consider the following circuit:



For all $|\boldsymbol{x}\rangle$ but $|N-1\rangle$, i.e., when every line is set to $|1\rangle$ nothing happens on the bottom line, so the state of the computer remains $|\boldsymbol{x}\rangle\otimes|1\rangle$. But if all control lines become $|1\rangle$ then the bottom line changes its sign and the state of the computer becomes: $-|N-1\rangle\otimes|1\rangle$. Now if you place a NOT gate, i.e., a $\boldsymbol{\sigma}_x$ gate before and after the control connection on every line then you invert the control, so that $|\mathbf{0}\rangle$ triggers the change of sign, whereas all else leaves the sign unchanged. This is the opposite to what we want, but this is still OK, because the difference is in the overall phase factor.

Recall that we have actually implemented a controlled-$\boldsymbol{\sigma}_z$ operator in the section about the controlled-NOT gate. But assuming that you have constructed the controlled-NOT gate by some other means and that you have plain single qubit NOT gates too you can construct the controlled-$\boldsymbol{\sigma}_z$ gate by placing $\boldsymbol{\sigma}_x$ between two Hadamard gates.

What does Grover circuit look like then? The whole iteration is shown in the figure below:



5.4.4 The Optimality of Grover Algorithm

Grover algorithm searches an N elements data base in $\mathcal{O}\left(\sqrt{N}\right)$ iterations. This is not an exponential speed up, which we would normally expect of a quantum algorithm. The obvious question to ask then is if a better algorithm can be found.

It turns out that Grover algorithm is optimal, i.e., no other algorithm can be found that would be any better. This is not to say that it is impossible to find an algorithm that would run, say, 2 times faster. The proof of the optimality of Grover algorithm that we are going to discuss below, shows merely that the problem cannot be solved in less than $\mathcal{O}\left(\sqrt{N}\right)$ iterations. Since Grover algorithm is already in this category it is optimal in this sense.

The Grover algorithm as well as the proof of its optimality constitute a remarkable accomplishment of this relatively new and budding area of quantum computing. It also shows that it is not necessary to focus on exponential speed ups for quantum algorithms only. A $\mathcal{O}\left(\sqrt{N}\right)$ speed up is sufficiently significant and interesting in its own right.

The remainder of this section, i.e., the proof itself, is somewhat lengthy and tedious – as most proofs of this nature are. It is OK to skip it if you're not interested in details.

Assume that we have a search problem, as specified in the introduction to this section, and that we start with some vector $\mid \boldsymbol{\Psi}\rangle$. We shall also assume that the problem has just one solution given by some vector $\mid \boldsymbol{x}\rangle$. In quantum terms we're going to move over the sphere of radius 1 in the Hilbert space to which $\mid \boldsymbol{\Psi}\rangle$ belongs, until we find the solution to the problem, i.e., until we stumble upon $\mid \boldsymbol{x}\rangle$. Our task is to minimize the number of moves required to reach the solution.

We also assume that we can use the circuit that gives a phase shift of -1 to $\mid \boldsymbol{x}\rangle$ and leaves all other states unchanged. As we have already demonstrated before, the circuit implements $\boldsymbol{O}_x = \mathbf{1} - 2 \mid \boldsymbol{x}\rangle\langle\boldsymbol{x} \mid$

The generic form of the algorithm is

$$\mid \boldsymbol{\Psi}_k^x\rangle = \boldsymbol{U}_k\boldsymbol{O}_x\boldsymbol{U}_{k-1}\boldsymbol{O}_x \ldots \boldsymbol{U}_1\boldsymbol{O}_x \mid \boldsymbol{\Psi}\rangle$$

Let us also introduce another vector called $\mid \boldsymbol{\Psi}_k\rangle$ defined by:

$$\mid \boldsymbol{\Psi}_k\rangle = \boldsymbol{U}_k\boldsymbol{U}_{k-1} \ldots \boldsymbol{U}_1 \mid \boldsymbol{\Psi}\rangle$$

And let $\mid \boldsymbol{\Psi}_0\rangle = \mid \boldsymbol{\Psi}\rangle$. What is the difference between $\mid \boldsymbol{\Psi}_k\rangle$ and $\mid \boldsymbol{\Psi}_k^x\rangle$? The former is a vector that is reached upon a unitary walk, which does not include any references to the search problem itself. The latter is a result of a walk interrupted upon every step with invoking the search circuitry.

Let us define an auxiliary expression, which is a sum of squared distances between $\mid \boldsymbol{\Psi}_k\rangle$ and $\mid \boldsymbol{\Psi}_k^x\rangle$ over all $x$es:

$$D_k = \sum_{x=0}^{N-1} \mid\mid \boldsymbol{\Psi}_k^x\rangle - \mid \boldsymbol{\Psi}_k\rangle\mid^2$$

We are now going to prove two interesting statements about D_k. The first one is that as we keep moving around the sphere, D_k does not grow faster than $\mathcal{O}(k^2)$. The second statement is that in order to make it possible for us to distinguish the final solution $\mid \boldsymbol{x}\rangle$ from the sea of other states, once we get to it after having made k steps, D_k, *as a function of* N must be of the order of N.

These two statements together imply that $k^2 \approx N$ and therefore $k \approx \sqrt{N}$, which is the basic tenet of this section.

We are going to prove the first statement, i.e., that $D_k \approx \mathcal{O}(k^2)$ by induction. To accomplish this we must express D_{k+1} in terms of D_k:

$$\begin{aligned} D_{k+1} &= \sum_x \left|\boldsymbol{U}_{k+1}\boldsymbol{O}_x \mid \boldsymbol{\Psi}_k^x\rangle - \boldsymbol{U}_{k+1} \mid \boldsymbol{\Psi}_k\rangle\right|^2 \\ &= \sum_x \left|\boldsymbol{U}_{k+1}\left(\boldsymbol{O}_x \mid \boldsymbol{\Psi}_k^x\rangle - \mid \boldsymbol{\Psi}_k\rangle\right)\right|^2 \\ &= \ldots \end{aligned}$$

Since $\boldsymbol{U}_{k+1}$ is a unitary operator, it doesn't change the norm. Hence we can drop it altogether:

$$\begin{aligned} \ldots &= \sum_x |\boldsymbol{O}_x \mid \boldsymbol{\Psi}_k^x\rangle - \mid \boldsymbol{\Psi}_k\rangle|^2 \\ &= \sum_x |\boldsymbol{O}_x \mid \boldsymbol{\Psi}_k^x\rangle - \boldsymbol{O}_x \mid \boldsymbol{\Psi}_k\rangle - \mid \boldsymbol{\Psi}_k\rangle + \boldsymbol{O}_x \mid \boldsymbol{\Psi}_k\rangle|^2 \\ &= \sum_x |\boldsymbol{O}_x \left(\mid \boldsymbol{\Psi}_k^x\rangle - \mid \boldsymbol{\Psi}_k\rangle\right) + (\boldsymbol{O}_x - \boldsymbol{1}) \mid \boldsymbol{\Psi}_k\rangle|^2 \\ &= \ldots \end{aligned}$$

But $\boldsymbol{O}_x = \boldsymbol{1} - 2 \mid x\rangle\langle x \mid$ therefore

$$(\boldsymbol{O}_x - \boldsymbol{1}) \mid \boldsymbol{\Psi}_k\rangle = -2 \mid x\rangle\langle x \mid \boldsymbol{\Psi}_k\rangle$$

hence

$$\ldots = \sum_x |\mathcal{O}_x \left(\mid \boldsymbol{\Psi}_k^x\rangle - \mid \boldsymbol{\Psi}_k\rangle\right) - 2 \mid x\rangle\langle x \mid \boldsymbol{\Psi}_k\rangle|^2$$

We are now going to make use of the Schwarz inequality, which says that

$$|a - b|^2 \leq |a|^2 + |b|^2 + 2|a||b|$$

Substituting this into what we have evaluated so far yields:

$$\begin{aligned} D_{k+1} \quad \leq \quad & \sum_x |\boldsymbol{O}_x \left(\mid \boldsymbol{\Psi}_k^x\rangle - \mid \boldsymbol{\Psi}_k\rangle\right)|^2 + 4\,|\langle x \mid \boldsymbol{\Psi}_k\rangle|^2 \\ & + 2\,|\boldsymbol{O}_x \left(\mid \boldsymbol{\Psi}_k^x\rangle - \mid \boldsymbol{\Psi}_k\rangle\right)|\,2\,|\langle x \mid \boldsymbol{\Psi}_k\rangle| \end{aligned}$$

We again make use of the fact that $\boldsymbol{O}_x$ being unitary does not change the norm of the vector, so we can drop it from the above expression altogether. Also notice that $\sum_x |\langle x \mid \boldsymbol{\Psi}_k\rangle|^2 = 1$ because $\mid \boldsymbol{\Psi}_k\rangle$ is a normalized state. And so

$$D_{k+1} \leq 4 + \sum_x || \boldsymbol{\Psi}_k^x\rangle - \mid \boldsymbol{\Psi}_k\rangle|^2 + 4\,|| \boldsymbol{\Psi}_k^x\rangle - \mid \boldsymbol{\Psi}_k\rangle|\,|\langle x \mid \boldsymbol{\Psi}_k\rangle|$$

First, observe that $\sum_x || \boldsymbol{\Psi}_k^x\rangle - \mid \boldsymbol{\Psi}_k\rangle|^2 = D_k$. This leaves us with

$$4\,|| \boldsymbol{\Psi}_k^x\rangle - \mid \boldsymbol{\Psi}_k\rangle|\,|\langle x \mid \boldsymbol{\Psi}_k\rangle|$$

in the expression for the D_{k+1} estimate that is still somewhat uncomfortable. But even here we can make use of the fact that

$$\left(\sum_x |a_x|\right)^2 \leq \sum_x |a_x|^2$$

so that the uncomfortable terms can now be estimated from below by

$$4\sqrt{\sum_x || \boldsymbol{\Psi}_k^x\rangle - \mid \boldsymbol{\Psi}_k\rangle|^2}\sqrt{\sum_x |\langle x \mid \boldsymbol{\Psi}_k\rangle|^2} = 4\sqrt{D_k}$$

In summary we have shown that:

$$D_{k+1} \leq 4 + D_k + 4\sqrt{D_k}$$

Observe that this growth is like that of $D_k = 4k^2$:

$$D_{k+1} = 4(k+1)^2 = 4(k^2+2k+1) = 4 + 4k^2 + 4(2k) = 4 + D_k + 4\sqrt{D_k}$$

Thus

$$D_k \leq \mathcal{O}(4k^2)$$

And so, we have demonstrated the first leg of our proof that Grover's algorithm is optimal. The second leg now talks about scaling of D_k with the size of the data base N. Here we want to show that there is a certain condition imposed on D_k that has to do with our ability to distinguish between $\mid \Psi_k^x \rangle$ and other states.

The condition is that as we keep stepping, or iterating, we should get to see $\mid x \rangle$ itself more clearly with every step. Since the state of our search after k steps is described by $\mid \Psi_k^x \rangle$ this implies that the probability of finding $\mid x \rangle$ upon measuring $\mid \Psi_k^x \rangle$ should be high. How high? Well, let us say at least $1/2$. This is then our criterion:

$$|\langle x \mid \Psi_k^x \rangle|^2 \geq \frac{1}{2} \tag{5.3}$$

If the algorithm is to be generally valid, this should hold for every $\mid x \rangle$.

Now consider the distance between $\mid x \rangle$ and $\mid \Psi_k^x \rangle$:

$$\begin{aligned} || \Psi_k^x \rangle - \mid x \rangle|^2 &= \langle \Psi_k^x \mid \Psi_k^x \rangle - 2\mathrm{Re}\,\langle x \mid \Psi_k^x \rangle + \langle x \mid x \rangle \\ &= 2 - 2\mathrm{Re}\,\langle x \mid \Psi_k^x \rangle \\ &\ldots \end{aligned}$$

Resorting to polar notation for $\langle x \mid \Psi_k^x \rangle$ and making use of the fact that $|\langle x \mid \Psi_k^x \rangle| \geq \sqrt{1/2}$ we can replace the above with

$$\ldots \leq 2 - 2\mathrm{Re}\, e^{i\theta} \frac{1}{\sqrt{2}}$$

The largest value of $\mathrm{Re}\, e^{i\theta} = 1$, therefore the sharpest form of this inequality is

$$|| \Psi_k^x \rangle - \mid x \rangle|^2 \leq 2 - \sqrt{2}$$

This is how we are going to replace our original somewhat weaker criterion (5.3)[3].

Let us now introduce two auxiliary quantities E_k and F_k, with which we are going to estimate D_k. E_k is simply the sum of distances between $\mid \Psi_k^x \rangle$ and $\mid x \rangle$ over all x-es, and we already have an estimate for it:

$$E_k = \sum_x || \Psi_k^x \rangle - \mid x \rangle|^2 \leq (2 - \sqrt{2})N$$

[3]Nielsen and Chuang [80] use the following argument here. They say that "Replacing $\mid x \rangle$ by $e^{i\theta} \mid x \rangle$ does not change the probability of success, so without loss of generality we may assume that $\langle x \mid \Psi_k^x \rangle = |\langle x \mid \Psi_k^x \rangle|$." As we can clearly see this replacement results in a new criterion, which is stronger than the original (5.3), i.e., we want $\mid \Psi_k^x \rangle$ to be even closer to $\mid x \rangle$. This is quite OK, because the original criterion (5.3) is of a hand-waving variety anyway. Replacing it with a stronger one makes the following calculations easier and this, of course, is the real reason why we do this.

F_k is a sum of distances between $\mid x\rangle$ and $\mid \Psi_k\rangle$:

$$F_k = \sum_x || \, x\rangle - \mid \Psi_k\rangle|^2$$

As before we can easily see that:

$$|| \, x\rangle - \mid \Psi_k\rangle|^2 = 2 - 2\mathrm{Re}\,\langle x \mid \Psi_k\rangle$$

Summing it up over x yields:

$$\begin{aligned} F_k &= \sum_x (2 - 2\mathrm{Re}\,\langle x \mid \Psi_k\rangle) \\ &= 2N - 2\sum_x \mathrm{Re}\,\langle x \mid \Psi_k\rangle \end{aligned}$$

The sum $\sum_x \mathrm{Re}\,\langle x \mid \Psi_k\rangle$ attains maximum for $\mid \Psi_k\rangle = \sum_y \mid y\rangle/\sqrt{N}$, i.e., when $\mid \Psi_k\rangle$ is an equally weighted superposition of all $\mid y\rangle$. This can be shown by using constrained extremization with Lagrange multipliers [87]. So let us make this substitution, which lets us estimate F_k from below:.

$$F_k \geq 2N - 2\sum_x \sum_y \frac{1}{\sqrt{N}} \langle x \mid y\rangle = 2N - 2\sqrt{N}$$

So now, let us see how we can estimate D_k with a combination of E_k and F_k.

$$\begin{aligned} D_k &= \sum_x || \, \Psi_k^x\rangle - \mid \Psi_k\rangle|^2 \\ &= \sum_x || \, \Psi_k^x\rangle - \mid x\rangle + \mid x\rangle - \mid \Psi_k\rangle|^2 \\ &= \sum_x || \, \Psi_k^x\rangle - \mid x\rangle|^2 + || \, x\rangle - \mid \Psi_k\rangle|^2 + 2\mathrm{Re}\,(\langle x \mid - \langle \Psi_k \mid)(\mid \Psi_k^x\rangle - \mid x\rangle) \\ &\geq \sum_x || \, \Psi_k^x\rangle - \mid x\rangle|^2 + \sum_x || \, x\rangle - \mid \Psi_k\rangle|^2 + 2\sum_x \mathrm{Re}\,(\langle x \mid - \langle \Psi_k \mid)(\mid \Psi_k^x\rangle - \mid x\rangle) \\ &\geq E_k + F_k - 2\sqrt{E_k F_k} \\ &= \left(\sqrt{F_k} - \sqrt{E_k}\right)^2 \end{aligned}$$

Observe the change of sign in front of $2\sqrt{E_k F_k}$, which correponds to the worst possible case of all phase factors $e^{i\theta_x}$ aligning at -1.

Substituting the estimates for E_k and F_k we have obtained above yields:

$$\begin{aligned} D_k &\geq \left(\sqrt{F_k} - \sqrt{E_k}\right)^2 \\ &\geq \left(\sqrt{2N - 2\sqrt{N}} - \sqrt{(2-\sqrt{2})N}\right)^2 \end{aligned}$$

For large N we can neglect the $\sqrt{N}$ in the first square root, compared to N and so we get:

$$D_k \geq \left(\sqrt{2} - \sqrt{2-\sqrt{2}}\right)^2 N = 0.421N$$

Let us now sum up what we have accomplished here. First we showed that $D_k \leq \mathcal{O}(4k^2)$. Then we also showed that $D_k \geq 0.421N$. Together these two inequalities imply:

$$0.421N \leq D_k \leq \mathcal{O}(4k^2)$$

Hence

$$k \geq \mathcal{O}\left(\sqrt{\frac{0.421N}{4}}\right)$$

In other words, the number of iterations required to solve the problem must be of the order of $\sqrt{N}$. Grover algorithm is therefore optimal.

The reason why Grover's problem is classically hard is because its search space has no structure. Hence we have no other choice but to resort to brute force and search the whole space through. Quantum mechanics offers improvements due to quantum parallelism: we can pass the whole search space all at once to the characteristic function and get the searched element marked in just a single step. Yet it still takes $\mathcal{O}(\sqrt{N})$ iterations to zoom on this marked element. Still, the saving is very substantial for large N.

On the other hand, if a hidden structure exists in the search space, we can make use of this and then further improvements to algorithm efficiency are possible. This is what we have seen in the hidden subgroup problem examples.

Chapter 6

Quantum Error Correction

Quantum computers, like classical computers, are subject to errors. The sources of errors are in some cases similar, e.g., thermal fluctuations, interaction with cosmic rays. In other cases errors are specific to quantum systems, for example beta decay or K-capture, decoherence, dissipation.

As we have already emphasized in section 4.6 (page 147), a description of a quantum system in terms of rays in a Hilbert space subjected to unitary operations is an idealization. In real life states are *not* rays, measurements are *not* orthogonal projections and evolution is *not* unitary.

> If the above statement shocks you then think about a well known classical idealization of a planet as a *material point*. Of course planets are *not* material points. Their trajectories are *not* elliptical. And, to make things worse, they are *not* exactly where we see them, because of the finite speed of light. The latter was used by Ole Christensen Rømer in 1676 to estimate, for the first time, the speed of light, which he did with about 25% error – not bad at all for the first shot!

Idealizations of this nature are favourite with some physicists, especially theoretical types (mostly because they lead to easily solvable models), and with mathematicians (for the same reason). But they are not looked very kindly upon by experimental physicists and they are definitely out of favour with engineers. The reason is obvious: engineers and experimentalists have to deal with real world. Engineers have to design and build devices for the real world. Imagine designing a car and completely ignoring friction. Or imagine designing a computer, and completely ignoring the need for shielding and fault-tolerant memory.

Consequently, ever since its conception, the budding area of quantum computing has been concerned with managing errors that may affect computational procedures.

Errors may creep into these procedures at various stages. Initial state preparation may be laden with errors. Gates applied to the system may be inaccurate.

qubit	τ (in s)	time per gate (in s)	number of steps
GaAs electron	10^{-10}	10^{-13}	10^{3}
Au electron	10^{-8}	10^{-14}	10^{6}
trapped In ion	10^{-1}	10^{-14}	10^{13}
optical cavity	10^{-5}	10^{-14}	10^{9}
electron spin	10^{-3}	10^{-7}	10^{4}
electron quantum dot	10^{-3}	10^{-6}	10^{3}
nuclear spin	10^{4}	10^{-3}	10^{7}

Table 6.1: Decoherence time τ versus time per gate for various qubit implementations. The last column show the number of gates that may be executed before the register decoheres [30].

As we have seen in section 4.7.3 (page 166) the register may be subjected to depolarization (spin or phase flip), decoherence (phase damping), and dissipation (spontaneous emmission or amplitude damping) during or between gate traversals. Finally, we may end up with measurement errors.

In section 4.7.3 (page 169) we have investigated a simple model of decoherence. We have also described decoherence observed in NMR measurements in section 4.6.7 (page 156). The characteristic feature of decoherence is exponential vanishing of off-diagonal terms in the density matrix of the register, which corresponds to the collapse of the wave function and the resulting replacement of a superposition of states with a mixture.

NMR measurements yield a time scale τ for this process. Measurements performed for other systems that are candidates for quantum registers also give us an idea about the time scale of decoherence in those systems. Although the simple model we have looked at in section 4.7.3 gives us some idea about how decoherence takes place, our understanding of this process is still rather sketchy and we do not have good theoretical predictions for τ. You will find a very good discussion of decoherence presented by the master of this subject, Wojciech Hubert Zurek from Los Alamos, in [107].

Table 6.1 shows estimated decoherence times for various potential qubit candidates. In the same table we also list the time it takes to execute a single gate on that qubit, and in the last column we print a number of gates that can be traversed by the register before it decoheres. Observe that what matters is two factors: the decoherence time and the time per gate. For example, the decoherence time for nuclear spins is 10^4 s, which is really very long. But it also takes about a ms to apply a gate to a nuclear spin. Consequently the total number of gates that one can hope to apply to a nuclear spin based register before it decoheres is about ten million. In principle, we should be able to do much better with, e.g., trapped Indium ions, where the decoherence time is 0.1 s, but gates can be applied in only 10^{-14} s. So we can apply 10^{13} gates before the register decoheres.

Decoherence can be controlled. In section 4.7.3 we attributed decoherence

to scattering of light low energy particles against a heavy particle. The light particles did not have enough energy or momentum to change the state of the heavy particle, but they sponged away information contained in the quantum state of the heavy particle. The decoherence time was related to the number of scattering events per second and the probability that such an event would result in the entanglement between the heavy particle and the light particle scattering off it. This has the following implications:

- *The denser, the hotter, and the larger the medium in which a register is embedded, the shorter is the coherence time of the register.*
- *The lighter (i.e., the more susceptible to the entanglement) the constituent particles of the register the shorter the coherence time of the register*

This explains why it is going to be so difficult to maintain coherence of a quantum register in a crystal. Here a high density of the medium, and the rich spectrum of various excitations such as phonons, polarons, photons and what not work against the long coherence time. If such a register is to be based on electron spin or electron energy levels we have additional problem caused by the small mass-to-charge ratio for the electron. In other words, structures based on electrons in solids (e.g., quantum dots) are perhaps the worst possible candidates for quantum registers, with the exception, maybe, of systems based on collective excitations, such as anyons.

On the other hand, systems based on nuclear spins attached to molecules (which provide structuring and binding needed to deliver individually addressable qubits and couplings within the register) suspended in a diluting medium, be it a magnetically neutral liquid or a rarified gas, are excellent candidates for quantum computing, at least as far as their resistance to decoherence goes. A proton is 1,836 times heavier than an electron while having the same electric charge. A proton is therefore that much less likely to entangle with light particles that try to sponge information away from its quantum state. A look at table 6.1 confirms this assessment: the decoherence time for electrons in GaAs crystals is only 10^{-10} s, whereas decoherence time for nuclear spins is 10^4 s.

Decoherence time is also going to depend on the type and temperature of the environment as well as on the size of the environment. Table 6.2 shows estimates for decoherence time in seconds if the register is exposed to interaction with (1) cosmic background radiation, (2) room temperature, (3) sunlight on earth, (4) vacuum, (5) air – all in an enclosure of various sizes.

Decoherence is an example of an error that cannot be described by a unitary operator. Decoherence, as we have seen in section 4.7.3, is a non-unitary phenomenon. So is dissipation, i.e., an act of emitting a photon or a phonon or some other quantum of energy, and dropping to a lower energy state.

> Although we have discussed dissipation in section 4.7.3 (page 171), you may possibly harbour some doubts as to the validity of this statement,

size (cm)	cosmic background radiation	room temperature	sunlight on earth	vacuum 10^6 particles/cm^3	air
10^{-3}	10^{-7}	10^{-14}	10^{-16}	10^{-18}	10^{-35}
10^{-5}	10^{15}	10^{-3}	10^{-8}	10^{-10}	10^{-23}
10^{-6}	10^{24}	10^{5}	10^{-2}	10^{-6}	10^{-19}

Table 6.2: Decoherence time, in seconds, for various types of the environment in function of the environment's size [54].

thinking of a dissipation as a case of a bit-flip. Consider the following description of this process:

$$\begin{aligned} \boldsymbol{D} \mid 1\rangle &= \mid 0\rangle \\ \boldsymbol{D} \mid 0\rangle &= \mid 0\rangle \end{aligned}$$

The matrix that corresponds to $\boldsymbol{D}$ in this basis is:

$$\boldsymbol{D} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$

It is easy to see that this matrix is not unitary, because

$$\boldsymbol{D} \cdot \boldsymbol{D}^\dagger = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

But there is also a group of errors that are quite unitary: spin and phase flips. These can be described by Pauli matrices, and they are collectively referred to as *depolarization* errors. We have discussed these in section 4.7.3 (page 167). To remind you the errors are as follows:

bit flip error described by $\boldsymbol{\sigma}_x$:

$$\boldsymbol{\sigma}_x \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} b \\ a \end{pmatrix}$$

phase shift error described by $\boldsymbol{\sigma}_z$:

$$\boldsymbol{\sigma}_z \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ -b \end{pmatrix}$$

phase and bit flip error described by $\boldsymbol{\sigma}_y$:

$$\boldsymbol{\sigma}_y \begin{pmatrix} a \\ b \end{pmatrix} = -i \begin{pmatrix} b \\ -a \end{pmatrix}$$

no error described by $\mathbf{1}$:

$$\mathbf{1} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$$

These depolarization errors, all of which are *large and discrete* are most tractable of all quantum errors. Here procedures exist that are, to a degree, derived from procedures that handle similar problems in classical computers. Of course, we cannot apply classical error correction procedures directly. For example some classical error correction procedures are based on replication. But in the quantum domain we cannot replicate, because of the *no-cloning* theorem. Yet, there is enough commonality here so that some ideas used in the construction of classical linear codes, for example, can be re-used to develop Calderbank-Shor-Steane codes for quantum registers. Applying these lets us encode *logical* qubits in groups of *physical* qubits, in analogy to the way that *logical* bits are encoded in groups of *physical* bits in classical computing. The encodings are such that errors can be detected and corrected, while at the same time preserving the coherence of the quantum state of the register.

Quantum circuits were developed and implemented even (to a degree... [61] [45]), which demonstrated such procedures.

But then quantum computers are also subject to *small errors*. For example:

$$\begin{pmatrix} a \\ b \end{pmatrix} \rightarrow \begin{pmatrix} a + \epsilon_a \\ b + \epsilon_b \end{pmatrix}$$

where ϵ_a and ϵ_b are small. Classical bits, of course, cannot do this: they can be only 0 or 1. But even here procedures were developed that can correct such errors without ever measuring them explicitly, which would destroy the quantum state of the register.

The ultimate in quantum error correction are the so called *concatenated codes*, which let us carry out quantum computations with arbitrary precision, sic!, assuming that the probability of error occuring can be pushed below a certain threashold.

Last, but not least, fault tolerant gates and measurement procedures have been developed.

The theory of fault-tolerant quantum computation has become a sizeable industry. Unfortunately, as the procedures concocted by this industry often require a fairly large number of additional qubits, well in excess of what's currently possible, there is little here in terms of experimental demonstration yet.

On the other hand, in *some* situations we don't really have to worry about quantum error correction procedures too much. Consider a case of NMR computation. Assume that the sample is cooled to reduce thermal fluctuations and to improve signal-to-noise ratio. Because of a large mass-to-charge ratio of nucleons depolarization errors are not going to be all that common. But more importantly, because NMR computation is carried out in parallel on trillions of quantum registers, when measurements are made we end up with very well defined distributions, in which any quantum errors that may have occurred during computations would either average away, or they would simply join the background noise. It should therefore be possible to carry out sizeable computations using NMR on quite a large number of *raw* qubits without explicit qubit-level error correction procedures.

We can think of this process as an error correction procedure of a kind. The procedure is based on replication. But it is not qubits that we replicate. Instead we replicate whole independent registers. Then we replicate the whole computation on all the registers, and finally we sum up answers from all the registers. The summation averages errors away. This is similar to *majority voting* schemes in classical computing.

6.1 Decoherence-Free Subspace

Apart from cooling and isolating the system, there is another way to deal with decoherence. Decoherence can be dealt with algorithmically. It is possible to encode *logical qubits* in physical qubits in such a way that even though decoherence affects physical qubits in some ways, logical qubits remain unaffected.

Consider the following encoding:

$$\begin{aligned} \mid 0\rangle_L &= \frac{1}{\sqrt{2}}(\mid 0\rangle \mid 1\rangle - i \mid 1\rangle \mid 0\rangle) \\ \mid 1\rangle_L &= \frac{1}{\sqrt{2}}(\mid 0\rangle \mid 1\rangle + i \mid 1\rangle \mid 0\rangle) \end{aligned}$$

First observe that $\mid 0\rangle_L^* = \mid 1\rangle_L$. You can also check easily that ${}_L\langle 0 \mid 0\rangle_L = {}_L\langle 1 \mid 1\rangle_L = 1$ and that ${}_L\langle 0 \mid 1\rangle_L = 0$. Therefore the states $\mid 0\rangle_L$ and $\mid 1\rangle_L$ constitute an orthonormal basis.

Decoherence can be modelled by an operation called *collective dephasing*. The operation transforms $\mid 1\rangle$ into $e^{i\theta} \mid 1\rangle$ for both physical qubits at the same time and leaves $\mid 0\rangle$ unchanged for both of them. You can see that this results in

$$\begin{aligned} \mid 0\rangle_L &\rightarrow e^{i\theta} \mid 0\rangle_L \\ \mid 1\rangle_L &\rightarrow e^{i\theta} \mid 1\rangle_L \\ \alpha \mid 0\rangle_L + \beta \mid 1\rangle_L &\rightarrow e^{i\theta}(\alpha \mid 0\rangle_L + \beta \mid 1\rangle_L) \end{aligned}$$

If all operations are carried out on qubits encoded in the same way, all that happens when collective dephasing takes place is that the whole system gets multiplied by $e^{i\theta}$. But this doesn't change anything because no probabilities get affected by it. What really matters in quantum mechanics are phase *differences* between qubits.

We can say that the encoding maps qubits onto a *Decoherence Free Subspace* of Hilbert space.

How well does this model correspond to what really happens when decoherence takes place? First, the model is *collective*, i.e., it assumes that all qubits in the register are affected by the same error at the same time. This is very different from an *independent* error model. Quoting from [68], the collective dephasing model derives from

> Dicke's quantum optics work on superradiance of atoms coupled to a radiation field, where it arose in the consideration of systems confined to a

region whose linear dimensions are small compared to the shortest wavelength of the field.

Recall that in our simple decoherence model in section 4.7.3 we made the assumption that particles scattering off the heavy particle subjected to the decoherence superoperator all had low energy. Low energy means a long wavelength – possibly longer than the physical dimensions of the register. This, in term, implies that all qubits in the register may get affected in the same way.

But the final arbiter of truth in physics is always the experiment. In February 2001 Kielpinski, Meyer, Rowe, Sackett, Itano, Monroe and Wineland reported in Science an experiment in which they encoded a qubit into a decoherence free subspace of a pair of trapped $^9Be^+$ ions. They used encoding exactly like the one shown above. Then they measured the storage time under ambient conditions and under interaction with an engineered noisy environment and observed that the encoding increased the storage time by up to an order of magnitude [59].

Mathematically the description of decoherence free spaces is as follows. We assume a Hamiltonian $\boldsymbol{H}_{SB}$ that describes interaction of a system S with a heat bath B of the following form:

$$\boldsymbol{H}_{SB} = \sum_i \boldsymbol{S}_i \otimes \boldsymbol{B}_i$$

The decoherence free states are *those, and only those* states which are simultaneous degenerate eigenvectors, i.e., eigenvectors with the same energy, of all *system* S operators appearing in $\boldsymbol{H}_{SB}$, i.e.,

$$\boldsymbol{S}_i \mid \psi\rangle = s_i \mid \psi\rangle$$

where the eigenvalues (energies) s_i do not depend on ψ. It turns out that the subspace spanned by these states is decoherence free, i.e., the $\boldsymbol{H}_{SB}$ evolution in this space is *unitary*, whereas in general, as you should remember, it isn't. Consequently, there is no decoherence in this subspace.

This results in a passive protection against errors – as opposed to the active protection of the quantum error correction codes, about which more below.

6.2 Linear Codes

Classical error protection codes are just as important as quantum error correction codes, perhaps even more so, because classical computers exist, whereas quantum ones don't (yet). Unprotected physical bits are seldom used in computing. They are vulnerable to thermal fluctuations, cosmic rays, power fluctuations and other environmental hazards. Yet bit protection can be implemented quite cheaply in classical computing. For example parity checking can be used to protect with just one or two additional bits the whole 8-bit register. Another classical method of error correction is replication combined with *majority*

voting, mentioned in the preamble to this chapter. All these methods can be described in terms of simple linear operations.

Assume that $\boldsymbol{x}$ corresponds to a classical *logical* register of k *logical* bits. The encoding of this register into n physical bits can be described by a *generator* matrix $\boldsymbol{G}$ with k columns and n rows, where $n > k$, and where each matrix entry is either 0 or 1. For example a replication code for 1 logical bit, to be replicated 3 times, can be described as follows:

$$\boldsymbol{G} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad \boldsymbol{x} = (x_0) \qquad \boldsymbol{Gx} = \begin{pmatrix} x_0 \\ x_0 \\ x_0 \end{pmatrix}$$

The following generator matrix, in turn, encodes a pair of two logical bits each into a triple of physical bits:

$$\boldsymbol{G} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \qquad \boldsymbol{x} = \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} \qquad \boldsymbol{Gx} = \begin{pmatrix} x_0 \\ x_0 \\ x_0 \\ x_1 \\ x_1 \\ x_1 \end{pmatrix}$$

The columns of $\boldsymbol{G}$ must be linearly independent so that there is a unique encoding for every vector $\boldsymbol{x}$. Observe that you can replace the first column in the $\boldsymbol{G}$ above with the sum of both columns, and this will produce a code, which, although not identical, is very similar nevertheless:

$$\boldsymbol{G} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \qquad \boldsymbol{x} = \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} \qquad \boldsymbol{Gx} = \begin{pmatrix} x_0 \\ x_0 \\ x_0 \\ x_0 +_2 x_1 \\ x_0 +_2 x_1 \\ x_0 +_2 x_1 \end{pmatrix}$$

If $x_0 = 0$ the bottom half of the resulting vector evaluates to three times x_1, so in this case the result is identical. If $x_0 = 1$ the bottom half of the resulting vector evaluates to three times $1 +_2 x_1 = \neg\, x_1$. We end up with another variant of a replication code.

The above formulation of encoding, in terms of the generator matrix $\boldsymbol{G}$, doesn't tell us how to detect and correct errors. One can associate another matrix $\boldsymbol{H}$, which is called the *parity check* matrix, with $\boldsymbol{G}$, which can be used to do this. Let $\boldsymbol{H}$ be orthogonal to $\boldsymbol{G}$, i.e., let

$$\boldsymbol{HG} = \boldsymbol{0}$$

Matrix $\boldsymbol{G}$, which encodes a k-bit register into n physical bits, where $n > k$, is $n \times k$, i.e., it has n rows and k columns. Consequently, to make the above orthogonality relation valid, matrix $\boldsymbol{H}$ must have n columns. We also require

that it should have $n-k$ rows. Thus matrix $\mathbf{0}$ in the orthogonality relation above is $(n-k)\times k$. Observe that if you were to put $\boldsymbol{H}$ on its side and place it to the right of $\boldsymbol{G}$ you would obtain an orthogonal matrix $n\times n$. Matrix $\boldsymbol{H}$ can therefore be thought of as an *orthogonal complement of matrix* $\boldsymbol{G}$.

In order to build $\boldsymbol{H}$, and assuming that you have $\boldsymbol{G}$ you need to find $n-k$ linearly independent vectors $\boldsymbol{y}_i$ orthogonal to the columns of $\boldsymbol{G}$ and then set the rows of $\boldsymbol{H}$ be $\boldsymbol{y}_1^T, \boldsymbol{y}_2^T, \ldots, \boldsymbol{y}_{n-k}^T$.

> Be aware that there are many surprises in binary arithmetic. In particular being orthogonal in the binary world does not in itself imply linear independence. Every vector with an even number of ones, for example, is orthogonal to itself. Consequently the rows of $\boldsymbol{H}$ while orthogonal to the columns of $\boldsymbol{G}$ and linearly independent of each other may be linear combinations of the columns of $\boldsymbol{G}$. You will see in the next section about the CSS codes that this is indeed the case for the Hamming code, and, in fact, we are going to make use of it in order to construct the so called Steane code.

In order to build $\boldsymbol{G}$, and assuming that you have $\boldsymbol{H}$ you need to find k linearly independent vectors orthogonal to the rows of $\boldsymbol{H}$. Matrix $\boldsymbol{G}$ is then built by making the vectors the columns of $\boldsymbol{G}$.

Matrix $\boldsymbol{H}$ maps n-dimensional vectors onto an $n-k$ dimensional space. All vectors that $\boldsymbol{H}$ maps onto zero of this $n-k$ dimensional space form a k-dimensional subspace, which is called the *kernel* of $\boldsymbol{H}$. This subspace corresponds to the k-dimensional vectors encoded by $\boldsymbol{G}$. This is easy to see because:

$$\boldsymbol{HGx} = \mathbf{0}\boldsymbol{x} = \mathbf{0}$$

Consider our simple example, which encoded a single bit by replicating it three times. The generator matrix for this encoding was:

$$\boldsymbol{G} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

We can choose the following 2×3 matrix $\boldsymbol{H}$:

$$\boldsymbol{H} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

Remember that all our additions are modulo 2, so this matrix is indeed perpendicular to $\boldsymbol{G}$:

$$\boldsymbol{HG} = \begin{pmatrix} 1\cdot 1 +_2 1\cdot 1 +_2 0\cdot 1 \\ 0\cdot 1 +_2 1\cdot 1 +_2 1\cdot 1 \end{pmatrix} = \begin{pmatrix} 1 +_2 1 \\ 1 +_2 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

To check if a given vector (x_0) has been correctly encoded into $\boldsymbol{y} = (x_0, x_0, x_0)$ we calculate:

$$\boldsymbol{Hy} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} y_0 \\ y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} y_0 +_2 y_1 \\ y_1 +_2 y_2 \end{pmatrix} = \begin{pmatrix} x_0 +_2 x_0 \\ x_0 +_2 x_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

because $0 +_2 0 = 0$ and $1 +_2 1 = 0$ too.

Observe that if any of the bits in $\boldsymbol{y}$ is different from the other two bits $\boldsymbol{Hy} \neq \boldsymbol{0}$. The value returned by $\boldsymbol{Hy}$ is called the *error syndrome*. If the first bit of $\boldsymbol{y}$ has flipped then, remembering that classical bit flip is simply $+_2 1$ we get $x_0 +_2 1 +_2 x_0 = 1$ and the error syndrome is $\binom{1}{0}$. If the third bit of $\boldsymbol{y}$ has flipped then the error syndrome is $\binom{0}{1}$. And if the second bit of $\boldsymbol{y}$ has flipped the error syndrome is going to be $\binom{1}{1}$. Observe that by looking at the error syndrome we can point to the bad bit and we can repair it. This, of course, is based on the assumption that the other two bits are good. This procedure is called the *majority voting*. But observe that if two bits flip at the same time then majority is going to be wrong and minority is going to be right. This, of course, never happens in politics.

In general we can describe the error syndrome finding procedure as follows. If $\boldsymbol{Gx} = \boldsymbol{y}$, then if an error occurs, we can represent it by $\boldsymbol{e}$ such that $\boldsymbol{y}' = \boldsymbol{y} + \boldsymbol{e}$. The error syndrome then returns:

$$\boldsymbol{Hy}' = \boldsymbol{H}(\boldsymbol{y} + \boldsymbol{e}) = \boldsymbol{He}$$

An important category of linear codes are the so called *Hamming* codes. The parity check matrix $\boldsymbol{H}$ for Hamming codes is constructed by making its columns be numbers in binary representation from 1 to $2^r - 1$, where r is the number of rows in the matrix. For example for $r = 3$ the parity check matrix for the Hamming code looks as follows:

$$\boldsymbol{H} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

The Hamming codes have this nice property that the error syndrome points to the location of the error right away. If, for example, $\boldsymbol{e} = (0, 0, 1, 0, 0, 0, 0)$ then

$$\boldsymbol{He} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

which, counting from the bottom to the top, is binary for 3. This tells us right away that the third bit has flipped.

The generator matrix $\boldsymbol{G}$ that corresponds to the Hamming parity check matrix $\boldsymbol{H}$ shown above is

$$\boldsymbol{G} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

Because matrices $\boldsymbol{G}$ and $\boldsymbol{H}$ complement each other their roles can be reversed. This means that you can use $\boldsymbol{H}^T$ as a generator matrix and $\boldsymbol{G}^T$ as a parity check matrix. This new encoding is different from the original one given by $\boldsymbol{G}$ and $\boldsymbol{H}$, but both codes are related. We say that they are *dual* to each other. The notation that accompanies this duality is as follows. If the pair $\{\boldsymbol{G}, \boldsymbol{H}\}$ defines code $C[n,k]$ that encodes k logical bits in n physical bits $(n > k)$ then the pair $\{\boldsymbol{H}^T, \boldsymbol{G}^T\}$ defines code $C^{\perp}[n, n-k]$ which encodes $n-k$ logical bits in n physical bits.

We can also identify C with $\ker \boldsymbol{H}$ and $C^{\perp}$ with $\ker \boldsymbol{G}^T$, i.e., we can think of C and $C^{\perp}$ as sets and apply set equality and set inclusion relations to them. Making use of these set relations, if $C = C^{\perp}$ we call such a code *self-dual* and if $C \subseteq C^{\perp}$ we call it *weakly* self-dual.

6.3 Calderbank-Shor-Steane Codes

Calderbank-Shor-Steane codes, or CSS codes for short, are quantum codes, which let us identify and correct large qubit errors, i.e., errors described by Pauli matrices. CSS codes derive from classical linear codes.

In order to construct a CSS code you need to have two classical linear codes, $C_1[n, k_1]$ and $C_2[n, k_2]$ such that $C_2 \subset C_1$. The resulting code is a quantum code called $\mathrm{CSS}(C_1/C_2)$ (this is read "CSS of C_1 over C_2"), which encodes $k_1 - k_2$ logical qubits in n physical qubits, so this code is $[n, k_1 - k_2]$.

> Codes C_1 and C_2 can be dual as long as $C_2 \subset C_1$. In this case $C_1 = C_1[n,k]$ and $C_2 = C_2[n, n-k]$. The resulting CSS code is $\mathrm{CSS}[n, 2k-n]$. For example, if we were to take the $(7,4)$ Hamming code discussed in the previous section, its $C^{\perp}$ would be a $[7,3]$ code. You will see below that in this case $C_2 \subset C_1$. The resulting CSS code would be a $[7, 2 \times 4 - 7] = [7,1]$ code, i.e., a code that can correct errors on a single logical qubit by encoding it in 7 physical qubits.
>
> The smallest quantum code that can be constructed this way is
>
> $$\mathrm{CSS}(C_1[5,3]/C_2[5,2]) = \mathrm{CSS}[5,1]$$
>
> This smallest code was first discovered by Laflamme, Miguel, Paz and Zurek in 1996 [63].

The encoding is a vector space spanned by *all* states constructed by taking a codeword $\boldsymbol{x} \in C_1$ and then adding to it the whole of C_2:

$$|\, \boldsymbol{x} +_2 C_2 \rangle = \frac{1}{\sqrt{|C_2|}} \sum_{\boldsymbol{y} \in C_2} |\, \boldsymbol{x} +_2 \boldsymbol{y} \rangle$$

where $|C_2|$ is the number of elements in C_2. Observe that if $\boldsymbol{x} \in C_1$ and $\boldsymbol{x}' \in C_1$ but $(\boldsymbol{x}' -_2 \boldsymbol{x}) \in C_2$ then $\boldsymbol{x} +_2 C_2 = \boldsymbol{x} +_2 \boldsymbol{x}' -_2 \boldsymbol{x} +_2 C_2 = \boldsymbol{x}' +_2 C_2$. This means that $\boldsymbol{x} +_2 C_2$ for all $\boldsymbol{x} \in C_1$ depends only on division of C_1 into layers parallel to C_2, each of which is an image of C_2 shifted by $\boldsymbol{x}$ (or any other vector of the form of $\boldsymbol{x} +_2 \boldsymbol{e}$ where $\boldsymbol{e} \in C_2$). This layering is what we mean by C_1/C_2.

Let us again take the $C_1[7,4]$ Hamming code and its $C_2[7,3]$ dual and let us try to construct the corresponding $[7,1]$ CSS(C_1/C_2).

The codewords of $C_1[7,4]$ are spanned by the columns of the generator matrix

$$\boldsymbol{G} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

The codewords of $C_2[7,3]$ are spanned by the rows of the parity check matrix

$$\boldsymbol{H} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

First observe that each row of $\boldsymbol{H}$ can be constructed by adding rows of $\boldsymbol{G}^T$, e.g.,

$$\begin{array}{rccccccc} & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ = & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ +_2 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \end{array}$$

The same holds for the other two rows of $\boldsymbol{H}$, which means that C_2 is indeed contained in C_1, as we have already asserted above. Now we need to generate every possible vector of C_2. Apart from the ones already listed in the $\boldsymbol{H}$ matrix we also have the following sums:

$$\begin{array}{rccccccc} & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ +_2 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ = & 0 & 1 & 1 & 1 & 1 & 0 & 0 \end{array}$$

$$\begin{array}{rccccccc} & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ +_2 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ = & 1 & 0 & 1 & 1 & 0 & 1 & 0 \end{array}$$

$$\begin{array}{rccccccc} & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ +_2 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ = & 1 & 1 & 0 & 0 & 1 & 1 & 0 \end{array}$$

$$\begin{array}{rccccccc} & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ +_2 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ +_2 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ = & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ +_2 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ = & 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{array}$$

An element that obviously must belong to C_1 is $|\,0000000\rangle$. This element belongs to C_2 too. It is obtained by adding any other C_1 or C_2 vector to itself. Therefore the first element of C_1/C_2, which we are going to identify with $|\,0\rangle_L$ is going to be $\sum_{y \in C_2}(|\,0000000\rangle +_2 |\,\boldsymbol{y}\rangle)$ where $|\,\boldsymbol{y}\rangle$ are all the vectors we have found above *including* $|\,0000000\rangle$ which belongs to C_2 too:

$$|\,0\rangle_L \quad = \quad (|\,0000000\rangle$$

$$
\begin{aligned}
&+ \mid 1010101\rangle \\
&+ \mid 0110011\rangle \\
&+ \mid 0001111\rangle \\
&+ \mid 0111100\rangle \\
&+ \mid 1011010\rangle \\
&+ \mid 1100110\rangle \\
&+ \mid 1101001\rangle)/\sqrt{8}
\end{aligned}
$$

In order to find the second element of C_1/C_2, which we are going to identify with $\mid 1\rangle_L$, we need to find a vector in C_1 that *does not belong to* C_2. Such a vector is, for example, $\mid 1111111\rangle$. Indeed, we have already listed in $\mid 0\rangle_L$ all vectors of C_2 and this one wasn't amongst them. You can see that this element belongs to C_1 if you replace the last row of $\boldsymbol{G}^T$ with the sum of the last two rows and then add all the rows of this new matrix together.

Now we need to form 8 sums of this vector and every vector listed inside $\mid 0\rangle_L$. But remember that adding 1 modulo 2 to a binary number means to negate this number. And so our second vector in C_1/C_2 is:

$$
\begin{aligned}
\mid 1\rangle_L \quad = \quad & (\mid 1111111\rangle \\
& + \mid 0101010\rangle \\
& + \mid 1001100\rangle \\
& + \mid 1110000\rangle \\
& + \mid 1000011\rangle \\
& + \mid 0100101\rangle \\
& + \mid 0011001\rangle \\
& + \mid 0010110\rangle)/\sqrt{8}
\end{aligned}
$$

This seven qubit encoding is due to Steane.

Having constructed a CSS code to protect a qubit we now need to demonstrate *how* the qubit is protected.

Suppose the bit flip errors, which correspond to $\boldsymbol{\sigma}_x$, are described by a binary vector $\boldsymbol{e}_1$. Positions of ones in the vector correspond to the qubits affected by the error. Suppose also the phase errors, which correspond to $\boldsymbol{\sigma}_z$ are described by a binary vector $\boldsymbol{e}_2$. Positions of ones in he vector correspond to the qubits affected by the phase error. Assume that the originally encoded register was of the form

$$\frac{1}{\sqrt{|C_2|}} \sum_{\boldsymbol{y} \in C_2} \mid \boldsymbol{x} +_2 \boldsymbol{y}\rangle$$

After the errors have occured the state of the register changes to

$$\frac{1}{\sqrt{|C_2|}} \sum_{\boldsymbol{y} \in C_2} (-1)^{(\boldsymbol{x}+_2\boldsymbol{y})\cdot\boldsymbol{e}_2} \mid \boldsymbol{x} +_2 \boldsymbol{y} +_2 \boldsymbol{e}_1\rangle$$

Remember that the phase error, $\boldsymbol{\sigma}_z$ negates $\mid 1\rangle$, but leaves $\mid 0\rangle$ alone. This is why we have the $(-1)^{(\boldsymbol{x}+_2\boldsymbol{y})\cdot\boldsymbol{e}_2}$ in front of $\mid \boldsymbol{x} +_2 \boldsymbol{y} +_2 \boldsymbol{e}_1\rangle$. Where $\boldsymbol{x}_k +_2 \boldsymbol{y}_k = 0$

the phase flip does not occur. Where it is 1, the phase flip *may* occur and if the corresponding entry in $\boldsymbol{e}_2$ is 1, it will.

In order to detect and process the error we need to add an ancilla register to the system, in which we are going to store the flip error syndrome. We will use $\boldsymbol{H}_1$ to generate this syndrome. Because $\boldsymbol{x} +_2 \boldsymbol{y} \in C_1$ acting with $\boldsymbol{H}_1$ on $\mid \boldsymbol{x} +_2 \boldsymbol{y} +_2 \boldsymbol{e}_1\rangle$ leaves $\mid \boldsymbol{H}_1\boldsymbol{e}_1\rangle$ in the ancilla register. The state of the whole computer, including the ancilla register, becomes

$$\frac{1}{\sqrt{|C_2|}} \sum_{\boldsymbol{y}\in C_2} (-1)^{(\boldsymbol{x}+_2\boldsymbol{y})\cdot\boldsymbol{e}_2} \mid \boldsymbol{x} +_2 \boldsymbol{y} +_2 \boldsymbol{e}_1\rangle \mid \boldsymbol{H}_1\boldsymbol{e}_1\rangle$$

> We should now stop and ponder on the number of additional qubits we will need in order to construct the ancilla register. The Steane code derives from the $[7,4]$ and $[7,3]$ dual linear codes. In this case matrix $\boldsymbol{H}_1$ is a 3×7 matrix, which means that 3 additional qubits will be required to store syndrome $\boldsymbol{H}_1\boldsymbol{e}_1$.

Having extracted $\boldsymbol{e}_1$ from the ancilla register we can apply NOT gates on lines pointed to by $\boldsymbol{e}_1$, which will convert

$$\mid \boldsymbol{x} +_2 \boldsymbol{y} +_2 \boldsymbol{e}_1\rangle \rightarrow \mid \boldsymbol{x} +_2 \boldsymbol{y} +_2 \boldsymbol{e}_1 +_2 \boldsymbol{e}_1\rangle = \mid \boldsymbol{x} +_2 \boldsymbol{y}\rangle$$

thus eliminating the bit-flip error.

Having repaired the bit-flip error we are now going to work on the phase error. To this effect we are going to apply Hadamard gates to each line of:

$$\frac{1}{\sqrt{|C_2|}} \sum_{\boldsymbol{y}\in C_2} (-1)^{(\boldsymbol{x}+_2\boldsymbol{y})\cdot\boldsymbol{e}_2} \mid \boldsymbol{x} +_2 \boldsymbol{y}\rangle$$

Remember our Hadamard formula:

$$\otimes\boldsymbol{H} \mid \boldsymbol{x}\rangle = \frac{1}{\sqrt{2^n}} \sum_{\boldsymbol{z}=0}^{2^n-1} (-1)^{\boldsymbol{x}\cdot\boldsymbol{z}} \mid \boldsymbol{z}\rangle$$

Applying this formula to our register yields:

$$\begin{aligned} &\frac{1}{\sqrt{|C_2|}}\frac{1}{\sqrt{2^n}} \sum_{\boldsymbol{z}=0}^{2^n-1} \sum_{\boldsymbol{y}\in C_2} (-1)^{(\boldsymbol{x}+_2\boldsymbol{y})\cdot\boldsymbol{z}} (-1)^{(\boldsymbol{x}+_2\boldsymbol{y})\cdot\boldsymbol{e}_2} \mid \boldsymbol{z}\rangle \\ = \quad &\frac{1}{\sqrt{|C_2|}}\frac{1}{\sqrt{2^n}} \sum_{\boldsymbol{z}=0}^{2^n-1} \sum_{\boldsymbol{y}\in C_2} (-1)^{(\boldsymbol{x}+_2\boldsymbol{y})\cdot(\boldsymbol{z}+_2\boldsymbol{e}_2)} \mid \boldsymbol{z}\rangle \end{aligned}$$

Let us now introduce a new variable $\boldsymbol{z}' = \boldsymbol{z} +_2 \boldsymbol{e}_2$. Because of binary arithmetic $\boldsymbol{z} = \boldsymbol{z}' +_2 \boldsymbol{e}_2$, and our state can now be described as:

$$\begin{aligned} &\frac{1}{\sqrt{|C_2|}}\frac{1}{\sqrt{2^n}} \sum_{\boldsymbol{z}'=0}^{2^n-1} \sum_{\boldsymbol{y}\in C_2} (-1)^{(\boldsymbol{x}+_2\boldsymbol{y})\cdot\boldsymbol{z}'} \mid \boldsymbol{z}' +_2 \boldsymbol{e}_2\rangle \\ &= \frac{1}{\sqrt{|C_2|}}\frac{1}{\sqrt{2^n}} \sum_{\boldsymbol{z}'=0}^{2^n-1} \left(\sum_{\boldsymbol{y}\in C_2} (-1)^{\boldsymbol{y}\cdot\boldsymbol{z}'} \right) (-1)^{\boldsymbol{x}\cdot\boldsymbol{z}'} \mid \boldsymbol{z}' +_2 \boldsymbol{e}_2\rangle \end{aligned}$$

As $\boldsymbol{z}'$ runs through 0 to $2^n - 1$ it will run through the dual of C_2, i.e., through $C_2^{\perp}$. As $\boldsymbol{z}'$ stays in $C_2^{\perp}$ it is orthogonal to all $\boldsymbol{y} \in C_2$, so that for these $\boldsymbol{z}'$ we have $(-1)^{\boldsymbol{y}\cdot\boldsymbol{z}'} = (-1)^0 = 1$. Hence for $\boldsymbol{z}' \in C_2^{\perp}$:

$$\sum_{\boldsymbol{y}\in C_2} (-1)^{\boldsymbol{y}\cdot\boldsymbol{z}'} = \sum_{\boldsymbol{y}\in C_2} 1 = |C_2|$$

Now let us assume that $\boldsymbol{z}' \notin C_2^{\perp}$. In this case

$$\sum_{\boldsymbol{y}\in C_2} (-1)^{\boldsymbol{y}\cdot\boldsymbol{z}'} = 0$$

because as $\boldsymbol{y}$ runs through C_2 half of $\boldsymbol{y}\cdot\boldsymbol{z}'$ will be 0 and half will be 1.

> In order to see this better consider again the Steane code. For this code we have that $C_2^{\perp} = C_1$. We also have that $C_2 \subset C_1$. Hence as $\boldsymbol{z}'$ runs through $[0, \ldots, 2^7 - 1]$ it will run through $C_2^{\perp} = C_1$ and then it will also run through all the remaining integers. In the meantime $\boldsymbol{y}$ runs through C_2. We know what are the elements of C_2. They are all the vectors we used to construct $|\ 0\rangle_L$. These vectors also belong to C_1. The other vectors that belong to C_1, but not to C_2 form a layer that is offset from C_2 by $|\ 1111111\rangle$, i.e., they are the vectors that form $|\ 1\rangle_L$. In summary, C_1 is made of vectors, which we have used to form both $|\ 0\rangle_L$ and $|\ 1\rangle_L$. Observe that every vector from C_2 is orthogonal to every vector from C_2 *and* from C_1 as well including itself. Hence for $\boldsymbol{y} \in C_2$ and for $\boldsymbol{z}' \in C_1 = C_2^{\perp}$ we have that $\boldsymbol{y}\cdot\boldsymbol{z}' = 0$.
>
> Now let us take a vector that is *not* in C_1. For example $|\ 0000001\rangle \notin C_1 = C_2^{\perp}$. Taking scalar products of this vector with vectors of C_2 we can see that it is going to be 0 in four cases (all the $\boldsymbol{y}$ vectors that have 0 in the rightmost position) and 1 in the other four cases (the remaining $\boldsymbol{y}$ vectors have 1 in the rightmost position). The same will hold for any other vector $\boldsymbol{z}'$ with 1 in just one position and zeros everywhere else, and, indeed for all vectors not in C_1. So this is how $\sum_{\boldsymbol{y}\in C_2}(-1)^{\boldsymbol{y}\cdot\boldsymbol{z}'} = 0$ for $\boldsymbol{z}' \notin C_2^{\perp}$.

Consequently we can rewrite the state of the register as:

$$\frac{|C_2|}{\sqrt{|C_2|}}\frac{1}{\sqrt{2^n}} \sum_{\boldsymbol{z}'\in C_2^{\perp}} (-1)^{\boldsymbol{x}\cdot\boldsymbol{z}'}\ |\ \boldsymbol{z}' +_2 \boldsymbol{e}_2\rangle$$

But this looks like the bit-flip problem, for which we know the correction procedure already: use the ancilla register and the parity check matrix $\boldsymbol{H}_2$ to find $\boldsymbol{e}_2$ and apply NOT gates to lines pointed to by $\boldsymbol{e}_2$.

> The parity check matrix $\boldsymbol{H}_2$ for the Steane code is going to be a 4×7 matrix. It is simply the generator matrix for C_1 turned sideways. So this will call for additional 4 qubits in the ancilla register.

Having generated

$$\frac{|C_2|}{\sqrt{|C_2|}}\frac{1}{\sqrt{2^n}}\sum_{\boldsymbol{z}'\in C_2^{\perp}}(-1)^{\boldsymbol{x}\cdot\boldsymbol{z}'}\mid \boldsymbol{z}'\rangle$$

we now apply $\otimes\boldsymbol{H}$ again in order to come back to

$$\frac{1}{\sqrt{|C_2|}}\sum_{\boldsymbol{y}\in C_2}\mid \boldsymbol{x}+\boldsymbol{y}\rangle$$

which completes the error detection and correction procedure.

The total number of auxiliary qubits needed by this routine would be 7 in case of Steane code, unless, the same 4 qubit register can be reused for both syndrome measurement operations.

But certain economies are possible, and the size of the ancilla register can be reduced to 6.

The easiest way to construct the relevant error syndrome measurement circuit is to resort to the so called *stabilizer formalizm*. We can't go into the details of this formalizm here, because we don't have enough time for it in this course. You can read more about it in Nielsen and Chuang [80].

The result of applying this formalizm to the Steane code is the so called *check matrix*, which describes the stabilizer *generators* for the Steane code. In its *standard* form this matrix looks as follows:

$$\begin{pmatrix}
1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0
\end{pmatrix}$$

The columns of the matrix correspond to 7 qubits The 1-s in the left half of this 6×14 matrix show the location of $\boldsymbol{\sigma}_x$ operators on the 7 qubit lines if the corresponding entries in the right half of the matrix are 0. The 1-s in the right half of the matrix show the location of $\boldsymbol{\sigma}_z$ operators on the 7 qubit lines if the corresponding entries in the left side of the matrix are 0. If there are 1-s in the same locations in the left and in the right halves of the matrix, these show the location of $\boldsymbol{\sigma}_y$ matrices.

First, you can see that there are no $\boldsymbol{\sigma}_y$ matrices in the error syndrome measurement circuit for the Steane code. $\boldsymbol{\sigma}_x$ and $\boldsymbol{\sigma}_z$ matrices are distributed on the 7 qubit lines as shown in Figure 6.1. Observe the similarity of the distribution to the Steane code check matrix.

The circuit shown in Figure 6.1 is amongst the most complex we have encountered so far. In total we end up using 13 physical qubits to encode and measure error syndrome on just one logical qubit. And we haven't implemented the error correction part of the circuit yet.

It should be quite clear already that quantum error correction procedures, although possible in principle, are much more costly than is the case in classical computing. In the situation in which even a small register

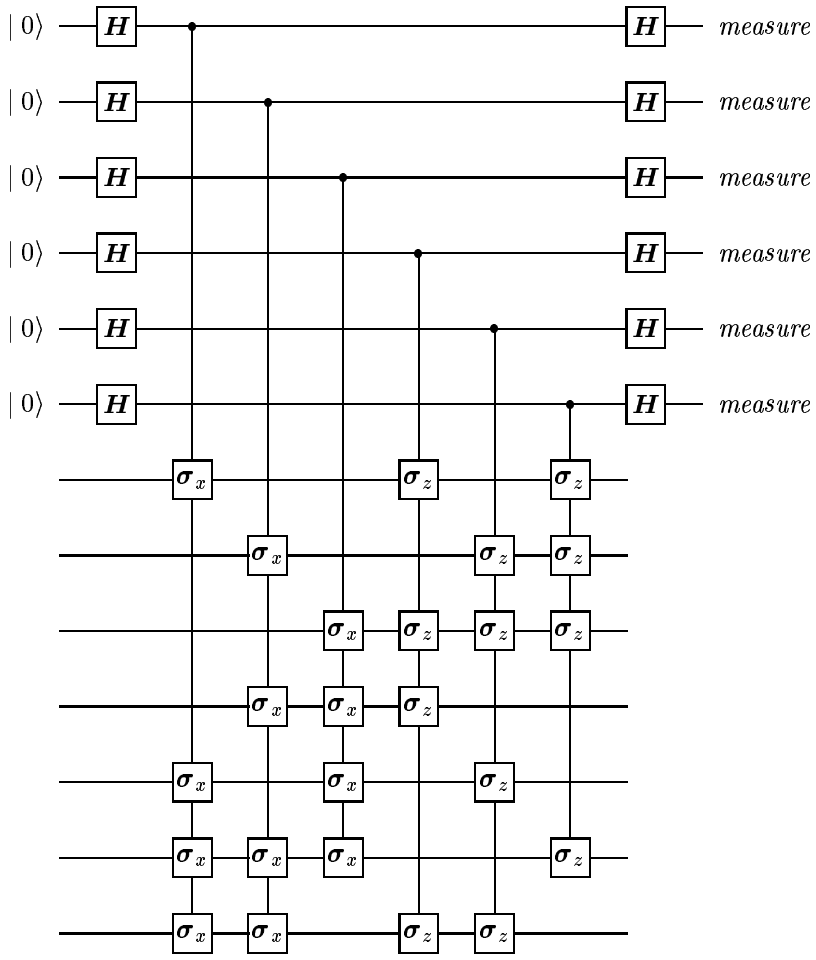


Figure 6.1: A circuit for measuring the error syndrome for the Steane code. The bottom 7 lines represent the 7-qubit register, which encodes a single logical qubit. The top 6 lines represent the ancilla.

of individually controllable qubits with individually controllable couplings is very difficult to construct and operate on (the largest quantum computation carried out so far, at the time of this writing, operated on just five physical qubits) such largesse seems very wasteful.

You will see in the next section that it is possible to carry out quantum computations with arbitrary precision, by stacking up quantum circuits hierarchically. But this will result in even higher consumption of physical qubits.

The obvious question for today therefore is if these procedures are worth the bother. After all, if computations are carried out on a sufficiently large statistical ensemble of raw quantum registers, and if quantum errors are truly random, they should average away in final measurement. The cost here would be gradual loss of signal as the computation proceeds and as errors accumulate. Methods other than software error control could be employed to slow down this process.

6.4 Concatenated Codes

In the previous section we have learnt how to encode a single logical qubit in $7+6 = 13$ physical qubits and how to carry out a procedure checking for possible encoding errors and correcting them.

This is all purely hypothetical, of course, at this stage, because there are no known implementations of any of these procedures – the largest quantum computation carried out so far being a 5 physical qubit one.

But *assuming* that the technology has progressed sufficiently and that we can have as many individually controllable qubits and couplings as we wish, what can we do next with our encoded qubit?

The strategy is to use the encoded qubits in *all* computations without ever having to decode them. This implies that the gates have to be redesigned to work on encoded qubits. Furthermore, gates themselves may have to be designed with fault tolerance in mind too. And so, whereas a normal raw-qubit controlled-NOT gate is a 2×2 gate, its logical qubit counterpart would be a 14×14 gate, not counting the additional 6-qubit ancillas for every logical qubit line.

It turns out that for the Steane code the controlled-NOT gate can be implemented simply as seven controlled-NOT gates applied pair-wise between the two logical qubit blocks, as shown in Figure 6.2.

This gate is *not* fault tolerant as such. But fault tolerance of gates can be implemented by qubit error correcting *after* the application of the gate. This can be seen quite easily as follows. Imagine that an error occurs *before* the application of the gate. Let $\oplus$ stands for the controlled-NOT gate and let the error be described by some unitary operator $\boldsymbol{U}$. The combined action of the error and the gate on the system of two logically encoded qubits is given by

$$\oplus \boldsymbol{U}$$

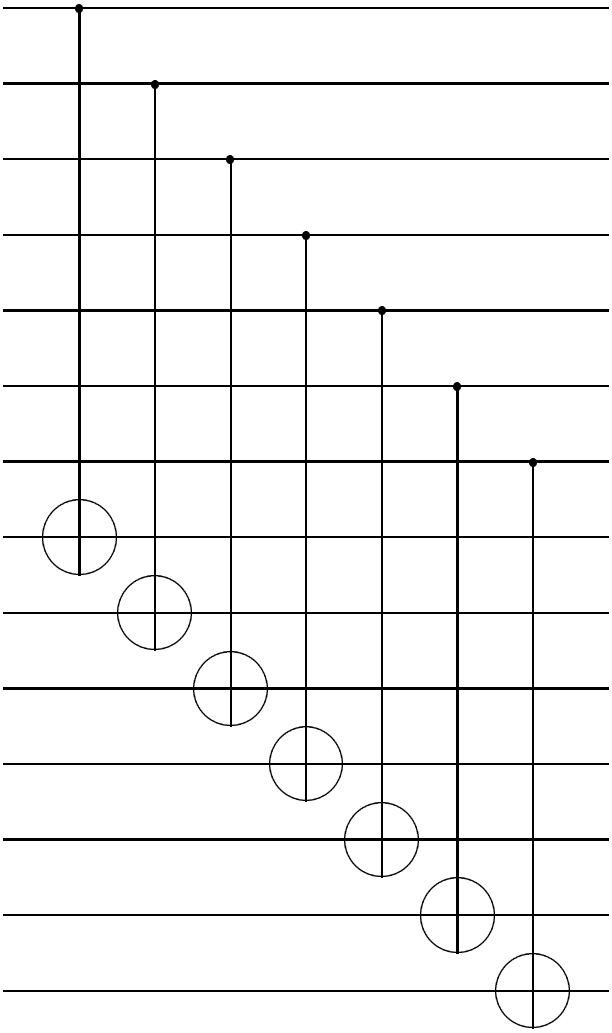
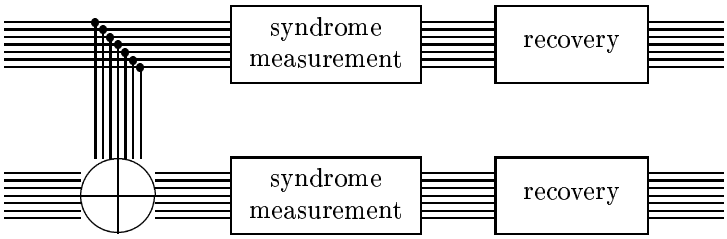


Figure 6.2: A controlled-NOT gate applied to two logical qubits encoded using Steane code.

But this can be rewritten as:

$$\oplus \boldsymbol{U} = \oplus \boldsymbol{U} \oplus^{\dagger} \oplus = \boldsymbol{U}' \oplus$$

In other words, whether the error occurs *before* or *after* the application of the gate makes no difference. If we apply an error correction procedure *after* the gate we'll have, in effect, a fault tolerant gate. A schematic diagram of such a gate is shown below:



Ancilla qubits aren't drawn in this diagram for clarity.

Now, observe that Steane code can detect and correct one error (on 7 physical qubits) only. This means that we cannot have more than one error on input to the upper line and one error on input to the lower line.

> Note that if there is a single error in the upper block of 7 qubits the error will carry to the lower block when the gate is traversed. So we'll end up with two errors on output. But a recovery procedure, having found an error in the upper block should be able to correct the propagated error in the lower block too and then attend to any independent errors in the lower block.

Suppose the probability of a failure on an individual physical qubit in the circuit is p.

Now we have the following seven scenarios.

1. The first scenario is that we have faulty qubits entering the gate in both input blocks. What is the probability of something like this happening? A probability of a failed qubit entering the controlled-NOT gate is $c_1 p$ for the upper block and the same for the lower block, where c_1 is the total number of places at which a failure could have occurred during syndrome measurement and recovery in the previous stage of the circuit (remember that every gate preceding our controlled-NOT is going to have a syndrome measurement and recovery procedure following it). Assuming that both errors in the upper and lower block occur independently the probability of two errors entering the gate is

$$P_1 = (c_1 p)^2 = c_1^2 p^2$$

The value of c_1 can be estimated to be about 10 locations times 7 qubits, i.e.,

$$c_1 \approx 10 \times 7 = 70 \qquad \text{and} \qquad c_1^2 = 4,900$$

2. The second scenario is that only a single error enters the gate. But then when the gate processes the input and then when the output of the gate is processed by the syndrome measurement and recovery procedures another error occurs. The probability of this happening is

$$P_2 = c_2 p^2$$

where c_2 is the number of pairs of points within the circuit shown above where a failure may occur. The value of c_2 is 140 locations times 7 qubits, i.e.,

$$c_2 \approx 140 \times 7 = 980$$

3. The third scenario is that there are no errors on input, but two failures occur during the traversal of the controlled-NOT gate. The probability of this occuring is

$$P_3 = c_3 p^2$$

where c_3 is the number of pairs of points where a failure can occur. This is about 100, i.e.,

$$c_3 \approx 100$$

4. The fourth scenario is that there are no errors on input, but a failure occurs during the traversal of the controlled-NOT gate and then another failure occurs during the syndrome measurement. The probability of this event taking place is

$$P_4 = c_4 p^2$$

where

$$c_4 \approx 100$$

5. The fifth scenario is that two or more failures occur during syndrome measurement. The probability of this happening is

$$P_5 = c_5 p^2$$

where c_5 is the number of pairs of points at which a failure may occur

$$c_5 \approx 70^2 = 4,900$$

6. The sixth scenario is that a failure occurs during syndrome measurement and another failure occurs during recovery. The probability of such an event is

$$P_6 = c_6 p^2$$

where c_6 is the number of pairs of points where a failure may occur. For the Steane code we're considering here it is

$$c_6 \approx 70 \times 7 = 490$$

7. The last scenario is that two or more failures occur during recovery. This is again given by:

$$P_7 = c_7 p^2$$

where

$$c_7 \approx 7^2 = 49$$

In summary the probability that the circuit introduces two or more errors into the encoded block of qubits is

$$\begin{aligned} P &= P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7 \\ &= p^2 \left(c_1^2 + c_2 + c_3 + c_4 + c_5 + c_6 + c_7\right) \\ &= p^2 \left(4,900 + 980 + 100 + 100 + 4,900 + 490 + 49\right) \\ &= p^2 11,519 \\ &= cp^2 \end{aligned}$$

where

$$c = 11,519$$

This means that if the probability of a single error occurring is very small, e.g., $p < 10^{-4}$ then every time we traverse our combination of a gate with error syndrome measurement and recovery procedures we end up reducing the probability of an incorrect output by p (since the first p in p^2 cancels c).

One can carry out similar reasoning for other gates and the value of c that comes out is going to be similar as long as we stick to the Steane code, i.e.,

$$c \approx 10^4$$

Because $P = cp^2 \ll 1$ for $p \ll 1/\sqrt{c}$ (we will actually show later that we really need to have $p < 1/c$), we can consider hierarchical encodings. What if, say, we apply the Steane 7-qubit code to every qubit that is used in a higher level 7-qubit encoding? I.e., instead of encoding one logical qubit in 7 physical qubits we would encode one logical qubit in 7 logical qubits, each of which, in turn would be encoded in 7 physical qubits. We would then have 49 physical qubits encoding a single logical qubit. But why stop at these two levels of encoding? We could have

0. one logical qubit
1. encoded using 7 logical qubits, each of which is
2. encoded using 7 logical qubits, each of which is
3. encoded using 7 physical qubits

So now we'd have $7 \times 7 \times 7 = 343$ qubits encoding a single logical qubit. Are we on the right track here? Is this going to improve things sufficiently to justify this enormous expense in terms of qubits?

Of course, it is not enough to just encode qubits so. We will have to construct fault tolerant quantum gates that operate on hierarchically encoded qubits. But assuming that we have all this in place, if the probability that an error occurs after a gate and correction procedure traversal using a single level encoding is

$$P = cp^2$$

then the probability of an error occurring after a gate and correction procedure traversal for a two level encoding is going to be

$$P = c(cp^2)^2 = (cp)^{2^2}/c$$

and the probability of an error occurring for the 3 level 343 physical qubit encoding is

$$P = c(c(cp^2)^2)^2 = (cp)^{2^3}/c$$

In general then, for a k-th level encoding

$$P = (cp)^{2^k}/c$$

The size of the circuit increases too. The increase is purely exponential, i.e., if a given encoding required d gates then the k-th level encoding will require d^k gates.

Now suppose we have a 0-level encoding circuit with a certain number of logical gates, say, the number of gates is N. Suppose we want to reduce a probability of error in our computation to less than ϵ. The error per each gate is thus ϵ/N. How many levels of concatenation do we have to use in order to achieve this level of accuracy? The answer is

$$\frac{(cp)^{2^k}}{c} \leq \frac{\epsilon}{N}$$

This inequality has a solution for k if

$$p < \frac{1}{c} = p_{\text{th}}$$

This is a *threashold condition* for quantum computation.

> *Provided that the threashold condition is satisfied we can achieve, in theory, arbitrary accuracy in any quantum computation.*

The solution to the threashold equation can be obtained by multiplying both sides by c and then taking logarithm. This yields:

$$2^k \log cp = \log \frac{c\epsilon}{N} \quad \text{or} \quad 2^k = \frac{\log c\epsilon/N}{\log cp}$$

Since

$$2 = d^x$$

we have

$$\log 2 = x \log d \quad \text{hence} \quad x = \log 2 / \log d \quad \text{and} \quad 2 = d^{\log 2 / \log d}$$

We can therefore replace 2 with $d^{\log 2 / \log d}$, which yields

$$d^{k \log 2 / \log d} = \frac{\log c\epsilon / N}{\log cp}$$

or

$$d^k = \left(\frac{\log (c\epsilon / N)}{\log cp} \right)^{\log d / \log 2} = \left(\frac{\log (N / c\epsilon)}{\log (1 / cp)} \right)^{\log d / \log 2}$$

We can thus summarize this section by stating the *threshold theorem* for quantum computation:

> *A logical quantum circuit, which contains N logical gates, can be implemented so that the probability of computational error is less than ϵ by using*
>
> $$\left(\frac{\log (N / c\epsilon)}{\log (1 / cp)} \right)^{\log d / \log 2}$$
>
> *physical gates on hardware whose components fail with probability $p_{\text{th}} = 1/c$.*

Chapter 7

Conclusions

What are the main challenges and focal points of quantum computing today?

The first and perhaps the most important challenge is to deliver a sufficiently large quantum register with individually addressable qubits and individually controllable couplings. The decoherence time for the register should be sufficiently long to allow for a sizeable computation of the order of at least thousands or tens of thousands of gates.

Our discussion in the previous chapter showed that in order to make such a register a basis for fault-tolerant computation, it would have to be very large indeed, at least of the order of 13 physical qubits per one logical qubit. Preferably even more, if we were to use contatenated encodings (49 plus ancillas?). Assuming 32 logical qubits, comparable to present day PCs, and $7 + 6$ qubit encodings, the register would have to comprise 416 physical qubits.

This, as we have already remarked, just about eliminates the possibility of working with fault tolerant registers in the near future, other than for simple demonstrations of a single qubit encoding perhaps. What we may see in the near future then will be quantum computations carried out on raw qubits and on increasingly large registers. Error control in this case would have to be based on working with statistical ensembles of registers. Methods such as NMR and its possible future improved variants seem quite adequate at this stage.

In order to implement fault tolerant quantum computation, we would need registers with a very large number of physical qubits. For a 64-bit qubit system and two levels of concatenation the register would have to have more than 3,000 physical qubits, not counting ancillas. A very large molecule, such as DNA, could perhaps be used in this role, but here the question is how to control individual qubits. The more repetitive the molecule, and DNA is very repetitive, the harder to implement individual qubit control. What I am driving at is that techniques such as NMR probably will not scale to this point. But there is still a lot of life left in NMR computation and it will continue to serve as a fine exploration vehicle in these early years.

The challenge for NMR computing will be to *design and synthesize* a molecule specially for quantum computation and then to develop techniques to automate

the computations. Amongst these may be the development of compilers, which, given a circuit diagram would generate the pulses required to implement the circuit. The next challenge will be to extend the computation time, for example by cooling the sample, and to improve the accuracy of the gates, while still working with raw, unencoded qubits. I think it should be possible to get to tens of raw qubits and thousands, perhaps even tens of thousands of gates along this path.

Whereupon we're going to hit a brick wall.

But this is OK, because by that time other implementations of quantum computers may become a reality, and in the *meantime* we will acquire a lot of practical experience with quantum computing.

What may the other implementations be? Systems based on printing millions of quantum dots and/or Josephson junctions on a chip are current favourites. But it will be very difficult to implement couplings on such systems, and even more difficult to ensure a sufficiently long decoherence time. Solids are dirty and noisy and lightweight particles such as electrons are very susceptible to all that noise.

There are two interesting alternatives to quantum dots and Josephson junctions. One is to use nuclear spins embedded in crystal lattice. This is the Kane computer, which we have mentioned right at the beginning of this course. The other alternative is to use systems based on anyons. In the next lecture series, M744, you will learn more about anyons and, in particular, about their extraordinary stability, which derives from the topology of anyon physics. Today people look at anyons, scratch their chins and mumble "hmmm... this *may* possibly work, but, it's a very exotic and a very risky path to take..." and so it is, but, perhaps with the notable exception of NMR, just about everything else in quantum computing is risky and exotic, so... what the hell, let's tackle it!

This is exactly what we have set to work on here at Indiana University, and if you would like to join us in this effort you should write to Prof. Zhenghan Wang.

The challenges I have outlined so far are hardware challenges, and, of course, these are central, because quantum computing will remain forever in the realm of fantasy, if we are not going to have quantum computers to carry out the computations on.

But the other challenge, which is also very interesting, is what to do with quantum computers, once we are going to have them. Today we know just a handful of quantum algorithms, which are quite unusual and interesting in their own right, and in some cases may have important applications, e.g., to code breaking or data base searches, but there are just a few of them, and they tend to be very similar to each other. So the challenge is to seek new areas of applicability for quantum computing and to seek new forms of quantum algorithms, that would be quite different from what we've seen so far.

Early this year (2001) the National Science Foundation announced a program called "Quantum and Biologically Inspired Computing". Quoting from the Program Description:

This program will try to emphasize two more fundamental, long-term issues in [Quantum Information Science]:

1. research which probes the physical foundations which are relevant not only to QIS but to other areas of future possible technology;
2. strategies to develop quantum computing principles for general-purpose computing and systems-level computing design, and special-purpose algorithms that transcend the limitations of special purpose algorithms now available for niche applications such as cryptography and number theory.

In physical foundations, the areas of interest include (but are not limited to) topics such as:

- Empirically-driven understanding of fundamental decoherence effects, particularly at low temperatures
- Better operational understanding of measurement and temporal effects in measurements of entangled states of all kinds
- Better understanding of novel types of entanglement, such as double entanglement or positional entanglement or $N > 2$ entanglement
- Use of QIS experiments to address issues in the foundations of physics
- Developing a broad and general collection of quantum algorithms
- Extending concepts of information theory to the realm of quantum foundations and experiments
- Strategies to use stable attractors or self-stabilization effects to reduce error rate in QIS

These points hark back to what we have said early on in the preamble to this course. Namely that once you have quantum computers, you can turn the table and use quantum computers, or more generally, quantum information science to explore fundamental physics.

The program statement then continues:

In general-purpose computing and systems-level computing design, there is interest in topics such as (but not limited to):

- Quantum simulation of quantum systems (e.g. molecular modeling)
- Use of learning rather than programming to achieve general-purpose capability (e.g. quantum neural networks, including quantum associative memory and quantum-based stochastic search)

- Use of computer science theory to address broader ranges of computational tasks
- Concepts like quantum fast Fourier transforms and similar approaches to reach a large user base
- Novel approaches to fault tolerance and to managing the stochastic errors in quantum systems.

Here we again see the reference to Feynman's original idea: use quantum computers to attack problems of quantum physics. Although the idea is there, we have not seen many specific quantum algorithms[1] that would address this quite central issue for quantum computing.

The issue of fault tolerance is mentioned last, but it is also a central issue. The existing fault tolerance control methods for quantum computers are clearly too costly. Anyonic computation again may help here and this would certainly be a *novel* and different approach to managing the accuracy of computations in quantum computers. But this would be a hardware solution. There may be alternative software approaches, such as decoherence free spaces, for example, which are much less costly than CSS codes.

All these are open areas for research in quantum computing. The National Science Foundation, in its wisdom, selected these as candidates for special funding this year. This opens numerous opportunities for researchers interested in this topic. The field is new, challenging and exciting. There is a lot here to discover, and a lot that may have implications for fundamental physics even.

[1] I actually haven't seen *any*, but this does not mean that they don't exist. If there is anyone out there who knows of such, I'd be delighted to include them and discuss in these notes.

Bibliography

[1] A. Aspect, J. Dalibard, G. Roger, "Experimental Test of Bell's Inequalities Using Time-Varying Analyzers", Physical Review Letters, Vol. 49, 1982, pp. 1804–1807

[2] D. V. Averin, "Solid-state qubits under control", Nature, Vol. 398, 29 April 1999, pp. 748-749

[3] D. V. Averin and V. J. Goldman, "Quantum computation with quasiparticles of the Fractional Quantum Hall Effect", `arXiv:cond-mat/0110193`, v1, 10 Oct. 2001

[4] John Baez and Javier P. Muniain, "Gauge Fields, Knots and Gravity", Series on Knots and Everything, Vol. 4, World Scientific Publishing, Singapore, 1994, ISBN 9810220340

[5] Ted Bastin (ed.), "Quantum Theory and Beyond: essays and discussions arising from a colloquium", Cambridge [Eng.] University Press, 1971

[6] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, H. Weinfurter, "Elementary gates for quantum computation", Physical Review A, Vol. 52, No. 5, November 1995, pp. 3457-3467

[7] J. Bell, "On the Einstein-Podolsky-Rosen Paradox", Physics, Vol. 1, 1964, pp. 195–200

[8] P. Benioff, "The Computer as a Physical System: A Microscopic Quantum Mechanical Hamiltonian Model of Computers as Represented by Turing Machines", Journal of Statistical Physics, Vol. 22, 1980, pp. 563-591

[9] Charles Bennett, "Logical Reversibility of Computation", IBM Journal of Research and Development, Vol. 17, 1973, pp. 525–532

[10] C. H. Bennett, G. Brassard, C. Crepeau, R. Jozsa, A. Peres, and W. Wootters, "Teleporting an Unknown Quantum State via Dual Classical and EPR Channels", Physical Review Letters, Vol. 70, 1993, pp. 1895–1899

[11] Ethan Bernstein and Umesh Vazirani, "Quantum Complexity Theory", Proceedings of the 25th Annual ACM Symposium on the Theory of Computing, 1993, pp. 11–20

[12] M. V. Berry, "Quantal phase factors accompanying adiabatic changes", Proceedings of the Royal Society of London, Vol. A 392, pp. 45-57, 1984

[13] A. Berthiaume and G. Brassard, "The Quantum Challenge to Complexity Theory", Proceedings of the 7th IEEE Conference on Structure in Complexity Theory", 1992, pp. 132-137

[14] David Bohm and B. J. Hiley, "The Undivided Universe: An Ontological Interpretation of Quantum Theory", Routledge, 1993, 397 pp.

[15] Arno Böhm, "Quantum Mechanics", Springer Verlag, 1979, 522 pp.

[16] Dik Bouwmeester, Jian-Wei Pan, Klaus Mattle, Manfred Eibl, Harald Weinfurter, Anton Zeilinger, "Experimental quantum teleportation", Nature, Vol. 390, No. 6660, pp. 575-579, 11-Dec-1997

[17] Dirk Bouwmeester, Artur Ekert, Anton Zeilinger (editors), "The Physics of Quantum Information", Springer, 2000

[18] G. Brassard, "Teleportation as a Quantum Computation", Proceedings of the 4th Workshop on Physics and Computation, PhysComp96, extended abstract, Los Alamos preprint archive, http://xxx.lanl.gov/archive/quant-ph/9605035, 1996

[19] Samuel L. Braunstein, A. Mann, M. Revzen, "Maximal Violation of Bell Inequalities for Mixed States", Physical Review Letters, Vol. 68, 1992, pp. 3259-3261

[20] Samuel L. Braunstein, H. J. Kimble, "A posteriori teleportation", Nature, Vol. 394, No. 6696, pp. 840-841, 27-Aug-1998

[21] James J. Brophy, "Basic Electronics for Scientists", McGraw-Hill Kogakusha, 1977, 430 pp.

[22] Cristian S. Calude and Gheorghe Paun, "Computing with Cells and Atoms: An Introduction to Quantum, DNA and Membrane Computing", Taylor and Francis, 2001, 242 pp.

[23] Tapash Chakraborty and Pekka Pietiläinen, "The Quantum Hall Effects: Integral and Fractional", Springer Series in Solid-State Sciences, vol. 85, Springer, 1995

[24] Kenneth Chang, "Spin Could Be Quantum Boost for Computers", The New York Times, 21st of August 2001.

[25] Kenneth Chang, "I.B.M. Creates a Tiny Circuit Out of Carbon" The New York Times, 27th of August 2001.

[26] J. I. Cirac and P. Zoller, "Quantum Computations with Cold Trapped Ions", Physical Review Letters, Vol. 74, 1995, pp. 4091-4094

[27] Richard Dalven, "Introduction to Applied Solid State Physics", Plenum Press, 1981, 330 pp.

[28] David Deutsch, "Quantum Theory, the Church-Turing Principle, and the Universal Quantum Computer", Proceedings of the Royal Society of London, Vol. A400, 1985, pp. 97–117

[29] David Deutsch and Richard Jozsa, "Rapid Solution of Problems by Quantum Computation", Proceedings of the Royal Society of London, Vol. 439A, 1992, pp. 553–558

[30] D. DiVincenzo, "Quantum Computation", Science, Vol. 270, 13 October 1995, pp. 255-261

[31] Gerald Dunne, "Self-Dual Chern-Simons Theories", Springer Lecture Notes in Physics, vol. M36, Springer 1995

[32] C. Durr and P. Hoyer, "A Quantum Algorithm for Finding the Minimum", Los Alamos preprint archive, http://xxx.lanl.gov/archive/quant-ph/9607014, 1996

[33] A. Einstein, B. Podolsky and N. Rosen, Physical Review, vol. 47, pp. 777-780, 1935

[34] Richard P. Feynman, "Simulating Physics with Computers", International Journal of Theoretical Physics, Vol. 21, Nos. 6/7, 1982, pp. 467–488

[35] Richard P. Feynman, Robert B. Leighton and Matthew L. Sands, "The Feynman Lectures on Physics", Addison-Wesley, 1989, 3 Volumes

[36] M. H. Freedman, A. Kitaev, and Z. Wang, "Simulation of topological field theories by quantum computers", 17th of March 2000, Physics e-Print archive, http://arXiv.org/abs/quant-ph/0001071

[37] M. H. Freedman, "Quantum Computation and the localization of Modular Functors", 12th of May 2000, Physics e-Print archive, http://arXiv.org/abs/quant-ph/0003128

[38] Linda Geppert, "Quantum transistors: toward nanoelectronics", IEEE Spectrum, September 2000, pp. 46–51

[39] L. Grover, "A Fast Quantum Mechanical Algorithm for Data Base Search", Proceedings of the 28th Annual ACM Symposium on the theory of Copmputing, 1996, pp. 212-219

[40] L. Grover, "A Fast Quantum Mechanical Algorithm for Estimating the Median", AT&T Bell Labs preprint, 1996

[41] Daniel Gottesman and Isaac L. Chuang, "Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations", Nature, Vol. 402, 1999, pp. 390–393

[42] HPCwire 15349, "NEC Researchers Take Major Step Toward Quantum Computing", Science and Engineering News, 30th of April 1999

[43] HPCwire 16803, "Bell Labs Demonstrates Sub-50-nanometer MOSFET", Science and Engineering News, 17th of December 1999

[44] HPCwire 16812, "Intel to Debut 1 GHz Chip in February", Commercial News, 17th of December 1999

[45] HPCwire 17320, "Los Alamos Scientists Make Seven Bit Quantum Leap", Features and Commentary, 24th of March 2000

[46] HPCwire 18307, "Team Uses Quantum Computer to Solve Problem", News Flash, 15th of August 2000

[47] HPCwire 18353, "New Insight into Quantum Superconductivity", Science and Engineering News, 29th of August 2000

[48] HPCwire 18359, "Scientists Advance to Make Electronics Tinier", Science and Engineering News, 25th of August 2000

[49] HPCwire 18405, "Dead times between photon emissions prove theory", Science and Engineering News, 1st of September 2000

[50] R. Hughes, D. James, E. Knill, R. Laflamme, A. Petschek, "Decoherence Bounds on Quantum Computation with Trapped Ions", Physical Review Letters, Vol. 77, 1996, pp. 3240-3243

[51] Lev B. Ioffe, Vadim B. Geshkenbein, Mikhail V. Feigel'man, Alban L. Fauchère and Gianni Blatter, "Environmentally decoupled *sds*-wave Josephson junctions for quantum computing", Nature, Vol. 398, 22 April 1999, pp. 679-681

[52] Josef M. Jauch, "Foundation of Quantum Mechanics", Addison-Wesley, 1968, 299 pp.

[53] Jonathan A. Jones, Vlatko Vedral, Artur Ekert, Giuseppe Castagnoli, "Geometric quantum computation using nuclear magnetic resonance", Nature, Vol. 403, pp. 869-871, February 2000

[54] E. Joos and H. Zeh, "The Emergence of Classical Properties Through Interaction with the Environment", Zeitschrift für Physik B, Vol. 59, 1985, pp. 223-243

[55] Richard Jozsa, "Characterizing Classes of Functions Computable by Quantum Parallelism", Proceedings of the Royal Society of London, Vol. A435, pp. 563–574, 1991

[56] B. E. Kane, "A silicon-based nuclear spin quantum computer", Nature, Vol. 393, p. 133, 1998

[57] Louis H. Kauffman, "Quantum Topology and Quantum Computing", AMS Short Course on Quantum Computation, Washington, D. C., January 17-18, 2000

[58] Avinash Khare, "Fractional Statistics and Quantum Theory", World Scientific, 1997

[59] D. Kielpinski, V. Meyer, M. A. Rowe, C. A. Sackett, W. M. Itano, C. Monroe, D. J. Wineland, "A Decoherence-Free Quantum Memory Using Trapped Ions", Science, Vol. 291, 9 February 2001

[60] Alexei Kitaev, "Quantum Measurements and the Abelian Stabiliser Problem", Los Alamos preprint archive, http://xxx.lanl.gov/archive/quant-ph/9511026, 1995

[61] E. Knill, R. Laflamme, R. Martinez, C. H. Tseng, "An algorithmic benchmark for quantum information processing", Nature, Vol. 404, 23 March 2000

[62] R. Laflamme, E. Knill, W. H. Zurek, P. Catasti, S. V. S. Mariappan, "NMR Greenberger-Horne-Zeilinger states", Philosophical Transactions of the Royal Society of London, Vol. A 356, pp. 1941–1947, 1998

[63] R. Laflamme, C. Miguel, P. Paz and W. Zurek, "Perfect Quantum Error Correcting Code", Physical Review Letters, Vol. 77, 1996, pp. 198-201

[64] Rolf Landauer, "Information is Physical", Physics Today, Vol. 44, 1991, pp. 23–29

[65] R. B. Laughlin, Physical Review Letters, vol. 50, p. 1395, 1983

[66] J. M. Leinaas and J. Myrheim, Nuovo Cimento, B37, 1977

[67] Debbie W. Leung, Isaac L. Chuang, Fumiko Yamaguchi, Yoshihisa Yamamoto, "Efficient implementation of coupled logic gates for quantum computation", Physical Review A, Vol. 61, 16 March 2000

[68] Daniel A. Lidar, David Bacon, Julia Kempe and K. Brigitta Whaley, "Protecting Quantum Information Encoded in Decoherence Free States Against Exchange Errors", arXiv:quant-ph/9907096 v3, 2 Dec 1999

[69] S. Lloyd, "A Potentially Realizable Quantum Computer", Science, Vol. 261, 17 September 1993, pp. 1569–1571

[70] Hoi-Kwong Lo, Sandu Popescu and Tim Spiller (ed.), "Introduction to Quantum Computation and Information", World Scientific, 1999

[71] Daniel Loss and David P. DiVincenzo, "Quantum computation with quantum dots", Physical Review A, Vol. 57, No. 1, January 1998, pp. 120–126

[72] J. Lubanski, Acta Physica Polonica, Vol. VI, Fasc. IV, 1937

[73] I. Malajovich, J. J. Berry, N. Samarth, D. D. Awschalom, "Persistent sourcing of coherent spins for multifunctional semiconductor spintronics", Nature, vol. 411, pp. 770-772, 2001

[74] M. Mathisson, Acta Physica Polonica, Vol. VI, Fasc. III, 1937

[75] P. Michler, A. Imamoglu, M. D. Mason, P. J. Carson, G. F. Strouse, S. K. Buratto, "Quantum correlation among photons from a single quantum dot at room temperature", Nature, Vol. 406, pp. 968–970, 31st August 2000

[76] C. W. Misner, K. S. Thorne and J. A. Wheeler, "Gravitation", W. H. Freeman and Company, San Francisco, 1973

[77] C. Monroe, D. Meekhof, B. King, W. Itano, and D. Wineland, "Demonstration of a fundamental Quantum Logic Gate", Physical Review Letters, Vol. 75, No. 25, 1995, pp. 4714-4717

[78] Christopher Moore, "Predictability and Undecidability in Dynamical Systems", Physical Review Letters, Vol. 64, 1990, pp. 2354–2357

[79] Y. Nakamura, Yu. A. Pashkin, and J. S. Tsai, "Coherent control of macroscopic quantum states in a single-Cooper-pair box", Nature, 398, 1999, pp. 786–788

[80] M. A. Nielsen and Isaac L. Chuang, "Quantum Computation and Quantum Information", Cambridge University Press, 2000, ISBN 0521635039, 700 pages

[81] M. A. Nielsen, E. Knill, and R. LaFlamme, "Complete quantum teleportation using nuclear magnetic resonance", Nature, Vol. 396, No. 6706, 1998, pp. 52-55

[82] Jian-Wei Pan, Dik Bouwmeester, Matthew Daniell, Harald Weinfurter, Anton Zeilinger, "Experimental test of quantum nonlocality in three-photon Greenberger-Horne-Zeilinger entanglement", Nature, Vol. 403, 2000, pp. 515-519

[83] A. Papapetrou, Proceedings of the Royal Society, A 209, 1951

[84] Asher Peres and Wojciech Zurek, "Is Quantum Theory Universally Valid?", American Journal of Physics, Vol. 50, September 1982, pp. 807-810

[85] Asher Peres, "Einstein, Gödel, Bohr", Foundations of Physics, Vol. 15, 1985, pp. 201-205

[86] Asher Peres, "Quantum Theory: Concepts and Methods", Kluwer Academic, Dordrecht/Boston, 1993, 446 pp., ISBN 0792325494

[87] John Preskill and Alexei Kitaev, "Lecture Notes for Physics 229, Quantum Information and Computation", on-line notes, http://www.theory.caltech.edu/people/preskill/ph229

[88] F. Rohrlich, "Classical Charged Particles; Foundations of Their Theory", Addison-Wesley, Reading, Mass., 1965

[89] B. Schwarzschild, "Labs Demonstrate Logic Gates for Quantum Computation", Physics Today, March 1996, pp. 21-23

[90] Semiconductor Nanofabrication Facility, University of New South Wales, Sydney, Australia, "Quantum Computation", http://www.snf.unsw.edu.au/

[91] Mark S. Sherwin, Atac Imamoglu, and Thomas Montroy, "Quantum computation with quantum dots and terahertz cavity quantum electrodynamics", Physical Review A, Vol. 60, No. 5, November 1999, pp. 3508-3514

[92] Peter Shor, "Algorithms for Quantum Computation, Discrete Logarithms and Factoring", Proceedings of the 35th Annual Symposium on Foundations of Computer Science, 1994, pp. 124–134

[93] D. Simon, "On the Power of Quantum Computation", Proceedings of the 35th Annual IEEE Symposium on Foundations of Computer Science, 1994, pp. 116-123

[94] G. L. Snider, A. O. Orlov, I. Amlani, X. Zuo, G. H. Bernstein, C. S. Lent, J. L. Merz, and W. Porod, "Quantum-dot cellular automata: Review and recent experiments (invited)", Journal of Applied Physics, Vol. 85, Issue 8, pp. 4283–4285, 15th of April 1999

[95] Thomas Sterling, "Achieving Petaflops-scale Performance through a Synthesis of Advanced Device Technologies and Adaptive Latency Tolerant Architecture", Proceedings of the SC99 Conference, Portland, Oregon, November 13-19, 1999, Invited Talks, http://www.sc99.org/proceedings/invtalk.htm#sterling

[96] W. Teich, K. Obermayer, and G. Mahler, "Structural Basis of Multistationary Quantum Systems. II. Effective Few-Particle Dynamics", Physical Review B, Vol. 37, No. 14, 1988, pp. 8111-8120

[97] Paul A. Tipler and Ralph A. Llewellyn, "Modern Physics", Third Edition, W. H. Freeman and Co., 1999, 14 chapters and appendices

[98] T. Toffoli, "Bicontinuous extensions of invertible combinatorial functions", Mathematical Systems Theory, Vol 14, 1981, pp. 13–23

[99] A. Tonomura, H. Kasai, O. Kamimura, T. Matsuda, K. Harada, Y. Nakayama, J. Shimoyama, K. Kishio, T. Hanaguri, K. Kitazawa, M. Sasase, S. Okayasu, "Observation of individual vortices trapped along columnar defects in high-temperature superconductors", Nature, vol. 412, pp. 620-622, 2001

[100] Q. Turchette, C. Hood, C. Lange, W. Mabuchi, H. Kimble, "Measurement of Conditional Phase Shifts for Quantum Logic", Physical Review Letter, Vol. 75, No. 25, 1995, pp. 4710-4713

[101] Alan Turing, "On Computable Numbers with an Application to the Entscheidungsproblem", Proceedings of the London Mathematical Society, vol. 42, 1937, pp. 230-265, erratum in 43, 1937, pp. 544-546

[102] Lieven M. K. Vendersypen, Matthias Steffen, Gregory Breyta, Constantino S. Yannoni, Richard Cleve, Isaac L. Chuang, "5 qubit 215 Hz Quantum Processor", 12th Annual Hot Chips Conference in Palo Alto, Stanford University, 2000

[103] F. R. Waugh, M. J. Berry, C. H. Crouch, C. Livermore, D. J. Mar, R. M. Wetervelt, K. L. Campman, C. Gossard, "Measuring interactions between tunnel-coupled quantum dots", Physical Review B, Vol. 53, No. 3, January 15, 1996, pp. 1413-1420

[104] Colin P. Williams and Scott H. Clearwater, "Explorations in Quantum Computing", Springer Verlag, 1998, ISBN 0-387-94768-X, 307 pp.

[105] A. Yao, "Quantum Circuit Complexity", Proceedings of the 34th IEEE Symposium on Foundations of Computer Science, IEEE Computer Society Press, Los Alamitos, CA, 1993, pp. 352-360

[106] Wojciech Hubert Zurek, "Sub-Planck structure in phase space and its relevance for quantum decoherence", Nature 412, pp. 712-717, 2001

[107] Wojciech Hubert Zurek, "Decoherence, einselection, and the quantum origins of the classical", arXiv:quant-ph/0105127, v1, 24 May 2001

Index