An Introduction to Quantum Computing for Non-Physicists

ELEANOR RIEFFEL
FX Palo Alto Laboratory
AND
WOLFGANG POLAK

Richard Feynman's observation that certain quantum mechanical effects cannot be simulated efficiently on a computer led to speculation that computation in general could be done more efficiently if it used these quantum effects. This speculation proved justified when Peter Shor described a polynomial time quantum algorithm for factoring integers.

In quantum systems, the computational space increases exponentially with the size of the system, which enables exponential parallelism. This parallelism could lead to exponentially faster quantum algorithms than possible classically. The catch is that accessing the results, which requires measurement, proves tricky and requires new nontraditional programming techniques.

The aim of this paper is to guide computer scientists through the barriers that separate quantum computing from conventional computing. We introduce basic principles of quantum mechanics to explain where the power of quantum computers comes from and why it is difficult to harness. We describe quantum cryptography, teleportation, and dense coding. Various approaches to exploiting the power of quantum parallelism are explained. We conclude with a discussion of quantum error correction.

Categories and Subject Descriptors: A.1 [**Introductory and Survey**]

General Terms: Algorithms, Security, Theory

Additional Key Words and Phrases: Quantum computing, complexity, parallelism

1. INTRODUCTION

Richard Feynman observed in the early 1980s [Feynman 1982] that certain quantum mechanical effects cannot be simulated efficiently on a classical computer. This observation led to speculation that perhaps computation in general could be done more efficiently if it made use of these quantum effects. But building quantum computers, computational machines that use such quantum effects, proved tricky, and as no one was sure how to use the quantum effects to speed up computation, the field developed slowly. It wasn't until 1994, when Peter Shor surprised the world by describing a polynomial time quantum algorithm for factoring integers [Shor 1994; 1997], that the field of quantum computing came into its own. This discovery prompted a flurry of activity among experimentalists trying to build quantum computers and theoreticians trying to find other quantum algorithms.

Authors' address: E. Rieffel, FX Palo Alto Laboratory, 3400 Hillview Av., Palo Alto, CA 94304; W. Polak, Consultant.

Additional interest in the subject has been created by the invention of quantum key distribution and, more recently, popular press accounts of experimental successes in quantum teleportation and the demonstration of a 3-bit quantum computer.

The aim of this paper is to guide computer scientists and other nonphysicists through the conceptual and notational barriers that separate quantum computing from conventional computing and to acquaint them with this new and exciting field. It is important for the computer science community to understand these new developments since they may radically change the way we have to think about computation, programming, and complexity.

Classically, the time it takes to do certain computations can be decreased by using parallel processors. To achieve an exponential decrease in time requires an exponential increase in the number of processors, and hence an exponential increase in the amount of physical space needed. However, in quantum systems the amount of parallelism increases exponentially with the size of the system. Thus, an exponential increase in parallelism requires only a linear increase in the amount of physical space needed. This effect is called quantum parallelism [Deutsch and Jozsa 1992].

There is a catch, and a big catch at that. While a quantum system can perform massive parallel computation, access to the results of the computation is restricted. Accessing the results is equivalent to making a measurement, which disturbs the quantum state. This problem makes the situation, on the face of it, seem even worse than the classical situation; we can only read the result of one parallel thread, and because measurement is probabilistic, we cannot even choose which one we get.

But in the past few years, various people have found clever ways of finessing the measurement problem to exploit the power of quantum parallelism. This sort of manipulation has no classical analog and requires nontraditional programming techniques. One technique manipulates the quantum state so that a common property of all of the output values such as the symmetry or period of a function can be read off. This technique is used in Shor's factorization algorithm. Another technique transforms the quantum state to increase the likelihood that output of interest will be read. Grover's search algorithm makes use of such an amplification technique. This paper describes quantum parallelism in detail, and the techniques currently known for harnessing its power.

Section 2, following this introduction, explains of the basic concepts of quantum mechanics that are important for quantum computation. This section cannot give a comprehensive view of quantum mechanics. Our aim is to provide the reader with tools in the form of mathematics and notation with which to work with the quantum mechanics involved in quantum computation. We hope that this paper will equip readers well enough that they can freely explore the theoretical realm of quantum computing.

Section 3 defines the quantum bit, or qubit. Unlike classical bits, a quantum bit can be put in a superposition state that encodes both 0 and 1. There is no good classical explanation of superpositions: a quantum bit representing 0 and 1 can neither be viewed as "between" 0 and 1 nor can it be viewed as a hidden unknown state that represents either 0 or 1 with a certain probability. Even single quantum bits enable interesting applications. We describe the use of a single quantum bit for secure key distribution.

But the real power of quantum computation derives from the exponential state spaces of multiple quantum bits: just as a single qubit can be in a superposition of 0 and 1, a register of n qubits can be in a superposition of all 2^n possible values. The "extra" states that have no classical analog and lead to the exponential size of the quantum state space are the entangled states, like the state leading to the famous EPR[1] paradox (see Section 3.4).

We discuss the two types of operations a quantum system can undergo:

[1] EPR = Einstein, Podolsky, and Rosen

measurement and quantum state transformations. Most quantum algorithms involve a sequence of quantum state transformations followed by a measurement. For classical computers there are sets of gates that are universal in the sense that any classical computation can be performed using a sequence of these gates. Similarly, there are sets of primitive quantum state transformations, called quantum gates, that are universal for quantum computation. Given enough quantum bits, it is possible to construct a universal quantum Turing machine.

Quantum physics puts restrictions on the types of transformations that can be done. In particular, all quantum state transformations, and therefore all quantum gates and all quantum computations, must be reversible. Yet all classical algorithms can be made reversible and can be computed on a quantum computer in comparable time. Some common quantum gates are defined in Section 4.

Two applications combining quantum gates and entangled states are described in Section 4.2: teleportation and dense coding. Teleportation is the transfer of a quantum state from one place to another through classical channels. That teleportation is possible is surprising, since quantum mechanics tells us that it is not possible to clone quantum states or even measure them without disturbing the state. Thus, it is not obvious what information could be sent through classical channels that could possibly enable the reconstruction of an unknown quantum state at the other end. Dense coding, a dual to teleportation, uses a single quantum bit to transmit two bits of classical information. Both teleportation and dense coding rely on the entangled states described in the EPR experiment.

It is only in Section 5 that we see where an exponential speed-up over classical computers might come from. The input to a quantum computation can be put in a superposition state that encodes all possible input values. Performing the computation on this initial state will result in superposition of all of the corresponding output values. Thus, in the same time it takes to compute the output for a single input state on a classical computer, a quantum computer can compute the values for all input states. This process is known as quantum parallelism. However, measuring the output states will randomly yield only one of the values in the superposition, and at the same time destroy all of the other results of the computation. Section 5 describes this situation in detail. Sections 6 and 7 describe techniques for taking advantage of quantum parallelism inspite of the severe constraints imposed by quantum mechanics on what can be measured.

Section 6 describes the details of Shor's polynomial time factoring algorithm. The fastest known classical factoring algorithm requires exponential time, and it is generally believed that there is no classical polynomial time factoring algorithm. Shor's is a beautiful algorithm that takes advantage of quantum parallelism by using a quantum analog of the Fourier transform.

Lov Grover developed a technique for searching an unstructured list of n items in $O(\sqrt{n})$ steps on a quantum computer. Classical computers can do no better than $O(n)$, so unstructured search on a quantum computer is provably more efficient than search on a classical computer. However, the speed-up is only polynomial, not exponential, and it has been shown that Grover's algorithm is optimal for quantum computers. It seems likely that search algorithms that could take advantage of some problem structure could do better. Tad Hogg, among others, has explored such possibilities. We describe various quantum search techniques in Section 7.

It is as yet unknown whether the power of quantum parallelism can be harnessed for a wide variety of applications. One tantalizing open question is whether quantum computers can solve NP-complete problems in polynomial time.

Perhaps the biggest open question is whether useful quantum computers can be built. There are a number of proposals for building quantum computers using ion traps, nuclear magnetic resonance (NMR), and optical and solid-state techniques. All of the current proposals have

scaling problems, so a breakthrough will be needed to go beyond tens of qubits to hundreds of qubits. While both optical and solid-state techniques show promise, NMR and ion trap technologies are the most advanced so far.

In an ion trap quantum computer [Circ and Zoller 1995; Steane 1996] a linear sequence of ions representing the qubits are confined by electric fields. Lasers are directed at individual ions to perform single-bit quantum gates. Two-bit operations are realized by using a laser on one qubit to create an impulse that ripples through a chain of ions to the second qubit, where another laser pulse stops the rippling and performs the 2-bit operation. The approach requires that the ions be kept in extreme vacuum and at extremely low temperatures.

The NMR approach has the advantage that it will work at room temperature and that NMR technology in general is already fairly advanced. The idea is to use macroscopic amounts of matter and encode a quantum bit in the average spin state of a large number of nuclei. The spin states can be manipulated by magnetic fields, and the average spin state can be measured with NMR techniques. The main problem with the technique is that it doesn't scale well; the measured signal scales as $1/2^n$ with the number of qubits n. However, a recent proposal [Schulman and Vazirani 1998] has been made that may overcome this problem. NMR computers with three qubits have been built successfully [Cory et al. 1998; Gershenfeld and Chuang 1997; Laflamme et al. 1997; Vandersypen et al. 1999]. This paper will not discuss further the physical and engineering problems of building quantum computers.

The greatest problem for building quantum computers is decoherence, the distortion of the quantum state due to interaction with the environment. For some time it was feared that quantum computers could not be built because it would be impossible to isolate them sufficiently from the external environment. The breakthrough came from the algorithmic rather than the physical side, through the invention of quantum error correction techniques. Initially people thought quantum error correction might be impossible because of the impossibility of reliably copying unknown quantum states, but it turns out that it is possible to design quantum error correcting codes that detect certain kinds of errors and enable the reconstruction of the exact error-free quantum state. Quantum error correction is discussed in Section 8.

Appendices provide background information on tensor products and continued fractions.

2. QUANTUM MECHANICS

Quantum mechanical phenomena are difficult to understand, since most of our everyday experiences are not applicable. This paper cannot provide a deep understanding of quantum mechanics (see Feynman et al. [1965], Liboff [1997], and Greenstein and Zajonc [1997] for expositions of quantum mechanics). Instead, we will give some feeling as to the nature of quantum mechanics and some of the mathematical formalisms needed to work with quantum mechanics to the extent needed for quantum computing.

Quantum mechanics is a theory in the mathematical sense: it is governed by a set of axioms. The consequences of the axioms describe the behavior of quantum systems. The axioms lead to several apparent paradoxes: in the Compton effect it appears as if an action precedes its cause; the EPR experiment makes it appear as if action over a distance faster than the speed of light is possible. We will discuss the EPR experiment in detail in Section 3.4. Verification of most predictions is indirect, and requires careful experimental design and specialized equipment. We will begin, however, with an experiment that requires only readily available equipment and that will illustrate some of the key aspects of quantum mechanics needed for quantum computation.

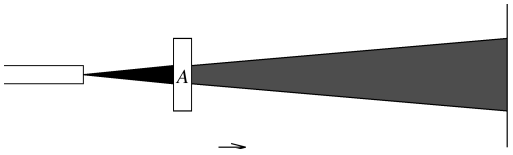
2.1. Photon Polarization

Photons are the only particles that we can observe directly. The following simple

experiment can be performed with minimal equipment: a strong light source, such as a laser pointer, and three polaroids (polarization filters), which can be picked up at any camera supply store. The experiment demonstrates some of the principles of quantum mechanics through photons and their polarization.

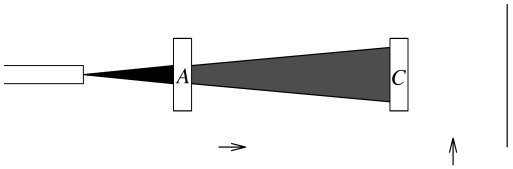
2.1.1 The Experiment. A beam of light shines on a projection screen. Filters A, B, and C are polarized horizontally, at 45°, and vertically, respectively, and can be placed so as to intersect the beam of light.

First, insert filter A. Assuming the incoming light is randomly polarized, the intensity of the output will have half of the intensity of the incoming light. The outgoing photons are now all horizontally polarized.

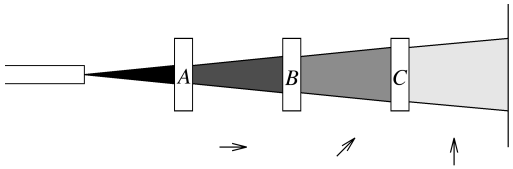


The function of filter A cannot be explained as a "sieve" that only lets those photons pass that happen to be already horizontally polarized. If that were the case, few of the randomly polarized incoming electrons would be horizontally polarized, so we would expect a much larger attenuation of the light as it passes through the filter.

Next, when filter C is inserted, the intensity of the output drops to zero. None of the horizontally polarized photons can pass through the vertical filter. A sieve model could explain this behavior.



Finally, after filter B is inserted between A and C, a small amount of light will be visible on the screen, exactly one eighth of the original amount of light.



Here we have a nonintuitive effect. Classical experience suggests that adding a filter should only be able to decrease the number of photons getting through. How can it increase it?

2.1.2 The Explanation. A photon's polarization state can be modeled by a unit vector pointing in the appropriate direction. Any arbitrary polarization can be expressed as a linear combination $a|\uparrow\rangle + b|\rightarrow\rangle$ of the two basis vectors[2] $|\rightarrow\rangle$ (horizontal polarization) and $|\uparrow\rangle$ (vertical polarization).

Since we are only interested in the direction of the polarization (the notion of "magnitude" is not meaningful), the state vector will be a unit vector (i.e., $|a|^2+|b|^2 = 1$). In general, the polarization of a photon can be expressed as $a|\uparrow\rangle + b|\rightarrow\rangle$ where a and b are complex numbers[3] such that $|a|^2 + |b|^2 = 1$. Note, the choice of basis for this representation is completely arbitrary: any two orthogonal unit vectors will do (e.g., $\{|\nwarrow\rangle, |\nearrow\rangle\}$).

The measurement postulate of quantum mechanics states that any device measuring a two-dimensional system has an associated orthonormal basis with respect to which the quantum measurement takes place. Measurement of a state transforms the state into one of the measuring device's associated basis vectors. The probability that the state is measured as basis vector $|u\rangle$ is the square of the norm of the amplitude of the component of the original state in the direction of the basis vector $|u\rangle$. For example, given a device

[2] The notation $|\rightarrow\rangle$ is explained in Section 2.2.

[3] Imaginary coefficients correspond to circular polarization.

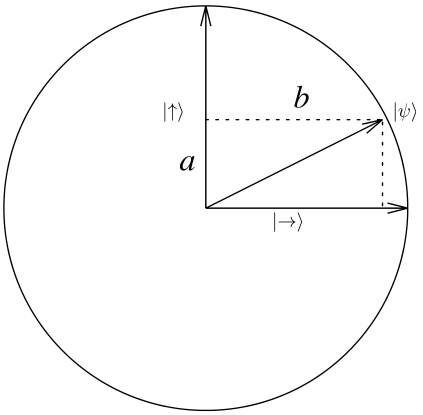


Fig. 1. Measurement is a projection onto the basis.

for measuring the polarization of photons with associated basis $\{|\uparrow\rangle, |\rightarrow\rangle\}$, the state $|\psi\rangle = a|\uparrow\rangle + b|\rightarrow\rangle$ is measured as $|\uparrow\rangle$ with probability $|a|^2$ and as $|\rightarrow\rangle$ with probability $|b|^2$ (see Figure 1). Note that different measuring devices will have different associated bases, and measurements using these devices will have different outcomes. As measurements are always made with respect to an orthonormal basis, throughout the rest of this paper all bases will be assumed to be orthonormal.

Furthermore, measurement of the quantum state will change the state to the result of the measurement. That is, if measurement of $|\psi\rangle = a|\uparrow\rangle + b|\rightarrow\rangle$ results in $|\uparrow\rangle$, then the state ψ changes to $|\uparrow\rangle$ and a second measurement with respect to the same basis will return $|\uparrow\rangle$ with probability 1. Thus, unless the original state happened to be one of the basis vectors, measurement will change that state, and it is not possible to determine what the original state was.

Quantum mechanics can explain the polarization experiment as follows. A polaroid measures the quantum state of photons with respect to the basis consisting of the vector corresponding to its polarization together with a vector orthogonal to its polarization. The photons that, after being measured by the filter, match the filter's polarization are let through. The others are reflected and now have a polarization perpendicular to that of the filter. For example, filter A measures the photon polarization with respect to the basis vector $|\rightarrow\rangle$, corresponding to its polarization. The photons that pass through filter A all have polarization $|\rightarrow\rangle$. Those that are reflected by the filter all have polarization $|\uparrow\rangle$.

Assuming that the light source produces photons with random polarization, filter A will measure 50% of all photons as horizontally polarized. These photons will pass through the filter and their state will be $|\rightarrow\rangle$. Filter C will measure these photons with respect to $|\uparrow\rangle$. But the state $|\rightarrow\rangle = 0|\uparrow\rangle + 1|\rightarrow\rangle$ will be projected onto $|\uparrow\rangle$ with probability 0, and no photons will pass filter C.

Finally, filter B measures the quantum state with respect to the basis

$$\left\{\frac{1}{\sqrt{2}}(|\uparrow\rangle + |\rightarrow\rangle), \frac{1}{\sqrt{2}}(|\uparrow\rangle - |\rightarrow\rangle)\right\}$$

which we write as $\{|\nearrow\rangle, |\nwarrow\rangle\}$. Note that $|\rightarrow\rangle = \frac{1}{\sqrt{2}}(|\nearrow\rangle - |\nwarrow\rangle)$ and $|\uparrow\rangle = \frac{1}{\sqrt{2}}(|\nearrow\rangle + |\nwarrow\rangle)$. Those photons that are measured as $|\nearrow\rangle$ pass through the filter. Photons passing through A with state $|\rightarrow\rangle$ will be measured by B as $|\nearrow\rangle$ with probability 1/2, and so 50% of the photons passing through A will pass through B and be in state $|\nearrow\rangle$. As before, these photons will be measured by filter C as $|\uparrow\rangle$ with probability 1/2. Thus only one eighth of the original photons manage to pass through the sequence of filters A, B, and C.

2.2. State Spaces and Bra/Ket Notation

The state space of a quantum system, consisting of the positions, momentums, polarizations, spins, and so on of the various particles, is modeled by a Hilbert space of wave functions. We will not look at the details of these wave functions. For quantum computing we need only deal with finite quantum systems and it suffices to consider finite dimensional complex vector spaces with an inner product that are spanned by abstract wave functions such as $|\rightarrow\rangle$.

Quantum state spaces and the tranformations acting on them can be described

in terms of vectors and matrices or in the more compact bra/ket notation invented by Dirac [1958]. Kets like $|x\rangle$ denote column vectors and are typically used to describe quantum states. The matching bra, $\langle x|$, denotes the conjugate transpose of $|x\rangle$. For example, the orthonormal basis $\{|0\rangle, |1\rangle\}$ can be expressed as $\{(1,0)^T, (0,1)^T\}$. Any complex linear combination of $|0\rangle$ and $|1\rangle$, $a|0\rangle + b|1\rangle$, can be written $(a,b)^T$. Note that the choice of the order of the basis vectors is arbitrary. For example, representing $|0\rangle$ as $(0,1)^T$ and $|1\rangle$ as $(1,0)^T$ would be fine as long as this is done consistently.

Combining $\langle x|$ and $|y\rangle$ as in $\langle x||y\rangle$, also written as $\langle x|y\rangle$, denotes the inner product of the two vectors. For instance, since $|0\rangle$ is a unit vector we have $\langle 0|0\rangle = 1$ and since $|0\rangle$ and $|1\rangle$ are orthogonal we have $\langle 0|1\rangle = 0$.

The notation $|x\rangle\langle y|$ is the outer product of $|x\rangle$ and $\langle y|$. For example, $|0\rangle\langle 1|$ is the transformation that maps $|1\rangle$ to $|0\rangle$ and $|0\rangle$ to $(0,0)^T$, since

$$|0\rangle\langle 1||1\rangle = |0\rangle\langle 1|1\rangle = |0\rangle$$
$$|0\rangle\langle 1||0\rangle = |0\rangle\langle 1|0\rangle = 0|0\rangle = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Equivalently, $|0\rangle\langle 1|$ can be written in matrix form, where $|0\rangle = (1,0)^T$, $\langle 0| = (1,0)$, $|1\rangle = (0,1)^T$, and $\langle 1| = (0,1)$. Then

$$|0\rangle\langle 1| = \begin{pmatrix} 1 \\ 0 \end{pmatrix}(0,1) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

This notation gives us a convenient way of specifying transformations on quantum states in terms of what happens to the basis vectors (see Section 4). For example, the transformation that exchanges $|0\rangle$ and $|1\rangle$ is given by the matrix

$$X = |0\rangle\langle 1| + |1\rangle\langle 0|.$$

In this paper we prefer the slightly more intuitive notation

$$X : \begin{array}{rcl} |0\rangle & \rightarrow & |1\rangle \\ |1\rangle & \rightarrow & |0\rangle, \end{array}$$

which explicitly specifies the result of a transformation on the basis vectors.

3. QUANTUM BITS

A quantum bit, or qubit, is a unit vector in a two-dimensional complex vector space for which a particular basis, denoted by $\{|0\rangle, |1\rangle\}$, has been fixed. The orthonormal basis $|0\rangle$ and $|1\rangle$ may correspond to the $|\uparrow\rangle$ and $|\rightarrow\rangle$ polarizations of a photon respectively, or to the polarizations $|\nearrow\rangle$ and $|\nwarrow\rangle$. Or $|0\rangle$ and $|1\rangle$ could correspond to the spin-up and spin-down states of an electron. When talking about qubits, and quantum computations in general, a fixed basis with respect to which all statements are made has been chosen in advance. In particular, unless otherwise specified, all measurements will be made with respect to the standard basis for quantum computation, $\{|0\rangle, |1\rangle\}$.

For the purposes of quantum computation, the basis states $|0\rangle$ and $|1\rangle$ are taken to represent the classical bit values 0 and 1 respectively. Unlike classical bits however, qubits can be in a superposition of $|0\rangle$ and $|1\rangle$ such as $a|0\rangle + b|1\rangle$, where a and b are complex numbers such that $|a|^2 + |b|^2 = 1$. Just as in the photon polarization case, if such a superposition is measured with respect to the basis $\{|0\rangle, |1\rangle\}$, the probability that the measured value is $|0\rangle$ is $|a|^2$ and the probability that the measured value is $|1\rangle$ is $|b|^2$.

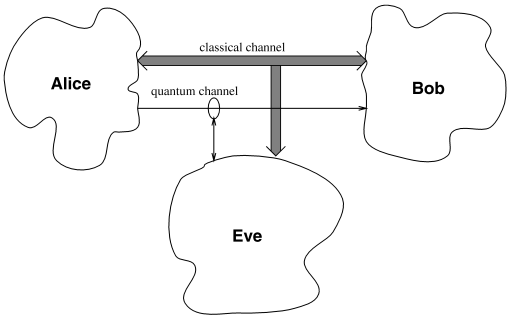
Even though a quantum bit can be put in infinitely many superposition states, it is only possible to extract a single classical bit's worth of information from a single quantum bit. The reason that no more information can be gained from a qubit than in a classical bit is that information can only be obtained by measurement. When a qubit is measured, the measurement changes the state to one of the basis states in the way seen in the photon polarization experiment. As every measurement can result in only one of two states, one of the basis vectors associated to the given measuring device, so, just as in the classical case, there are only two possible results. As measurement changes the state, one cannot measure the state of a

qubit in two different bases. Furthermore, as we shall see in Section 4.1.2, quantum states cannot be cloned, so it is not possible to measure a qubit in two ways, even indirectly by, say, copying the qubit and measuring the copy in a different basis from the original.

3.1. Quantum Key Distribution

Sequences of single qubits can be used to transmit private keys on insecure channels. In 1984 Bennett and Brassard described the first quantum key distribution scheme [Bennett and Brassard 1987; Bennett et al. 1992]. Classically, public key encryption techniques (e.g., RSA) are used for key distribution.

Consider the situation in which Alice and Bob want to agree on a secret key so that they can communicate privately. They are connected by an ordinary bidirectional open channel and a unidirectional quantum channel, both of which can be observed by Eve, who wishes to eavesdrop on their conversation. This situation is illustrated in the figure that follows. The quantum channel allows Alice to send individual particles (e.g., photons) to Bob who can measure their quantum state. Eve can attempt to measure the state of these particles and can resend the particles to Bob.



To begin the process of establishing a secret key, Alice sends a sequence of bits to Bob by encoding each bit in the quantum state of a photon as follows. For each bit, Alice randomly uses one of the following two bases for encoding each bit:

$$0 \rightarrow |\uparrow\rangle$$
$$1 \rightarrow |\rightarrow\rangle$$

or

$$0 \rightarrow |\nwarrow\rangle$$
$$1 \rightarrow |\nearrow\rangle.$$

Bob measures the state of the photons he receives by randomly picking either basis. After the bits have been transmitted, Bob and Alice communicate the basis they used for encoding and decoding of each bit over the open channel. With this information both can determine which bits have been transmitted correctly, by identifying those bits for which the sending and receiving bases agree. They will use these bits as the key and discard all the others. On average, Alice and Bob will agree on 50% of all bits transmitted.

Suppose that Eve measures the state of the photons transmitted by Alice and re-sends new photons with the measured state. In this process she will use the wrong basis approximately 50% of the time, in which case she will re-send the bit with the wrong basis. So when Bob measures a re-sent qubit with the correct basis, there will be a 25% probability that he measures the wrong value. Thus any eavesdropper on the quantum channel is bound to introduce a high error rate that Alice and Bob can detect by communicating a sufficient number of parity bits of their keys over the open channel. So, not only is it likely that Eve's version of the key is 25% incorrect, but the fact that someone is eavesdropping will be apparent to Alice and Bob.

Other techniques for exploiting quantum effects for key distribution have been proposed. See, for example, Ekert et al. [1992], Bennett [1992], and Lo and Chau [1999]. But none of the quantum key distribution techniques are substitutes for public key encryption schemes. Attacks by eavesdroppers other than the one described here are possible. Security against all such schemes is discussed in both Mayers [1998] and Lo and Chau [1999].

Quantum key distribution has been realized over a distance of 24 km using standard fiber optical cables [Hughes et al.

1997] and over 0.5 km through the atmosphere [Hughes et al. 1999].

3.2. Multiple Qubits

Imagine a macroscopic physical object breaking apart and multiple pieces flying off in different directions. The state of this system can be described completely by describing the state of each of its component pieces separately. A surprising and unintuitive aspect of the state space of an n-particle quantum system is that the state of the system cannot always be described in terms of the state of its component pieces. It is when examining systems of more than one qubit that one first gets a glimpse of where the computational power of quantum computers could come from.

As we saw, the state of a qubit can be represented by a vector in the two-dimensional complex vector space spanned by $|0\rangle$ and $|1\rangle$. In classical physics, the possible states of a system of n particles, whose individual states can be described by a vector in a two-dimensional vector space, form a vector space of $2n$ dimensions. However, in a quantum system the resulting state space is much larger; a system of n qubits has a state space of 2^n dimensions.[4] It is this exponential growth of the state space with the number of particles that suggests a possible exponential speed-up of computation on quantum computers over classical computers.

Individual state spaces of n particles combine classically through the cartesian product. Quantum states, however, combine through the tensor product. Details on properties of tensor products and their expression in terms of vectors and matrices are given in Appendix A. Let us look briefly at distinctions between the cartesian product and the tensor product that will be crucial to understanding quantum computation.

Let V and W be 2 two-dimensional complex vector spaces with bases $\{v_1, v_2\}$ and $\{w_1, w_2\}$ respectively. The cartesian product of these two spaces can take as its basis the union of the bases of its component spaces $\{v_1, v_2, w_1, w_2\}$. Note that the order of the basis was chosen arbitrarily. In particular, the dimension of the state space of multiple classical particles grows linearly with the number of particles, since $\dim(X \times Y) = \dim(X) + \dim(Y)$. The tensor product of V and W has basis $\{v_1 \otimes w_1, v_1 \otimes w_2, v_2 \otimes w_1, v_2 \otimes w_2\}$. Note that the order of the basis, again, is arbitrary.[5] So the state space for two qubits, each with basis $\{|0\rangle, |1\rangle\}$, has basis $\{|0\rangle \otimes |0\rangle, |0\rangle \otimes |1\rangle, |1\rangle \otimes |0\rangle, |1\rangle \otimes |1\rangle\}$, which can be written more compactly as $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. More generally, we write $|x\rangle$ to mean $|b_n b_{n-1} \ldots b_0\rangle$ where b_i are the binary digits of the number x.

A basis for a 3-qubit system is

$$\{|000\rangle, |001\rangle, |010\rangle, |011\rangle, \\ |100\rangle, |101\rangle, |110\rangle, |111\rangle\}$$

and in general an n-qubit system has 2^n basis vectors. We can now see the exponential growth of the state space with the number of quantum particles. The tensor product $X \otimes Y$ has dimension $dim(X) \times dim(Y)$.

The state $|00\rangle + |11\rangle$ is an example of a quantum state that cannot be described in terms of the state of each of its components (qubits) separately. In other words, we cannot find a_1, a_2, b_1, b_2 such that $(a_1|0\rangle + b_1|1\rangle) \otimes (a_2|0\rangle + b_2|1\rangle) = |00\rangle + |11\rangle$, since

$$(a_1|0\rangle + b_1|1\rangle) \otimes (a_2|0\rangle + b_2|1\rangle) = a_1 a_2|00\rangle \\ + a_1 b_2|01\rangle + b_1 a_2|10\rangle + b_1 b_2|11\rangle$$

and $a_1 b_2 = 0$ implies that either $a_1 a_2 = 0$ or $b_1 b_2 = 0$. States that cannot be decomposed in this way are called entangled states. These states represent situations that have no classical counterpart and for which we have no intuition. These are also the states that provide the exponential growth of quantum state spaces with the number of particles.

[4] Actually, as we shall see, the state space is the set of normalized vectors in this 2^n dimensional space, just as the state $a|0\rangle + b|1\rangle$ of a qubit is normalized so that $|a|^2 + |b|^2 = 1$.

[5] It is only when we use matrix notation to describe state transformations that the order of basis vectors becomes relevant.

Note that it would require vast resources to simulate even a small quantum system on traditional computers. The evolution of quantum systems is exponentially faster than their classical simulations. The reason for the potential power of quantum computers is the possibility of exploiting the quantum state evolution as a computational mechanism.

3.3. Measurement

The experiment in Section 2.1.2 illustrates how measurement of a single qubit projects the quantum state on to one of the basis states associated with the measuring device. The result of a measurement is probabilistic and the process of measurement changes the state to that measured.

Let us look at an example of measurement in a two-qubit system. Any two-qubit state can be expressed as $a|00\rangle + b|01\rangle + c|10\rangle + d|11\rangle$, where a, b, c, and d are complex numbers such that $|a|^2 + |b|^2 + |c|^2 + |d|^2 = 1$. Suppose we wish to measure the first qubit with respect to the standard basis $\{|0\rangle, |1\rangle\}$. For convenience we will rewrite the state as follows:

$$\begin{aligned} &a|00\rangle + b|01\rangle + c|10\rangle + d|11\rangle \\ &\quad = |0\rangle \otimes (a|0\rangle + b|1\rangle) + |1\rangle \otimes (c|0\rangle + d|1\rangle) \\ &\quad = u|0\rangle \otimes (a/u|0\rangle + b/u|1\rangle) + v|1\rangle \\ &\qquad \otimes (c/v|0\rangle + d/v|1\rangle). \end{aligned}$$

For $u = \sqrt{|a|^2 + |b|^2}$ and $v = \sqrt{|c|^2 + |d|^2}$ the vectors $a/u|0\rangle + b/u|1\rangle$ and $c/v|0\rangle + d/v|1\rangle$ are of unit length. Once the state has been rewritten as above, as a tensor product of the bit being measured and a second vector of unit length, the probabalistic result of a measurement is easy to read off. Measurement of the first bit will with probability $u^2 = |a|^2 + |b|^2$ return $|0\rangle$, projecting the state to $|0\rangle \otimes (a/u|0\rangle + b/u|1\rangle)$, or with probability $v = |c|^2 + |d|^2$ yield $|1\rangle$, projecting the state to $|1\rangle \otimes (c/v|0\rangle + d/v|1\rangle)$. As $|0\rangle \otimes (a/u|0\rangle + b/u|1\rangle)$ and $|1\rangle \otimes (c/v|0\rangle + d/v|1\rangle)$ are both unit vectors, no scaling is necessary. Measuring the second bit works similarly.

For the purposes of quantum computation, multibit measurement can be treated as a series of single-bit measurements in the standard basis. Other sorts of measurements are possible, such as measuring whether two qubits have the same value without learning the actual value of the two qubits. But such measurements are equivalent to unitary transformations followed by a standard measurement of individual qubits, and so it suffices to look only at standard measurements.

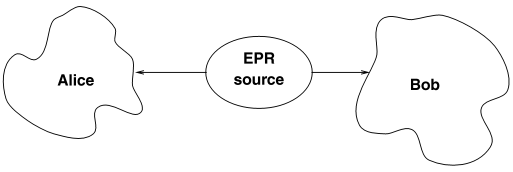
In the two-qubit example, the state space is a cartesian product of the subspace consisting of all states whose first qubit is in the state $|0\rangle$ and the orthogonal subspace of states whose first qubit is in the state $|1\rangle$. Any quantum state can be written as the sum of two vectors, one in each of the subspaces. A measurement of k qubits in the standard basis has 2^k possible outcomes m_i. Any device measuring k qubits of an n-qubit system splits of the 2^n-dimensional state space $\mathcal{H}$ into a cartesian product of orthogonal subspaces $S_1, \ldots, S_{2^k}$ with $\mathcal{H} = S_1 \times \cdots \times S_{2^k}$, such that the value of the k qubits being measured is m_i and the state after measurement is in space the space S_i for some i. The device randomly chooses one of the S_i's, with probability the square of the amplitude of the component of ψ in S_i, and projects the state into that component, scaling to give length 1. Equivalently, the probability that the result of the measurement is a given value is the sum of the squares of the the absolute values of the amplitudes of all basis vectors compatible with that value of the measurement.

Measurement gives another way of thinking about entangled particles. Particles are not entangled if the measurement of one has no effect on the other. For instance, the state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ is entangled, since the probability that the first bit is measured to be $|0\rangle$ is $1/2$ if the second bit has not been measured. However, if the second bit had been measured, the probability that the first bit is measured as $|0\rangle$ is either 1 or 0, depending on whether the second bit was measured as $|0\rangle$ or $|1\rangle$ respectively. Thus the probable result of measuring the first bit is changed by a measurement of the second bit. On

the other hand, the state $\frac{1}{\sqrt{2}}(|00\rangle + |01\rangle)$ is not entangled: since $\frac{1}{\sqrt{2}}(|00\rangle + |01\rangle) = |0\rangle \otimes \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, any measurement of the first bit will yield $|0\rangle$ regardless of whether the second bit was measured. Similarly, the second bit has a fifty-fifty chance of being measured as $|0\rangle$ regardless of whether the first bit was measured or not. Note that entanglement, in the sense that measurement of one particle has an effect on measurements of another particle, is equivalent to our previous definition of entangled states as states that cannot be written as a tensor product of individual states.

3.4. The EPR Paradox

Einstein, Podolsky, and Rosen proposed a gedanken experiment that uses entangled particles in a manner that seemed to violate fundamental principles of relativity. Imagine a source that generates two maximally entangled particles $\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$, called an EPR pair, and sends one each to Alice and Bob.



Alice and Bob can be arbitrarily far apart. Suppose that Alice measures her particle and observes state $|0\rangle$. This means that the combined state will now be $|00\rangle$, and if now Bob measures his particle he will also observe $|0\rangle$. Similarly, if Alice measures $|1\rangle$, so will Bob. Note that the change of the combined quantum state occurs instantaneously even though the two particles may be arbitrarily far apart. It appears that this would enable Alice and Bob to communicate faster than the speed of light. Further analysis, as we shall see, shows that even though there is a coupling between the two particles, there is no way for Alice or Bob to use this mechanism to communicate.

There are two standard ways that people use to describe entangled states and their measurement. Both have their positive aspects, but both are incorrect and can lead to misunderstandings. Let us examine both in turn.

Einstein, Podolsky, and Rosen proposed that each particle has some internal state that completely determines what the result of any given measurement will be. This state is, for the moment, hidden from us, and therefore the best we can currently do is to give probabilistic predictions. Such a theory is known as a local hidden variable theory. The simplest hidden variable theory for an EPR pair is that the particles are either both in state $|0\rangle$ or both in state $|1\rangle$, we just don't happen to know which. In such a theory no communication between possibly distant particles is necessary to explain the correlated measurements. However, this point of view cannot explain the results of measurements with respect to a different basis. In fact, Bell showed that any local hidden variable theory predicts that certain measurements will satisfy an inequality, known as Bell's inequality. However, the result of actual experiments performing these measurements show that Bell's inequality is violated. Thus quantum mechanics cannot be explained by any local hidden variable theory. See Greenstein and Zajonc [1997] for a highly readable account of Bell's theorem and related experiments.

The second standard description is in terms of cause and effect. For example, we said earlier that a measurement performed by Alice affects a measurement performed by Bob. However, this view is incorrect also, and results, as Einstein, Podolsky, and Rosen recognized, in deep inconsistencies when combined with relativity theory. It is possible to set up the EPR scenario so that one observer sees Alice measure first, then Bob, while another observer sees Bob measure first, then Alice. According to relativity, physics must equally well explain the observations of the first observer as the second. While our terminology of cause and effect cannot be compatible with both observers, the actual experimental values

are invariant under change of observer. The experimental results can be explained equally well by Bob's measuring first and causing a change in the state of Alice's particle, as the other way around. This symmetry shows that Alice and Bob cannot, in fact, use their EPR pair to communicate faster than the speed of light, and thus resolves the apparent paradox. All that can be said is that Alice and Bob will observe the same random behavior.

As we will see in the section on dense coding and teleportation, EPR pairs can be used to aid communication, albeit communication slower than the speed of light.

4. QUANTUM GATES

So far we have looked at static quantum systems, which change only when measured. The dynamics of a quantum system, when not being measured, are governed by Schrödinger's equation; the dynamics must take states to states in a way that preserves orthogonality. For a complex vector space, linear transformations that preserve orthogonality are unitary transformations, defined as follows. Any linear transformation on a complex vector space can be described by a matrix. Let M^* denote the conjugate transpose of the matrix M. A matrix M is unitary (describes a unitary transformation) if $MM^* = I$. Any unitary transformation of a quantum state space is a legitimate quantum transformation, and vice versa. One can think of unitary transformations as being rotations of a complex vector space.

One important consequence of the fact that quantum transformations are unitary is that they are reversible. Thus quantum gates must be reversible. Bennett, Fredkin, and Toffoli had already looked at reversible versions of standard computing models showing that all classical computations can be done reversibly. See Feynman's *Lectures on Computation* [Feynman 1996] for an account of reversible computation and its relation to the energy of computation and information.

4.1. Simple Quantum Gates

The following are some examples of useful single-qubit quantum state transformations. Because of linearity, the transformations are fully specified by their effect on the basis vectors. The associated matrix, with $\{|0\rangle, |1\rangle\}$ as the preferred ordered basis, is also shown.

$$\begin{aligned}
I &: |0\rangle \rightarrow |0\rangle \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\
&\quad |1\rangle \rightarrow |1\rangle \\
X &: |0\rangle \rightarrow |1\rangle \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\
&\quad |1\rangle \rightarrow |0\rangle \\
Y &: |0\rangle \rightarrow -|1\rangle \quad \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \\
&\quad |1\rangle \rightarrow |0\rangle \\
Z &: |0\rangle \rightarrow |0\rangle \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\
&\quad |1\rangle \rightarrow -|1\rangle
\end{aligned}$$

The names of these transformations are conventional. I is the identity transformation, X is negation, Z is a phase shift operation, and $Y = ZX$ is a combination of both. The X transformation was discussed previously in Section 2.2. It can be readily verified that these gates are unitary. For example

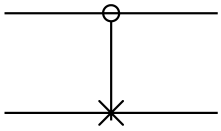
$$YY^* = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = I.$$

The controlled-NOT gate, C_{not}, operates on two qubits as follows: it changes the second bit if the first bit is 1 and leaves this bit unchanged otherwise. The vectors $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$ form an orthonormal basis for the state space of a two-qubit system, a four-dimensional complex vector space. In order to represent transformations of this space in matrix notation we need to choose an isomorphism between this space and the space of complex 4-tuples. There is no reason, other than convention, to pick one isomorphism over another. The one we use here associates $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$ to the standard 4-tuple basis $(1, 0, 0, 0)^T$, $(0, 1, 0, 0)^T$, $(0, 0, 1, 0)^T$, and $(0, 0, 0, 1)^T$, in that order. The C_{not} transformation has representations

$$C_{not}: \begin{array}{ccc} |00\rangle & \rightarrow & |00\rangle \\ |01\rangle & \rightarrow & |01\rangle \\ |10\rangle & \rightarrow & |11\rangle \\ |11\rangle & \rightarrow & |10\rangle \end{array} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

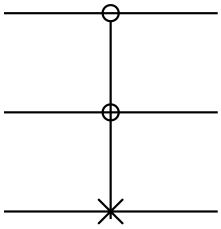
The transformation C_{not} is unitary since $C_{not}^* = C_{not}$ and $C_{not}C_{not} = I$. The C_{not} gate cannot be decomposed into a tensor product of two single-bit transformations.

It is useful to have graphical representations of quantum state transformations, especially when several transformations are combined. The controlled-NOT gate C_{not} is typically represented by a circuit of the form

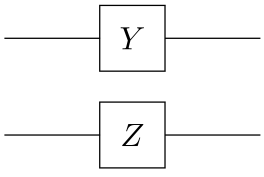


The open circle indicates the control bit, and the $\times$ indicates the conditional negation of the subject bit. In general there can be multiple control bits. Some authors use a solid circle to indicate negative control, in which the subject bit is toggled when the control bit is 0.

Similarly, the controlled-controlled-NOT, which negates the last bit of three if and only if the first two are both 1, has the following graphical representation.



Single bit operations are graphically represented by appropriately labeled boxes as shown.



4.1.1 The Walsh–Hadamard Transformation. Another important single-bit transformation is the Hadamard transformation, defined by

$$\begin{aligned} H: |0\rangle &\rightarrow \tfrac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \\ |1\rangle &\rightarrow \tfrac{1}{\sqrt{2}}(|0\rangle - |1\rangle). \end{aligned}$$

The transformation H has a number of important applications. When applied to $|0\rangle$, H creates a superposition state $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. Applied to n bits individually, H generates a superposition of all 2^n possible states, which can be viewed as the binary representation of the numbers from 0 to $2^n - 1$.

$$\begin{aligned} &(H \otimes H \otimes \cdots \otimes H)|00\ldots0\rangle \\ &\quad = \frac{1}{\sqrt{2^n}}((|0\rangle + |1\rangle) \otimes (|0\rangle + |1\rangle) \\ &\qquad \otimes \cdots \otimes (|0\rangle + |1\rangle)) \\ &\quad = \frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} |x\rangle. \end{aligned}$$

The transformation that applies H to n bits is called the Walsh, or Walsh–Hadamard, transformation W. It can be defined as a recursive decomposition of the form

$$W_1 = H, W_{n+1} = H \otimes W_n.$$

4.1.2 No Cloning. The unitary property implies that quantum states cannot be copied or cloned. The no cloning proof given here, originally due to Wootters and Zurek [1982], is a simple application of the linearity of unitary transformations.

Assume that U is a unitary transformation that clones, in that $U(|a0\rangle) = |aa\rangle$ for all quantum states $|a\rangle$. Let $|a\rangle$ and $|b\rangle$ be two orthogonal quantum states. Say $U(|a0\rangle) = |aa\rangle$ and $U(|b0\rangle) = |bb\rangle$. Consider $|c\rangle = (1/\sqrt{2})(|a\rangle + |b\rangle)$. By linearity,

$$\begin{aligned} U(|c0\rangle) &= \frac{1}{\sqrt{2}}(U(|a0\rangle) + U(|b0\rangle)) \\ &= \frac{1}{\sqrt{2}}(|aa\rangle + |bb\rangle). \end{aligned}$$

But if U is a cloning transformation then

$$U(|c0\rangle) = |cc\rangle = 1/2(|aa\rangle + |ab\rangle + |ba\rangle + |bb\rangle),$$

which is not equal to $(1/\sqrt{2})(|aa\rangle + |bb\rangle)$. Thus there is no unitary operation that can reliably clone unknown quantum states. It is clear that cloning is not possible by using measurement, since measurement is both probabalistic and destructive of states not in the measuring device's associated subspaces.

It is important to understand what sort of cloning is and isn't allowed. It is possible to clone a known quantum state. What the no cloning principle tells us is that it is impossible to reliably clone an unknown quantum state. Also, it is possible to obtain n particles in an entangled state $a|00\ldots0\rangle + b|11\ldots1\rangle$ from an unknown state $a|0\rangle + b|1\rangle$. Each of these particles will behave in exactly the same way when measured with respect to the standard basis for quantum computation $\{|00\ldots0\rangle, |00\ldots01\rangle, \ldots, |11\ldots1\rangle\}$, but not when measured with respect to other bases. It is not possible to create the n-particle state $(a|0\rangle + b|1\rangle) \otimes \ldots \otimes (a|0\rangle + b|1\rangle)$ from an unknown state $a|0\rangle + b|1\rangle$.

4.2. Examples

The use of simple quantum gates can be studied with two simple examples: dense coding and teleportation.

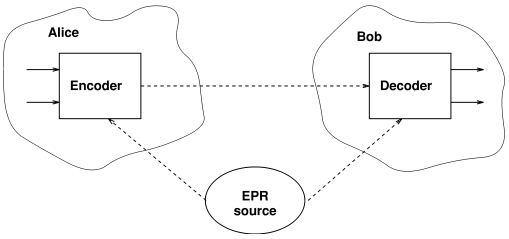
Dense coding uses one quantum bit together with an EPR pair to encode and transmit two classical bits. Since EPR pairs can be distributed ahead of time, only one qubit (particle) needs to be physically transmitted to communicate two bits of information. This result is surprising since, as was discussed in Section 3, only one classical bit's worth of information can be extracted from a qubit. Teleportation is the opposite of dense coding, in that it uses two classical bits to transmit a single qubit. Teleportation is surprising in light of the no cloning principle of quantum mechanics, in that it enables the transmission of an unknown quantum state.

The key to both dense coding and teleportation is the use of entangled particles. The initial set up is the same for both processes. Alice and Bob wish to communicate. Each is sent one of the entangled particles making up an EPR pair,

$$\psi_0 = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle).$$

Say Alice is sent the first particle, and Bob the second. Until a particle is transmitted, only Alice can perform transformations on her particle, and only Bob can perform transformations on his.

4.2.1 Dense Coding. Alice. Alice receives two classical bits, encoding the numbers 0 through 3. Depending on this number Alice performs one of the transformations $\{I, X, Y, Z\}$ on her qubit of the



entangled pair ψ_0. Transforming just one bit of an entangled pair means performing the identity transformation on the other bit. The resulting state is shown in the table.

Value	Transformation	New state
0	$\psi_0 = (I \otimes I)\psi_0$	$\frac{1}{\sqrt{2}}(\|00\rangle + \|11\rangle)$
1	$\psi_1 = (X \otimes I)\psi_0$	$\frac{1}{\sqrt{2}}(\|10\rangle + \|01\rangle)$
2	$\psi_2 = (Y \otimes I)\psi_0$	$\frac{1}{\sqrt{2}}(-\|10\rangle + \|01\rangle)$
3	$\psi_3 = (Z \otimes I)\psi_0$	$\frac{1}{\sqrt{2}}(\|00\rangle - \|11\rangle)$

Alice then sends her qubit to Bob.

Bob. Bob applies a controlled-NOT to the two qubits of the entangled pair.

Initial state	Controlled-NOT	First bit	Second bit
$\psi_0 = \frac{1}{\sqrt{2}}(\|00\rangle + \|11\rangle)$	$\frac{1}{\sqrt{2}}(\|00\rangle + \|10\rangle)$	$\frac{1}{\sqrt{2}}(\|0\rangle + \|1\rangle)$	$\|0\rangle$
$\psi_1 = \frac{1}{\sqrt{2}}(\|10\rangle + \|01\rangle)$	$\frac{1}{\sqrt{2}}(\|11\rangle + \|01\rangle)$	$\frac{1}{\sqrt{2}}(\|1\rangle + \|0\rangle)$	$\|1\rangle$
$\psi_2 = \frac{1}{\sqrt{2}}(-\|10\rangle + \|01\rangle)$	$\frac{1}{\sqrt{2}}(-\|11\rangle + \|01\rangle)$	$\frac{1}{\sqrt{2}}(-\|1\rangle + \|0\rangle)$	$\|1\rangle$
$\psi_3 = \frac{1}{\sqrt{2}}(\|00\rangle - \|11\rangle)$	$\frac{1}{\sqrt{2}}(\|00\rangle - \|10\rangle)$	$\frac{1}{\sqrt{2}}(\|0\rangle - \|1\rangle)$	$\|0\rangle$

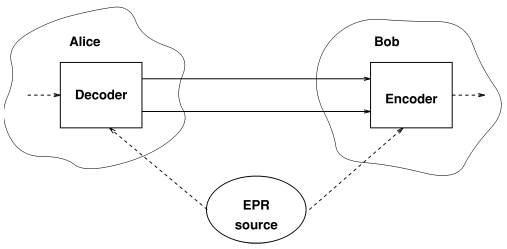
Note that Bob can now measure the second qubit without disturbing the quantum state. If the measurement returns $|0\rangle$ then the encoded value was either 0 or 3, if the measurement returns $|1\rangle$ then the encoded value was either 1 or 2.

Bob now applies H to the first bit:

Initial state	First bit	H(First bit)
ψ_0	$\frac{1}{\sqrt{2}}(\|0\rangle + \|1\rangle)$	$\frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}(\|0\rangle + \|1\rangle) + \frac{1}{\sqrt{2}}(\|0\rangle - \|1\rangle)\right) = \|0\rangle$
ψ_1	$\frac{1}{\sqrt{2}}(\|1\rangle + \|0\rangle)$	$\frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}(\|0\rangle - \|1\rangle) + \frac{1}{\sqrt{2}}(\|0\rangle + \|1\rangle)\right) = \|0\rangle$
ψ_2	$\frac{1}{\sqrt{2}}(-\|1\rangle + \|0\rangle)$	$\frac{1}{\sqrt{2}}\left(-\frac{1}{\sqrt{2}}(\|0\rangle - \|1\rangle) + \frac{1}{\sqrt{2}}(\|0\rangle + \|1\rangle)\right) = \|1\rangle$
ψ_3	$\frac{1}{\sqrt{2}}(\|0\rangle - \|1\rangle)$	$\frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}(\|0\rangle + \|1\rangle) - \frac{1}{\sqrt{2}}(\|0\rangle - \|1\rangle)\right) = \|1\rangle$

Finally, Bob measures the resulting bit, which allows him to distinguish between 0 and 3, and 1 and 2.

4.2.2 Teleportation. The objective is to transmit the quantum state of a particle using classical bits and reconstruct the exact quantum state at the receiver. Since quantum state cannot be copied, the quantum state of the given particle will necessarily be destroyed. Single-bit teleportation has been realized experimentally [Boschi et al. 1998; Bouwmeester et al. 1997; Nielsen et al. 1998].



Alice. Alice has a qubit whose state she doesn't know. She wants to send the state of ths qubit

$$\phi = a|0\rangle + b|1\rangle$$

to Bob through classical channels. As with dense coding, Alice and Bob each possess one qubit of an entangled pair

$$\psi_0 = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle).$$

Alice applies the decoding step of dense coding to the qubit ϕ to be transmitted and her half of the entangled pair. The starting state is quantum state

$$\begin{aligned}\phi \otimes \psi_0 &= \frac{1}{\sqrt{2}}(a|0\rangle \otimes (|00\rangle + |11\rangle) \\ &\quad + b|1\rangle \otimes (|00\rangle + |11\rangle)) \\ &= \frac{1}{\sqrt{2}}(a|000\rangle + a|011\rangle + b|100\rangle \\ &\quad + b|111\rangle),\end{aligned}$$

of which Alice controls the first two bits and Bob controls the last one. Alice now applies $C_{not} \otimes I$ and $H \otimes I \otimes I$ to this state:

$$
\begin{aligned}
&(H \otimes I \otimes I)(C_{not} \otimes I)(\phi \otimes \psi_0) \\
&\quad = (H \otimes I \otimes I)(C_{not} \otimes I)\frac{1}{\sqrt{2}}(a|000\rangle \\
&\qquad + a|011\rangle + b|100\rangle + b|111\rangle) \\
&\quad = (H \otimes I \otimes I)\frac{1}{\sqrt{2}}(a|000\rangle + a|011\rangle \\
&\qquad + b|110\rangle + b|101\rangle) \\
&\quad = \frac{1}{2}(a(|000\rangle + |011\rangle + |100\rangle + |111\rangle) \\
&\qquad + b(|010\rangle + |001\rangle - |110\rangle - |101\rangle)) \\
&\quad = \frac{1}{2}(|00\rangle(a|0\rangle + b|1\rangle) + |01\rangle(a|1\rangle \\
&\qquad + b|0\rangle) + |10\rangle(a|0\rangle - b|1\rangle) \\
&\qquad + |11\rangle(a|1\rangle - b|0\rangle))
\end{aligned}
$$

Alice measures the first two qubits to get one of $|00\rangle$, $|01\rangle$, $|10\rangle$, or $|11\rangle$ with equal probability. Depending on the result of the measurement, the quantum state of Bob's qubit is projected to $a|0\rangle + b|1\rangle$, $a|1\rangle + b|0\rangle$, $a|0\rangle - b|1\rangle$, or $a|1\rangle - b|0\rangle$ respectively. Alice sends the result of her measurement as two classical bits to Bob.

Note that when she measured it, Alice irretrievably altered the state of her original qubit ϕ, whose state she is in the process of sending to Bob. This loss of the original state is the reason teleportation does not violate the no cloning principle.

Bob. When Bob receives the two classical bits from Alice he knows how the state of his half of the entangled pair compares to the original state of Alice's qubit.

Bits received	State	Decoding
00	$a\|0\rangle + b\|1\rangle$	I
01	$a\|1\rangle + b\|0\rangle$	X
10	$a\|0\rangle - b\|1\rangle$	Z
11	$a\|1\rangle - b\|0\rangle$	Y

Bob can reconstruct the original state of Alice's qubit, ϕ, by applying the appropriate decoding transformation to his part of the entangled pair. Note that this is the encoding step of dense coding.

5. QUANTUM COMPUTERS

This section discusses how quantum mechanics can be used to perform computations and how these computations are qualitatively different from those performed by a conventional computer. Recall from Section 4 that all quantum state transformations have to be reversible. While the classical NOT gate is reversible, AND, OR, and NAND gates are not. Thus it is not obvious that quantum transformations can carry out all classical computations. The first subsection describes complete sets of reversible gates that can perform any classical computation on a quantum computer. Furthermore, it describes sets of gates with which all quantum computations can be done. The second subsection discusses quantum parallelism.

5.1. Quantum Gate Arrays

The bra/ket notation is useful in defining complex unitary operations. For two arbitrary unitary transformations U_1 and U_2, the "conditional" transformation $|0\rangle\langle 0| \otimes U_1 + |1\rangle\langle 1| \otimes U_2$ is also unitary. The controlled-NOT gate can be defined by

$$C_{not} = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes X.$$

The three-bit controlled-controlled-NOT gate or Toffoli gate of Section 4 is also an instance of this conditional definition:

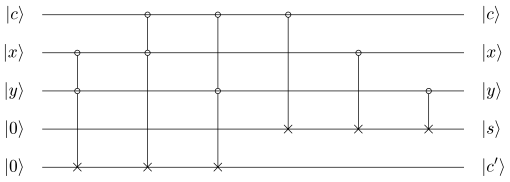
$$T = |0\rangle\langle 0| \otimes I \otimes I + |1\rangle\langle 1| \otimes C_{not}.$$

The Toffoli gate T can be used to construct complete set of boolean connectives, as can be seen from the fact that it can be used to construct the AND and NOT operators in the following way:

$$
\begin{aligned}
T|1, 1, x\rangle &= |1, 1, \neg x\rangle \\
T|x, y, 0\rangle &= |x, y, x \wedge y\rangle
\end{aligned}
$$

The T gate is sufficient to construct arbitrary combinatorial circuits.

The following quantum circuit, for example, implements a 1 bit full adder using Toffoli and controlled-NOT gates:



where x and y are the data bits, s is their sum (modulo 2), c is the incoming carry bit, and c' is the new carry bit. Vedral, Barenco, and Ekert [1996] define more complex circuits that include in-place addition and modular addition.

The Fredkin gate is a "controlled swap" and can be defined as

$$F = |0\rangle\langle 0| \otimes I \otimes I + |1\rangle\langle 1| \otimes S$$

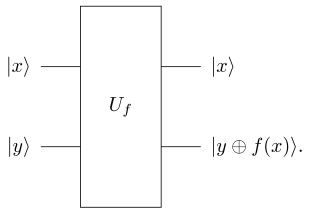
where S is the swap operation

$$S = |00\rangle\langle 00| + |01\rangle\langle 10| + |10\rangle\langle 01| + |11\rangle\langle 11|.$$

The reader can verify that F, like T, is complete for combinatorial circuits.

Deutsch has shown [1985] that it is possible to construct reversible quantum gates for any classically computable function. In fact, it is possible to conceive of a universal quantum Turing machine [Bernstein and Vazirani 1997]. In this construction we must assume a sufficient supply of bits that correspond to the tape of a Turing machine.

Knowing that an arbitrary classical function f with m input and k output bits can be implemented on quantum computer, we assume the existence of a *quantum gate array* U_f that implements f. U_f is a $m+k$–bit transformation of the form $U_f : |x, y\rangle \rightarrow |x, y \oplus f(x)\rangle$, where $\oplus$ denotes the bitwise exclusive-OR.[6] Quantum gate arrays U_f, defined in this way, are unitary for any function f. To compute $f(x)$ we apply U_f to $|x\rangle$ tensored with k zeros $|x, 0\rangle$. Since $f(x) \oplus f(x) = 0$ we have $U_f U_f = I$. Graphically the transformation $U_f : |x, y\rangle \rightarrow |x, y \oplus f(x)\rangle$ is depicted as



While the T and F gates are complete for combinatorial circuits, they cannot achieve arbitrary quantum state transformations. In order to realize arbitrary unitary transformations,[7] single-bit rotations need to be included. Barenco et al. [1995] show that C_{not} together with all 1-bit quantum gates is a universal gate set. It suffices to include the following 1-bit transformations

$$\begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}, \begin{pmatrix} e^{i\alpha} & 0 \\ 0 & e^{-i\alpha} \end{pmatrix}$$

for all $0 \leq \alpha \leq 2\pi$ together with the C_{not} to obtain a universal set of gates. As we shall see, such nonclassical transformations are crucial for exploiting the power of quantum computers.

5.2. Quantum Parallelism

What happens if U_f is applied to input that is in a superposition? The answer is easy but powerful: since U_f is a linear transformation, it is applied to all basis vectors in the superposition simultaneously and will generate a superposition of the results. In this way, it is possible to compute $f(x)$ for n values of x in a single application of U_f. This effect is called quantum parallelism.

The power of quantum algorithms comes from taking advantage of quantum parallelism and entanglement. So most quantum algorithms begin by computing a function of interest on a superposition of all values as follows. Start with an n-qubit state $|00\ldots0\rangle$. Apply the

[6] $\oplus$ is not the direct sum of vectors.

[7] More precisely, we mean arbitrary unitary transformations up to a constant phase factor. A constant phase shift of the state has no physical, and therefore no computational, significance.

Walsh–Hadamard transformation W of Section 4.1.1 to get a superposition

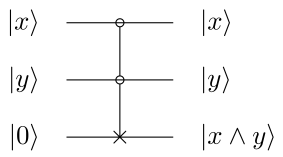
$$\frac{1}{\sqrt{2^n}}(|00\ldots0\rangle + |00\ldots1\rangle + \cdots + |11\ldots1\rangle) = \frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} |x\rangle$$

which should be viewed as the superposition of all integers $0 \le x < 2^n$. Add a k-bit register $|0\rangle$ then by linearity

$$U_f\left(\frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} |x, 0\rangle\right) = \frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} U_f(|x, 0\rangle) = \frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} |x, f(x)\rangle$$

where $f(x)$ is the function of interest. Note that since n qubits enable working simultaneously with 2^n states, quantum parallelism circumvents the time/space tradeoff of classical parallelism through its ability to provide an exponential amount of computational space in a linear amount of physical space.

Consider the trivial example of a controlled-controlled-NOT (Toffoli) gate, T, that computes the conjunction of two values:



Now take as input a superposition of all possible bit combinations of x and y together with the necessary 0:

$$H|0\rangle \otimes H|0\rangle \otimes |0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes |0\rangle = \frac{1}{2}(|000\rangle + |010\rangle + |100\rangle + |110\rangle).$$

Apply T to the superposition of inputs to get a superposition of the results, namely

$$T(H|0\rangle \otimes H|0\rangle \otimes |0\rangle) = \frac{1}{2}(|000\rangle + |010\rangle + |100\rangle + |111\rangle).$$

The resulting superposition can be viewed as a truth table for the conjunction, or more generally as the graph of a function. In the output the values of x, y, and $x \wedge y$ are entangled in such a way that measuring the result will give one line of the truth table, or more generally one point of graph of the function. Note that the bits can be measured in any order: measuring the result will project the state to a superposition of the set of all input values for which f produces this result, and measuring the input will project the result to the corresponding function value.

Measuring at this point gives no advantage over classical parallelism because only one result is obtained, and worse still one cannot even choose which result one gets. The heart of any quantum algorithm is the way in which it manipulates quantum parallelism so that desired results will be measured with high probability. This sort of manipulation has no classical analog and requires nontraditional programming techniques. We list a couple of the techniques currently known.

- Amplify output values of interest. The general idea is to transform the state in such a way that values of interest have a larger amplitude and therefore have a higher probability of being measured. Examples of this approach will be described in Section 7.
- Find common properties of all the values of $f(x)$. This idea is exploited in Shor's algorithm, which uses a quantum Fourier transformation to obtain the period of f.

6. SHOR'S ALGORITHM

In 1994, inspired by work of Daniel Simon (later published in Simon [1997]), Peter Shor found a bounded probability polynomial time algorithm for factoring n-digit numbers on a quantum computer. Since the 1970s people have searched for efficient algorithms for factoring integers. The most efficient classical

algorithm known today is that of Lenstra and Lenstra [1993], which is exponential in the size of the input. The input is the list of digits of M, which has size $n \sim \log M$. People were confident enough that no efficient algorithm existed, that the security of cryptographic systems, like the widely used RSA algorithm, depend on the difficulty of this problem. Shor's result surprised the community at large, prompting widespread interest in quantum computing.

Most factoring algorithms, including Shor's, use a standard reduction of the factoring problem to the problem of finding the period of a function. Shor uses quantum parallelism in the standard way to obtain a superposition of all the values of the function in one step. He then computes the quantum Fourier transform of the function, which, like classical Fourier transforms, puts all the amplitude of the function into multiples of the reciprocal of the period. With high probability, measuring the state yields the period, which in turn is used to factor the integer M.

This description captures the essence of the quantum algorithm but is something of an oversimplification. The biggest complication is that the quantum Fourier transform is based on the fast Fourier transform and thus gives only approximate results in most cases. Thus extracting the period is trickier than outlined here, but the techniques for extracting the period are classical.

We will first describe the quantum Fourier transform and then give a detailed outline of Shor's algorithm.

6.1. The Quantum Fourier Transform

Fourier transforms in general map from the time domain to the frequency domain. So Fourier transforms map functions of period r to functions that have nonzero values only at multiples of the frequency $\frac{2\pi}{r}$. The discrete Fourier transform (DFT) operates on N equally spaced samples in the interval $[0, 2\pi)$ for some N and outputs a function whose domain is the integers between 0 and $N-1$. The discrete Fourier transform of a (sampled) function of period r is a function concentrated near multiples of $\frac{N}{r}$. If the period r divides N evenly, the result is a function that has nonzero values only at multiples of $\frac{N}{r}$. Otherwise, the result will approximate this behavior, and there will be nonzero terms at integers close to multiples of $\frac{N}{r}$.

The Fast Fourier transform (FFT) is a version of DFT where N is a power of 2. The quantum Fourier transform (QFT) is a variant of the discrete Fourier transform, which, like FFT, uses powers of 2. The quantum Fourier transform operates on the amplitude of the quantum state, by sending

$$\sum_x g(x)|x\rangle \rightarrow \sum_c G(c)|c\rangle,$$

where $G(c)$ is the discrete Fourier transform of $g(x)$, and x and c both range over the binary representations for the integers between 0 and $N-1$. If the state were measured after the Fourier transform was performed, the probability that the result was $|c\rangle$ would be $|G(c)|^2$. Note that the quantum Fourier transform does not output a function the way the U_f transformation does; no output appears in an extra register.

Applying the quantum Fourier transform to a periodic function $g(x)$ with period r, we would expect to end up with $\sum_c G(c)|c\rangle$, where $G(c)$ is zero except at multiples of $\frac{N}{r}$. Thus, when the state is measured, the result would be a multiple of $\frac{N}{r}$, say $j\frac{N}{r}$. But as described above, the quantum Fourier transform only gives approximate results for periods that are not a power of two (i.e., do not divide N). However the larger the power of two used as a base for the transform, the better the approximation. The quantum Fourier transform U_{QFT} with base $N = 2^m$ is defined by

$$U_{QFT} : |x\rangle \rightarrow \frac{1}{\sqrt{2^m}} \sum_{c=0}^{2^m-1} \exp\left(\frac{2\pi icx}{2^m}\right)|c\rangle.$$

In order for Shor's algorithm to be a polynomial algorithm, the quantum Fourier transform must be efficiently

computable. Shor shows that the quantum Fourier transform with base 2^m can be constructed using only $\frac{m(m+1)}{2}$ gates. The construction makes use of two types of gates. One is a gate to perform the familiar Hadamard transformation H. We will denote by H_j the Hadamard transformation applied to the jth bit. The other type of gate performs 2-bit transformations of the form

$$S_{j,k} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\theta_{k-j}} \end{pmatrix},$$

where $\theta_{k-j} = \pi/2^{k-j}$. This transformation acts on the kth and jth bits of a larger register. The quantum Fourier transform is given by

$$H_0 S_{0,1} \ldots S_{0,m-1} H_1 \ldots H_{m-3} \\ S_{m-3,m-2} S_{m-3,m-1} H_{m-2} S_{m-2,m-1} H_{m-1}$$

followed by a bit reversal transformation. If FFT is followed by measurement, as in Shor's algorithm, the bit reversal can be performed classically. See Shor [1997] for more details.

6.2. A Detailed Outline of Shor's algorithm

The detailed steps of *Shor's* algorithm are illustrated with a running example where we factor $M = 21$.

Step 1. Quantum parallelism. Choose an integer a arbitrarily. If a is not relatively prime to M, we have found a factor of M. Otherwise apply the rest of the algorithm.

Let m be such that $M^2 \leq 2^m < 2M^2$. [This choice is made so that the approximation used in Step 3 for functions whose period is not a power of 2 will be good enough for the rest of the algorithm to work.] Use quantum parallelism as described in Section 5.2 to compute $f(x) = a^x \bmod M$ for all integers from 0 to $2^m - 1$. The function is thus encoded in the quantum state

$$\frac{1}{\sqrt{2^m}} \sum_{x=0}^{2^m-1} |x, f(x)\rangle. \qquad (1)$$

Example. Suppose $a = 11$ were randomly chosen. Since $M^2 = 441 \leq 2^9 < 882 = 2M^2$, we find $m = 9$. Thus, a total of 14 quantum bits, 9 for x and 5 for $f(x)$, are required to compute the superposition of equation 1.

Step 2. A state whose amplitude has the same period as f. The quantum Fourier transform acts on the amplitude function associated with the input state. In order to use the quantum Fourier transform to obtain the period of f, a state is constructed whose amplitude function has the same period as f.

To construct such a state, measure the last $\lceil \log_2 M \rceil$ qubits of the state of Eq. 1 that encode $f(x)$. A random value u is obtained. The value u is not of interest in itself; only the effect the measurement has on our set of superpositions is of interest. This measurement projects the state space onto the subspace compatible with the measured value, so the state after measurement is

$$C \sum_x g(x)|x, u\rangle,$$

for some scale factor C where

$$g(x) = \begin{cases} 1 & \text{if } f(x) = u \\ 0 & \text{otherwise.} \end{cases}$$

Note that the x's that actually appear in the sum, those with $g(x) \neq 0$, differ from each other by multiples of the period; thus $g(x)$ is the function we are looking for. If we could measure two successive x's in the sum, we would have the period. Unfortunately the laws of quantum physics permit only one measurement.

Example. Suppose that random measurement of the superposition of Eq. 1 produces 8. The state after this measurement[8] (Figure 2) clearly shows the periodicity of f.

Step 3. Applying a quantum Fourier transform. The $|u\rangle$ part of the state will not be

[8] Only the 9 bits of x are shown in Figure 2; the bits of $f(x)$ are known from the measurement.