Quantum Information Science

Seth Lloyd
Professor of Quantum-Mechanical Engineering
Director, WM Keck Center for Extreme Quantum Information Theory (xQIT)
Massachusetts Institute of Technology

Article Outline:

Glossary

Algorithm: A systematic procedure for solving a problem, frequently implemented as a computer program.

Bit: The fundamental unit of information, representing the distinction between two possible states, conventionally called 0 and 1. The word 'bit' is also used to refer to a physical system that registers a bit of information.

Boolean Algebra: The mathematics of manipulating bits using simple operations such as AND, OR, NOT, and COPY.

Communication Channel: A physical system that allows information to be transmitted from one place to another.

Computer: A device for processing information. A digital computer uses Boolean algebra (*q.v.*) to processes information in the form of bits.

Cryptography: The science and technique of encoding information in a secret form. The process of encoding is called encryption, and a system for encoding and decoding is called a cipher. A key is a piece of information used for encoding or decoding. Public-key cryptography operates using a public key by which information is encrypted, and a separate private key by which the encrypted message is decoded.

Decoherence: A peculiarly quantum form of noise that has no classical analog. Decoherence destroys quantum superpositions and is the most important and ubiquitous form of noise in quantum computers and quantum communication channels.

Error-Correcting Code: A technique for encoding information in a form that is resistant to errors. The syndrome is the part of the code that allows the error to be detected and that specifies how it should be corrected.

Entanglement: A peculiarly quantum form of correlation that is responsible for many types of quantum weirdness. Entanglement arises when two or more quantum systems exist in a superposition of correlated states.

Entropy: Information registered by the microscopic motion of atoms and molecules. The second law of thermodynamics (*q.v.*) states that entropy does not decrease over time.

Fault-Tolerant Computation: Computation that uses error-correcting codes to perform algorithms faithfully in the presence of noise and errors. If the rate of errors falls below a certain threshold, then computations of any desired length can be performed in a fault-tolerant fashion. Also known as robust computation.

Information: When used in a broad sense, information is data, messages, meaning, knowledge, etc. Used in the more specific sense of information theory, information is a quantity that can be measured in bits.

Logic Gate: A physical system that performs the operations of Boolean algebra (*q.v.*) such as AND, OR, NOT, and COPY, on bits.

Moore's Law: The observation, first made by Gordon Moore, that the power of computers increases by a factor of two every year and a half or so.

Quantum Algorithm: An algorithm designed specifically to be performed by a quantum computer using quantum logic. Quantum algorithms exploit the phenomena of superposition and entanglement to solve problems more rapidly than classical computer algorithms can. Examples of quantum algorithms include Shor's algorithm for factoring large numbers and breaking public-key cryptosystems, Grover's algorithm for searching databases, quantum simulation, the adiabatic algorithm, etc.

Quantum Bit: A bit registered by a quantum-mechanical system such as an atom, photon, or nuclear spin. A quantum bit, or 'qubit,' has the property that it can exist in a quantum superposition of the states 0 and 1.

Qubit: A quantum bit.

Quantum Communication Channel: A communication channel that transmits quantum bits. The most common communication channel is the bosonic channel, which transmits information using light, sound, or other substances whose elementary excitations consist of bosons (photons for light, phonons for sound).

Quantum Computer: A computer that operates on quantum bits to perform quantum algorithms. Quantum computers have the feature that they can preserve quantum superpositions and entanglement.

Quantum Cryptography: A cryptographic technique that encodes information on quantum bits. Quantum cryptography uses the fact that measuring quantum systems typically disturbs them to implement cryptosystems whose security is guaranteed by the laws of physics. Quantum key distribution (QKD) is a quantum cryptographic technique for distributing secret keys.

Quantum Error-Correcting Code: An error-correcting code that corrects for the effects of noise on quantum bits. Quantum error-correcting codes can correct for the effect of decoherence (*q.v.*) as well as for conventional bit-flip errors.

Quantum Information: Information that is stored on qubits rather than on classical bits.

Quantum Mechanics: The branch of physics that describes how matter and energy behave at their most fundamental scales. Quantum mechanics is famously weird and counterinuitive.

Quantum Weirdness: A catch-all term for the strange and counterintuitive aspects of quantum mechanics. Well-known instances of quantum weirdness include Schrödinger's cat (*q.v.*), the Einstein-Podolsky-Rosen thought experiment, violations of Bell's inequalities, and the Greenberger-Horne-Zeilinger experiment.

Reversible Logic: Logical operations that do not discard information. Quantum computers operate using reversible logic.

Schrödinger's Cat: A famous example of quantum weirdness. A thought experiment proposed by Erwin Schrödinger, in which a cat is put in a quantum superposition of being alive and being dead. Not sanctioned by the Society for Prevention of Cruelty to Animals.

Second Law of Thermodynamics: The second law of thermodynamics states that entropy does not increase. An alternative formulation of the second law states that it is not possible to build an eternal motion machine.

Superposition: The defining feature of quantum mechanics which allows particles such as electrons to exist in two or more places at once. Quantum bits can exist in superpositions of 0 and 1 simultaneously.

Teleportation: A form of quantum communication that uses pre-existing entanglement and classical communication to send quantum bits from one place to another.

I. Definition of the Subject and its Importance

Quantum mechanics is the branch of physics that describes how systems behave at their most fundamental level. The theory of information processing studies how information can be transferred and transformed. Quantum information science, then, is the theory of communication and computation at the most fundamental physical level. Quantum computers store and process information at the level of individual atoms. Quantum communication systems transmit information on individual photons.

Over the past half century, the wires and logic gates in computers have halved in size every year and a half, a phenomenon known as Moore's law. If this exponential rate of miniaturization continues, then the components of computers should reach the atomic scale within a few decades. Even at current (2008) length scales of a little larger than one hundred nanometers, quantum mechanics plays a crucial role in governing the behavior of these wires and gates. As the sizes of computer components press down toward the atomic scale, the theory of quantum information processing becomes increasingly important for characterizing how computers operate. Similarly, as communication systems become more powerful and efficient, the quantum mechanics of information transmission becomes the key element in determining the limits of their power.

Miniaturization and the consequences of Moore's law are not the primary reason for studying quantum information, however. Quantum mechanics is weird: electrons, photons, and atoms behave in strange and counterintuitive ways. A single electron can exist in two places simultaneously. Photons and atoms can exhibit a bizarre form of correlation called entanglement, a phenomenon that Einstein characterized as *spukhafte Fernwirkung*, or 'spooky action at a distance.' Quantum weirdness extends to information processing. Quantum bits can take on the values of 0 and 1 simultaneously. Entangled photons can be used to teleport the states of matter from one place to another. The essential goal of quantum information science is to determine how quantum weirdness can be used to enhance the capabilities of computers and communication systems. For example, even a moderately sized quantum computer, containing a few tens of thousands of bits, would be able to factor large numbers and thereby break cryptographic systems that have until now resisted the attacks of even the largest classical supercomputers [1]. Quantum computers could search databases faster than classical computers. Quantum communication systems allow information to be transmitted in a manner whose security against eavesdropping is guaranteed by the laws of physics.

Prototype quantum computers that store bits on individual atoms and quantum communication systems that transmit information using individual photons have been built and operated. These prototypes have been used to confirm the predictions of quantum information theory and to explore the behavior of information processing at the most microscopic scales. If larger, more powerful versions of quantum computers and communication systems become readily available, they will offer considerable enhancements over existing computers and communication systems. In the meanwhile, the field of quantum information processing is constructing a unified theory of how information can be registered and transformed at the fundamental limits imposed by physical law.

The remainder of this article is organized as follows:

II A review of the history of ideas of information, computation, and the role of information in quantum mechanics is presented.

III The formalism of quantum mechanics is introduced and applied to the idea of quantum information.

IV Quantum computers are defined and their properties presented.

V The effects of noise and errors are explored.

VI The role of quantum mechanics in setting limits to the capacity of communication channels is delineated. Quantum cryptography is explained.

VII Implications are discussed.

This review of quantum information theory is mathematically self-contained in the sense that all the necessary mathematics for understanding the quantum effects treated in detail here are contained in the introductory sectiion on quantum mechanics. By necessity, not all topics in quantum information theory can be treated in detail within the confines of this article. We have chosen to treat a few key subjects in more detail: in the case of other topics we supply references to more complete treatments. The standard reference on quantum information theory is the text by Nielsen and Chuang [1], to which the reader may turn for in depth treatments of most of the topics covered here. One topic that is left largely uncovered is the broad field of quantum technologies and techniques for actually building quantum computers and quantum communication systems. Quantum technologies are rapidly changing, and no brief review like the one given here could adequately cover both the theoretical and the experimental aspects of quantum information processing.

II. Introduction: History of Information and Quantum Mechanics

Information

Quantum information processing as a distinct, widely recognized field of scientific inquiry has arisen only recently, since the early 1990s. The mathematical theory of information and information processing dates to the mid-twentieth century. Ideas of quantum mechanics, information, and the relationships between them, however, date back more than a century. Indeed, the basic formulae of information theory were discovered in the second half of the nineteenth century, by James Clerk Maxwell, Ludwig Boltzmann, and J. Willard Gibbs [2]. These statistical mechanicians were searching for the proper mathematical characterization of the physical quantity known as entropy. Prior to Maxwell, Boltzmann, and Gibbs, entropy was known as a somewhat mysterious quantity that reduced the amount of work that steam engines could perform. After their work established the proper formula for entropy, it became clear that entropy was in fact a form of information — the information required to specify the actual microscopic state of the atoms in a substance such as a gas. If a system has W possible states, then it takes $\log_2 W$ bits to specify one state. Equivalently, any system with distinct states can be thought of as registering information, and a system that can exist in one out of W equally likely states can register $\log_2 W$ bits of information. The formula, $S = k \log W$, engraved on Boltzmann's tomb, means that entropy S is proportional to the number of bits of information registered by the microscopic state of a system such as a gas. (Ironically, this formula was first written down not by Boltzmann, but by Max Planck [3], who also gave the first numerical value 1.38×10^{-23} joule/K for the constant k. Consequently, k is called Planck's constant in early works on statistical mechanics [2]. As the fundamental constant of quantum mechanics, $h = 6.6310^{-34}$ joule seconds, on which more below, is also called Planck's constant, k was renamed Boltzmann's constant and is now typically written k_B.)

Although the beginning of the information processing revolution was still half a century away, Maxwell, Boltzmann, Gibbs, and their fellow statistical mechanicians were well aware of the connection between information and entropy. These researchers established that if the probability of the i'th microscopic state of some system is p_i, then the entropy of the system is $S = k_B(-\sum_i p_i \ln p_i)$. The quantity $\sum_i p_i \ln p_i$ was first introduced by Boltzmann, who called it H. Boltzmann's famous H-theorem declares that H never increases [2]. The H-theorem is an expression of the second law of thermodynamics, which declares that $S = -k_B H$ never decreases. Note that this formula for S reduces to that on

Boltzmann's tomb when all the states are equally likely, so that $p_i = 1/W$.

Since the probabilities for the microscopic state of a physical system depend on the knowledge possessed about the system, it is clear that entropy is related to information. The more certain one is about the state of a system–the more information one possesses about the system– the lower its entropy. As early as 1867, Maxwell introduced his famous 'demon' as a hypothetical being that could obtain information about the actual state of a system such as a gas, thereby reducing the number of states W compatible with the information obtained, and so decreasing the entropy [4]. Maxwell's demon therefore apparently contradicts the second law of thermodynamics. The full resolution of the Maxwell's demon paradox was not obtained until the end of the twentieth century, when the theory of the physics of information processing described in this review had been fully developed.

Quantum Mechanics

For the entropy, S, to be finite, a system can only possess a finite number W of possible states. In the context of classical mechanics, this feature is problematic, as even the simplest of classical systems, such as a particle moving along a line, possesses an infinite number of possible states. The continuous nature of classical mechanics frustrated attempts to use the formula for entropy to calculate many physical quantities such as the amount of energy and entropy in the radiation emitted by hot objects, the so-called 'black body radiation.' Calculations based on classical mechanics suggested the amount of energy and entropy emitted by such objects should be infinite, as the number of possible states of a classical oscillator such as a mode of the electromagnetic field was infinite. This problem is known as 'the ultraviolet catastrophe.' In 1901, Planck obtained a resolution to this problem by suggesting that such oscillators could only possess discrete energy levels [3]: the energy of an oscillator that vibrates with frequency ν can only come in multiples of $h\nu$, where h is Planck's constant defined above. Energy is *quantized*. In that same paper, as noted above, Planck first wrote down the formula $S = k \log W$, where W referred to the number of discrete energy states of a collection of oscillators. In other words, the very first paper on quantum mechanics was about information. By introducing quantum mechanics, Planck made information/entropy finite. Quantum information as a distinct field of inquiry may be young, but its origins are old: the origin of quantum information coincides with the origin of quantum mechanics.

Quantum mechanics implies that nature is, at bottom, discrete. Nature is digital. After Planck's advance, Einstein was able to explain the photo-electric effect using quantum

mechanics [5]. When light hits the surface of a metal, it kicks off electrons. The energy of the electrons kicked off depends only on the frequency ν of the light, and not on its intensity. Following Planck, Einstein's interpretation of this phenomenon was that the energy in the light comes in chunks, or *quanta*, each of which possesses energy $h\nu$. These quanta, or particles of light, were subsequently termed photons. Following Planck and Einstein, Niels Bohr used quantum mechanics to derive the spectrum of the hydrogen atom [6].

In the mid nineteen-twenties, Erwin Schrödinger and Werner Heisenberg put quantum mechanics on a sound mathematical footing [7-8]. Schrödinger derived a wave equation – the Schrödinger equation – that described the behavior of particles. Heisenberg derived a formulation of quantum mechanics in terms of matrices, matrix mechanics, which was subsequently realized to be equivalent to Schrödinger's formulation. With the precise formulation of quantum mechanics in place, the implications of the theory could now be explored in detail.

It had always been clear that quantum mechanics was strange and counterintuitive: Bohr formulated the phrase 'wave-particle duality' to capture the strange way in which waves, like light, appeared to be made of particles, like photons. Similarly, particles, like electrons, appeared to be associated with waves, which were solutions to Schrödinger's equation. Now that the mathematical underpinnings of quantum mechanics were in place, however, it became clear that quantum mechanics was downright weird. In 1935, Einstein, together with his collaborators Boris Podolsky and Nathan Rosen, came up with a thought experiment (now called the EPR experiment after its originators) involving two photons that are correlated in such a way that a measurement made on one photon appears instantaneously to affect the state of the other photon [9]. Schrödinger called this form of correlation 'entanglement.' Einstein, as noted above, referred to it as 'spooky action at a distance.' Although it became clear that entanglement could not be used to transmit information faster than the speed of light, the implications of the EPR thought experiment were so apparently bizarre that Einstein felt that it demonstrated that quantum mechanics was fundamentally incorrect. The EPR experiment will be discussed in detail below. Unfortunately for Einstein, when the EPR experiment was eventually performed, it confirmed the counterintuitive predictions of quantum mechanics. Indeed, every experiment ever performed so far to test the predictions of quantum mechanics has confirmed them, suggesting that, despite its counterintuitive nature, quantum mechanics is fundamentally

correct.

At this point, it is worth noting a curious historical phenomenon, which persists to the present day, in which a famous scientist who received his or her Nobel prize for work in quantum mechanics, publicly expresses distrust or disbelief in quantum mechanics. Einstein is the best known example of this phenomenon, but more recent examples exist, as well. The origin of this phenomenon can be traced to the profoundly counterintuitive nature of quantum mechanics. Human infants, by the age of a few months, are aware that objects – at least, large, classical objects like toys or parents – cannot be in two places simultaneously. Yet in quantum mechanics, this intuition is violated repeatedly. Nobel laureates typically possess a powerful sense of intuition: if Einstein is not allowed to trust his intuition, then who is? Nonetheless, quantum mechanics contradicts their intuition just as it does everyone else's. Einstein's intuition told him that quantum mechanics was wrong, and he trusted that intuition. Meanwhile, scientists who are accustomed to their intuitions being proved wrong may accept quantum mechanics more readily. One of the accomplishments of quantum information processing is that it allows quantum weirdness such as that found in the EPR experiment to be expressed and investigated in precise mathematical terms, so we can discover exactly how and where our intuition goes wrong.

In the 1950's and 60's, physicists such as David Bohm, John Bell, and Yakir Aharonov, among others, investigated the counterintuitive aspects of quantum mechanics and proposed further thought experiments that threw those aspects in high relief [10-12]. Whenever those thought experiments have been turned into actual physical experiments, as in the well-known Aspect experiment that realized Bell's version of the EPR experiment [13], the predictions of quantum mechanics have been confirmed. Quantum mechanics is weird and we just have to live with it.

As will be seen below, quantum information processing allows us not only to express the counterintuitive aspects of quantum mechanics in precise terms, it allows us to exploit those strange phenomena to compute and to communicate in ways that our classical intuitions would tell us are impossible. Quantum weirdness is not a bug, but a feature.

Computation

Although rudimentary mechanical calculators had been constructed by Pascal and Leibnitz, amongst others, the first attempts to build a full-blown digital computer also lie in the nineteenth century. In 1822, Charles Babbage conceived the first of a series of mechanical computers, beginning with the fifteen ton Difference Engine, intended to

calculate and print out polynomial functions, including logarithmic tables. Despite considerable government funding, Babbage never succeeded in building a working difference. He followed up with a series of designs for an Analytical Engine, which was to have been powered by a steam engine and programmed by punch cards. Had it been constructed, the analytical engine would have been the first modern digital computer. The mathematician Ada Lovelace is frequently credited with writing the first computer program, a set of instructions for the analytical engine to compute Bernoulli numbers.

In 1854, George Boole's *An investigation into the laws of thought* laid the conceptual basis for binary computation. Boole established that any logical relation, no matter how complicated, could be built up out of the repeated application of simple logical operations such as AND, OR, NOT, and COPY. The resulting 'Boolean logic' is the basis for the contemporary theory of computation.

While Schrödinger and Heisenberg were working out the modern theory of quantum mechanics, the modern theory of information was coming into being. In 1928, Ralph Hartley published an article, 'The Transmission of Information,' in the Bell System Technical Journal [14]. In this article he defined the amount of information in a sequence of n symbols to be $n \log S$, where S is the number of symbols. As the number of such sequences is S^n, this definition clearly coincides with the Planck-Boltzmann formula for entropy, taking $W = S^n$.

At the same time as Einstein, Podolsky, and Rosen were exploring quantum weirdness, the theory of computation was coming into being. In 1936, in his paper "On Computable Numbers, with an Application to the *Entscheidungsproblem*," Alan Turing extended the earlier work of Kurt Gödel on mathematical logic, and introduced the concept of a Turing machine, an idealized digital computer [15]. Claude Shannon, in his 1937 master's thesis, "A Symbolic Analysis of Relay and Switching Circuits," showed how digital computers could be constructed out of electronic components [16]. (Howard Gardner called this work, "possibly the most important, and also the most famous, master's thesis of the century.")

The Second World War provided impetus for the development of electronic digital computers. Konrad Zuse's Z3, built in 1941, was the first digital computer capable of performing the same computational tasks as a Turing machine. The Z3 was followed by the British Colossus, the Harvard Mark I, and the ENIAC. By the end of the 1940s, computers had begun to be built with a stored program or 'von Neumann' architecture (named after the pioneer of quantum mechanics and computer science John von Neumann), in which the

set of instructions – or program – for the computer were stored in the computer's memory and executed by a central processing unit.

In 1948, Shannon published his groundbreaking article, "A Mathematical Theory of Communication," in the Bell Systems Journal [17]. In this article, perhaps the most influential work of applied mathematics of the twentieth century (following the tradition of his master's thesis), Shannon provided the full mathematical characterization of information. He introduced his colleague, John Tukey's word, 'bit,' a contraction of 'binary digit,' to describe the fundamental unit of information, a distinction between two possibilities, True or False, Yes or No, 0 or 1. He showed that the amount of information associated with a set of possible states i, each with probability p_i, was uniquely given by formula $-\sum_i p_i \log_2 p_i$. When Shannon asked von Neumann what he should call this quantity, von Neumann is said to have replied that he should call it H, 'because that's what Boltzmann called it.' (Recalling the Boltzmann's orginal definition of H, given above, we see that von Neumann had evidently forgotten the minus sign.)

It is interesting that von Neumann, who was one of the pioneers both of quantum mechanics and of information processing, apparently did not consider the idea of processing information in a uniquely quantum-mechanical fashion. Von Neumann had many things on his mind, however – game theory, bomb building, the workings of the brain, etc. – and can be forgiven for failing to make the connection. Another reason that von Neumann may not have thought of quantum computation was that, in his research into computational devices, or 'organs,' as he called them, he had evidently reached the impression that computation intrinsically involved dissipation, a process that is inimical to quantum information processing [18]. This impression, if von Neumann indeed had it, is false, as will now be seen.

Reversible computation

The date of Shannon's paper is usually taken to be the beginning of the study of information theory as a distinct field of inquiry. The second half of the twentieth century saw a huge explosion in the study of information, computation, and communication. The next step towards quantum information processing took place in the early 1960s. Until that point, there was an impression, fostered by von Neumann amongst others, that computation was intrinsically irreversible: according to this view, information was necessarily lost or discarded in the course of computation. For example, a logic gate such as an AND gate takes in two bits of information as input, and returns only one bit as output: the

output of an AND gate is 1 if and only if both inputs are 1, otherwise the output is 0. Because the two input bits cannot be reconstructed from the output bits, an AND gate is irreversible. Since computations are typically constructed from AND, OR, and NOT gates (or related irreversible gates such as $NAND$, the combination of an AND gate and a NOT gate), computations were thought to be intrinsically irreversible, discarding bits as they progress.

In 1960, Rolf Landauer showed that because of the intrinsic connection between information and entropy, when information is discarded in the course of a computation, entropy must be created [19]. That is, when an irreversible logic gate such as an AND gate is applied, energy must be dissipated. So far, it seems that von Neumann could be correct. In 1963, however, Yves Lecerf showed that Turing Machines could be constructed in such a way that all their operations were logically reversible [20]. The trick for making computation reversible is record-keeping: one sets up logic circuits in such a way that the values of all bits are recorded and kept. To make an AND gate reversible, for example, one adds extra circuitry to keep track of the values of the input to the AND gate. In 1973, Charles Bennett, unaware of Lecerf's result, rederived it, and, most importantly, constructed physical models of reversible computation based on molecular systems such as DNA [21]. Ed Fredkin, Tommaso Toffoli, Norman Margolus, and Frank Merkle subsequently made significant contributions to the study of reversible computation [22].

Reversible computation is important for quantum information processing because the laws of physics themselves are reversible. It's this underlying reversibility that is responsible for Landauer's principle: whenever a logically irreversible process such as an AND gate takes place, the information that is discarded by the computation has to go somewhere. In the case of an conventional, transistor-based AND gate, the lost information goes into entropy: to operate such an AND gate, electrical energy must be dissipated and turned into heat. That is, once the AND gate has been performed, then even if the logical circuits of the computer no longer record the values of the inputs to the gate, the microscopic motion of atoms and electrons in the circuit effectively 'remember' what the inputs were. If one wants to perform computation in a uniquely quantum-mechanical fashion, it is important to avoid such dissipation: to be effective, quantum computation should be reversible.

Quantum computation

In 1980, Paul Benioff showed that quantum mechanical systems such as arrays of spins or atoms could perform reversible computation in principle [23]. Benioff mapped the operation of a reversible Turing machine onto the a quantum system and thus exhibited the first quantum-mechanical model of computation. Benioff's quantum computer was no more computationally powerful than a conventional classical Turing machine, however: it did not exploit quantum weirdness. In 1982, Richard Feynman proposed the first non-trivial application of quantum information processing [24]. Noting that quantum weirdness made it hard for conventional, classical digital computers to simulate quantum systems, Feynman proposed a 'universal quantum simulator' that could efficiently simulate other quantum systems. Feynman's device was not a quantum Turing machine, but a sort of quantum analog computer, whose dynamics could be tuned to match the dynamics of the system to be simulated.

The first model of quantum computation truly to embrace and take advantage of quantum weirdness was David Deutsch's quantum Turing machine of 1985 [25]. Deutsch pointed out that a quantum Turing machine could be designed in such a way as to use the strange and counterintuitive aspects of quantum mechanics to perform computations in ways that classical Turing machines or computers could not. In particular, just as in quantum mechanics it is acceptable (and in many circumstances, mandatory) for an electron to be in two places at once, so in a quantum computer, a quantum bit can take on the values 0 and 1 simultaneously. One possible role for a bit in a computer is as part a program, so that 0 instructs the computer to 'do this' and 1 instructs the computer to 'do that.' If a quantum bit that takes on the values 0 and 1 at the same time is fed into the quantum computer as part of a program, then the quantum computer will 'do this' and 'do that' simultaneously, an effect that Deutsch termed 'quantum parallelism.' Although it would be years before applications of quantum parallelism would be presented, Deutsch's paper marks the beginning of the formal theory of quantum computation.

For almost a decade after the work of Benioff, Feynman, and Deutsch, quantum computers remained a curiosity. Despite the development of a few simple algorithms (described in greater detail below) that took advantage of quantum parallelism, no compelling application of quantum computation had been discovered. In addition, the original models of quantum computation were highly abstract: as Feynman noted [24], no one had the slightest notion of how to build a quantum computer. Absent a 'killer ap,' and a physical

implementation, the field of quantum computation languished.

That languor dissipated rapidly with Peter Shor's discovery in 1994 that quantum computers could be used to factor large numbers [26]. That is, given the product r of two large prime numbers, a quantum computer could find the factors p and q such that $pq = r$. While it might not appear so instantaneously, solving this problem is indeed a 'killer ap.' Solving the factoring problem is the key to breaking 'public-key' cryptosystems. Public-key cryptosystems are a widely used method for secure communication. Suppose that you wish to buy something from me over the internet, for example. I openly send you a public key consisting of the number r. The public key is not a secret: anyone may know it. You use the public key to encrypt your credit card information, and send me that encrypted information. To decrypt that information, I need to employ the 'private keys' p and q. The security of public-key cryptography thus depends on the factoring problem being hard: to obtain the private keys p and q from the public key r, one must factor the public key.

If quantum computers could be built, then public-key cryptography was no longer secure. This fact excited considerable interest among code breakers, and some consternation within organizations, such as security agencies, whose job it is to keep secrets. Compounding this interest and consternation was the fact that the year before, in 1993, Lloyd had shown how quantum computers could be built using techniques of electromagnetic resonance together with 'off-the shelf' components such as atoms, quantum dots, and lasers [27]. In 1994, Ignacio Cirac and Peter Zoller proposed a technique for building quantum computers using ion traps [28]. These designs for quantum computers quickly resulted in small prototype quantum computers and quantum logic gates being constructed by David Wineland [29], and Jeff Kimble [30]. In 1996, Lov Grover discovered that quantum computers could search databases significantly faster than classical computers, another potentially highly useful application [31]. By 1997, simple quantum algorithms had been performed using nuclear magnetic resonance based quantum information processing [32-34]. The field of quantum computation was off and running.

Since 1994, the field of quantum computation has expanded dramatically. The decade between the discovery of quantum computation and the development of the first applications and implementations saw only a dozen or so papers published in the field of quantum computation. As of the date of publication of this article, it is not uncommon for a dozen papers on quantum computation to be posted on the Los Alamos preprint archive (ArXiv)

every day.

Quantum communication

While the idea of quantum computation was not introduced until 1980, and not fully exploited until the mid-1990s, quantum communication has exhibited a longer and steadier advance. By the beginning of the 1960s, J.P. Gordon [35] and Lev Levitin [36] had begun to apply quantum mechanics to the analysis of the capacity of communication channels. In 1973, Alexander Holevo derived the capacity for quantum mechanical channels to transmit classical information [37] (the Holevo-Schumacher-Westmoreland theorem [38-39]). Because of its many practical applications, the so-called 'bosonic' channel has received a great deal of attention over the years [40]. Bosonic channels are quantum communication channels in which the medium of information exchange consists of bosonic quantum particles, such as photons or phonons. That is, bosonic channels include communication channels that use electromagnetic radiation, from radio waves to light, or sound.

Despite many attempts, it was not until 1993 that Horace Yuen and Masanao Ozawa derived the capacity of the bosonic channel, and their result holds only in the absence of noise and loss [41] The capacity of the bosonic channel in the presence of loss alone was not derived until 2004 [42], and the capacity of this most important of channels in the presence of noise and loss is still unknown [43].

A second use of quantum channels is to transmit *quantum* information, rather than classical information. The requirements for transmitting quantum information are more stringent than those for transmatting classical information. To transmit a classical bit, one must end up sending a 0 or a 1. To transmit a quantum bit, by contrast, one must also faithfully transmit states in which the quantum bit registers 0 and 1 simultaneously. The quantity which governs the capacity of a channel to transmit quantum information is called the coherent information [44-45]. A particularly intriguing method of transmitting quantum information is *teleportation* [46]. Quantum teleportation closely resembles the teleportation process from the television series Star Trek. In Star Trek, entities to be teleported enter a special booth, where they are measured and dematerialized. Information about the composition of the entities is then sent to a distant location, where the entities rematerialize.

Quantum mechanics at first seems to forbid Trekkian teleportation, for the simple reason that it is not possible to make a measurement that reveals an arbitrary unknown quantum state. Worse yet, any attempt to reveal that state is likely to destroy it. Nonetheless,

if one adds just one ingredient to the protocol, quantum teleportation is indeed possible. That necessary ingredient is entanglement.

In quantum teleportation, an entity such as a quantum bit is to be teleported from Alice at point A to Bob at point B. For historical reasons, in communication protocols the sender of information is called Alice and the receiver is called Bob; an eavesdropper on the communication process is called Eve. Alice and Bob possess prior entanglement in the form of a pair of Einstein-Podolsky-Rosen particles. Alice performs a suitable measurement (described in detail below) on the qubit to be teleported together with her EPR particle. This measurement destroys the state of the particle to be teleported ('dematerializing' it), and yields two classical bits of information, which Alice sends to Bob over a conventional communication channel. Bob then performs a transformation on his EPR particle. The transformation Bob performs is a function of the information he receives from Alice: there are four possible transformations, one for each of the four possible values of the two bits he has received. After the Bob has performed his transformation of the EPR particle, the state of this particle is now guaranteed to be the same as that of the original qubit that was to be teleported.

Quantum teleportation forms a integral part of quantum communication and of quantum computation. Experimental demonstrations of quantum teleportation have been performed with photons and atoms as the systems whose quantum states are to be teleported [47-48]. At the time of the writing of this article, teleportation of larger entities such as molecules, bacteria, or human beings remains out of reach of current quantum technology.

Quantum cryptography

A particularly useful application of the counterintuitive features of quantum mechanics is quantum cryptography [49-51]. Above, it was noted that Shor's algorithm would allow quantum computers to crack public-key cryptosystems. In the context of code breaking, then, quantum information processing is a disruptive technology. Fortunately, however, if quantum computing represents a cryptographic disease, then quantum communication represents a cryptographic cure. The feature of quantum mechanics that no measurement can determine an unknown state, and that almost any measurement will disturb such a state, can be turned into a protocol for performing quantum cryptography, a method of secret communication whose security is guaranteed by the laws of physics.

In the 1970s, Stephen Wiesner developed the concept of quantum conjugate coding, in which information can be stored on two conjugate quantum variables, such as position

and momentum, or linear or helical polarization [49]. In 1984, Charles Bennett and Gilles Brassard turned Wiesner's quantum coding concept into a protocol for quantum cryptography [50]: by sending suitable states of light over a quantum communication channel, Alice and Bob can build up a shared secret key. Since any attempt of Eve to listen in on their communication must inevitably disturb the states sent, Alice and Bob can determine whether Eve is listening in, and if so, how much information she has obtained. By suitable privacy amplification protocols, Alice and Bob can distill out secret key that they alone share and which the laws of physics guarantee is shared by no one else. In 1990 Artur Ekert, unaware of Wiesner, Bennett, and Brassard's work, independently derived a protocol for quantum cryptography based on entanglement [51].

Commercial quantum cryptographic systems are now available for purchase by those who desire secrecy based on the laws of physics, rather than on how hard it is to factor large numbers. Such systems represent the application of quantum information processing that is closest to every day use.

The future

Quantum information processing is currently a thriving scientific field, with many open questions and potential applications. Key open questions include,

• Just what can quantum computers do better than classical computers? They can apparently factor large numbers, search databases, and simulate quantum systems better than classical computers. That list is quite short, however. What is the full list of problems for which quantum computers offer a speed up?

• How can we build large scale quantum computers? Lots of small scale quantum computers, with up to a dozen bits, have been built and operated. Building large scale quantum computers will require substantial technological advances in precision construction and control of complex quantum systems. While advances in this field have been steady, we're still far away from building a quantum computer that could break existing public-key cryptosystems.

• What are the ultimate physical limits to communication channels? Despite many decades of effort, fundamental questions concerning the capacity of quantum communication channels remain unresolved.

Quantum information processing is a rich stream with many tributaries in the fields of engineering, physics, and applied mathematics. Quantum information processing investigates the physical limits of computation and communication, and it devises methods for

reaching closer to those limits, and someday perhaps to attain them.

III Quantum Mechanics

In order to understand quantum information processing in any non-trivial way, some math is required. As Feynman said, " ... it is impossible to explain honestly the beauties of the laws of nature in a way that people can feel, without their having some deep understanding of mathematics. I am sorry, but this seems to be the case." [52] The counterintuitive character of quantum mechanics makes it even more imperative to use mathematics to understand the subject. The strange consequences of quantum mechanics arise directly out of the underlying mathematical structure of the theory. It is important to note that every bizarre and weird prediction of quantum mechanics that has been experimentally tested has turned out to be true. The mathematics of quantum mechanics is one of the most trustworthy pieces of science we possess.

Luckily, this mathematics is also quite simple. To understand quantum information processing requires only a basic knowledge of linear algebra, that is, of vectors and matrices. No calculus is required. In this section a brief review of the mathematics of quantum mechanics is presented, along with some of its more straightforward consequences. The reader who is familiar with this mathematics can safely skip to the following sections on quantum information. Readers who desire further detail are invited to consult reference [1].

Qubits

The states of a quantum system correspond to vectors. In a quantum bit, the quantum logic state 0 corresponds to a two-dimensional vector, $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, and the quantum logic state 1 corresponds to the vector $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. It is customary to write these vectors in the so-called 'Dirac bracket' notation:

$$|0\rangle \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix}, |1\rangle \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{1}$$

A general state for a qubit, $|\psi\rangle$, corresponds to a vector $\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \alpha|0\rangle + \beta|1\rangle$, where α and β are complex numbers such that $|\alpha|^2 + |\beta|^2 = 1$. The requirement that the amplitude squared of the components of a vector sum to one is called 'normalization.' Normalization arises because amplitudes squared in quantum mechanics are related to probabilities. In particular, suppose that one prepares a qubit in the state $|\psi\rangle$, and then performs a measurement whose purpose is to determine whether the qubit takes on the value 0 or 1 (such measurments will be discussed in greater detail below). Such a measurement will give the

outcome 0 with probability $|\alpha|^2$, and will give the outcome 1 with probability $|\beta|^2$. These probabilities must sum to one.

The vectors $|0\rangle, |1\rangle, |\psi\rangle$ are column vectors: we can also define the corresponding row vectors,

$$\langle 0| \equiv \begin{pmatrix} 1 & 0 \end{pmatrix}, \langle 1| \equiv \begin{pmatrix} 0 & 1 \end{pmatrix}, \langle \psi| \equiv \begin{pmatrix} \bar{\alpha} & \bar{\beta} \end{pmatrix}. \tag{2}$$

Note that creating the row vector $\langle\psi|$ involves both transposing the vector and taking the complex conjugate of its entries. This process is called Hermitian conjugation, and is denoted by the superscript †, so that $\langle\psi| = |\psi\rangle^\dagger$.

The two-dimensional, complex vector space for a qubit is denoted C^2. The reason for introducing Dirac bracket notation is that this vector space, like all the vector spaces of quantum mechanics, possesses a natural inner product, defined in the usual way by the product of row vectors and column vectors. Suppose $|\psi\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ and $|\phi\rangle = \begin{pmatrix} \gamma \\ \delta \end{pmatrix}$, so that $\langle\phi| = \begin{pmatrix} \bar{\gamma} & \bar{\delta} \end{pmatrix}$. The row vector $\langle\phi|$ is called a 'bra' vector, and the column vector $|\psi\rangle$ is called a 'ket' vector. Multiplied together, these vectors form the inner product, or 'bracket,'

$$\langle\phi|\psi\rangle \equiv \begin{pmatrix} \bar{\gamma} & \bar{\delta} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \alpha\bar{\gamma} + \beta\bar{\delta}. \tag{3}$$

Note that $\langle\psi|\psi\rangle = |\alpha|^2 + |\beta^2| = 1$. The definition of the inner product (3) turns the vector space for qubits C^2 into a 'Hilbert space,' a complete vector space with inner product. (Completeness means that any convergent sequence of vectors in the space attains a limit that itself lies in the space. Completeness is only an issue for infinite-dimensional Hilbert spaces and will be discussed no further here.)

We can now express probabilities in terms of brackets: $|\langle 0|\psi\rangle|^2 = |\alpha|^2 \equiv p_0$ is the probability that a measurement that distinguishes 0 and 1, made on the state $|\psi\rangle$, yields the output 0. Similarly, $|\langle 1|\psi\rangle|^2 = |\beta|^2 \equiv p_1$ is the probability that the same measurement yields the output 1. Another way to write these probabilities is to define the two 'projectors'

$$\begin{aligned} P_0 &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} = |0\rangle\langle 0| \\ P_1 &= \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \end{pmatrix} = |1\rangle\langle 1|. \end{aligned} \tag{4}$$

Note that

$$P_0^2 = |0\rangle\langle 0|0\rangle\langle 0| = |0\rangle\langle 0| = P_0. \tag{5}$$

Similarly, $P_1^2 = P_1$. A projection operator or projector P is defined by the condition $P^2 = P$. Written in terms of these projectors, the probabilities p_0, p_1 can be defined as

$$p_0 = \langle\psi|P_0|\psi\rangle, \quad p_1 = \langle\psi|P_1|\psi\rangle. \tag{6}$$

Note that $\langle 0|1\rangle = \langle 1|0\rangle = 0$: the two states $|0\rangle$ and $|1\rangle$ are orthogonal. Since any vector $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ can be written as a linear combination, or superposition, of $|0\rangle$ and $|1\rangle$, $\{|0\rangle, |1\rangle\}$ make up an orthornormal basis for the Hilbert space C^2. From the probabilistic interpretation of brackets, we see that orthogonality implies that a measurement that distinguishes between 0 and 1, made on the state $|0\rangle$, will yield the output 0 with probability 1 ($p_0 = 1$), and will never yield the output 1 ($p_1 = 0$). In quantum mechanics, orthogonal states are reliably distinguishable.

Higher dimensions

The discussion above applied to qubits. More complicated quantum systems lie in higher dimensional vector spaces. For example, a 'qutrit' is a quantum system with three distinguishable states $|0\rangle, |1\rangle, |2\rangle$ that live in the three-dimensional complex vector space C^3. All the mechanisms of measurement and definitions of brackets extend to higher dimensional systems as well. For example, the distinguishability of the three states of the qutrit implies $\langle i|j\rangle = \delta_{ij}$. Many of the familiar systems of quantum mechanics, such as a free particle or a harmonic oscillator, have states that live in *infinite* dimensional Hilbert spaces. For example, the state of a free particle corresponds to a complex valued *function* $\psi(x)$ such that $\int_{-\infty}^{\infty} \bar{\psi}(x)\psi(x)dx = 1$. The probability of finding the particle in the interval between $x = a$ and $x = b$ is then $\int_a^b \bar{\psi}(x)\psi(x)dx$. Infinite dimensional Hilbert spaces involve subtleties that, fortunately, rarely impinge upon quantum information processing except in the use of bosonic systems as in quantum optics [40].

Matrices

Quantum mechanics is an intrinsically *linear* theory: transformations of states are represented by matrix multiplication. (Nonlinear theories of quantum mechanics can be constructed, but there is no experimental evidence for any intrinsic nonlinearity in quantum mechanics.) Consider the set of matrices U such that $U^\dagger U = Id$, where Id is the identity matrix. Such a matrix is said to be 'unitary.' (For matrices on infinite-dimensional Hilbert spaces, i.e., for linear operators, unitarity also requires $UU^\dagger = Id$.) If we take a normalized

vector $|\psi\rangle$, $\langle\psi|\psi\rangle = 1$, and transform it by multiplying it by U, so that $|\psi'\rangle = U|\psi\rangle$, then we have

$$\langle\psi'|\psi\rangle = \langle\psi|U^\dagger U|\psi\rangle = \langle\psi|\psi\rangle = 1. \tag{7}$$

That is, unitary transformations U preserve the normalization of vectors. Equation (7) can also be used to show that any U that preserves the normalization of all vectors $|\psi\rangle$ is unitary. Since to be given a physical interpretation in terms of probabilities, the vectors of quantum mechanics must be normalized, the set of unitary transformations represents the set of 'legal' transformations of vectors in Hilbert space. (Below, we'll see that when one adds an environment with which qubits can interact, then the set of legal transformations can be extended.) Unitary transformations on a single qubit make up the set of two-by-two unitary matrices $U(2)$.

Spin and other observables

A familiar quantum system whose state space is represented by a qubit is the spin 1/2 particle, such as an electron or proton. The spin of such a particle along a given axis can take on only two discrete values, 'spin up,' with angular momentum $\hbar/2$ about that axis, or 'spin down,' with angular momentum $-\hbar/2$. Here, $\hbar$ is Planck's reduced constant: $\hbar \equiv h/2\pi = 1.05457\ 10^{-34} joule - sec.$ It is conventional to identify the state $|\uparrow\rangle$, spin up along the z-axis, with $|0\rangle$, and the state $|\downarrow\rangle$, spin up along the z-axis, with $|1\rangle$. In this way, the spin of an electron or proton can be taken to register a qubit.

Now that we have introduced the notion of spin, we can introduce an operator or matrix that corresponds to the measurement of spin. Let $P_\uparrow = |\uparrow\rangle\langle\uparrow|$ be the projector onto the state $|\uparrow\rangle$, and let $P_\downarrow = |\downarrow\rangle\langle\downarrow|$ be the projector onto the state $|\downarrow$. The matrix, or 'operator' corresponding to spin 1/2 along the z-axis is then

$$I_z = \frac{\hbar}{2}(P_\uparrow - P_\downarrow) = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \frac{\hbar}{2}\sigma_z, \tag{8}$$

where $\sigma_z \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ is called the z Pauli matrix. In what way does I_z correspond to spin along the z-axis? Suppose that one starts out in the state $|\psi\rangle = \alpha|\uparrow\rangle + \beta|\downarrow\rangle$ and then measures spin along the z-axis. Just as in the case of measuring 0 or 1, with probability $p_\uparrow = |\alpha|^2$ one obtains the result $\uparrow$, and with probability $p_\downarrow = |\beta|^2$ one obtains the result $\downarrow$. The *expectation value* for the angular momentum along the z-axis is then

$$\langle I_z\rangle = p_\uparrow(\hbar/2) + p_\downarrow(-\hbar/2) = \langle\psi|I_z|\psi\rangle. \tag{9}$$

That is, the expectation value of the observable quantity corresponding to spin along the z-axis is given by taking the bracket of the state $|\psi\rangle$ with the operator I_z corresponding to that observable.

In quantum mechanics, every observable quantity corresponds to an operator. The operator corresponding to an observable with possible outcome values $\{a\}$ is $A = \sum_a a|a\rangle\langle a| = \sum_a aP_a$, where $|a\rangle$ is the state with value a and $P_a = |a\rangle\langle a|$ is the projection operator corresponding to the outcome a. Note that since the outcomes of measurements are real numbers, $A^\dagger = A$: the operators corresponding to observables are Hermitian. The states $\{|a\rangle\}$ are, by definition, distinguishable and so make up an orthonormal set. From the definition of A one sees that $A|a\rangle = a|a\rangle$. That is, the different possible outcomes of the measurement are eigenvalues of A, and the different possible outcome states of the measurement are eigenvectors of A.

If more than one state $|a\rangle_i$ corresponds to the outcome a, then $A = \sum_a aP_a$, where $P_a = \sum_i |a\rangle_i\langle a|$ is the projection operator onto the eigenspace corresponding to the 'degenerate' eigenvalue a. Taking, for the moment, the case of non-degenerate eigenvalues, then the expectation value of an observable A in a particular state $|\chi\rangle = \sum_a \chi_a|a\rangle$ is obtained by bracketing the state about the corresponding operator:

$$\langle A\rangle \equiv \langle\chi|A|\chi\rangle = \sum_a |\chi_a|^2 a = \sum_a p_a a, \tag{10}$$

where $p_a = |\chi_a|^2$ is the probability that the measurement yields the outcome a.

Above, we saw that the operator corresponding to spin along the z-axis was $I_z = (\hbar/2)\sigma_z$. What then are the operators corresponding to spin along the x- and y-axes? They are given by $I_x = (\hbar/2)\sigma_x$ and $I_y = (\hbar/2)\sigma_y$, where σ_x and σ_y are the two remaining Pauli spin matrices out of the trio:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad . \tag{11}$$

By the prescription for obtaining expectation values (10), for an initial state $|\chi\rangle$ the expectation values of spin along the x-axis and spin along the y-axis are

$$\langle I_x\rangle = \langle\chi|I_x|\chi\rangle, \langle I_y\rangle = \langle\chi|I_y|\chi\rangle. \tag{12}$$

The eigenvectors of I_x, σ_x and I_y, σ_y are also easily described. The eigenvector of I_x, σ_x corresponding to spin up along the x-axis is $|\rightarrow\rangle = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$, while the eigenvector

of I_x, σ_x corresponding to spin down along the x-axis is $|\leftarrow\rangle = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$. Note that these eigenvectors are orthogonal and normalized – they make up an orthonormal set. It's easy to verify that, $\sigma_x|\uparrow\rangle = +1|\uparrow\rangle$, and $\sigma_x|\downarrow\rangle = -1|\downarrow\rangle$, so the eigenvalues of σ_x are ± 1. The eigenvalues of $I_x = (\hbar/2)\sigma_x$ are $\pm\hbar/2$, the two different possible values of angular momentum corresponding to spin up or spin down along the x-axis. Similarly, the eigenvector of I_y, σ_y corresponding to spin up along the y-axis is $|\otimes\rangle = \begin{pmatrix} 1/\sqrt{2} \\ i/\sqrt{2} \end{pmatrix}$, while the eigenvector of I_y, σ_y corresponding to spin down along the y-axis is $|\odot\rangle = \begin{pmatrix} i/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$. (Here, in deference to the right-handed coordinate system that we are implicitly adopting, $\otimes$ corresponds to an arrow heading away from the viewer, and $\odot$ corresponds to an arrow heading towards the viewer.)

Rotations and SU(2)

The Pauli matrices $\sigma_x, \sigma_y, \sigma_z$ play a crucial role not only in characterizing the measurement of spin, but in generating rotations as well. Because of their central role in describing qubits in general, and spin in particular, several more of their properties are elaborated here. Clearly, $\sigma_i = \sigma_i^\dagger$: Pauli matrices are Hermitian. Next, note that

$$\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = Id = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{13}$$

Since $\sigma_i = \sigma_i^\dagger$, and $\sigma_i^2 = Id$, it's also the case that $\sigma_i^\dagger \sigma_i = Id$: that is, the Pauli matrices are unitary. Next, defining the commutator of two matrices A and B to be $[A, B] = AB - BA$, it is easy to verify that $[\sigma_x, \sigma_y] = 2i\sigma_z$. Cyclic permutations of this identity also hold, e.g., $[\sigma_z, \sigma_x] = 2i\sigma_y$.

Now introduce the concept of a rotation. The operator $e^{-i(\theta/2)\sigma_x}$ corresponds to a rotation by an angle θ about the x-axis. The analogous operators with x replaced by y or z are expressions for rotations about the y- or z- axes. Exponentiating matrices may look at first strange, but exponentiating Pauli matrices is significantly simpler. Using the Taylor expansion for the matrix exponential, $e^A = Id + A + A^2/2! + A^3/3! + \ldots$, and employing the fact that $\sigma_j^2 = Id$, one obtains

$$e^{-i(\theta/2)\sigma_j} = \cos(\theta/2)Id - i\sin(\theta/2)\sigma_j. \tag{14}$$

It is useful to verify that the expression for rotations (14) makes sense for the states we have defined. For example, rotation by π about the x-axis should take the state $|\uparrow\rangle$,

spin z up, to the state $|\downarrow\rangle$, spin z down. Inserting $\theta = \pi$ and $j = x$ in equation (14), we find that the operator corresponding to this rotation is the matrix $-i\sigma_x$. Multiplying $|\uparrow\rangle$ by this matrix, we obtain

$$-i\sigma_x|\uparrow\rangle = -i\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = -i\begin{pmatrix} 0 \\ 1 \end{pmatrix} = -i|\downarrow\rangle. \tag{15}$$

The rotation does indeed take $|\uparrow\rangle$ to $|\downarrow\rangle$, but it also introduces an overall phase of $-i$.

What does this overall phase do? The answer is Nothing! Or, at least, nothing observable. Overall phases cannot change the expectation value of any observable. Suppose that we compare expectation values for the state $|\chi\rangle$ and for the state $|\chi'\rangle = e^{i\phi}|\chi\rangle$ for some observable corresponding to an operator A. We have

$$\langle\chi|A|\chi\rangle = \langle\chi|e^{-i\phi}Ae^{i\phi}|\chi\rangle = \langle\chi'|A|\chi'\rangle. \tag{16}$$

Overall phases are undetectable. Keeping the undetectability of overall phases in mind, it is a useful excercise to verify that other rotations perform as expected. For example, a rotation by $\pi/2$ about the x-axis takes $|\otimes\rangle$, spin up along the y-axis, to $|\uparrow\rangle$, together with an overall phase.

Once rotation about the x, y, and z axes have been defined, it is straightforward to construct rotations about any axis. Let $\hat{\iota} = (\iota_x, \iota_y, \iota_z)$, $\iota_x^2 + \iota_y^2 + \iota_z^2 = 1$, be a unit vector along the $\hat{\iota}$ direction in ordinary three-dimensional space. Define $\sigma_{\hat{\iota}} = \iota_x\sigma_x + \iota_y\sigma_y + \iota_z\sigma_z$ to be the *generalized Pauli matrix* associated with the unit vector $\hat{\iota}$. It is easy to verify that $\sigma_{\hat{\iota}}$ behaves like a Pauli matrix, e.g., $\sigma_{\hat{\iota}}^2 = Id$. Rotation by θ about the $\hat{\iota}$ axis then corresponds to an operator $e^{-i(\theta/2)\sigma_{\hat{\iota}}} = \cos(\theta/2)Id - i\sin(\theta/2)\sigma_{\hat{\iota}}$. Once again, it is a useful excercise to verify that such rotations behave as expected. For example, a rotation by π about the $(1/\sqrt{2}, 0, 1/\sqrt{2})$ axis should 'swap' $|\uparrow\rangle$ and $|\rightarrow\rangle$, up to some phase.

The set of rotations of the form $e^{-i\theta/2\sigma_{\hat{\iota}}}$ forms the *group* $SU(2)$, the set of complex 2 by 2 unitary matrices with determinant equal to 1. It is instructive to compare this group with the 'conventional' group of rotations in three dimensions, $SO(3)$. $SO(3)$ is the set of real 3 by 3 matrices with orthonormal rows/columns and determinant 1. In $SO(3)$, when one rotates a vector by 2π, the vector returns to its original state: a rotation by 2π corresponds to the 3 by 3 identity matrix. In $SU(2)$, rotating a vector by 2π corresponds to the transformation $-Id$: in rotating by 2π, the vector acquires an overall phase of -1. As will be seen below, the phase of -1, while unobservable for single qubit rotations, can be, and has been observed in two-qubit operations. To return to the original state, with

no phase, one must rotate by 4π. A macroscopic, classical version of this fact manifests itself when one grasps a glass of water firmly in the palm of one's hand and rotates one's arm and shoulder to rotate the glass without spilling it. A little experimentation with this problem shows that one must rotate glass and hand around twice to return them to their initial orientation.

Why quantum mechanics?

Why is the fundamental theory of nature, quantum mechanics, a theory of complex vector spaces? No one knows for sure. One of the most convincing explanations came from Aage Bohr, the son of Niels Bohr and a Nobel laureate in quantum mechanics in his own right [53]. Aage Bohr pointed out that the basic mathematical representation of *symmetry* consists of complex vector spaces. For example, while the apparent symmetry group of rotations in three dimensional space is the real group $SO(3)$, the actual underlying symmetry group of space, as evidenced by rotations of quantum-mechanical spins, is $SU(2)$: to return to the same state, one has to go around not once, but twice. It is a general feature of complex, continuous groups, called 'Lie groups' after Sophus Lie, that their fundamental representations are complex. If quantum mechanics is a manifestation of deep, underlying symmetries of nature, then it should come as no surprise that quantum mechanics is a theory of transformations on complex vector spaces.

Density matrices

The review of quantum mechanics is almost done. Before moving on to quantum information processing proper, two topics need to be covered. The first topic is how to deal with uncertainty about the underlying state of a quantum system. The second topic is how to treat two or more quantum systems together. These topics turn out to possess a strong connection which is the source of most counterintuitive quantum effects.

Suppose that don't know exactly what state a quantum system is in. Say, for example, it could be in the state $|0\rangle$ with probability p_0 or in the state $|1\rangle$ with probability p_1. Note that this state is not the same as a quantum superposition, $\sqrt{p_0}|0\rangle + \sqrt{p_1}|1\rangle$, which is a definite state with spin oriented in the $x-z$ plane. The expectation value of an operator A when the underlying state possesses the uncertainty described is

$$\langle A \rangle = p_0 \langle 0|A|0\rangle + p_1 \langle 1|A|1\rangle = \mathrm{tr}\rho A, \tag{17}$$

where $\rho = p_0|0\rangle\langle 0| + p_1|1\rangle\langle 1|$ is the *density matrix* corresponding to the uncertain state.

The density matrix can be thought of as the quantum mechanical analogue of a probability distribution.

Density matrices were developed to provide a quantum mechanical treatment of statistical mechanics. A famous density matrix is that for the canonical ensemble. Here, the energy state of a system is uncertain, and each energy state $|E_i\rangle$ is weighted by a probability $p_i = e^{-E_i/k_BT}/Z$, where $Z = \sum_i e^{-E_i/k_BT}$ is the partition function. Z is needed to normalize the proabilities $\{p_i\}$ so that $\sum_i p_i = 1$. The density matrix for the canonical ensemble is then $\rho_C = (1/Z)\sum_i e^{-E_i/k_BT}|E_i\rangle\langle E_i|$. The expectation value of any operator, e.g., the energy operator H (for 'Hamiltonian') is then given by $\langle H\rangle = \mathrm{tr}\rho_C H$.

Multiple systems and tensor products

To describe two or more systems requires a formalism called the tensor product. The Hilbert space for two qubits is the space $C^2 \otimes C^2$, where $\otimes$ is the tensor product. $C^2 \otimes C^2$ is a four-dimensional space spanned by the vectors $|0\rangle\otimes|0\rangle, |0\rangle\otimes|1\rangle, |1\rangle\otimes|0\rangle, |1\rangle\otimes|1\rangle$. (To save space these vectors are sometimes written $|0\rangle|0\rangle, |0\rangle|1\rangle, |1\rangle|0\rangle, |1\rangle|1\rangle$, or even more compactly, $|00\rangle, |01\rangle, |10\rangle, |11\rangle$. Care must be taken, however, to make sure that this notation is unambiguous in a particular situation.) The tensor product is *multilinear*: in performing the tensor product, the distributive law holds. That is, if $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, and $|\phi\rangle = \gamma|0\rangle + \delta|1\rangle$, then

$$\begin{aligned} |\psi\rangle \otimes |\phi\rangle =& (\alpha|0\rangle + \beta|1\rangle) \otimes (\gamma|0\rangle + \delta|1\rangle) \\ =& \alpha\gamma|0\rangle\otimes|0\rangle + \alpha\delta|0\rangle\otimes|1\rangle + \beta\gamma|1\rangle\otimes|0\rangle + \beta\delta|1\rangle\otimes|1\rangle. \end{aligned} \tag{18}$$

A tensor is a thing with slots: the key point to keep track of in tensor analysis is which operator or vector acts on which slot. It is often useful to label the slots, e.g., $|\psi\rangle_1 \otimes |\phi\rangle_2$ is a tensor product vector in which $|\psi\rangle$ occupies slot 1 and $|\phi\rangle$ occupies slot 2.

One can also define the tensor product of operators or matrices. For example, $\sigma_x^1 \otimes \sigma_z^2$ is a tensor product operator with σ_x in slot 1 and σ_z in slot 2. When this operator acts on a tensor product vector such as $|\psi\rangle_1 \otimes |\phi\rangle_2$, the operator in slot 1 acts on the vector in that slot, and the operator in slot 2 acts on the vector in that slot:

$$(\sigma_x^1 \otimes \sigma_z^2)(|\psi\rangle_1 \otimes |\phi\rangle_2) = (\sigma_x^1|\psi\rangle_1) \otimes (\sigma_z^2|\phi\rangle_2). \tag{19}$$

The no-cloning theorem

Now that tensor products have been introduced, one of the most famous theorems of quantum information – the no-cloning theorem – can immediately be proved [54]. Classical

information has the property that it can be copied, so that $0 \to 00$ and $1 \to 11$. How about quantum information? Does there exist a procedure that allows one to take an arbitrary, unknown state $|\psi\rangle$ to $|\psi\rangle \otimes |\psi\rangle$? Can you clone a quantum? As the title to this section indicates, the answer to this question is No.

Suppose that you could clone a quantum. Then there would exist a unitary operator U_C that would take the state

$$|\psi\rangle \otimes |0\rangle \to U_C|\psi\rangle \otimes |0\rangle = |\psi\rangle \otimes |\psi\rangle, \tag{20}$$

for any initial state $|\psi\rangle$. Consider another state $|\phi\rangle$. Since U_C is supposed to clone any state, we have then we would also have $U_C|\phi\rangle \otimes |0\rangle = |\phi\rangle \otimes |\phi\rangle$. If U_C exists, then, the following holds for any states $|\psi\rangle$, $|\phi\rangle$:

$$\begin{aligned}
\langle\phi|\psi\rangle &= ({}_1\langle\phi| \otimes {}_2\langle 0|)(|\psi\rangle_1 \otimes |0\rangle_2) \\
&= ({}_1\langle\phi| \otimes {}_2\langle 0|)(U_C^\dagger U_C)(|\psi\rangle_1 \otimes |0\rangle_2) \\
&= ({}_1\langle\phi| \otimes {}_2\langle 0|U_C^\dagger)(U_C|\psi\rangle_1 \otimes |0\rangle_2) \\
&= ({}_1\langle\phi| \otimes_2 \langle\phi|)(|\psi\rangle_1 \otimes |\psi\rangle_2) \\
&({}_1\langle\phi|\psi\rangle_1)({}_2\langle\phi|\psi\rangle_2) \\
&= \langle\phi|\psi\rangle^2,
\end{aligned} \tag{21}$$

where we have used the fact that U_C is unitary so that $U_C^\dagger U_C = Id$. So if cloning is possible, then $\langle\phi|\psi\rangle = \langle\phi|\psi\rangle^2$ for *any* two vectors $|\psi\rangle$ and $|\phi\rangle$. But this is impossible, as it implies that $\langle\phi|\psi\rangle$ equals either 0 or 1 for all $|\psi\rangle$, $|\phi\rangle$, which is certainly not true. You can't clone a quantum.

The no-cloning theorem has widespread consequences. It is responsible for the efficacy of quantum cryptography, which will be discussed in greater detail below. Suppose that Alice sends a state $|\psi\rangle$ to Bob. Eve wants to discover what state this is, without Alice or Bob uncovering her eavesdropping. That is, she would like to make a copy of $|\psi\rangle$ and send the original state $|\psi\rangle$ to Bob. The no-cloning theorem prevents her from doing so: any attempt to copy $|\psi\rangle$ will necessarily perturb the state. An 'optimal cloner' is a transformation that does the best possible job of cloning, given that cloning is impossible [55].

Reduced density matrices

Suppose that one makes a measurement corresponding to an observable A_1 on the state in slot 1. What operator do we take the bracket of to get the expectation value? The answer is $A_1 \otimes Id_2$: we have to put the identity in slot 2. The expectation value for this measurement for the state $|\psi\rangle_1 \otimes |\phi\rangle_2$ is then

$$ {}_1\langle\psi| \otimes {}_2\langle\phi|A_1 \otimes Id_2|\psi\rangle_1 \otimes |\phi\rangle_2 = {}_1\langle\psi|A_1|\psi\rangle_1 \otimes {}_2\langle\phi|Id_2|\phi\rangle_2 = {}_1\langle\psi|A_1|\psi\rangle_1. \tag{22} $$

Here we have used the rule that operators in slot 1 act on vectors in slot 1. Similarly, the operators in slot 2 act on vectors in slot 2. As always, the key to performing tensor manipulations is to keep track of what is in which slot. (Note that the tensor product of two numbers is simply the product of those numbers.)

In ordinary probability theory, the probabilities for two sets of events labeled by i and j is given by a joint probability distribution $p(ij)$. The probabilities for the first set of events on their own is obtained by averaging over the second set: $p(i) = \sum_j p(ij)$ is the marginal distribution for the first set of events labeled by i. In quantum mechanics, the analog of a probability distribution is density matrix. Two systems 1 and 2 are described by a joint density matrix ρ_{12}, and system 1 on its own is described by a 'reduced' density matrix ρ_1.

Suppose that systems 1 and 2 are in a state described by a density matrix

$$ \rho_{12} = \sum_{ii'jj'} \rho_{ii'jj'} |i\rangle_1\langle i'| \otimes |j\rangle_2\langle j'|, \tag{23} $$

where $\{|i\rangle_1\}$ and $\{|j\rangle_2\}$ are orthonormal bases for systems 1 and 2 respectively. As in the previous paragraph, the expectation value of a measurement made on ρ_{12} alone is given by $\mathrm{tr}\rho_{12}(A_1 \otimes Id_2)$. Another way to write such expectation values is to define the *reduced density matrix*,

$$ \begin{aligned} \rho_1 = \mathrm{tr}_2\rho_{12} &\equiv \sum_{ii'jj'} \rho_{ii'jj'} |i\rangle_1\langle i'| \otimes_2 \langle j'|j\rangle_2 \\ &= \sum_{ii'j} \rho_{ii'jj} |i\rangle_1\langle i'|. \end{aligned} \tag{24} $$

Equation (24) defines the partial trace tr_2 over system 2. In other words, if ρ_{12} has components, $\{\rho_{ii'jj'}\}$, then reduced density matrix $\rho_1 = \mathrm{tr}_2\rho_{12}$ has components $\{\sum_j \rho_{ii'jj}\}$. The expectation value of a measurement A made on the first system alone is then simply $\langle A\rangle = \mathrm{tr}\rho_1 A$. Just as in ordinary probability theory, where the marginal distribution for system 1 is obtained by averaging over the state of system 2, so in quantum mechanics

the reduced density matrix that describes system 1 is obtained by tracing over the state of system 2.

Entanglement

One of the central features of quantum information processing is entanglement. Entanglement is a peculiarly quantum-mechanical form of correlation between quantum systems, that has no classical analogue. Entanglement lies at the heart of the various speedups and enhancements that quantum information processing offers over classical information processing.

A pure state $|\psi\rangle_{12}$ for two systems 1 and 2 is entangled if the reduced density matrix for either system taken on its own has non-zero entropy. In particular, the reduced density matrix for system 1 is $\rho_1 = \mathrm{tr}_2\rho_{12}$, where $\rho_{12} = |\psi\rangle_{12}\langle\psi|$. The entropy of this density matrix is $S(\rho_1) = -\mathrm{tr}\rho_1 \log_2 \rho_1$. For pure states, the entropy of ρ_1 is equal to the entropy of ρ_2 and is a good measure of the degree of entanglement between the two systems. $S(\rho_1) = S(\rho_2)$ measures the number of 'e-bits' of entanglement between systems 1 and 2.

A mixed state ρ_{12} for 1 and 2 is entangled if it is not separable. A density matrix is separable if it can be written $\rho_{12} = \sum_j p_j \rho_1^j \otimes \rho_2^j$. In other words, a separable state is one that can be written as a classical mixture of uncorrelated states. The correlations in a separable state are purely classical.

Entanglement can take a variety of forms and manifestations. The key to understanding those forms is the notion of Local Operations and Classical Communication (LOCC) [56]. Local operations such as unitary transformations and measurement, combined with classical communication, can not, on their own, create entanglement. If one state can be transformed into another via local operations and classical communication, then the first state is 'at least as entangled' as the second. LOCC can then be used to categorize the different forms of entanglement.

Distillable entanglement is a form of entanglement that can be transformed into pure-state entanglement [57]. Systems 1 and 2 posses d qubits worth of distillable entanglement if local operations and classical communication can transform their state into a pure state that contains d e-bits (possibly with some leftover 'junk' in a separate quantum register). Systems that are non-separable, but that possess no distillable entanglement are said to possess bound entanglement [58].

The entanglement of formation for a state ρ_{12} is equal to the minimum number of e-bits of pure-state entanglement that are required to create ρ_{12} using only local operations

and classical control [59]. The entanglement of formation of ρ_{12} is greater than or equal to ρ_{12}'s distillable entanglement. A variety of entanglement measures exist. Each one is useful for different purposes. Squashed entanglement, for example, plays an important role in quantum cryptography [60]. (Squashed entanglement is a notion of entanglement based on conditional information.)

One of the most interesting open questions in quantum information theory is the definition of entanglement for marti-partite systems consisting of more than two subsystems. Here, even in the case of pure states, no unique definition of entanglement exists.

Entanglement plays a key role in quantum computation and quantum communication. Before turning to those fields, however, it is worth while investigating the strange and counterintuitive features of entanglement.

Quantum weirdness

Entanglement is the primary source of what for lack of a better term may be called 'quantum weirdness.' Consider the two-qubit state

$$|\psi\rangle_{12} = \frac{1}{\sqrt{2}}(|0\rangle_1 \otimes |1\rangle_2 - |1\rangle_1|0\rangle_2). \tag{25}$$

This state is called the 'singlet' state: if the two qubits correspond to two spin 1/2 particles, as described above, so that $|0\rangle$ is the spin z up state and $|1\rangle$ is the spin z down state, then the singlet state is the state with zero angular momentum. Indeed, rewriting $|\psi\rangle_{12}$ in terms of spin as

$$|\psi\rangle_{12} = \frac{1}{\sqrt{2}}(|\uparrow\rangle_1 \otimes |\downarrow\rangle_2 - |\downarrow\rangle_1|\uparrow\rangle_2). \tag{26}$$

one sees that if one makes a measurement of spin z, then if the first spin has spin z up, then the second spin has spin z down, and *vice versa*.

If one decomposes the state in terms of spin along the x-axis, $|\rightarrow\rangle = (1/\sqrt{2})(|\uparrow\rangle + |\downarrow\rangle)$, $|\leftarrow\rangle = (1/\sqrt{2})(|\uparrow\rangle - |\downarrow\rangle)$, then $|\psi\rangle_{12}$ can be rewritten

$$|\psi\rangle_{12} = \frac{1}{\sqrt{2}}(|\rightarrow\rangle_1 \otimes |\leftarrow\rangle_2 - |\leftarrow\rangle_1 \otimes |\rightarrow\rangle_2). \tag{27}$$

Similarly, rewriting in terms of spin along the y-axis, we obtain

$$|\psi\rangle_{12} = \frac{1}{\sqrt{2}}(|\otimes\rangle_1|\odot\rangle_2 - |\odot\rangle_1|\otimes\rangle_2), \tag{28}$$

where $|\otimes\rangle$ is the state with spin up along the y-axis and $|\odot\rangle$ is the state with spin down along the y-axis. No matter what axis one decomposes the spin about, if the first spin has

spin up along that axis then the second spin has spin down along that axis, and *vice versa*. The singlet state has angular momentum zero about every axis.

So far, this doesn't sound too strange. The singlet simply behaves the way a state with zero angular momentum should: it is not hard to see that it is the unique two-spin state with zero angular momentum about every axis. In fact, the singlet state exhibits lots of quantum weirdness. Look at the reduced density matrix for spin 1:

$$\rho_1 = \mathrm{tr}_2 \rho_{12} = \mathrm{tr}_2 |\psi\rangle_{12}\langle\psi| = \frac{1}{2}(|\uparrow\rangle_1\langle\uparrow| + |\downarrow\rangle_1\langle\downarrow| = Id/2. \tag{29}$$

That is, the density matrix for spin 1 is in a completely indefinite, or 'mixed' state: nothing is known about whether it is spin up or spin down along any axis. Similarly, spin 2 is in a completely mixed state. This is already a little strange. The two spins together are in a definite, 'pure' state, the singlet state. Classically, if two systems are in a definite state, then each of the systems on its own is in a definite state: the only way to have uncertainty about one of the parts is to have uncertainty about the whole. In quantum mechanics this is not the case: two systems can be in a definite, pure state taken together, while each of the systems on its own is in an indefinite, mixed state. Such systems are said to be *entangled* with eachother.

Entanglement is a peculiarly quantum form of correlation. Two spins in a singlet state are highly correlated (or, more precisely, anticorrelated): no matter what axis one measures spin along, one spin will be found to have the opposite spin of the other. In itself, that doesn't sound so bad, but when one makes a measurement on one spin, something funny seems to happen. Both spins start out in a completely indefinite state. Now one chooses to make a measurement of spin 1 along the z-axis. Suppose that one gets the result, spin up. As a result of the measurement, spin 2 is now in a definite state, spin down along the z axis. If one had chosen to make a measurement of spin 1 along the x-axis, then spin 2 would also be put in a definite state along the x-axis. Somehow, it seems as if one can affect the state of spin 2 by making a measurement of spin 1 on its own. This is what Einstein called 'spooky action at a distance.'

In fact, such measurements involve no real action at a distance, spooky or otherwise. If one could really act on spin 2 by making a measurement on spin 1, thereby changing spin 2's state, then one could send information instantaneously from spin 1 to spin 2 by measuring spin 1 alone. Such instantaneous transmission of information would violate special relativity and give rise to all sorts of paradoxical capabilities, such as the ability to travel backwards in time. Luckily, it is easy to see that it is impossible to send information

superluminally using entanglement: no matter what one does to spin 1, the outcomes of measurements on spin 2 are unaffected by that action. In particular, operations on spin 1 correspond to operators of the form $A_1 \otimes Id_2$, while operations on spin 2 correspond to operators of the form $Id_1 \otimes B_2$. The commutator between such operators is

$$[A_1 \otimes Id_2, Id_1 \otimes B_2] = A_1 \otimes B_2 - A_1 \otimes B_2 = 0. \tag{30}$$

Since they commute, it doesn't matter if one does something to spin 1 first, and then measures spin 2, or if one measures spin 2 first and then does something to spin 1: the results of the measurement will be the same. That is, nothing one does to spin 1 on its own can effect spin 2.

Nonetheless, entanglement is counterintuitive. One's classical intuition would like to believe that before the measurement, the system to be measured is in a definite state, even if that definite state is unknown. Such a definite state would constitute a 'hidden variable,' an unknown, classical value for the measured variable. Entanglement implies that such hidden variables can't exist in any remotely satisfactory form. The spin version of the EPR effect described above is due to David Bohm [61]. Subsequently, John Bell proposed a set of relations, the 'Bell inequalities,' that a hidden variable theory should obey [62]. Bell's inequalities are expressed in terms of the probabilities for the outcomes of measurements made on the two spins along different axes.

Suppose that each particle indeed has a particular value of spin along each axis before it is measured. Designate a particle that has spin up along the x-axis, spin down along the y axis, and spin up along the z-axis by $(x+, y-, z+)$. Designate other possible orientations similarly. In a collection of particles, let $N(x+, y-, z+)$ be the number of particles with orientations $(x+, y-, z+)$. Clearly, $N(x+, y-) = N(x+, y-, z+) + N(x+, y-, z-)$. Now, in a collection of measurements made on pairs of particles, originally in a singlet state, let $\#(x_1+, y_2-)$ be the number of measurements that give the result spin up along the x-axis for particle 1, and spin down along the y-axis for particle 2. Bell showed that for classical particles that actually possess definite values of spin along different axes before measurement, $\#(x_1+, y_2+) \leq \#(x_1+, z_2+) + \#(y_1-, z_2-)$, together with inequalities that are obtained by permuting axes and signs.

Quantum mechanics decisively violates these Bell inequalities: in entangled states like the singlet state, particles simply do not possess definite, but unknown, values of spin before they are measured. Bell's inequalities have been verified experimentally on numerous occasions [13], although not all exotic forms of hidden variables have been eliminated.

Those that are consistent with experiment are not very aesthetically appealing however (depending, of course, on one's aesthetic ideals). A stronger set of inequalities than Bell's are the CHSH inequalities (Clauser-Horne-Shimony-Holt), which have also been tested in numerous venues, with the predictions of quantum mechanics confirmed each time [63]. One of weirdest violation of classical intuition can be found in the so-called GHZ experiment, named after Daniel Greenberger, Michael Horne, and Anton Zeilinger [64].

To demonstrate the GHZ paradox, begin with the three-qubit state

$$|\chi\rangle = (1/\sqrt{2})(|\uparrow\uparrow\uparrow\rangle - |\downarrow\downarrow\downarrow\rangle) \tag{31}$$

(note that in writing this state we have suppressed the tensor product $\otimes$ signs, as mentioned above). Prepare this state four separate times, and make four distinct measurements. In the first measurement measure σ_x on the first qubit, σ_y on the second qubit, and σ_y on the third qubit. Assign the value $+1$ to the result, spin up along the axis measured, and -1 to spin down. Multiply the outcomes together. Quantum mechanics predicts that the result of this multplication will always be $+1$, as can be verified by taking the expectation value $\langle\chi|\sigma_x^1 \otimes \sigma_y^2 \otimes \sigma_y^3|\chi\rangle$ of the operator $\sigma_x^1 \otimes \sigma_y^2 \otimes \sigma_y^3$ that corresponds to making the three individual spin measurements and multiplying their results together.

In the second measurement measure σ_y on the first qubit, σ_x on the second qubit, and σ_y on the third qubit. Multiply the results together. Once again, quantum mechanics predicts that the result will be $+1$. Similarly, in the third measurement measure σ_y on the first qubit, σ_y on the second qubit, and σ_x on the third qubit. Multiply the results together to obtain the predicted result $+1$. Finally, in the fourth measurement measure σ_x on all three qubits and multiply the results together. Quantum mechanics predicts that this measurement will give the result $\langle\chi|\sigma_x^1 \otimes \sigma_x^2 \otimes \sigma_x^3|\chi\rangle = -1$.

So far, these predictions may not seem strange. A moment's reflection, however, will reveal that the results of the four GHZ experiments are completely incompatible with any underlying assignment of values of ± 1 to the spin along the x- and y-axes before the measurement. Suppose that such pre-measurement values existed, and that these are the values revealed by the measurements. Looking at the four measurements, each consisting of three individual spin measurements, one sees that each possible spin measurement appears twice in the full sequence of twelve individual spin measurements. For example, measurement of spin 1 along the x-axis occurs in the first of the four three-fold measurements, and in the last one. Similarly, measurement of spin 3 along the 3-axis occurs in the first and second three-fold measurements. The classical consequence of each individual

measurement occurring twice is that the product of all twelve measurements should be $+1$. That is, if measurement of σ_x^1 in the first measurement yields the result -1, it should also yield the result -1 in the fourth measurement. The product of the outcomes for σ_x^1 then gives $(-1) \times (-1) = +1$; similarly, if σ_x^1 takes on the value $+1$ in both measurements, it also contributes $(+1) \times (+1) = +1$ to the overall product. So if each spin possesses a definite value before the measurement, classical mechanics unambiguously predicts that the product of all twelve individual measurements should be $+1$.

Quantum mechanics, by contrast, unambiguously predicts that the product of all twelve individual measurements should be -1. The GHZ experiment has been performed in a variety of different quantum-mechanical systems, ranging from nuclear spins to photons [65-66]. The result: the predictions of classical mechanics are wrong and those of quantum mechanics are correct. Quantum weirdness triumphs.

IV. Quantum computation

Quantum mechanics has now been treated in sufficient detail to allow us to approach the most startling consequence of quantum weirdness: quantum computation. The central counterintuitive feature of quantum mechanics is quantum superposition: unlike a classical bit, which either takes on the value 0 or the value 1, a quantum bit in the superposition state $\alpha|0\rangle + \beta|1\rangle$ takes on the values 0 and 1 simultaneously. A quantum computer is a device that takes advantage of quantum superposition to process information in ways that classical computers can't. A key feature of any quantum computation is the way in which the computation puts entanglement to use: just as entanglement plays a central role in the quantum paradoxes discussed above, it also lies at the heart of quantum computation.

A classical digital computer is a machine that can perform arbitrarily complex logical operations. When you play a computer game, or operate a spread sheet, all that is going on is that your computer takes in the information from your joy stick or keyboard, encodes that information as a sequence of zeros and ones, and then performs sequences of simple logical operations one that information. Since the work of George Boole in the first half of the nineteenth century, it is known that any logical expression, no matter how involved, can be broken down into sequences of elementary logical operations such as NOT, AND, OR and $COPY$. In the context of computation, these operations are called 'logic gates': a logic gates takes as input one or more bits of information, and produces as output one or more bits of information. The output bits are a function of the input bits. A NOT gate, for example, takes as input a single bit, X, and returns as output the flipped bit, $NOT\ X$, so that $0 \rightarrow 1$ and $1 \rightarrow 0$. Similarly, an AND gate takes in two bits X, Y as input, and returns the output $X\ AND\ Y$. $X\ AND\ Y$ is equal to 1 when both X and Y are equal to 1; otherwise it is equal to 0. That is, an AND gate takes $00 \rightarrow 0$, $01 \rightarrow 0$, $10 \rightarrow 0$, and $11 \rightarrow 1$. An OR gate takes X, Y to 1 if either X or Y is 1, and to 0 if both X and Y are 0, so that $00 \rightarrow 0$, $01 \rightarrow 1$, $10 \rightarrow 1$, and $11 \rightarrow 1$. A $COPY$ gate takes a single input, X, and returns as output two bits X that are copies of the input bit, so that $0 \rightarrow 00$ and $1 \rightarrow 11$.

All elementary logical operations can be built up from NOT, AND, OR, and $COPY$. For example, implication can be written $A \rightarrow B \equiv A\ OR\ (NOT\ B)$, since $A \rightarrow B$ is false if and only if A is true and B is false. Consequently, any logical expression, e.g.,

$$\Big(\big(A\ AND\ (NOT\ B)\big)\ OR\ \big(C\ AND\ (NOT\ A)\big)\ AND\ \big(NOT(C\ OR\ B)\big) \Big), \qquad (32)$$

can be evaluated using NOT, AND, OR, and $COPY$ gates, where $COPY$ gates are used to supply the different copies of A,B and C that occur in different places in the expres-

sion. Accordingly, $\{NOT, AND, OR, COPY\}$ is said to form a 'computationally universal' set of logic gates. Simpler computationally universal sets of logic gates also exist, e.g. $\{NAND, COPY\}$, where $X\ NAND\ Y = NOT(X\ AND\ Y)$.

Reversible logic

A logic gate is said to be *reversible* if its outputs are a one-to-one function of its inputs. NOT is reversible, for example: since $X = NOT(NOT\ X)$, NOT is its own inverse. AND and OR are not reversible, as the value of their two input bits cannot be inferred from their single output. $COPY$ is reversible, as its input can be inferred from either of its outputs.

Logical reversibility is important because the laws of physics, at bottom, are reversible. Above, we saw that the time evolution of a closed quantum system (i.e., one that is not interacting with any environment) is given by a unitary transformation: $|\psi\rangle \rightarrow |\psi'\rangle = U|\psi\rangle$. All unitary transformations are invertible: $U^{-1} = U^{\dagger}$, so that $|\psi\rangle = U^{\dagger}U|\psi\rangle = U^{\dagger}|\psi'\rangle$. The input to a unitary transformation can always be obtained from its output: the time evolution of quantum mechanical systems is one-to-one. As noted in the introduction, in 1961, Rolf Landauer showed that the underlying reversibility of quantum (and also of classical) mechanics implied that logically irreversible operations such as AND necessarily required physical dissipation [19]: any physical device that performs an AND operation must possess additional degrees of freedom (i.e., an environment) which retain the information about the actual values of the inputs of the AND gate after the irreversible logical operation has discarded those values. In a conventional electronic computer, those additional degrees of freedom consist of the microscopic motions of electrons, which, as Maxwell and Boltzmann told us, register large amounts of information.

Logic circuits in contemporary electronic circuits consist of field effect transistors, or FETs, wired together to perform NOT, AND, OR and $COPY$ operations. Bits are registered by voltages: a FET that is charged at higher voltage registers a 1, and an uncharged FET at Ground voltage registers a 0. Bits are erased by connecting the FET to Ground, discharging them and restoring them to the state 0. When such an erasure or resetting operation occurs, the underlying reversibility of the laws of physics insure that the microscopic motions of the electrons in the Ground still retain the information about whether the FET was charged or not, i.e., whether the bit before the erasure operation registered 1 or 0. In particular, if the bit registered 1 initially, the electrons in Ground will be slightly more energetic than if it registered 0. Landauer argued that any such operation

that erased a bit required dissipation of energy $k_B T \ln 2$ to an environment at temperature T, corresponding to an increase in the environment's entropy of $k_B \ln 2$.

Landauer's principle can be seen to be a straightforward consequence of the microscopic reversibility of the laws of physics, together with the fact that entropy is a form of information – information about the microscopic motions of atoms and molecules. Because the laws of physics are reversible, any information that resides in the logical degrees of freedom of a computer at the beginning of a computation (i.e., in the charges and voltages of FETs) must still be present at the end of the computation in some degrees of freedom, either logical or microscopic. Note that physical reversibility also implies that if information can flow from logical degrees of freedom to microscopic degrees of freedom, then it can also flow back again: the microscopic motions of electrons cause voltage fluctuations in FETs which can give rise to logical errors. Noise is necessary.

Because $AND, OR, NAND$ are not logically reversible, Landauer initially concluded that computation was necessarily dissipative: entropy had to increase. As is often true in the application of the second law of thermodynamics, however, the appearance of irreversibility does not always imply the actual fact of irreversibility. In 1963, Lecerf showed that digital computation could always be performed in a reversible fashion [20]. Unaware of Lecerf's work, in 1973 Bennett rederived the possibility of reversible computation [21]. Most important, because Bennett was Landauer's colleague at IBM Watson laboratories, he realized the physical significance of embedding computation in a logically reversible context. As will be seen, logical reversibility is essential for quantum computation.

A simple derivation of logically reversible computation is due to Fredkin, Toffoli, and Margolus [22]. Unaware of Bennett's work, Fredkin constructed three-input, three-output reversible logic gates that could perform NOT, AND, OR, and $COPY$ operations. The best-known example of such a gate is the Toffoli gate. The Toffoli gate takes in three inputs, X, Y, and Z, and returns three outputs, X', Y' and Z'. The first two inputs go through unchanged, so that $X' = X$, $Y' = Y$. The third output is equal to the third input, unless both X and Y are equal to 1, in which case the third output is the NOT of the third input. That is, when either X or Y is 0, $Z' = Z$, and when both X and Y are 1, $Z' = NOT\ Z$. (Another way of saying the same thing is to say that $Z' = Z\ XOR\ (X\ AND\ Y)$, where XOR is the exclusive OR operation whose output is 1 when either one of its inputs is 1, but not both. That is, XOR takes $00 \rightarrow 0$, $01 \rightarrow 1$, $10 \rightarrow 1$, $11 \rightarrow 0$.) Because it performs a NOT operation on Z controlled on whether both X and Y are 1, a Toffoli gate is often

called a controlled-controlled-NOT ($CCNOT$) gate.

Figure 1: a Toffoli gate

To see that $CCNOT$ gates can be wired together to perform NOT, AND, OR, and $COPY$ operations, note that when one sets the first two inputs X and Y both to the value 1, and allows the input Z to vary, one obtains $Z' = NOT\ Z$. That is, supplying additional inputs allows a $CCNOT$ to perform a NOT operation. Similarly, setting the input Z to 0 and allowing X and Y to vary yields $Z' = X\ AND\ Y$. OR and $COPY$ (not to mention $NAND$) can be obtained by similar methods. So the ability to set inputs to predetermined values, together with ability to apply $CCNOT$ gates allows one to perform any desired digital computation.

Because reversible computation is intrinsically less dissipative than conventional, irreversible computation, it has been proposed as a paradigm for constructing low power electronic logic circuits, and such low power circuits have been built and demonstrated [67]. Because additional inputs and wires are required to perform computation reversibly, however, such circuits are not yet used for commercial application. As the miniaturization of the components of electronic computers proceeds according to Moore's law, however, dissipation becomes an increasingly hard problem to solve, and reversible logic may become commercially viable.

Quantum computation

In 1980, Benioff proposed a quantum-mechanical implementation of reversible computation [23]. In Benioff's model, bits corresponded to spins, and the time evolution of those spins was given by a unitary transformation that performed reversible logic operations. (In 1986, Feynman embedded such computation in a local, Hamiltonian dynamics, corresponding to interactions between groups of spins [68].) Benioff's model did not take into account the possibility of putting quantum bits into superpositions as an integral part of the computation, however. In 1985, Deutsch proposed that the ordinary logic gates of reversible computation should be supplemented with intrinsically quantum-mechanical single qubit operations [25]. Suppose that one is using a quantum-mechanical system to implement reversible computations using $CCNOT$ gates. Now add to the ability to prepare qubits in desired states, and to perform $CCNOT$ gates, the ability to perform single-qubit rotations of the form $e^{-i\theta\hat{\sigma}/2}$ as described above. Deutsch showed that the resulting set of operations allowed *universal quantum computation.* Not only could such a

computer perform any desired classical logical transformation on its quantum bits; it could perform any desired unitary transformation U whatsoever.

Deutsch pointed out that a computer endowed with the ability to put quantum bits into superpositions and to perform reversible logic on those superpositions could compute in ways that classical computers could not. In particular, a classical reversible computer can evaluate any desired function of its input bits: $(x_1 \ldots x_n, 0 \ldots 0) \rightarrow (x_1 \ldots x_n, f(x_1 \ldots x_n))$, where x_i represents the logical value, 0 or 1, of the ith bit, and f is the desired function. In order to preserve reversibility, the computer has been supplied with an 'answer' register, initially in the state $00 \ldots 0$, into which to place the answer $f(x_1 \ldots x_n)$. In a quantum computer, the input bits to any transformation can be in a quantum superposition. For example, if each input bit is in an equal superposition of 0 and 1, $(1/\sqrt{2})(|0\rangle + |1\rangle)$, then all n qubits taken together are in the superposition

$$\frac{1}{2^{n/2}}(|00 \ldots 0\rangle + |00 \ldots 1\rangle + \ldots + |11 \ldots 1\rangle = \frac{1}{2^{n/2}} \sum_{x_1,\ldots,x_n=0,1} |x_1 \ldots x_n\rangle. \qquad (33)$$

If such a superposition is supplied to a quantum computer that performs the transformation $x_1 \ldots x_n \rightarrow f(x_1 \ldots x_n)$, then the net effect is to take the superposition

$$\frac{1}{2^{n/2}} \sum_{x_1,\ldots,x_n=0,1} |x_1 \ldots x_n\rangle|00 \ldots 0\rangle \rightarrow \frac{1}{2^{n/2}} \sum_{x_1,\ldots,x_n=0,1} |x_1 \ldots x_n\rangle|f(x_1 \ldots x_n)\rangle. \qquad (34)$$

That is, even though the quantum computer evaluates the function f only once, it evaluates it on every term in the superposition of inputs simultaneously, an effect which Deutsch termed 'quantum parallelism.'

At first, quantum parallelism might seem to be spectacularly powerful: with only one function 'call,' one performs the function on 2^n different inputs. The power of quantum parallelism is not so easy to tease out, however. For example, suppose one makes a measurement on the output state in equation (33) in the $\{|0\rangle, |1\rangle\}$ basis. The result is a *randomly selected* input-output pair, $(x_1 \ldots x_n, f(x_1 \ldots x_n))$. One could have just as easily obtained such a pair by feeding a random input string into a classical computer that evaluates f. As will now be seen, the secret to orchestrating quantum computations that are more powerful than classical compuations lies in arranging quantum interference between the different states in the superposition of equation (34).

The word 'orchestration' in the previous sentence was used for a reason. In quantum mechanics, states of physical systems correspond to waves. For example, the state of an

electron is associated with a wave that is the solution of the Schrödinger equation for that electron. Similarly, in a quantum computer, a state such as $|x_1 \ldots x_n\rangle|f(x_1 \ldots x_n)\rangle$ is associated with a wave that is the solution of the Schrödinger equation for the underlying quantum degrees of freedom (e.g., electron spins or photon polarizations) that make up the computers quantum bits. The waves of quantum mechanics, like waves of water, light, or sound, can be superposed on eachother to construct composite waves. A quantum computer that performs a conventional reversible computation, in which its qubits only take on the values 0 or 1 and are never in superpositions $\alpha|0\rangle + \beta|1\rangle$, can be thought of as an analogue of a piece of music like a Gregorian chant, in which a single, unaccompanied voice follows a prescribed set of notes. A quantum computer that performs many computations in quantum parallel is analogous to a symphony, in which many lines or voices are superposed to create chords, counterpoint, and harmony. The quantum computer programmer is the composer who writes and orchestrates this quantum symphony: her job is to make that counterpoint reveal meaning that is not there in each of the voices taken separately.

Deutsch-Jozsa algorithm

Let's examine a simple example, due to David Deutsch and Richard Jozsa, in which the several 'voices' of a quantum computer can be orchestrated to solve a problem more rapidly than a classical computer [69]. Consider the set of functions f that take one bit of input and produce one bit of output. There are four such functions:

$$f(x) = 0, f(x) = 1, f(x) = x, f(x) = NOT\ x. \tag{35}$$

The first two of these functions are constant functions; the second two are 'balanced' in the sense that half of their inputs yield 0 as output, while the other half yield 1. Suppose that one is presented with a 'black box' that implements one of these functions. The problem is to query this black box and to determine whether the function the box contains is constant or balanced.

Classically, it clearly takes exactly two queries to determine whether the function in the box is constant or balanced. Using quantum information processing, however, one query suffices. The following quantum circuit shows how this is accomplished.

Figure 2: Deutsch-Jozsa Circuit

Quantum circuit diagrams are similar in character to their classical counterparts: qubits enter on the left, undergo a series of transformations effected by quantum logic

gates, and exit at the right, where they are measured. In the circuit above, the first gate, represented by H is called a Hadamard gate. The Hadamard gate is a single-qubit quantum logic gate that effects the transformation

$$|0\rangle \rightarrow (1/\sqrt{2})(|0\rangle + |1\rangle), |1\rangle \rightarrow (1/\sqrt{2})(|0\rangle - |1\rangle). \tag{36}$$

In other words, the Hadamard performs a unitary transformation $U_H = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$ on its single-qubit input. Note that the Hadamard transformation is its own inverse: $U_H^2 = Id$.

The second logic gate implements the unknown, black-box function f. It takes two binary inputs, x, y, and gives two binary outputs. The gate leaves the first input unchanged, and adds $f(x)$ to the second input (modulo 2), so that $x \rightarrow x$ and $y \rightarrow y + f(x)$ $(mod 2)$. Such gates can be implemented using the controlled-NOT operation introduced above. Recall that the controlled-NOT or CNOT leaves its first input bit unchanged, and flips the second if and only if the first input is 1. In the symbol for a controlled-NOT operation, the $\bullet$ part represents the control bit and the $\oplus$ part represents the bit that can be flipped. The circuits required to implement the four different functions from one bit to one bit are as follows:

$$f(x) = 0: \quad f(x) = 1: \quad f(x) = x: \quad f(x) = NOT\ x: \tag{37}$$

The black box in the Deutsch-Jozsa algorithm contains one of these circuits. Note that the black-box circuits are 'classical' in the sense that they map input combinations of 0's and 1's to output combinations of 0's and 1's: the circuits of equation (37) make sense as classical circuits as well as quantum circuits.

Any classical circuit that can determine whether f is constant or balanced requires at least two uses of the f gate. By contrast, the Deutsch-Jozsa circuit above requires only one use of the f gate. Going through the quantum logic circuit, one finds that a constant function yields the output $|0\rangle$ on the first output line, while a balanced function yields the output $|1\rangle$ (up to an overall, unobservable phase). That is, only a single function call is required to reveal whether f is constant or balanced.

Several comments on the Deutsch-Jozsa algorithm are in order. The first is that, when comparing quantum algorithms to classical algorithms, it is important to compare apples to apples: that is, the gates used in the quantum algorithm to implement the black-box circuits should be the same as those used in any classical algorithms. The difference, of course, is that the quantum gates preserve quantum coherence, a concept

which is meaningless in the classical context. This requirement has been respected in the Deutsch-Jozsa circuit above.

The second comment is that the Deutsch-Jozsa algorithm is decidedly odd and counterintuitive. The f gates and the controlled-NOT gates from which they are constructed both have the property that the first input passes through unchanged $|0\rangle \rightarrow |0\rangle$ and $|1\rangle \rightarrow |1\rangle$. Yet somehow, when the algorithm is identifying balanced functions, the first bit flips. How can this be? This is the part where quantum weirdness enters. Even though the f and controlled-NOT gates leave their first input unchanged in the logical basis $\{|0\rangle, |1\rangle\}$, the same property does not hold in other bases. For example, let $|+\rangle = (1/\sqrt{2})(|0\rangle + |1\rangle) = U_H|0\rangle$, and let $|1\rangle = (1/\sqrt{2})(|0\rangle - |1\rangle) = U_H|1\rangle$. Straightforward calculation shows that, when acting on the basis $\{|+\rangle, |-\rangle\}$, the CNOT still behaves like a CNOT, but with the roles of its inputs reversed: now the second qubit passes through unchanged, while the first qubit gets flipped. It is this quantum role reversal that underlies the efficacy of the Deutsch-Jozsa algorithm. Pretty weird.

It is important to note that the Deutsch-Jozsa algorithm is not just a theoretical point. The algorithm has been implemented using techniques from nuclear magnetic resonance (NMR) [33]. The results are exactly as predicted by quantum mechanics: a single function call suffices to determine whether that function is constant or balanced.

The two-qubit algorithm was first described by David Deutsch. Later, with Richard Jozsa, he extended this algorithm to a multi-qubit algorithm. Now consider functions f from n qubits to a single qubit. Once again, the problem is to determine whether or not f is constant or balanced. That is, the function f in the black box is either constant: $f(x) = 0$ for all n-bit inputs x, or $f(x) = 1$ for all x, or balanced: $f(x) = 0$ for exactly half of its 2^n possible input strings, and $f(x) = 1$ for the other half. (If this problem statement seems somewhat artificial, note that the algorithm works equally well for distinguishing between constant functions and 'typical' functions, which are *approximately* balanced.)

On average, a classical algorithm takes a little more than two function calls to distinguish between a constant or a balanced function. However, in the worst case, it takes $2^{n-1} + 1$ calls, as more than half the inputs have to be sampled. As before, the quantum algorithm takes but a single function call, as the following circuit shows.

Figure 3: full Deutsch-Jozsa circuit

To determine whether the f is constant or balanced, one measures the first n output

bits: if they are all 0, then the function is constant; if one or more is 1, then the function is balanced.

Other algorithms: the Quantum Fourier Transform

While it conclusively demonstrates that quantum computers are strictly more powerful than classical computers for certain problems, the Deutsch-Jozsa algorithm does not solve a problem of burning interest to applied computer scientists. Once it was clear that quantum computers could offer a speedup over classical algorithms, however, other algorithms began to be developed. Simon's algorithm [70], for example, determines whether a function f from n bits to n bits is (a) one-to-one, or (b) two-to-one with a large period s, so that $f(x+s) = f(x)$ for all x. (In Simon's algorithm the addition is bitwise modulo 2, with no carry bits.)

Simon's algorithm has a similar 'flavor' to the Deutsch-Jozsa algorithm: it is intriguing but does not obviously admit wide application. A giant step towards constructing more useful algorithms was Coppersmith's introduction [71] of the Quantum Fourier Transform (QFT). The fast Fourier transform maps a function of n bits to its discrete Fourier transform function:

$$f(x) \rightarrow g(y) = \sum_{x=0}^{2^n-1} e^{2\pi i x y/2^n} f(x). \tag{39}$$

The fast Fourier transform takes $O(n2^n)$ steps. The quantum Fourier transform takes a *wave function* over n qubits to a Fourier transformed wave function:

$$\sum_{x=0}^{2^n-1} f(x)|x\rangle \rightarrow 2^{-n/2} \sum_{x,y=0}^{2^n-1} e^{2\pi i x y/2^n} f(x)|y\rangle. \tag{40}$$

It is not difficult to show that the quantum Fourier transform is a unitary.

To obtain a quantum logic circuit that accomplishes the QFT, it is convenient to express states in a binary representation. In the equations above, x and y are n-bit numbers. Write x as $x_n \ldots x_1$, where $x_n, \ldots x_1$ are the bits of x. This is just a more concise way of saying that $x = x_1 2^0 + \ldots + x_n 2^{n-1}$. Similarly, the expression $0.y_1 \ldots y_m$ represents the number $y_1/2 + \ldots y_m/2^{-m}$. Using this binary notation, it is not hard to show that the quantum Fourier transform can be written:

$$|x_1 \ldots x_n\rangle \rightarrow 2^{-n/2}(|0\rangle + e^{2\pi i 0.x_1}|1\rangle)(|0\rangle + e^{2\pi i 0.x_2 x_1}|1\rangle) \ldots (|0\rangle + e^{2\pi i 0.x_n \ldots x_1}|1\rangle). \tag{41}$$

When the quantum Fourier transform is written in this form, it is straightforward to construct a circuit that implements it:

Figure 4: Quantum Fourier transform circuit.

Note that the QFT circuit for wave functions over n qubits takes $O(n^2)$ steps: it is exponentially faster than the FFT for functions over n bits, which takes $O(n2^n)$ steps. This exponential speedup of the quantum Fourier transform is what guarantees the efficacy of many quantum algorithms.

The quantum Fourier transform is a potentially powerful tool for obtaining exponential speedups for quantum computers over classical computers. The key is to find a way of posing the problem to be solved in terms of finding periodic structure in a wave function. This step is the essence of the best known quantum algorithm, Shor's algorithm for factoring large numbers [26].

Shor's algorithm

The factoring problem can be stated as follows: Given $N = pq$, where p, q are prime, find p and q. For large p and q, this problem is apparently hard for classical computers. The fastest known algorithm (the 'number sieve') takes $O(N^{1/3})$ steps. The apparent difficulty of the factoring problem for classical computers is important for cryptography. The commonly used RSA public-key cryptosystem relies on the difficulty of factoring to guarantee security. Public-key cryptography addresses the following societally important situation. Alice wants to send Bob some secure information (e.g., a credit card number). Bob sends Alice the number N, but does not reveal the identity of p or q. Alice then uses N to construct an encrypted version of the message she wishes to send. Anyone who wishes to decrypt this message must know what p and q are. That is, encryption can be performed using the public key N, but decryption requires the private key p, q.

In 1994, Peter Shor showed that quantum computers could be used to factor large numbers and so crack public-key cryptosystems that whose security rests on the difficulty of factoring [26]. The algorithm operates by solving the so-called 'discrete logarithm' problem. This problem is, given N and some number x, find the smallest r such that $x^r \equiv 1 \ (mod \ N)$. Solving the discrete logarithm allows N to be factored by the following procedure. First, pick $x < N$ at random. Use Euclid's algorithm to check that the greatest common divisor of x and N is 1. (Euclid's algorithm is to divide N by x; take the remainder r_1 and divide x by r_1; take the remainder of that division, r_2 and divide r_1 by that, etc. The final remainder in this procedure is the greatest common divisor, or g.c.d., of x and N.) If the g.c.d. of x and N is not 1, then it is either p or q and we are done.

If the greatest common divisor of x and N is 1, suppose that we can solve the discrete logarithm problem to find the smallest r such that $x^r \equiv 1 \ (mod\ N)$. As will be seen, if r is even, we will be able to find the factors of N easily. If r turns out to be odd, just pick a new x and start again: continue until you obtain an even r (since this occurs half the time, you have to repeat this step no more than twice on average). Once an even r has been found, we have $(x^{r/2}-1)(x^{r/2}+1) \equiv 1 \ (mod\ N)$. In other words, $(x^{r/2}-1)(x^{r/2}+1) = bN = bpq$ for some b. Finding the greatest common divisor of $x^{r/2}-1$, $x^{r/2}+1$ and N now reveals p and q. The goal of the quantum algorithm, then, is to solve the discrete logarithm problem to find the smallest r such that $x^r \equiv 1\ mod\ N$. If r can be found, then N can be factored.

In its discrete logarithm guise, factoring possesses a periodic structure that the quantum Fourier transform can reveal. First, find an x whose g.c.d. with N is 1, as above, and pick n so that $N^2 < 2^n < 2N^2$. The quantum algorithm uses two n-qubit registers. Begin by constructing a uniform superposition $2^{-n/2}\sum_{k=0}^{2^n-1}|k\rangle|0\rangle$. Next, perform exponentiation modulo N to construct the state,

$$2^{-n/2}\sum_{k=0}^{2^n-1}|k\rangle|x^k\ mod\ N\rangle. \tag{42}$$

This modular exponentiation step takes $O(n^3)$ operations (note that $x^{2^k}\ mod\ N$ can be evaluated by first constructing $x^2\ mod\ N$, then constructing $(x^2)^2\ mod\ N$, etc.). The periodic structure in equation (42) arises because if $x^k \equiv a\ mod\ N$, for some a, then $x^{k+r} \equiv a\ mod\ N$, $x^{k+2r} \equiv a\ mod\ N$, ..., $x^{k+mr} \equiv a\ mod\ N$, up to the largest m such that $k+mr < N^2$. The same periodicity holds for any a. That is, the wave function (42) is periodic with with period r. So if we apply the quantum Fourier transform to this wave function, we can reveal that period and find r, thereby solving the discrete logarithm and factoring problems.

To reveal the hidden period and find r apply the QFT to the *first* register in the state (42). The result is

$$2^{-n}\sum_{jk=0}^{2^n-1}e^{2\pi ijk/2^n}|j\rangle|x^k\ mod\ N\rangle. \tag{43}$$

Because of the periodic structure, positive interference takes place when $j(k+\ell r)$ is close to a multiple of 2^n. That is, measuring the first register now yields a number j such that $jr/2^n$ is close to an integer: only for such j does the necessary positive interference take place. In other words, the algorithm reveals a j such that $j/2^n = s/r$ for some

integer s. That is, to find r, we need to find fractions s/r that approximate $j/2^n$. Such fractions can be obtained using a continued fraction approximation. With reasonably high probability, the result of the continued fraction approximation can be shown to yield the desired answer r. (More precisely, repetition of this procedure $O(2 \log N)$ times suffices to identify r.) Once r is known, then the factors p and q of N can be recovered by the reduction of factoring to discrete logarithm given above.

The details of Shor's algorithm reveal considerable subtlety, but the basic idea is straightforward. In its reduction to discrete logarithm, factoring possesses a hidden periodic structure. This periodic structure can be revealed using a quantum Fourier transform, and the period itself in turn reveals the desired factors.

More recent algorithms also put the quantum Fourier transform to use to extract hidden periodicities. Notably, the QFT can be used to find solutions to Pell's equation ($x^2 - ny^2 = 1$, for non-square n) [72]. Generalizations of the QFT to transforms over groups (the dihedral group and the permutation group on n objects S_n) have been applied to other problems such as the shortest vector on a lattice [73] (dihedral group, with some success) and the graph isomorphism problem (S_n, without much success [74]).

The phase-estimation algorithm

One of the most useful applications of the quantum Fourier transform is finding the eigenvectors and eigenvalues of unitary transformations. The resulting algorithm is called the 'phase-estimation' algorithm: its original form is due to Kitaev [75]. Suppose that we have the ability to apply a 'black box' unitary transformation U. U can be written $U = \sum_j e^{i\phi_j} |j\rangle\langle j|$, where $|j\rangle$ are the eigenvectors of U and $e^{i\phi_j}$ are the corresponding eigenvalues. The goal of the algorithm is to estimate the $e^{i\phi_j}$ and the $|j\rangle$. (The goal of the original Kitaev algorithm was only to estimate the eigenvalues $e^{i\phi_j}$. However, Abrams and Lloyd showed that the algorithm could also be used to construct and estimate the eigenvectors $|j\rangle$, as well [76]. The steps of the phase estimation algorithm are as follows.

(0) Begin with the inital state $|0\rangle|\psi\rangle$, where $|0\rangle$ is the n-qubit state $00\ldots0\rangle$, and $|\psi\rangle$ is the state that one wishes to decompose into eigenstates: $|\psi\rangle = \sum_j \psi_j |j\rangle$.

(1) Using Hadamards or a QFT, put the first register into a uniform superposition of all possible states:

$$\rightarrow 2^{-n/2} \sum_{k=0}^{2^n-1} |k\rangle|\psi\rangle.$$

(2) In the k'th component of the superposition, apply U^k to $|\psi\rangle$:

$$\begin{aligned}
&\rightarrow 2^{-n/2} \sum_{k=0}^{2^n-1} |k\rangle U^k |\psi\rangle \\
&= 2^{-n/2} \sum_{j,k=0}^{2^n-1} |k\rangle U^k \psi_j |j\rangle \\
&= 2^{-n/2} \sum_{j,k=0}^{2^n-1} \psi_k e^{ik\phi_j} |k\rangle |j\rangle.
\end{aligned}$$

(3) Apply inverse QFT to first register:

$$\rightarrow 2^{-n} \sum_{j,k,l=0}^{2^n-1} \psi_k e^{ik\phi_j} e^{-2\pi ikl/2^n} |l\rangle |j\rangle.$$

(4) Measure the registers. The second register contains the eigenvector $|j\rangle$. The first register contains $|l\rangle$ where $2\pi l/2^n \approx \phi_j$. That is, the first register contains an n-bit approximation to ϕ_j.

By repeating the phase-estimation algorithm many times, one samples the eigenvectors and eigenvalues of U. Note that to obtain n-bits of accuracy, one must possess the ability to apply U 2^n times. This feature limits the applicability of the phase-estimation algorithm to a relatively small number of bits of accuracy, or to the estimation of eigenvalues of Us that can easily be applied an exponentially large number of times. We've already seen such an example of a process in modular exponentiation. Indeed, Kitaev originally identified the phase estimation algorithm as an alternative method for factoring.

Even when only a relatively small number of applications of U can be performed, however, the phase-estimation algorithm can provide an exponential improvement over classical algorithms for problems such as estimating the ground state of some physical Hamiltonian [76-77], as will now be seen.

Quantum simulation

One of the earliest uses for a quantum computer was suggested by Richard Feynman [24]. Feynman noted that simulating quantum systems on a classical computer was hard: computer simulations of systems such as lattice gauge theories take up a substantial fraction of all supercomputer time, and, even then, are often far less effective than their

programmers could wish them to be. The reason why it's hard to simulate a quantum system on a classical computer is straightforward: in the absence of any sneaky tricks, the only known way to simulate a quantum system's time evolution is to construct a representation of the full state of the system, and to evolve that state forward using the system's quantum equation of motion. To represent the state of a quantum system on a classical computer is typically exponentially hard, however: an n-spin system requires 2^n complex numbers to represent its state. Evolving that state forward is even harder: it requires exponentiation of a 2^n by 2^n matrix. Even for a small quantum system, for example, one containing fifty spins, this task lies beyond the reach of existing classical supercomputers. True, supercomputers are also improving exponentially in time (Moore's law). No matter how powerful they become, however, they will not be able to simulate more than 300 spins directly, for the simple reason that to record the 2^{300} numbers that characterize the state of the spins would require the use of all 2^{300} particles in the universe within the particle horizon.

Feynman noted that if one used qubits instead of classical bits, the state of an n-spin system can be represented using just n qubits. Feynman proposed a class of systems called 'universal quantum simulators' that could be programmed to simulate any other quantum system. A universal quantum simulator has to possess a flexible dynamics that can be altered at will to mimic the dynamics of the system to be simulated. That is, the dynamics of the universal quantum simulator form an *analog* to the dynamics of the simulated system. Accordingly, one might also call quantum simulators, 'quantum analog computers.'

In 1996, Lloyd showed how Feynman's proposal could be turned into a quantum algorithm [78]. For each degree of freedom of the system to be simulated, allocate a quantum register containing a sufficient number of qubits to approximate the state of that degree of freedom to some desired accuracy. If one wishes to simulate the system's interaction with the environment, a number of registers should also be allocated to simulate the environment (for a d-dimensional system, up to d^2 registers are required to simulate the environment). Now write the Hamiltonian of the system an environment as $H = \sum_{\ell=1}^{m} H_\ell$, where each H_ℓ operates on only a few degrees of freedom. The Trotter formula implies that

$$e^{-iH\delta t} = e^{-iH_1\Delta t} \ldots e^{-iH_m\Delta t} - \frac{1}{2}\sum_{jk}[H_j, H_k]\Delta t^2 + O(\Delta t^3). \tag{44}$$

Each $e^{-iH_\ell\Delta t}$ can be simulated using quantum logic operations on the quantum bits in the

registers corresponding to the degrees of freedom on which H_ℓ acts. To simulate the time evolution of the system over time $t = n\Delta t$, we simply apply $e^{-iH\Delta t}$ n times, yielding

$$e^{-iHt} = (e^{-iH\Delta t})^n = (\Pi_\ell e^{-iH_\ell \Delta t})^n - \frac{n}{2}\sum_{jk}[H_j, H_k]\Delta t^2 + O(\Delta t^3). \tag{45}$$

The quantum simulation takes $O(mn)$ steps, and reproduces the original time evolution to an accuracy $h^2t^2m^2/n$, where h is the average size of $\|[H_j, H_k]\|$ (note that for simulating systems with local interactions, most of these terms are zero, because most of the local interactions commute with eachother).

A second algorithm for quantum simulation takes advantage of the quantum Fourier transform [79-80]. Suppose that one wishes to simulate the time evolution of a quantum particle whose Hamiltonian is of the form $H = P^2/2m + V(X)$, where $P = -i\partial/\partial x$ is the momentum operator for the particle, and $V(X)$ is the potential energy operator for the particle expressed as a function of the position operator X. Using an n-bit discretization for the state we identify the x eigenstates with $|x\rangle = |x_n \ldots x_1\rangle$. The momentum eigenstates are then just the quantum Fourier transform of the position eigenstates: $|p\rangle = 2^{-n/2}\sum_{x=0}^{2^n-1} e^{2\pi i x p/2^n}|x\rangle$. That is, $P = U_{QFT} X U_{QFT}^\dagger$.

By the Trotter formula, the infinitesimal time evolution operator is

$$e^{-iH\Delta t} = e^{-iP^2\Delta t/2m} e^{-iV(X)\Delta t} + O(\delta t^2). \tag{46}$$

To enact this time evolution operator one proceeds as above. Write the state of the particle in the x-basis: $|\psi\rangle = \sum_x \psi_x |x\rangle$. First apply the infinitesimal $e^{-iV(X)\Delta t}$ operator:

$$\sum_x \psi_x |x\rangle \rightarrow \sum_x \psi_x e^{-iV(x)}|x\rangle. \tag{47}$$

To apply the infinitesimal $e^{-iP^2\delta t/2m}$ operator, first apply an inverse quantum Fourier transform on the state, then apply the unitary transformation $|x\rangle \rightarrow e^{-ix^2\Delta t/2m}|x\rangle$, and finally apply the regular QFT. Because X and P are related by the quantum Fourier transform, these three steps effectively apply the transformation $e^{-iP^2\Delta t/2m}$. Applying first $e^{-iV(X)\Delta t}$ then $e^{-iP^2\Delta t/2m}$ yields the full infinitesimal time evolution (46). The full time evolution operator e^{-iHt} can then be built up by repeating the infinitesimal operator $t/\Delta t$ times. As before, the accuracy of the quantum simulation can be enhanced by slicing time ever more finely.

Quantum simulation represents one of the most powerful uses of quantum computers. It is probably the application of quantum computers that will first give an advantage over classical supercomputers, as only one hundred qubits or fewer are required to simulate, e.g., molecular orbitals or chemical reactions, more accurately than the most powerful classical supercomputer. Indeed, special purpose quantum simulators have already been constructed using nuclear magnetic resonance techniques [81]. These quantum analog computers involve interactions between many hundreds of nuclear spins, and so are already performing computations that could not be performed by any classical computer, even one the size of the entire universe.

Quantum search

The algorithms described above afford an exponential speedup over the best classical algorithms currently known. Such exponential speedups via quantum computation are hard to find, and are currently limited to a few special problems. There exists a large class of quantum algorithms afford a *polynomial* speedup over the best possible classical algorithms, however. These algorithms are based on Grover's quantum search algorithm.

Grover's algorithm [31] allows a quantum computer to search an unstructured database. Suppose that this database contains N items, one of which is 'marked,' and the remainder of which are unmarked. Call the marked item w, for 'winner.' Such a database can be represented by a function $f(x)$ on the items in the database, such that f of the marked item is 1, and f of any unmarked item is 0. That is, $f(w) = 1$, and $f(x \neq w) = 0$. A classical search for the marked item must take $N/2$ database calls, on average. By contrast, a quantum search for the marked item takes $O(\sqrt{N})$ calls, as will now be shown.

Unstructured database search is an 'oracle' problem. In computer science, an oracle is a 'black box' function: one can supply the black box with an input x, and the black box then provides an output $f(x)$, but one has no access to the mechanism inside the box that computes $f(x)$ from x. For the quantum case, the oracle is represented by a function on two registers, one containing x, and the other containing a single qubit. The oracle takes $|x\rangle|y\rangle \rightarrow |x\rangle|y + f(x)\rangle$, where the addition takes place modulo 2.

Grover originally phrased his algorithm in terms of a 'phase' oracle U_w, where $|x\rangle U_w|x\rangle = (-1)^{f(x)}|x\rangle$. In other words, the 'winner' state acquires a phase of -1: $|w\rangle \rightarrow -|w\rangle$, while the other states remain unchanged: $|x \neq w\rangle \rightarrow |x\rangle$. Such a phase oracle can be constructed from the original oracle in several ways. The first way involves two oracle calls. Begin with the state $|x\rangle|0\rangle$ and call the oracle once to construct the state $|x\rangle|f(x)\rangle$. Now

apply a σ_z transformation to the second register. The effect of this is to take the state to $(-1)^{f(x)}|x\rangle|f(x)\rangle$. Applying the oracle for a second time yields the desired phase-oracle state $(-1)^{f(x)}|x\rangle|0\rangle$. A second, sneakier way to construct a phase oracle is to initialize the second qubit in the state $(1/\sqrt{2})(|0\rangle - |1\rangle)$. A *single* call of the original oracle on the state $|x\rangle(1/\sqrt{2})(|0\rangle - |1\rangle)$ then transforms this state into $(-1)^{f(x)}|x\rangle((1/\sqrt{2})(|0\rangle - |1\rangle))$. In this way a phase oracle can be constructed from a single application of the original oracle.

Two more ingredients are needed to perform Grover's algorithm. Let's assume that $N = 2^n$ for some n, so that the different states $|j\rangle$ can be written in binary form. Let U_0 be the unitary transformation that takes $|0\ldots0\rangle \to -|0\ldots0\rangle$, that takes $|j\rangle \to |j\rangle$ for $j \neq 0$. That is, U_0 acts in the same way as U_w, but applies a phase of -1 to $|0\ldots0\rangle$ rather than to $|w\rangle$. In addition, let H be the transformation that performs Hadamard transformations on all of the qubits individually.

Grover's algorithm is performed as follows. Prepare all qubits in the state $|0\rangle$ and apply the global Hadamard transformation H to create the state $|\psi\rangle = (1/\sqrt{N})\sum_{j=0}^{N-1}|j\rangle$. Apply, in succession, U_w, then H, then U_0, then H again. These four transformations make up the composite transformation $U_G = HU_0HU_w$. Now apply U_G again, and repeat for a total of $\approx (\pi/4)\sqrt{N}$ times (that is, the total number of times U_G is applied is equal to the integer closest to $(\pi/4)\sqrt{N}$). The system is now, with high probability, in the state $|w\rangle$. That is, $U_G^{\sqrt{N}}|0\ldots0\rangle \approx |w\rangle$. Since each application of U_G contains a single call to the phase oracle U_w, the winner state $|w\rangle$ has now been identified with $O(\sqrt{N})$ oracle calls, as promised.

The quantum algorithm succeeds because the transformation U_G acts as a *rotation* in the two-dimensional subspace defined by the states $|\psi\rangle$ and $|w\rangle$. The angle of the rotation effected by each application of U_G can be shown to be given by $\sin\theta = 2/\sqrt{N}$. Note that $|\psi\rangle$ and $|w\rangle$ are approximately orthogonal, $\langle\psi|w\rangle = 1/\sqrt{N}$, and that after the initial Hadamard transformation the system begins in the state $|\psi\rangle$. Each successive application of U_G moves it an angle θ closer to $|w\rangle$. Finally, after $\approx (\pi/4)\sqrt{N}$ interations, the state has rotated the full $\approx \pi/2$ distance to $|w\rangle$.

Grover's algorithm can be shown to be optimal [82]: no black-box algorithm can find $|w\rangle$ with fewer than $O(\sqrt{N})$ iterations of the oracle. The algorithm also works for oracles where there are M winners, so that $f(x) = 1$ for M distinct inputs. In this case, the angle of rotation for each iteration of U_G is given by $\sin\theta = (2/N)\sqrt{M(N-M}$, and the algorithm takes $\approx (\pi/4)\sqrt{N/M}$ steps to identify a winner.

The adiabatic algorithm

Many classically hard problems take the form of optimization problems. In the well-known travelling salesman problem, for example, one aims to find the shortest route connecting a set of cities. Such optimization problems can be mapped onto a physical system, in which the function to be optimized is mapped onto the energy function of the system. The ground state of the physical system then represents a solution to the optimization problem. A common classical technique for solving such problems is simulated annealing: one simulates the process of gradually cooling the system in order to find its ground state [83]. Simulated annealing is bedeviled by the problem of local minima, states of the system that are close to the optimal states in terms of energy, but very far away in terms of the particular configuration of the degrees of freedom of the state. To avoid getting stuck in such local minima, one must slow the cooling process to a glacial pace in order to insure that the true ground state is reached in the end.

Quantum computing provides a method for getting around the problem of local minima. Rather than trying to reach the ground state of the system by cooling, one uses a purely quantum-mechanical technique for finding the state [84]. One starts the system with a Hamiltonian dynamics whose ground state is simple to prepare (e.g., 'all spins sideways'). Then one gradually deforms the Hamiltonian from the simple dynamics to the more complex dynamics whose ground state encodes the answer to the problem in question. If the deformation is sufficiently gradual, then the adiabatic theorem of quantum mechanics guarantees that the system remains in its ground state throughout the deformation process. When the adiabatic deformation is complete, then the state of the system can be measured to reveal the answer.

Adiabatic quantum computation (also called 'quantum annealing') represents a purely quantum way to find the answer to hard problems. How powerful is adiabatic quantum computation? The answer is, 'nobody knows for sure.' The key question is, what is 'sufficently gradual' deformation? That is, how slowly does the deformation have to be to guarantee that the transformation is adiabatic? The answer to this question lies deep in the heart of quantum matter. As one performs the transformation from simple to complex dynamics, the adiabatic quantum computer goes through a quantum phase transition. The maximum speed at which the computation can be performed is governed by the size of the minimum energy gap of this quantum phase transition. The smaller the gap, the slower the computation. The scaling of gaps during phase transitions ('Gapology') is one of the

key disciplines in the study of quantum matter [85]. While the scaling of the gap has been calculated for many familiar quantum systems such as Ising spin glasses, calculating the gap for adiabatic quantum computers that are solving hard optimization problems seems to be just about as hard as solving the problem itself.

While few quantum computer scientists believe that adiabatic quantum computation can solve the travelling salesman problem, there is good reason to believe that adiabatic quantum computation can outperform simulated annealing on a wide variety of hard optimization problems. In addition, it is known that adiabatic quantum computation is neither more nor less powerful than quantum computation itself: a quantum computer can simulate a physical system undergoing adiabatic time evolution using the quantum simulation techniques described above; in addition, it is possible to construct devices that perform conventional quantum computation in an adiabatic fashion [86].

Quantum walks

A final, 'physics-based,' type of algorithm is the quantum walk [87-90]. Quantum walks are coherent versions of classical random walks. A classical random walk is a stochastic Markov process, the random walker steps between different states, labelled by j, with a probability w_{ij} for making the transition from state j to state i. Here w_{ij} is a stochastic matrix, $w_{ij} \geq 0$ and $\sum_j w_{ij} = 1$. In a quantum walk, the stochastic, classical process is replaced by a coherent, quantum process: the states $|j\rangle$ are quantum states, and the transition matrix U_{ij} is unitary.

By exploiting quantum coherence, quantum walks can be shown typically to give a square root speed up over classical random walks. For example, in propagation along a line, a classical random walk is purely diffusive, with the expectation value of displacement along the line going as the square root of the number of steps in the walk. By contrast, a quantum walk can be set up as a coherent, propagating wave, so that the expectation value of the displacement is proportional to the number of steps [88]. A particularly elegant example of a square root speed up in a quantum walk is the evaluation of a NAND tree [90]. A NAND tree is a binary tree containing a NAND gate at each vertex. Given inputs on the leaves of the tree, the problem is to evaluate the outcome at the root of the tree: is it zero or one? NAND trees are ubiquitous in, e.g., game theory: the question of who wins at chess, checkers, or Go, is determined by evaluating a suitable NAND tree. Classically, a NAND tree can be evaluated with a minimum of steps. A quantum walk, by contrast, can evaluate a NAND tree using only $2^{n/2}$ steps.

For some specially designed problems, such as propagation along a random tree, quantum walks can give exponential speedups over classical walks [89]. The question of what problems can be evaluated more rapidly using quantum walks than classical walks remains open.

The future of quantum algorithms

The quantum algorithms described above are potentially powerful, and, if large-scale quantum computers can be constructed, could be used to solve a number of important problems for which no efficient classical algorithms exist. Many questions concerning quantum algorithms remain open. While the majority of quantum computer scientists would agree that quantum algorithms are unlikely to provide solutions to NP-complete problems, it is not known whether or not quantum algorithms could provide solutions to such problems as graph isomorphism or shortest vector on a lattice. Such questions are an active field of research in quantum computer science.

(V) Noise and Errors

The picture of quantum computation given in the previous section is an idealized picture that does not take into account the problems that arise when quantum computers are built in practice. Quantum computers can be built using nuclear magnetic resonance, ion traps, trapped atoms in cavities, linear optics with feedback of nonlinear measurements, superconducting systems, quantum dots, electrons on the surface of liquid helium, and a variety of other standard and exotic techniques. Any system that can be controlled in a coherent fashion is a candidate for quantum computation. Whether a coherently controllable system can actually be made to computer depends primarily on whether it is possible to deal effectively with the noise intrinsic to that system. Noise induces errors in computation. Every type of quantum information processor is subject to its own particular form of noise.

A detailed discussion of the various technologies for building quantum computers lies beyond the scope of this article. While the types of noise differ from quantum technology to quantum technology, however, the methods for dealing with that noise are common between technologies. This section presents a general formalism for characterizing noise and errors, and discusses the use of quantum error-correcting codes and other techniques for coping with those errors.

Open-system operations

The time evolution of a closed quantum-mechanical system is given by unitary transformation: $\rho \rightarrow U\rho U^\dagger$, where U is unitary, $U^\dagger = U^{-1}$. For discussing quantum communications, it is necessary to look at the time evolution of open quantum systems that can exchange quantum information with their environment. The discussion of open quantum systems is straightforward: simply adjoin the system's environment, and consider the coupled system and environment as a closed quantum system. If the joint density matrix for system and environment is

$$\rho_{SE}(0) \rightarrow \rho_{SE}(t) = U_{SE}\rho_{SE}(0)U_{SE}^\dagger. \tag{48}$$

The state of the system on its own is obtained by taking the partial trace over the environment, as described above: $\rho_S(t) = \mathrm{tr}_E \rho_{SE}(t)$.

A particularly useful case of system and environmental interaction is one in which the system and environment are initially uncorrelated, so that $\rho_{SE}(0) = \rho_S(0) \otimes \rho_E(0)$. In this case, the time evolution of the system on its own can always be written as $\rho_S(t) =$

$\sum_k A_k \rho_S(0) A_k^\dagger$. Here the A_k are operators that satisfy the equation $\sum_k A_k^\dagger A_k = Id$: the A_k are called Kraus operators, or effects. Such a time evolution for the system on its own is called a completely positive map. A simple example of such a completely positive map for a qubit is $A_0 = Id/\sqrt{2}$, $A_1 = \sigma_x/\sqrt{2}$. $\{A_0, A_1\}$ can easily be seen to obey $A_0^\dagger A_0 + A_1^\dagger A_1 = Id$. This completely positive map for the qubit corresponds to a time evolution in which the qubit has a 50% chance of being flipped about the x-axis (the effect A_1), and a 50% chance of remaining unchanged (the effect A_0).

The infinitesimal version of any completely positive map can be obtained by taking $\rho_{SE}(0) = \rho_S(0) \otimes \rho_E(0)$, and by expanding equation (49) to second order in t. The result is the Lindblad master equation:

$$\frac{\partial \rho_S}{\partial t} = -i[\tilde{H}_S, \rho_S] - \sum_k (L_k^\dagger L_k \rho_S - 2 L_k \rho_S \text{Ł}_k + \rho_S L_k^\dagger L_k). \tag{49}$$

Here $\tilde{H}_S$ is the effective system Hamiltonian: it is equal to the Hamiltonian H_S for the system on its own, plus a perturbation induced by the interaction with the environment (the so-called 'Lamb shift'). The L_k correspond to open system effects such as noise and errors.

Quantum error-correcting codes

One of the primary effects of the environment on quantum information is to cause errors. Such errors can be corrected using quantum error-correcting codes. Quantum error-correcting codes are quantum analogs of classical error-correcting codes such as Hamming codes or Reed-Solomon codes [91]. Like classical error-correcting codes, quantum error-correcting codes involve first encoding quantum information in a redundant fashion; the redundant quantum information is then subjected to noise and errors; then the code is decoded, at which point the information needed to correct the errors lie in the code's syndrome.

More bad things can happen to quantum information than to classical information. The only error that can occur to a classical bit is a bit-flip. By contrast, a quantum bit can either be flipped about the x-axis (the effect σ_x), flipped about the y-axis (the effect σ_y), flipped about the z-axis (the effect σ_z), or some combination of these effects. Indeed, an error on a quantum bit could take the form of a rotation by an unknown angle θ about an unknown axis. Since specifying that angle and axis precisely could take an infinite number of bits of information, it might at first seem impossible to detect and correct such an error.

In 1996, however, Peter Shor [92] and Andrew Steane [93] independently realized that if an error correcting code could detect and correct bit-flip errors (σ_x) and phase-flip errors (σ_z), then such a code would in fact correct any single-qubit error. The reasoning is as follows. First, since $\sigma_y = i\sigma_x\sigma_z$, a code that detects and corrects first σ_x errors, then σ_z errors will also correct σ_y errors. Second, since any single-qubit rotation can be written as a combination of σ_x, σ_y and σ_z rotations, the code will correct arbitrary single qubit errors. The generalization of such quantum error-correcting codes to multiple qubit errors are called Calderbank-Shor-Steane (CSS) codes [94]. A powerful technique for identifying and characterizing quantum codes is Gottesman's stabilizer formalism [95].

Concatenation is a useful method for constructing codes, both classical and quantum. Concatenation combines two codes, with the second code acting on bits that have been encoded using the first code. Quantum error-correcting codes can be combined with quantum computation to perform fault-tolerant quantum computation. Fault-tolerant quantum computation allows quantum computation to be performed accurately even in the presence of noise and errors, as long as those errors occur at a rate below some threshold [96-98]. For restricted error models [99], this rate can be as high as $1\% - 3\%$. For realistic error models, however, the rate is closer to $10^{-3} - 10^{-4}$.

Re-focusing

Quantum error-correcting codes are not the only technique available for dealing with noise. If, as is frequently the case, environmentally induced noise possesses some identifiable structure in terms of correlations in space and time, or obeys some set of symmetries, then powerful techniques come into play for coping with noise.

First of all, suppose that noise is correlated in time. The simplest such correlation is a static imperfection: the Hamiltonian of the system is supposed to be H, but the actual Hamiltonian is $H + \Delta H$, where ΔH is some unknown perturbation. For example, an electron spin could have the Hamiltonian $H = -(\hbar/2)(\omega + \Delta\omega)\sigma_z$, where $\Delta\omega$ is an unknown frequency shift. If not attended to, such a frequency shift will introduce unknown phases in a quantum computation, which will in turn cause errors.

Such an unknown perturbation can be dealt with quite effectively simply by flipping the electron back and forth. Let the electron evolve for time T; flip it about the x-axis; let it evolve for time T; finally, flip it back about the x-axis. The total time-evolution operator for the system is then

$$\sigma_x e^{i(\omega+\Delta\omega)T\sigma_z}\sigma_x e^{i(\omega+\Delta\omega)T\sigma_z} = Id. \tag{50}$$

That is, this simple refocusing technique cancels out the effect of the unknown frequency shift, along with the time evolution of the unperturbed Hamiltonian.

Even if the environmental perturbation varies in time, refocusing can be used significantly to reduce the effects of such noise. For time-varying noise, refocusing effectively acts as a filter, suppressing the effects of noise with a correlation time longer than the refocusing timescale T. More elaborate refocusing techniques can be used to cope with the effect of couplings between qubits. Refocusing requires no additional qubits or syndromes, and so is a simpler (and typically much more effective) technique for dealing with errors than quantum error-correcting codes. For existing experimental systems, refocusing typically makes up the 'first line of defence' against environmental noise. Once refocusing has dealt with time-correlated noise, quantum error correction can then be used to deal with any residual noise and errors.

Decoherence-free subspaces and noiseless subsystems

If the noise has correlations in space, then quantum information can often be encoded in such a way as to be resistant to the noise even in the absence of active error correction. A common version of such spatial correlation occurs when each qubit is subjected to the *same* error. For example, suppose that two qubits are subjected to noise of the form of a fluctuating Hamiltonian $H(t) = (\hbar/2)\gamma(t)(\sigma_z^1 + \sigma_z^2)$. This Hamiltonian introduces a time-varying phase $\gamma(t)$ between the states $|\uparrow\rangle_i, |\downarrow\rangle_i$. The key point to note here is that this phase is the *same* for both qubits. A simple way to compensate for such a phase is to encode the logical state $|0\rangle$ as the two-qubit state $|\uparrow\rangle_1|\downarrow\rangle_2$, and the logical state $|1\rangle$ as the two-qubit state $|\downarrow\rangle_1|\uparrow\rangle_2$. It is simple to verify that the two-qubit encoded states are now invariant under the action of the noise: any phase acquired by the first qubit is cancelled out by the equal and opposite phase acquired by the second qubit. The subspace spanned by the two-qubit states $|0\rangle, |1\rangle$ is called a decoherence-free subspace: it is invariant under the action of the noise.

Decoherence-free subspaces were first discovered by Zanardi [100] and later popularized by Lidar [101]. Such subspaces can be found essentially whenever the generators of the noise possess some symmetry. The general form that decoherence-free subspaces take arises from the following observation concerning the relationship between noise and symmetry.

Let $\{E_k\}$ be the effects that generate the noise, so that the noise takes $\rho \rightarrow \sum_k E_k \rho E_k^\dagger$, and let $\mathcal{E}$ be the algebra generated by the $\{E_k\}$. Let G be a symmetry of this algebra, so

that $[g, E] = 0$ for all $g \in G, E \in \mathcal{E}$. The Hilbert space for the system then decomposes into irreducible representation of $\mathcal{E}$ and G in the following well-known way:

$$\mathcal{H} = \sum_j \mathcal{H}_E^j \otimes \mathcal{H}_G^j, \tag{51}$$

where $\mathcal{H}_E^j$ are the irreducible representations of $\mathcal{E}$, and $\mathcal{H}_G^j$ are the irreducible representations of G.

The decomposition (51) immediately suggests a simple way of encoding quantum information in a way that is immune to the effects of the noise. Look at the effect of the noise on states of the form $|\phi\rangle_j \otimes |\psi\rangle_j$ where $|\phi\rangle_j \in \mathcal{H}_E^j$, and $|\psi\rangle_j \in \mathcal{H}_G^j$ for some j. The effect E_k acts on this state as $(E_k^j|\phi\rangle_j) \otimes |\psi\rangle_j$, where E_k^j is the effect corresponding to E_k within the representation $\mathcal{H}_E^j$. In other words, if we encode quantum information in the state $|\psi\rangle_j$, then the noise has *no effect* on $|\psi\rangle_j$. A decoherence-free subspace corresponds to an $\mathcal{H}_G^j$ where the corresponding representation of $\mathcal{E}$, $\mathcal{H}_E^j$, is one-dimensional. The case where $\mathcal{H}_E^j$, is higher dimensional is called a noiseless subsystem [102].

Decoherence-free subspaces and noiseless subsystems represent highly effective methods for dealing with the presence of noise. Like refocusing, these methods exploit symmetry to encode quantum information in a form that is immune to noise that possesses that symmetry. Where refocusing exploits temporal symmetry, decoherence-free subspaces and noiseless subsystems exploit spatial symmetry. All such symmetry-based techniques have the advantage that no error-correcting process is required. Like refocusing, therefore, decoherence-free subspaces and noiseless subsystems form the first line of defense against noise and errors.

The tensor product decomposition of irreducible representations in equation (52) lies behind all known error-correcting codes [103]. A general quantum-error correcting code begins with a state $|00\ldots0\rangle_A|\psi\rangle$, where $|00\ldots0\rangle_A$ is the initial state of the ancilla. An encoding transformation U_{en} is then applied; an error E_k occurs; finally a decoding transformation U_{de} is applied to obtain the state

$$|\mathrm{e_k}\rangle_A|\psi\rangle = U_{de}E_kU_{en}|00\ldots0\rangle_A|\psi\rangle. \tag{52}$$

Here, $|e_k\rangle_A$ is the state of the ancilla that tells us that the error corresponding to the effect E_k has occurred. Equation (52) shows that an error-correcting code is just a noiseless subsystem for the 'dressed errors' $\{U_{de}E_kU_{en}\}$. At bottom, all quantum error-correcting codes are based on symmetry.

Topological quantum computing

A particularly interesting form of quantum error correction arises when the underlying symmetry is a topological one. Kitaev [104] has shown how quantum computation can be embedded in a topological context. Two-dimensional systems with the proper symmetries exhibit topological excitations called anyons. The name, 'anyon,' comes from the properties of these excitations under exchange. Bosons, when exchanged, obtain a phase of 1; fermions, when exchanged, obtain a phase of -1. Anyons, by contrast, when exchanged, can obtain an arbitrary phase $e^{i\phi}$. For example, the anyons that underlie the fractional quantum Hall effect obtain a phase $e^{2\pi i/3}$ when exchanged. Fractional quantum Hall anyons can be used for quantum computation in a way that makes two-qubit quantum logic gates intrinsically resistant to noise [105].

The most interesting topological effects in quantum computation arise when one employs non-abelian anyons [104]. Non-abelian anyons are topological excitations that possess internal degrees of freedom. When two non-abelian anyons are exchanged, those internal degrees of freedom are subjected not merely to an additional phase, but to a general unitary transformation U. Kitaev has shown how in systems with the proper symmetries, quantum computation can be effected simply by exchanging anyons. The actual computation takes place by dragging anyons around eachother in the two-dimensional space. The resulting transformation can be visualized as a braid in two dimensional space plus the additional dimension of time.

Topological quantum computation is intrinisically fault tolerant. The topological excitations that carry quantum information are impervious to locally occurring noise: only a global transformation that changes the topology of the full system can create a error. Because of their potential for fault tolerance, two-dimensional systems that possess the exotic symmetries required for topelogical quantum computation are being actively sought out.

VI. Quantum Communication

Quantum mechanics provides the fundamental limits to information processing. Above, quantum limits to computation were investigated. Quantum mechanics also provides the fundamental limits to communication. This section discusses those limits. The session closes with a section on quantum cryptography, a set of techniques by which quantum mechanics guarantees the privacy and security of cryptographic protocols.

Multiple uses of channels

Each quantum communication channel is characterized by its own open-system dynamics. Quantum communication channels can possess memory, or be memoryless, depending on their interaction with their environment. Quantum channels with memory are a difficult topic, which will be discussed briefly below. Most of the discussion that follows concerns the memoryless quantum channel. A single use of such a channel corresponds to a completely positive map, $\rho \rightarrow \sum_k A_k \rho A_k^\dagger$, and n uses of the channel corresponds to a transformation

$$\rho_{1\ldots n} \rightarrow \sum_{k_1 \ldots k_n} A_{k_n} \otimes \ldots \otimes A_{k_1} \rho_{1\ldots n} A_{k_1}^\dagger \otimes \ldots \otimes A_{k_n}^\dagger \equiv \sum_K A_K \rho_{1\ldots n} A_K^\dagger, \qquad (53)$$

where we have used the capital letter K to indicate the n uses of the channel $k_1 \ldots k_n$. In general, the input state $\rho_{1\ldots n}$ may be entangled from use to use of the channel. Many outstanding questions in quantum communication theory remain unsolved, including, for example, the question of whether entangling inputs of the channel helps for communicating classical information.

Sending quantum information

Let's begin with using quantum channels to send quantum information. That is, we wish to send some quantum state $|\psi\rangle$ from the input of the channel to the output. To do this, we encode the state as some state of n inputs to the channel, send the encoded state down the channel, and then apply a decoding procedure at the output to the channel. It is immediately seen that such a procedure is equivalent to employing a quantum error-correcting code.

The general formula for the capacity of such quantum channels is known [44-45]. Take some input or 'signal' state $\rho_{1\ldots n}$ for the channel. First, construct a purification of this state. A purification of a density matrix ρ for the signal is a pure state $|\psi\rangle_{AS}$ for the signal together with an ancilla, such that the state $\rho_S = \mathrm{tr}_A |\psi\rangle_{AS}\langle\psi|$ is equal to the original

density matrix ρ. There are many different ways to purify a state: a simple, explicit way is to write $\rho = \sum_j p_j |j\rangle\langle j|$ in diagonal form, where $\{|j\rangle\}$ is the eigenbasis for ρ. The state $|\psi\rangle_{AS} = \sum_j \sqrt{p_j}|j\rangle_A|j\rangle_S$, where $\{|j\rangle_A\}$ is an orthonormal set of states for the ancilla, then yields a purification of ρ.

To obtain the capacity of the channel for sending quantum information, proceed as follows. Construct a purification for the signal $\rho_{1\ldots n}$: $|\psi_n\rangle = \sum_J \sqrt{p_J}|J\rangle_A^n|J\rangle_S^n$, where we have used an index J instead of j to indicate that these states are summed over n uses of the channel. Now send the signal state down the channel, yielding the state

$$\rho_{AS} = \sum_{JJ'} \sqrt{p_J p_{J'}}|J\rangle_A^n\langle J'| \otimes \sum_K A_K|J\rangle_S^n\langle J'|A_K^\dagger, \tag{54}$$

where as above $K = k_1 \ldots k_n$ indicates k uses of the channel. ρ_{AS} is the state of output signal state together with the ancilla. Similarly, $\rho_S = \mathrm{tr}_A\rho_{AS}$ is the state of the output signal state on its own.

Let $I(AS) = -\mathrm{tr}\rho_{AS}\log_2\rho_{AS}$ be the entropy of ρ_{AS}, measured in bits. Similarly, let $I_(S) = -\mathrm{tr}\rho_S\log_2\rho_S$ be the entropy of the output state ρ_S, taken on its own. Define $I(S/A) \equiv I(S) - I(AS)$ if this quantity is positive, and $I(S/A) \equiv 0$ otherwise. The quantity $I(S/A)$ is a measure of the capacity of the channel to send quantum information if the signals being sent down the channel are described by the density matrix $\rho_{1\ldots n}$. It can be shown using either CSS codes [106] or random codes [45,107] that encodings exist that allow quantum information to be sent down the channel and properly decoded at the output at a rate of $I(S/A)/n$ qubits per use.

$I(S/A)$ is a function only of the properties of the channel and the input signal state $\rho_{1\ldots n}$. The bigger $I(S/A)$ is, the less coherence the channel has managed to destroy. For example, if the channel is just a unitary transformation of the input, which destroys no quantum information, then $I(AS) = 0$ and $I(S/A) = I(S)$: the state of the signal and ancilla after the signal has passed through the channel is pure, and all quantum information passes down the channel unscathed. By contrast, a completely decohering channel takes an the input $\sum_j \sqrt{p_j}|j\rangle_A|j\rangle_S$ to the output $\sum_j p_j|j\rangle_A\langle j| \otimes |j\rangle_S\langle j|$. In this case, $I(AS) = I(S)$ and $I(S/A) = 0$: the channel has completely destroyed all quantum information sent down the channel.

In order to find the absolute capacity of the channel to transmit quantum information, we must maximize the quantity $I(S/A)/n$ over all n-state inputs $\rho_{1\ldots n}$ to the channel and take the limit as $n \to \infty$. More precisely, define

$$I_C = \lim_{n\to\infty}\min\ \sup I(S/A)/n, \tag{55}$$

where the supremum (sup) is taken over all n-state inputs $\rho_{1\ldots n}$. I_C is called the coherent information [44-45]: it is the capacity of the channel to transmit quantum information reliably.

Because the coherent information is defined only in the limit that the length of the input state goes to infinity, it has been calculated exactly in only a few cases. One might hope, in analogue to Shannon's theory of classical communication, that for memoryless channels one need only optimize over single inputs. That hope is mistaken, however: entangling the input states typically increases the quantum channel capacity even for memoryless channels [108].

Capacity of quantum channels to transmit classical information

One of the most important questions in quantum communications is the capacity of quantum channels to transmit classical information. All of our classical communication channels – voice, free space electromagnetic, fiber optic, etc. – are at bottom quantum mechanical, and their capacities are set using the laws of quantum mechanics. If quantum information theory can discover those limits, and devise ways of attaining them, it will have done humanity a considerable service.

The general picture of classical communication using quantum channels is as follows. The conventional discussion of communication channels, both quantum and classical, designates the sender of information as Alice, and the receiver of information as Bob. Alice selects an ensemble of input states ρ_J over n uses of the channel, and send the J'th input ρ_J with probability p_J. The channel takes the n-state input ρ_J to the output $\tilde{\rho}_J = \sum_K A_K \rho_J A_K^\dagger$. Bob then performs a generalized measurement $\{B_\ell\}$ with outcomes $\{\ell\}$ to try to reveal which state Alice sent. A generalized measurement is simply a specific form an open-system transformation. The $\{B_\ell\}$ are effects for a completely positive map: $\sum_\ell B_\ell^\dagger B_\ell = Id$. After making the generalized measurement on an output state $\tilde{\rho}_J$, Bob obtains the outcome ℓ with probability $p_{\ell|J} = \mathrm{tr} B_\ell \tilde{\rho}_J B_\ell^\dagger$, and the system is left in the state $(1/p_{\ell|J}) B_\ell \tilde{\rho}_J B_\ell^\dagger$.

Once Alice has chosen a particular ensemble of signal states $\{\rho_J, p_J\}$, and Bob has chosen a particular generalized measurement, then the amount of information that can be sent along the channel is determined by the input probabilities p_J and the output probabilities $p_{\ell|J}$ and $p_\ell = \sum_J p_J p_{\ell|J}$. In particular, the rate at which information can be sent through the channel and reliably decoded at the output is given by the mutual information $I(in:out) = I(out) - I(out|in)$, where $I(out) = -\sum_\ell p_\ell \log_2 p_\ell$ is the entropy

of the output and $I(out|in) = \sum_J p_J(-\sum_\ell p_{\ell|J} \log_2 p_{\ell|J})$ is the average entropy of the output conditioned on the state of the input.

To maximize the amount of information that can be sent down the channel, Alice and Bob need to maximize over both input states and over Bob's measurement at the output. The Schumacher-Holevo-Westmoreland theorem, however, considerably simplifies the problem of maximizing the information transmission rate of the channel by obviating the need to maximize over Bob's measurements [37-39]. Define the quantity

$$\mathcal{X} = S(\sum_J p_J \tilde{\rho}_J) - \sum_J p_J S(\tilde{\rho}_J), \tag{56}$$

where $S(\rho) \equiv -\mathrm{tr}\rho \log_2 \rho$. $\mathcal{X}$ is the difference between the entropy of the average output state and the average entropy of the output states. The Schumacher-Holevo-Westmoreland theorem then states that the capacity of the quantum channel for transmitting classical information is given by the limit as $\lim_{n\to\infty} \min \sup \mathcal{X}/n$, where the supremum is taken over all possible ensembles of input states $\{\rho_J, p_J\}$ over n uses of the channel.

For Bob to attain the channel capacity given by $\mathcal{X}$, he must in general make entangling measurements over the channel outputs, even when the channel is memoryless and when Alice does not entangle her inputs. (An entangling measurement is one the leaves the outputs in an entangled state after the measurement is made.) It would simplify the process of finding the channel capacity still further if the optimization over input states could be performed over a single use of the channel for memoryless channels, as is the case for classical communication channels, rather than having to take the limit as the number of inputs goes to infinity. If this were the case, then the channel capacity for memoryless channels would be attained for Alice sending unentangled states down the channel. Whether or not one is allowed to optimize over a single use for memoryless channels was for many years one of the primary unsolved conjectures of quantum information theory.

Let's state this conjecture precisely. Let $\mathcal{X}_n$ be the maximum of $\mathcal{X}$ over n uses of a memoryless channel. We then have the

Channel additivity conjecture: $\mathcal{X}_n = n\mathcal{X}_1$.

Shor has shown that the channel additivity conjecture is equivalent to two other additivity conjectures, the additivity of minimum output entropy and the additivity of entanglement of formation [109]. Entanglement of formation was discussed in the section on entanglement above. The minimum output entropy for n uses of a memoryless channel

is simply the minimum over input states ρ_n, for n uses of the channel, of $S(\tilde{\rho}_n)$, where $\tilde{\rho}_n$ is the output state arising from the input ρ_n. We then have the

Minimum output entropy additivity conjecture: The minimum over ρ_n of $S(\tilde{\rho}_n)$ is equal to n times the minimum over ρ_1 of $S(\tilde{\rho}_1)$.

Shor's result shows that the channel additivity conjecture and the minimum output entropy additivity conjecture are equivalent: each one implies the other. If these additivity conjectures could have been proved to be true, that would have resolved some of the primary outstanding problems in quantum channel capacity theory. Remarkably, however, Hastings recently showed that the minimum output entropy conjecture is *false*, by exhibiting a channel whose minimum output entropy for multiple uses is achieved for entangled inputs. As a result, the question of just how much classical information can be sent down a quantum channel, and just which quantum channels are additive and which are not, remains wide open.

Bosonic channels

The most commonly used quantum communication channel is the so-called bosonic channel with Gaussian noise and loss [40]. Bosonic channels are ones that use bosons such as photons or phonons to communicate. Gaussian noise and loss is the most common type of noise and loss for such channels, it includes the effect of thermal noise, noise from linear amplification, and leakage of photons or phonons out of the channel. It has been shown that the capacity for bosonic channels with loss alone is attained by sending coherent states down the channel [42]. Coherent states are the sort of states produced by lasers and are the states that are currently used in most bosonic channels.

It has been conjectured that coherent states also maximize the capacity of quantum communication channels with Gaussian noise as well as loss [43]. This conjecture, if true, would establish the quantum-mechanical equivalent of Shannon's theorem for the capacity of classical channels with Gaussian noise and loss. The resolution of this conjecture can be shown to be equivalent to the following, simpler conjecture:

Gaussian minimum output entropy conjecture: Coherent states minimize the output entropy of bosonic channels with Gaussian noise and no loss.

The Gaussian minimum output entropy is intuitively appealing: an equivalent statement is that the *vacuum* input state minimizes the output entropy for a channel with

Gaussian noise. In other words, to minimize the output entropy of the channel, send *nothing*. Despite its intuitive appeal, the Gaussian minimum output entropy conjecture has steadfastly resisted proof for decades.

Entanglement assisted capacity

Just as quantum bits possess greater mathematical structure than classical bits, so quantum channels possess greater variety than their classical counterparts. A classical channel has but a single capacity. A quantum channel has one capacity for transmitting quantum information (the coherent information), and another capacity for transmitting classical information (the Holevo quantity $\mathcal{X}$). We can also ask about the capacity of a quantum channel in the presence of prior entanglement.

The entanglement assisted capacity of a channel arises in the following situation. Suppose that Alice and Bob have used their quantum channel to build up a supply of entangled qubits, where Alice possesses half of the entangled pairs of qubits, and Bob possesses the other half of the pairs. Now Alice sends Bob some qubits over the channel. How much classical information can these qubits convey?

At first one might think that the existence of shared prior entanglement should have no effect on the amount of information that Alice can send to Bob. After all, entanglement is a form of correlation, and the existence of prior correlation between Alice and Bob in a classical setting has no effect on the amount of information sent. In the quantum setting, however, the situation is different.

Consider, for example, the case where Alice and Bob have a perfect, noiseless channel. When Alice and Bob share no prior entanglement, then a single qubit sent down the channel conveys exactly one bit of classical information. When Alice and Bob share prior entanglement, however, a single quantum bit can convey more than one bit of classical information. Suppose that Alice and Bob share an entangled pair in the singlet state $(1/\sqrt{2})(|0\rangle_A|1\rangle_B - |1\rangle_A|0\rangle_B)$. Alice then performs one of four actions on her qubit: either she does nothing (performs the identity Id on the qubit), or she flips the qubit around the x-axis (performs σ_x), or she flips the qubit around the y-axis (performs σ_y), she flips the qubit around the z-axis (performs σ_z).

Now Alice sends her qubit to Bob. Bob now possesses one of the four orthogonal states, $(1/\sqrt{2})(|0\rangle_A|1\rangle_B - |1\rangle_A|0\rangle_B)$, $(1/\sqrt{2})(|1\rangle_A|1\rangle_B - |0\rangle_A|0\rangle_B)$, $(i/\sqrt{2})(|1\rangle_A|1\rangle_B + |0\rangle_A|0\rangle_B)$, $(1/\sqrt{2})(|0\rangle_A|1\rangle_B + |1\rangle_A|0\rangle_B)$. By measuring which of these states he possesses, Bob can determine which of the four actions Alice performed. That is, when Alice and Bob share

prior entanglement, Alice can send *two* classical bits for each quantum bit she sends. This phenomenon is known as superdense coding [111].

In general, the quantum channel connecting Alice to Bob is noisy. We can then ask, given the form of the quantum channel, how much does the existence of prior entanglement help Alice in sending classical information to Bob? The answer to this question is given by the following theorem, due to Shor *et al.* The entanglement assisted capacity of a quantum channel is equal to the maximum of the quantum mutual information between the input and output of the channel [112]. The quantum mutual information is defined as follows. Prepare a purification $|\psi\rangle_{AS}$ of an input state ρ and send the signal state S down the channel, resulting the state ρ_{AS} as in equation (55) above. Defining $\rho_S = \mathrm{tr}_A \rho_{AS}$, $\rho_A = \mathrm{tr}_S \rho_{AS}$, as before, the quantum mutual information is defined to be $I_Q(A : S) = S(\rho_A) + S(\rho_S) - S(\rho_{AS})$. The entanglement assisted capacity of the channel is obtained by maximizing the quantum mutual information $I_Q(A : S)$ over input states ρ.

The entanglement assisted capacity of a quantum channel is greater than or equal to the channel's Holevo quantity, which is in turn greater than or equal to the channel's coherent information. Unlike the coherent information, which is known not to be additive over many uses of the channel, or the Holevo quantity, which is suspected to be additive but which has not been proved to be so, the entanglement assisted capacity is known to be additive and so can readily be calculated for memoryless channels.

Teleportation

As mentioned in the introduction, one of the most strange and useful effects in quantum computation is teleportation [46]. The traditional, science fiction picture of teleportation works as follows.

An object such as an atom or a human being is placed in a device called a teleporter. The teleporter makes complete measurements of the physical state of the object, destroying it in the process. The detailed information about that physical state is sent to a distant location, where a second teleporter uses that information to reconstruct an exact copy of the original object.

At first, quantum mechanics would seem to make teleportation impossible. Quantum measurements tend to disturb the object measured. Many identical copies of the object are required to obtain even a rough picture of the underlying quantum state of the object. In the presence of shared, prior entanglement, however, teleportation is in fact possible in principle, and simple instances of teleportation have been demonstrated experimentally.

A hint to the possibility of teleportation comes from the phenomenon of superdense coding described in the previous section. If one qubit can be used to convey two classical bits using prior entanglement, then maybe two classical bits might be used to convey one qubit. This hope turns out to be true. Suppose that Alice and Bob each possess one qubit out of an entangled pair of qubits (that is, they mutually possess one 'e-bit'). Alice desires to teleport the state $|\psi\rangle$ of another qubit. The teleportation protocol goes as follows.

First, Alice makes a Bell-state measurement on the qubit to be teleported together with her half of the entangled pair. A Bell-state measurement on two qubits is one that determines whether the two qubits are in one of the four states $|\phi_{00}\rangle = (1/\sqrt{2})(|01\rangle - |10\rangle)$, $|\phi_{01}\rangle = (1/\sqrt{2})(|00\rangle - |11\rangle)$, $|\phi_{10}\rangle = (1/\sqrt{2})(|00\rangle + |11\rangle)$, or $|\phi_{11}\rangle = (1/\sqrt{2})(|01\rangle + |10\rangle)$. Alice obtains two classical bits of information as a result of her measurement, depending on which $|\phi_{ij}\rangle$ the measurement revealed. She sends these two bits to Bob. Bob now performs a unitary transformation on his half of the entangled qubit pair. If he receives 00, then he does nothing. If he receives 01, then he applies σ_x to flip his bit about the x-axis. If he receives 10, then he applies σ_y to flip his bit about the y-axis. If he receives 11, then he applies σ_z to flip his bit about the z-axis. The result? After Bob has performed his transformation conditioned on the two classical bits he received from Alice, his qubit is now in the state $|\psi\rangle$, up to an overall phase. Alice's state has been teleported to Bob.

It might seem at first somewhat mysterious how this sequence of operations can teleport Alice's state to Bob. The mechanism of teleportation can be elucidated as follows. Write $|\psi\rangle = \alpha|0\rangle_i + \beta|1\rangle_i$. Alice and Bob's entangled pair is originally in the state $|\phi_{00}\rangle_{AB} = (1/\sqrt{2})(|0\rangle_A|1\rangle_B - |1\rangle_A|0\rangle_B)$. The full initial state of qubit to be teleported together with the entangled pair can then be written as

$$\begin{aligned}
&|\psi\rangle|phi_{00}\rangle_{AB} \\
&= (\alpha|0\rangle_i + \beta|1\rangle_i)\frac{1}{\sqrt{2}}(|0\rangle_A|1\rangle_B - |1\rangle_A|0\rangle_B) \\
&= \frac{1}{2\sqrt{2}}(|0\rangle_i|1\rangle_A - |1\rangle_i|0\rangle_A) \otimes (\alpha|0\rangle_B + \beta|1\rangle_B) \\
&+ \frac{1}{2\sqrt{2}}(|0\rangle_i|0\rangle_A - |1\rangle_i|1\rangle_A) \otimes (\alpha|1\rangle_B + \beta|0\rangle_B) \\
&+ \frac{1}{2\sqrt{2}}(|0\rangle_i|0\rangle_A + |1\rangle_i|1\rangle_A) \otimes (\alpha|1\rangle_B - \beta|0\rangle_B) \\
&+ \frac{1}{2\sqrt{2}}(|0\rangle_i|1\rangle_A + |1\rangle_i|0\rangle_A) \otimes (\alpha|0\rangle_B - \beta|1\rangle_B). \\
&= \frac{1}{2}(|\phi_{00}\rangle_{iA} \otimes |\psi\rangle_B + |\phi_{01}\rangle_{iA} \otimes \sigma_x|\psi\rangle_B + |\phi_{10}\rangle_{iA} \otimes i\sigma_y|\psi\rangle_B + |\phi_{11}\rangle_{iA} \otimes \sigma_z|\psi\rangle_B).
\end{aligned} \tag{57}$$

When the initial state is written in this form, one sees immediately how the protocol works: the measurement that Alice makes contains exactly the right information that Bob needs to reproduce the state $|\psi\rangle$ by performing the appropriate transformation on his qubit.

Teleportation is a highly useful protocol that lies at the center of quantum communication and fault tolerant quantum computation. There are several interesting features to note. The two bits of information that Alice obtains are completely random: $00, 01, 10, 11$ all occur with equal probability. These bits contain no information about $|\psi\rangle$ taken on their own: it is only when combined with Bob's qubit that those bits suffice to recreate $|\psi\rangle$. During the teleportation process, it is difficult to say just where the state $|\psi\rangle$ 'exists.' After Alice has made her measurement, the state $|\psi\rangle$ is in some sense 'spread out' between her two classical bits and Bob's qubit. The proliferation of quotation marks in this paragraph is a symptom of quantum weirdness: classical ways of describing things are inadequate to capture the behavior of quantum things. The only way to see what happens to a quantum system during a process like teleportation is to apply the mathematical rules of quantum mechanics.

Quantum cryptography

A common problem in communication is security. Suppose that Alice and Bob wish to communicate with each other with the secure knowledge that no eavesdropper (Eve) is listening in. The study of secure communication is commonly called cryptography, since to attain security Alice must encrypt her messages and Bob must decrypt them. The no-cloning theorem together with the fact that if one measures a quantum system, one typically disturbs it, implies that quantum mechanics can play a unique role in constructing cryptosystems. There are a wide variety of quantum cryptographic protocols [49-51]. The most common of these fall under the heading of quantum key distribution (QKD).

The most secure form of classical cryptographic protocols is the one-time pad. Here, Alice and Bob each possess a random string of bits. This string is called the key. If no one else possesses the key, then Alice and Bob can send messages securely as follows. Suppose that Alice's message has been encoded in bits in some conventional way (e.g., mapping characters to ASCII bit strings). Alice encrypts the message by adding the bits of the key to the bits of her message one by one, modulo 2 (i.e., without carrying). Because the key was random, the resulting string possesses no statistical order left over from the original message. Alice then sends the encrypted message to Bob, who decrypts it by adding the key to the bits of the encrypted message, modulo 2. As long as no one other than Alice

and Bob possess the shared key, this form of cryptography is completely secure. Alice and Bob must be careful not to use the key more than once. If they use it twice or more, then Eve can detect patterns in the encrypted messages.

The problem with the one-time pad is to distribute the keys to Alice and Bob and to no one else. Classically, someone who intercepts the key can copy it and pass it on without Alice and Bob detecting their copying. Quantum-mechanically, however, Alice and Bob can set up key-distribution protocols that can detect and foil any eavesdropping.

The idea of quantum cryptography was proposed, in embryonic form, by Stephen Wiesner in [49]. The first quantum cryptographic protocol was proposed by Bennett and Brassard in 1984 and is commonly called BB84 [50]. The BB84 protocol together with its variants is the one most commonly used by existing quantum cryptosystems.

In BB84, Alice sends Bob a sequence of qubits. The protocol is most commonly described in terms of qubits encoded on photon polarization. Here, we will describe the qubits in terms of spin, so that we can use the notation developed in section III. Spin 1/2 is isomorphic to photon polarization and so the quantum mechanics of the protocol remains the same.

Alice choses a sequence of qubits from the set $\{|\uparrow\rangle, |\downarrow\rangle, |\leftarrow\rangle, |\rightarrow\rangle\}$ at random, and sends that sequence to Bob. As he receives each qubit in turn, Bob picks at random either the z-axis or the x-axis and measures the received qubit along that axis. Half of the time, on average, Bob measures the qubit along the same axis along which it was prepared by Alice.

Alice and Bob now check to see if Eve is listening in. Eve can intercept the qubits Alice sends, make a measurement on them, and then send them on to Bob. Because she does not know the axis along which any individual qubit has been prepared, however, here measurement will inevitably disturb the qubits. Alice and Bob can then detect Eve's intervention by the following protocol.

Using an ordinary, insecure form of transmission, e.g., the telephone, Alice reveals to Bob the state of some of the qubits that she sent. On half of those qubits, on average, Bob measured them along the same axis along which they were sent. Bob then checks to see if he measured those qubits to be in the same state that Alice sent them. If he finds them all to be in the proper state, then he and Alice can be sure that Eve is not listening in. If Bob finds that some fraction of the qubits are not in their proper state, then he and Alice know that either the qubits have been corrupted by the environment in transit, or Eve is

listening in. The degree of corruption is related to the amount of information that Eve can have obtained: the greater the corruption, the more information Eve may have. From monitoring the degree of corruption of the received qubits, Alice and Bob can determine just how many bits of information Eve has obtained about their transmission.

Alice now reveals to Bob the axis along which she prepared the remainder of her qubits. On half of those, on average, Bob measured using the same axis. If Eve is not listening, those qubits on which Bob measured using the same axis along which Eve prepared them now constitute a string of random bits that is shared by Alice and Bob and by them only. This shared random string can then be used as a key for a one-time pad.

If Eve is listening in, then from their checking stage, Alice and Bob know just how many bits out of their shared random string are also known by Eve. Alice and Bob can now perform classical privacy amplification protocols [113] to turn their somewhat insecure string of shared bits into a shorter string of shared bits that is more secure. Once privacy amplification has been performed, Alice and Bob now share a key whose secrecy is guaranteed by the laws of quantum mechanics.

Eve could, of course, intercept *all* the bits sent, measure them, and send them on. Such a 'denial of service' attack prevents Alice and Bob from establishing a shared secret key. No cryptographic system, not even a quantum one, is immune to denial of service attacks: if Alice and Bob can exchange no information then they can exchange no secret information! If Eve lets enough information through, however, then Alice and Bob can always establish a secret key.

A variety of quantum key distribution schemes have been proposed [50-51]. Ekert suggested using entangled photons to distribute keys to Alice and Bob. In practical quantum key distribution schemes, the states sent are attenuated coherent states, consisting of mostly vacuum with a small amplitude of single photon states, and an even smaller amplitude of states with more than one photon. It is also possible to use continuous quantum variables such as the amplitudes of the electric and magnetic fields to distribute quantum keys [114-115]. To guarantee the full security of a quantum key distribution scheme requires a careful examination of all possible attacks given the actual physical implementation of the scheme.

VII. Implications and Conclusions

Quantum information theory is a rich and fundamental field. Its origins lie with the origins of quantum mechanics itself a century ago. The field has expanded dramatically since the mid 1990s, due to the discovery of practical applications of quantum information processing such as factoring and quantum cryptography, and because of the rapid development of technologies for manipulating systems in a way that preserves quantum coherence.

As an example of the rapid pace of development in the field of quantum information, while this article was in proof, a new algorithm for solving linear sets of equations was discovered [116]. Based on the quantum phase algorithm, this algorithm solves the following problem: given a sparse matrix A and a vector $\vec{b}$, find a vector $\vec{x}$ such that $A\vec{x} = \vec{b}$. That is, construct $\vec{x} = A^{-1}\vec{b}$. If A is an n by n matrix, the best classical algorithms for solving this problem run in time $O(n)$. Remarkably, the quantum matrix inversion algorithm runs in time $O(\log n)$, an exponential improvement: a problem that could take $10^{12} - 10^{15}$ operations to solve on a classical computer could be solved on a quantum computer in fewer than one hundred steps.

When they were developed in the mid twentieth century, the fields of classical computation and communication provided unifying methods and themes for all of engineering and science. So at the beginning of the twenty first century, quantum information is providing unifying concepts such as entanglement, and unifying techniques such as coherent information processing and quantum error correction, that have the potential to transform and bind together currently disparate fields in science and engineering.

Indeed, quantum information theory has perhaps even a greater potential to transform the world than classical information theory. Classical information theory finds its greatest application in the man-made systems such as electronic computers. Quantum information theory applies not only to man-made systems, but to all physical systems at their most fundamental level. For example, entanglement is a characteristic of virtually all physical systems at their most microscopic levels. Quantum coherence and the relationship between symmetries and the conservation and protection of information underlie not only quantum information, but the behavior of elementary particles, atoms, and molecules.

When or whether techniques of quantum information processing will become tools of mainstream technology is an open question. The technologies of precision measurement are already fully quantum mechanical: for example, the atomic clocks that lie at the heart

of the global positioning system (GPS) rely fundamentally on quantum coherence. Ubiquitous devices such as the laser and the transistor have their roots in quantum mechanics. Quantum coherence is relatively fragile, however: until such a time as we can construct robust, easily manufactured coherent systems, quantum information processing may have its greatest implications at the extremes of physics and technology.

Quantum information processing analyzes the universe in terms of information: at bottom, the universe is composed not just of photons, electrons, neutrinos and quarks, but of quantum bits or qubits. Many aspects of the behavior of those elemental qubits are independent of the particular physical system that registers them. By understanding how information behaves at the quantum mechanical level, we understand the fundamental behavior of the universe itself.

References

[1] M.A. Nielsen, I.L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge (2000).

[2] P. Ehrenfest, T. Ehrenfest, (1912) *The Conceptual Foundations of the Statistical Approach in Mechanics,* Cornell University Press (Ithaca, NY, 1959).

[3] M. Planck, *Ann. Phys.* **4**, 553 (1901).

[4] J.C. Maxwell, *Theory of Heat*, Appleton, London, (1871).

[5] A. Einstein *Ann. Phys.* **17**, 132 (1905).

[6] N. Bohr, *Phil. Mag.* **26**, 1-25, 476-502, 857-875 (1913).

[7] E. Schrödinger, *Ann. Phys.* **79**, 361-376, 489-527 (1926); *Ann. Phys.* **80**, 437-490 (1926); *Ann. Phys.* **81**, 109-139 (1926).

[8] W. Heisenberg, *Z. Phys.* **33**, 879-893 (1925); *Z. Phys.*a **34**, 858-888 (1925); *Z. Phys.* **35**, 557-615 (1925).

[9] A. Einstein, B. Podolsky, N. Rosen, *Phys. Rev.* **47**, 777 (1935).

[10] D. Bohm, *Phys. Rev.* **85**, 166-179, 180-193 (1952).

[11] J.S. Bell, *Physics* **1**, 195 (1964).

[12] Y. Aharonov, D. Bohm, *Phys. Rev.***115**, 485-491 (1959); *Phys. Rev.* **123**, 1511-1524 (1961).

[13] A. Aspect, P. Grangier, G. Roger, *Phys. Rev. Lett.* **49**, 91-94 (1982); A. Aspect, J. Dalibard, G. Roger, *Phys. Rev. Lett.* **49**, 1804-1807 (1982); A. Aspect, *Nature* **398**, 189 (1999).

[14] R.V.L. Hartley, *Bell System Technical Journal*, July 1928, 535-563.

[15] A.M. Turing, *Proc. Lond. Math. Soc.* **42**, 230-265 (1936).

[16] C.E. Shannon, *Symbolic Analysis of Relay and Switching Circuits,* Master's Thesis MIT (1937).

[17] C.E. Shannon, "A Mathematical theory of communication," *Bell Systems Technical Journal* **27**, 379–423, 623–656 (1948).

[18] J. von Neumann, in *Theory of Self-Reproducing Automata*, A. W. Burks, University of Illinois Press, (Urbana, IL, 1966). 1966]

[19] R. Landauer, *IBM J. Res. Develop.* **5**, 183–191 (1961).

[20] Y. Lecerf, *Compte Rendus* **257** 2597-2600 (1963).

[21] C.H. Bennett, *IBM J. Res. Develop.* **17**, 525–532 (1973); *Int. J. Theor. Phys.* **21**, 905–940 (1982).

[22] E. Fredkin, T. Toffoli, *Int. J. Theor. Phys.* **21**, 219-253 (1982).

[23] P. Benioff, *J. Stat. Phys.* **22**, 563-591 (1980); *Phys. Rev. Lett.* **48**, 1581-1585 (1982); *J. Stat. Phys.* **29**, 515-546 (1982); *Ann. N.Y. Acad. Sci.* **480**, 475-486 (1986).

[24] R.P. Feynman, *Int. J. Th. Ph.* **21**, 467 (1982).

[25] D. Deutsch, *Proc. Roy. Soc. Lond.* **A 400**, 97–117 (1985); *Proc. Roy. Soc. Lond.* **A 425**, 73–90 (1989).

[26] P. Shor, in *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*, S. Goldwasser, ed., (*IEEE* Computer Society, Los Alamitos, CA, 1994), pp. 124-134.

[27] S. Lloyd, *Science* **261**, 1569 (1993); *Science* **263**, 695 (1994).

[28] J.I. Cirac, P. Zoller, *Phys. Rev. Lett* **74**, 4091 (1995).

[29] C. Monroe, D.M. Meekhof, B.E. King, W.M. Itano, D.J. Wineland, *Phys. Rev. Lett.* **75**, 4714, (1995).

[30] Q.A. Turchette, C.J. Hood, W. Lange, H. Mabuchi, H.J. Kimble, *Phys. Rev. Lett.* **75**, 4710, (1995).

[31] L.K. Grover, in *Proceedings, 28th Annual ACM Symposium on the Theory of Computing,* (May 1996) p. 212.

[32] D. Cory, A.F. Fahmy, T.F. Havel, *Proc. Nat. Acad. Sci. USA* **94**, 1634-1639 (1997).

[33] I.L. Chuang, L.M.K. Vandersypen, X. Zhou, D.W. Leung, S. Lloyd, *Nature* **393**, 143-146 (1998).

[34] I.L. Chuang, N. Gershenfeld, M. Kubinec, *Phys. Rev. Lett.* **80**, 3408 - 3411 (1998).

[35] J.P. Gordon, *Proc. IRE* **50**, 1898-1908 (1962).

[36] D.S. Lebedev, L.B. Levitin, *Sov. Phys. Dok.* **8**, 377 (1963).

[37] A.S. Holevo, *Prob. Per. Inf.* **9**, 3 (1973); *Prob. Inf. Trans. (USSR)* **9**, 110 (1973).

[38] B. Schumacher, M. Westmoreland, *Phys. Rev. A* **55**, 2738 (1997).

[39] A.S. Holevo, *IEEE Trans. Inf. Th.* **44**, 69 (1998).

[40] C.M. Caves, P.D. Drummond, *Rev. Mod. Phys.* **66**, 481-537 (1994).

[41] H.P. Yuen, M. Ozawa, *Phys. Rev. Lett.* **70**, 363-366 (1993).

[42] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, J. H. Shapiro, H. P. Yuen, *Phys. Rev. Lett.* **92**, 027902 (2004).

[43] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, J. H. Shapiro, *Phys. Rev. A* **70**, 032315 (2004).

[44] M. Nielsen, B. Schumacher, *Phys. Rev. A* **54**, 2629-2635 (1996).

[45] S. Lloyd, *Phys. Rev. A* **55**, 1613-1622 (1997).

[46] C.H. Bennett, G.Brassard, C. Cripeau, R. Jozsa, A. Peres, W.K. Wootters, *Phys. Rev. Lett.* **70**, 1895-1899 (1993).

[47] J.W. Pan, M. Daniell, S. Gasparoni, G. Weihs, A. Zeilinger *Phys. Rev. Lett.* **86**, 4435-4438 (2001).

[48] T.C. Zhang, K.W. Goh, C.W. Chou, P. Lodahl, H.J. Kimble, *Phys. Rev. A* **67**, 033802 (2003).

[49] S. Wiesner, *SIGACT News* **15**, 78-88, (1983).

[50] C.H. Bennett, G. Brassard, *Proceedings of IEEE International Conference on Computers* **10-12**, 175-179 (1984).

[51] A. Ekert, *Phys. Rev. Lett.* **67**, 661-663 (1991).

[52] R.P. Feynman, *The Character of Physical Law*, MIT Press (1965).

[53] A. Bohr, O. Ulfbeck, *Rev.Mod.Phys.* **67**, 1-35 (1995).

[54] W.K. Wootters, W.H. Zurek, *Nature* **299**, 802-803 (1982).

[55] R.F. Werner, *Phys. Rev. A* **58**, 1827-1832 (1998).

[56] C.H. Bennett, H.J. Bernstein, S. Popescu, B. Schumacher, *Phys. Rev. A* **53**, 2046 - 2052 (1996).

[57] M. Horodecki, P. Horodecki, R. Horodecki, *Phys. Rev. Lett.* **80**, 5239-5242 (1998).

[58] M. Horodecki, P. Horodecki, R. Horodecki, *Phys. Rev. Lett.* **84**, 2014-2017 (2000).

[59] W.K. Wootters, *Phys. Rev. Lett.* **80**, 2245 - 2248 (1998).

[60] M. Christandl, A. Winter, *J. Math. Phys.* **45**, 829-840 (2004).

[61] D. Bohm, *Am. J. Phys.* **20**, 522-523 (1952).

[62] J.S. Bell, *Speakable and Unspeakable in Quantum Mechanics*, ed. A. Aspect, Cambridge University Press, Cambridge (2004).

[63] J.F. Clauser, M.A. Horne, A. Shimony, R.A. Holt, *Phys. Rev. Lett.* **23**, 880-884 (1969).

[64] D.M. Greenberger, M.A. Horne, A. Zeilinger, in *Bell's Theorem, Quantum Theory, and Conceptions of the Universe,* M. Kafatos, ed. Kluwer, Dordrecht (1989).

[65] J.W. Pan, M. Daniell, H. Weinfurter, A. Zeilinger Phys. Rev. Lett. 82, 1345 - 1349 (1999).

[66] R.J. Nelson, D.G. Cory, S. Lloyd, *Phys. Rev. A* **61**, 022106 (2000).

[67] M.P. Frank, TF Knight, *Nanotechnology* **9**, 162-176 (1998).

[68] R.P. Feynman, *Opt. News* **11**, 11 (1985); *Found. Phys.* **16**, 507 (1986).

[69] D. Deutsch, R. Jozsa, *Proc. R. Soc. London A* **439**, 553 (1992).

[70] D.R. Simon, in *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*, S. Goldwasser, ed., (*IEEE* Computer Society, Los Alamitos, CA, 1994), pp. 116-123.

[71] D. Coppersmith, *IBM Research Report* RC19642 (1994)

[72] S. Hallgren, *J. ACM* **54**, 1 (2007).

[73] G. Kuperberg, arXiv:quant-ph/0302112, (2003).

[74] S. Hallgren, C. Moore, M. Rötteler, A. Russell, P. Sen, *Proc. of 38th ann. ACM symposium th. comp.*, 604-617 (2006).

[75] A.Yu. Kitaev, A. Shen, M.N. Vyalyi, *Classical and Quantum Computation*, American Mathematical Society, Providence RI (2002).

[76] D.S. Abrams, S. Lloyd, *Phys. Rev. Lett.* **79**, 2586-2589 (1997).

[77] A. Aspuru-Guzik, A.D. Dutoi, P.J. Love, M. Head-Gordon, *Science* **309**, 1704-1707 (2005).

[78] S. Lloyd, *Science* **273**, 1073-8 (1996).

[79] S. Wiesner, arXiv:quant-ph/9603028.

[80] C. Zalka, *Proc.Roy.Soc.Lond. A* **454**, 313-322 (1998).

[81] C. Ramanathan, S. Sinha, J. Baugh, T.F. Havel, D.G. Cory, *Phys. Rev. A* **71**, 020303(R) (2005).

[82] M. Boyer, G. Brassard, P. Hoyer, A. Tapp, *Fortsch. Phys.* **46** 493-506, (1998).

[83] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, *Science* **220**, 671-680 (1983).

[84] E. Farhi, J. Goldstone, S. Gutmann, *Science* **292**, 472-47 6 (2001).

[85] S. Sachdev, *Quantum Phase Transitions*, Cambridge University Press, Cambridge (1999).

[86] D. Aharonov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd, O. Regev, *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science* (FOCS'04), 42-51 (2004); quant-ph/0405098.

[87] E. Farhi, S. Gutmann, *Phys. Rev. A* **58**, 915-928 (1998).

[88] D. Aharonov, A. Ambainis, J. Kempe, U. Vazirani, *Proc. 33th ACM Symposium on Theory of Computing (STOC 2001)*, 50-59 (2001).

[89] A.M. Childs, R. Cleve, E. Deotto, E. Farhi, S. Gutmann, D.A. Spielman *Proc. 35th ACM Symposium on Theory of Computing (STOC 2003)*, 59-68 (2003).

[90] E. Farhi, J. Goldstone, S. Gutmann, 'A quantum algorithm for the Hamiltonian NAND tree,' arXiv:quant-ph/0702144.

[91] R. Hamming, *Coding and Information Theory,* Prentice Hall, Upper Saddle River NJ (1980).

[92] P.W. Shor, *Physical Review A*, **52**, R2493-R2496 (1995)

[93] A.M. Steane, *Phys. Rev. Lett.* **77**, 793-797 (1996).

[94] A.R. Calderbank and P.W. Shor, *Phys. Rev.A* **54**, 1098-1106 (1996).

[95] D. Gottesman, *Phys. Rev. A* **54**, 1862 (1996).

[96] D. Aharonov, M. Ben-Or, *Proc. 29th ann. ACM symposium on th. comp*, 176-188 (1997).

[97] E. Knill, R. Laflamme, W. Zurek, arXiv:quant-ph/9610011.

[98] D. Gottesman, Phys. Rev. A 57, 127 - 137 (1998)

[99] E. Knill, *Nature* **434**, 39-44, (2005).

[100] P. Zanardi, M. Rasetti, *Phys. Rev. Lett.* **79**, 3306 - 3309 (1997).

[101] D.A. Lidar, I.L. Chuang, K.B. Whaley, *Phys. Rev. Lett.* **81**, 2594-2597 (1998).

[102] E. Knill., R. Laflamme, L. Viola, *Phys. Rev. Lett.* **84**, 2525-2528 (2000).

[103] P. Zanardi, S. Lloyd, *Phys. Rev. Lett.* **90**, 067902 (2003)

[104] A.Yu. Kitaev, *Ann. Phys.* **30**, 2-30 (2003).

[105] S. Lloyd, *Quant. Inf. Proc.* **1**, 15-18 (2004).

[106] I. Devetak, *IEEE Trans. Inf. Th.* **51**, 44-55 (2005).

[107] P. W. Shor, Lecture Notes, MSRI Workshop on Quantum Computation, Mathematical Sciences Research Institute, Berkeley CA, November 2002.

[108] D.P. DiVincenzo, P.W. Shor, J.A. Smolin, *Phys. Rev. A* **57**, 830-839 (1998).

[109] P.W. Shor, *Comm. Math. Phys.* **246**, 453-472 (2004).

[110] M.B. Hastings, 'A counterexample to additivity of minimum output entropy,' arXiv:0809.3972.

[111] C.H. Bennett, S.J. Wiesner, *Phys. Rev. Lett.* **69**, 2881 (1992).

[112] C.H. Bennett, P.W. Shor, J.A. Smolin, A.V. Thapliyal, *Phys. Rev. Lett.* **83**, 3081 - 3084 (1999).

[113] C. H. Bennett, G. Brassard, J. M. Robert, *SIAM Journal on Computing,* **17**, 210-229, (1988).

[114] T.C. Ralph, *Phys. Rev. A* **61**, 010303 (1999).

[115] S. Braunstein, A.K. Pati, *Quantum Information With Continuous Variables,* Springer (2003).

[116] A.W. Harrow, A. Hassidim, S. Lloyd, 'Quantum algorithm for solving linear sets of equations,' arXiv:0811.3171.