NMR quantum information processing faces three special difficulties that make it rather different from other quantum information processing schemes. First, the molecules are typically prepared by letting them equilibrate at room temperature, which is so much higher than typical spin flip energies that the spins become nearly completely randomly oriented. This fact makes the initial state rather more 'noisy' than is desirable for quantum information processing. How this noise may be overcome is an interesting story that we tell in Chapter 7. A second problem is that the class of measurements that may be performed in NMR falls well short of the most general measurements we would like to perform in quantum information processing. Nevertheless, for many instances of quantum information processing the class of measurements allowed in NMR is sufficient. Third, because molecules cannot be individually addressed in NMR you might ask how it is that individual qubits can be manipulated in an appropriate way. Fortunately, different nuclei in the molecule can have different properties that allow them to be individually addressed – or at least addressed at a sufficiently fine-grained scale to allow the operations essential for quantum computation.

Many of the elements required to perform large-scale quantum information processing can be found in existing proposals: superb state preparation and quantum measurements can be performed on a small number of qubits in the ion trap; superb dynamics can be performed in small molecules using NMR; fabrication technology in solid state systems allows designs to be scaled up tremendously. A single system having all these elements would be a long way down the road to a dream quantum computer. Unfortunately, all these systems are very different, and we are many, many years from having large-scale quantum computers. However, we believe that the existence of all these properties in existing (albeit different) systems does bode well for the long-term existence of large-scale quantum information processors. Furthermore, it suggests that there is a great deal of merit to pursuing *hybrid* designs which attempt to marry the best features of two or more existing technologies. For example, there is much work being done on trapping atoms inside *electromagnetic cavities*. This enables flexible manipulation of the atom inside the cavity via optical techniques, and makes possible real-time feedback control of single atoms in ways unavailable in conventional atom traps.

To conclude, note that it is important not to assess quantum information processing as though it were just another technology for information processing. For example, it is tempting to dismiss quantum computation as yet another technological fad in the evolution of the computer that will pass in time, much as other fads have passed – for example, the 'bubble memories' widely touted as the next big thing in memory during the early 1980s. This is a mistake, since quantum computation is an *abstract paradigm* for information processing that may have many *different* implementations in technology. One can compare two different proposals for quantum computing as regards their technological merits – it makes sense to compare a 'good' proposal to a 'bad' proposal – however even a very poor proposal for a quantum computer is of a different qualitative nature from a superb design for a classical computer.

1.6 Quantum information

The term 'quantum information' is used in two distinct ways in the field of quantum computation and quantum information. The first usage is as a broad catch-all for all manner of operations that might be interpreted as related to information processing using quantum mechanics. This use encompasses subjects such as quantum computation, quantum teleportation, the no-cloning theorem, and virtually all other topics in this book.

The second use of ‘quantum information’ is much more specialized: it refers to the study of *elementary* quantum information processing tasks. It does not typically include, for example, quantum algorithm design, since the details of specific quantum algorithms are beyond the scope of ‘elementary’. To avoid confusion we will use the term ‘quantum information theory’ to refer to this more specialized field, in parallel with the widely used term ‘(classical) information theory’ to describe the corresponding classical field. Of course, the term ‘quantum information theory’ has a drawback of its own – it might be seen as implying that theoretical considerations are all that matter! Of course, this is not the case, and experimental demonstration of the elementary processes studied by quantum information theory is of great interest.

The purpose of this section is to introduce the basic ideas of quantum information theory. Even with the restriction to elementary quantum information processing tasks, quantum information theory may look like a disordered zoo to the beginner, with many apparently unrelated subjects falling under the ‘quantum information theory’ rubric. In part, that’s because the subject is still under development, and it’s not yet clear how all the pieces fit together. However, we can identify a few fundamental goals uniting work on quantum information theory:

(1) **Identify elementary classes of static resources in quantum mechanics**. An example is the qubit. Another example is the *bit*; classical physics arises as a special case of quantum physics, so it should not be surprising that elementary static resources appearing in classical information theory should also be of great relevance in quantum information theory. Yet another example of an elementary class of static resources is a Bell state shared between two distant parties.

(2) **Identify elementary classes of dynamical processes in quantum mechanics**. A simple example is *memory*, the ability to store a quantum state over some period of time. Less trivial processes are quantum information transmission between two parties, Alice and Bob; copying (or trying to copy) a quantum state, and the process of protecting quantum information processing against the effects of noise.

(3) **Quantify resource tradeoffs incurred performing elementary dynamical processes.** For example, what are the minimal resources required to reliably transfer quantum information between two parties using a noisy communications channel?

Similar goals define classical information theory; however, quantum information theory is broader in scope than classical information theory, for quantum information theory includes all the static and dynamic elements of classical information theory, as well as *additional* static and dynamic elements.

The remainder of this section describes some examples of questions studied by quantum information theory, in each case emphasizing the fundamental static and dynamic elements under consideration, and the resource tradeoffs being considered. We begin with an example that will appear quite familiar to classical information theorists: the problem of sending classical information through a quantum channel. We then begin to branch out and explore some of the new static and dynamic processes present in quantum mechanics, such as quantum error-correction, the problem of distinguishing quantum states, and entanglement transformation. The chapter concludes with some reflections on how the tools of quantum information theory can be applied elsewhere in quantum computation and quantum information.

1.6.1 Quantum information theory: example problems

Classical information through quantum channels

The fundamental results of classical information theory are Shannon's *noiseless channel coding theorem* and Shannon's *noisy channel coding theorem*. The noiseless channel coding theorem quantifies how many bits are required to store information being emitted by a source of information, while the noisy channel coding theorem quantifies how much information can be reliably transmitted through a noisy communications channel.

What do we mean by an *information source*? Defining this notion is a fundamental problem of classical and quantum information theory, one we'll re-examine several times. For now, let's go with a provisional definition: a classical information source is described by a set of probabilities p_j, $j = 1, 2, \ldots, d$. Each use of the source results in the 'letter' j being emitted, chosen at random with probability p_j, independently for each use of the source. For instance, if the source were of English text, then the numbers j might correspond to letters of the alphabet and punctuation, with the probabilities p_j giving the relative frequencies with which the different letters appear in regular English text. Although it is not true that the letters in English appear in an independent fashion, for our purposes it will be a good enough approximation.

Regular English text includes a considerable amount of redundancy, and it is possible to exploit that redundancy to *compress* the text. For example, the letter 'e' occurs much more frequently in regular English text than does the letter 'z'. A good scheme for compressing English text will therefore represent the letter 'e' using fewer bits of information than it uses to represent 'z'. Shannon's noiseless channel coding theorem quantifies exactly how well such a compression scheme can be made to work. More precisely, the noiseless channel coding theorem tells us that a classical source described by probabilities p_j can be compressed so that on average each use of the source can be represented using $H(p_j)$ bits of information, where $H(p_j) \equiv -\sum_j p_j \log(p_j)$ is a function of the source probability distribution known as the *Shannon entropy*. Moreover, the noiseless channel coding theorem tells us that to attempt to represent the source using fewer bits than this will result in a high probability of error when the information is decompressed. (Shannon's noiseless channel coding theorem is discussed in much greater detail in Chapter 12.)

Shannon's noiseless coding theorem provides a good example where the goals of information theory listed earlier are all met. Two static resources are identified (goal number 1): the bit and the information source. A two-stage dynamic process is identified (goal 2), compressing an information source, and then decompressing to recover the information source. Finally a quantitative criterion for determining the resources consumed (goal 3) by an optimal data compression scheme is found.

Shannon's second major result, the noisy channel coding theorem, quantifies the amount of information that can be reliably transmitted through a noisy channel. In particular, suppose we wish to transfer the information being produced by some information source to another location through a noisy channel. That location may be at another point in space, or at another point in time – the latter is the problem of storing information in the presence of noise. The idea in both instances is to encode the information being produced using error-correcting codes, so that any noise introduced by the channel can be corrected at the other end of the channel. The way error-correcting codes achieve this is by introducing enough redundancy into the information sent through the channel so that even after some of the information has been corrupted it is still possible to recover the original message. For example, suppose the noisy channel is for the transmission of single bits, and the noise in the channel is such that to achieve reliable transmission each bit produced by the source must be encoded using two bits before being sent through the channel. We say that such a channel has a *capacity* of half a bit, since each use of the channel can be used to reliably convey roughly half a bit of information. Shannon's noisy channel coding theorem provides a general procedure for calculating the capacity of an arbitrary noisy channel.

Shannon's noisy channel coding theorem also achieves the three goals of information theory we stated earlier. Two types of static resources are involved (goal 1), the information source, and the bits being sent through the channel. Three dynamical processes are involved (goal 2). The primary process is the noise in the channel. To combat this noise we perform the dual processes of encoding and decoding the state in an error-correcting code. For a fixed noise model, Shannon's theorem tells us how much redundancy must be introduced by an optimal error-correction scheme if reliable information transmission is to be achieved (goal 3).

For both the noiseless and noisy channel coding theorems Shannon restricted himself to storing the output from an information source in classical systems – bits and the like. A natural question for quantum information theory is what happens if the storage medium is changed so that classical information is transmitted using quantum states as the medium. For example, it may be that Alice wishes to compress some classical information produced by an information source, transmitting the compressed information to Bob, who then decompresses it. If the medium used to store the compressed information is a quantum state, then Shannon's noiseless channel coding theorem cannot be used to determine the optimal compression and decompression scheme. One might wonder, for example, if using qubits allows a better compression rate than is possible classically. We'll study this question in Chapter 12, and prove that, in fact, qubits do not allow any significant saving in the amount of communication required to transmit information over a noiseless channel.

Naturally, the next step is to investigate the problem of transmitting classical information through a *noisy* quantum channel. Ideally, what we'd like is a result that quantifies the *capacity* of such a channel for the transmission of information. Evaluating the capacity is a very tricky job for several reasons. Quantum mechanics gives us a huge variety of noise models, since it takes place in a continuous space, and it is not at all obvious how to adapt classical error-correction techniques to combat the noise. Might it be advantageous, for example, to encode the classical information using *entangled* states, which are then transmitted one piece at a time through the noisy channel? Or perhaps it will be advantageous to decode using entangled measurements? In Chapter 12 we'll prove the *HSW (Holevo–Schumacher–Westmoreland) theorem*, which provides a lower bound on the capacity of such a channel. Indeed, it is widely believed that the HSW theorem provides an exact evaluation of the capacity, although a complete proof of this is not yet known! What remains at issue is whether or not encoding using entangled states can be used to raise the capacity beyond the lower bound provided by the HSW theorem. All evidence to date suggests that this doesn't help raise the capacity, but it is still a fascinating open problem of quantum information theory to determine the truth or falsity of this conjecture.

Quantum information through quantum channels

Classical information is, of course, not the only static resource available in quantum mechanics. Quantum states themselves are a natural static resource, even more natural than classical information. Let's look at a *different* quantum analogue of Shannon's coding theorems, this time involving the compression and decompression of quantum states.

To begin, we need to define some quantum notion of an information source, analogous to the classical definition of an information source. As in the classical case, there are several different ways of doing this, but for the sake of definiteness let's make the provisional definition that a quantum source is described by a set of probabilities p_j and corresponding quantum states $|\psi_j\rangle$. Each use of the source produces a state $|\psi_j\rangle$ with probability p_j, with different uses of the source being independent of one another.

Is it possible to compress the output from such a quantum mechanical source? Consider the case of a qubit source which outputs the state $|0\rangle$ with probability p and the state $|1\rangle$ with probability $1 - p$. This is essentially the same as a classical source emitting single bits, either 0 with probability p, or 1 with probability $1 - p$, so it is not surprising that similar techniques can be used to compress the source so that only $H(p, 1 - p)$ qubits are required to store the compressed source, where $H(\cdot)$ is again the Shannon entropy function.

What if the source had instead been producing the state $|0\rangle$ with probability p, and the state $(|0\rangle + |1\rangle)/\sqrt{2}$ with probability $1 - p$? The standard techniques of classical data compression no longer apply, since in general it is not possible for us to distinguish the states $|0\rangle$ and $(|0\rangle + |1\rangle)/\sqrt{2}$. Might it still be possible to perform some type of compression operation?

It turns out that a type of compression is still possible, even in this instance. What is interesting is that the compression may no longer be *error-free*, in the sense that the quantum states being produced by the source may be slightly distorted by the compression–decompression procedure. Nevertheless, we require that this distortion ought to become very small and ultimately negligible in the limit of large blocks of source output being compressed. To quantify the distortion we introduce a *fidelity* measure for the compression scheme, which measures the average distortion introduced by the compression scheme. The idea of quantum data compression is that the compressed data should be recovered with very good fidelity. Think of the fidelity as being analogous to the probability of doing the decompression correctly – in the limit of large block lengths, it should tend towards the no error limit of 1.

Schumacher's noiseless channel coding theorem quantifies the resources required to do quantum data compression, with the restriction that it be possible to recover the source with fidelity close to 1. In the case of a source producing orthogonal quantum states $|\psi_j\rangle$ with probabilities p_j Schumacher's theorem reduces to telling us that the source may be compressed down to but not beyond the classical limit $H(p_j)$. However, in the more general case of non-orthogonal states being produced by the source, Schumacher's theorem tells us how much a quantum source may be compressed, and the answer is *not* the Shannon entropy $H(p_j)$! Instead, a new entropic quantity, the *von Neumann* entropy, turns out to be the correct answer. In general, the von Neumann entropy agrees with the Shannon entropy if and only if the states $|\psi_j\rangle$ are orthogonal. Otherwise, the von Neumann entropy for the source $p_j, |\psi_j\rangle$ is in general strictly *smaller* than the Shannon entropy $H(p_j)$. Thus, for example, a source producing the state $|0\rangle$ with probability p

and $(|0\rangle + |1\rangle)/\sqrt{2}$ with probability $1 - p$ can be reliably compressed using fewer than $H(p, 1 - p)$ qubits per use of the source!

The basic intuition for this decrease in resources required can be understood quite easily. Suppose the source emitting states $|0\rangle$ with probability p and $(|0\rangle + |1\rangle)/\sqrt{2}$ with probability $1 - p$ is used a large number n times. Then by the law of large numbers, with high probability the source emits about np copies of $|0\rangle$ and $n(1 - p)$ copies of $(|0\rangle + |1\rangle)/\sqrt{2}$. That is, it has the form

$$|0\rangle^{\otimes np} \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}} \right)^{\otimes n(1-p)}, \tag{1.60}$$

up to re-ordering of the systems involved. Suppose we expand the product of $|0\rangle + |1\rangle$ terms on the right hand side. Since $n(1 - p)$ is large, we can again use the law of large numbers to deduce that the terms in the product will be roughly one-half $|0\rangle$s and one-half $|1\rangle$s. That is, the $|0\rangle + |1\rangle$ product can be well approximated by a superposition of states of the form

$$|0\rangle^{\otimes n(1-p)/2} |1\rangle^{\otimes n(1-p)/2}. \tag{1.61}$$

Thus the state emitted by the source can be approximated as a superposition of terms of the form

$$|0\rangle^{\otimes n(1+p)/2} |1\rangle^{\otimes n(1-p)/2}. \tag{1.62}$$

How many states of this form are there? Roughly n choose $n(1 + p)/2$, which by Stirling's approximation is equal to $N \equiv 2^{nH[(1+p)/2,(1-p)/2]}$. A simple compression method then is to label all states of the form (1.62) $|c_1\rangle$ through $|c_N\rangle$. It is possible to perform a unitary transform on the n qubits emitted from the source that takes $|c_j\rangle$ to $|j\rangle|0\rangle^{\otimes n - nH[(1+p)/2,(1-p)/2]}$, since j is an $nH[(1 + p)/2, (1 - p)/2]$ bit number. The compression operation is to perform this unitary transformation, and then drop the final $n - nH[(1+p)/2, (1-p)/2]$ qubits, leaving a compressed state of $nH[(1+p)/2, (1-p)/2]$ qubits. To decompress we append the state $|0\rangle^{\otimes n - nH[(1+p)/2,(1-p)/2]}$ to the compressed state, and perform the inverse unitary transformation.

This procedure for quantum data compression and decompression results in a storage requirement of $H[(1 + p)/2, (1 - p)/2]$ qubits per use of the source, which whenever $p \geq 1/3$ is an improvement over the $H(p, 1 - p)$ qubits we might naively have expected from Shannon's noiseless channel coding theorem. In fact, Schumacher's noiseless channel coding theorem allows us to do somewhat better even than this, as we will see in Chapter 12; however, the essential reason in that construction is the same as the reason we were able to compress here: we exploited the fact that $|0\rangle$ and $(|0\rangle + |1\rangle)/\sqrt{2}$ are not orthogonal. Intuitively, the states contain some redundancy since both have a component in the $|0\rangle$ direction, which results in more physical similarity than would be obtained from orthogonal states. It is this redundancy that we have exploited in the coding scheme just described, and which is used in the full proof of Schumacher's noiseless channel coding theorem. Note that the restriction $p \geq 1/3$ arises because when $p < 1/3$ this particular scheme doesn't exploit the redundancy in the states: we end up effectively *increasing* the redundancy present in the problem! Of course, this is an artifact of the particular scheme we have chosen, and the general solution exploits the redundancy in a much more sensible way to achieve data compression.

Schumacher's noiseless channel coding theorem is an analogue of Shannon's noiseless channel coding theorem for the compression and decompression of quantum states. Can we find an analogue of Shannon's noisy channel coding theorem? Considerable progress on this important question has been made, using the theory of quantum error-correcting codes; however, a fully satisfactory analogue has not yet been found. We review some of what is known about the quantum channel capacity in Chapter 12.

Quantum distinguishability

Thus far all the dynamical processes we have considered – compression, decompression, noise, encoding and decoding error-correcting codes – arise in both classical and quantum information theory. However, the introduction of new types of information, such as quantum states, enlarges the class of dynamical processes beyond those considered in classical information theory. A good example is the problem of distinguishing quantum states. Classically, we are used to being able to distinguish different items of information, at least in principle. In practice, of course, a smudged letter 'a' written on a page may be very difficult to distinguish from a letter 'o', but in principle it is possible to distinguish between the two possibilities with perfect certainty.

On the other hand, quantum mechanically it is *not* always possible to distinguish between arbitrary states. For example, there is no process allowed by quantum mechanics that will reliably distinguish between the states $|0\rangle$ and $(|0\rangle + |1\rangle)/\sqrt{2}$. Proving this rigorously requires tools we don't presently have available (it is done in Chapter 2), but by considering examples it's pretty easy to convince oneself that it is not possible. Suppose, for example, that we try to distinguish the two states by measuring in the computational basis. Then, if we have been given the state $|0\rangle$, the measurement will yield 0 with probability 1. However, when we measure $(|0\rangle + |1\rangle)/\sqrt{2}$ the measurement yields 0 with probability 1/2 and 1 with probability 1/2. Thus, while a measurement result of 1 implies that state must have been $(|0\rangle + |1\rangle)/\sqrt{2}$, since it couldn't have been $|0\rangle$, we can't infer anything about the identity of the quantum state from a measurement result of 0.

This indistinguishability of non-orthogonal quantum states is at the heart of quantum computation and quantum information. It is the essence of our assertion that a quantum state contains hidden information that is not accessible to measurement, and thus plays a key role in quantum algorithms and quantum cryptography. One of the central problems of quantum information theory is to develop measures quantifying how well non-orthogonal quantum states may be distinguished, and much of Chapters 9 and 12 is concerned with this goal. In this introduction we'll limit ourselves to pointing out two interesting aspects of indistinguishability – a connection with the possibility of faster-than-light communication, and an application to 'quantum money.'

Imagine for a moment that we could distinguish between arbitrary quantum states. We'll show that this implies the ability to communicate faster than light, using entanglement. Suppose Alice and Bob share an entangled pair of qubits in the state $(|00\rangle + |11\rangle)/\sqrt{2}$. Then, if Alice measures in the computational basis, the post-measurement states will be $|00\rangle$ with probability 1/2, and $|11\rangle$ with probability 1/2. Thus Bob's system is either in the state $|0\rangle$, with probability 1/2, or in the state $|1\rangle$, with probability 1/2. Suppose, however, that Alice had instead measured in the $|+\rangle, |-\rangle$ basis. Recall that $|0\rangle = (|+\rangle + |-\rangle)/\sqrt{2}$ and $|1\rangle = (|+\rangle - |-\rangle)/\sqrt{2}$. A little algebra shows that the initial state of Alice and Bob's system may be rewritten as $(|++\rangle + |--\rangle)/\sqrt{2}$. Therefore, if Alice measures in the $|+\rangle, |-\rangle$ basis, the state of Bob's system after the measurement will be $|+\rangle$ or $|-\rangle$ with probability $1/2$ each. So far, this is all basic quantum mechanics. But if Bob had access to a device that could distinguish the four states $|0\rangle, |1\rangle, |+\rangle, |-\rangle$ from one another, then he could tell whether Alice had measured in the computational basis, or in the $|+\rangle, |-\rangle$ basis. Moreover, he could get that information *instantaneously*, as soon as Alice had made the measurement, providing a means by which Alice and Bob could achieve faster-than-light communication! Of course, we know that it is not possible to distinguish non-orthogonal quantum states; this example shows that this restriction is also intimately tied to other physical properties which we expect the world to obey.

The indistinguishability of non-orthogonal quantum states need not always be a handicap. Sometimes it can be a boon. Imagine that a bank produces banknotes imprinted with a (classical) serial number, and a sequence of qubits each in either the state $|0\rangle$ or $(|0\rangle + |1\rangle)/\sqrt{2}$. Nobody but the bank knows what sequence of these two states is embedded in the note, and the bank maintains a list matching serial numbers to embedded states. The note is impossible to counterfeit exactly, because it is impossible for a would-be counterfeiter to determine with certainty the state of the qubits in the original note, without destroying them. When presented with the banknote a merchant (of certifiable repute) can verify that it is not a counterfeit by calling the bank, telling them the serial number, and then asking what sequence of states were embedded in the note. They can then check that the note is genuine by measuring the qubits in the $|0\rangle, |1\rangle$ or $(|0\rangle + |1\rangle)/\sqrt{2}, (|0\rangle - |1\rangle)/\sqrt{2}$ basis, as directed by the bank. With probability which increases exponentially to one with the number of qubits checked, any would-be counterfeiter will be detected at this stage! This idea is the basis for numerous other quantum cryptographic protocols, and demonstrates the utility of the indistinguishability of non-orthogonal quantum states.

Exercise 1.2: Explain how a device which, upon input of one of two non-orthogonal quantum states $|\psi\rangle$ or $|\varphi\rangle$ correctly identified the state, could be used to build a device which cloned the states $|\psi\rangle$ and $|\varphi\rangle$, in violation of the no-cloning theorem. Conversely, explain how a device for cloning could be used to distinguish non-orthogonal quantum states.

Creation and transformation of entanglement

Entanglement is another elementary static resource of quantum mechanics. Its properties are amazingly different from those of the resources most familiar from classical information theory, and they are not yet well understood; we have at best an incomplete collage of results related to entanglement. We don't yet have all the language needed to understand the solutions, but let's at least look at two information-theoretic problems related to entanglement.

Creating entanglement is a simple dynamical process of interest in quantum information theory. How many qubits must two parties exchange if they are to create a particular entangled state shared between them, given that they share no prior entanglement? A second dynamical process of interest is *transforming entanglement* from one form into another. Suppose, for example, that Alice and Bob share between them a Bell state, and wish to transform it into some other type of entangled state. What resources do they need to accomplish this task? Can they do it without communicating? With classical communication only? If quantum communication is required then how much quantum communication is required?

Answering these and more complex questions about the creation and transformation of entanglement forms a fascinating area of study in its own right, and also promises to give insight into tasks such as quantum computation. For example, a distributed quantum computation may be viewed as simply a method for generating entanglement between two or more parties; lower bounds on the amount of communication that must be done to perform such a distributed quantum computation then follow from lower bounds on the amount of communication that must be performed to create appropriate entangled states.

1.6.2 Quantum information in a wider context

We have given but the barest glimpse of quantum information theory. Part III of this book discusses quantum information theory in much greater detail, especially Chapter 11, which deals with fundamental properties of entropy in quantum and classical information theory, and Chapter 12, which focuses on pure quantum information theory.

Quantum information theory is the most abstract part of quantum computation and quantum information, yet in some sense it is also the most fundamental. The question driving quantum information theory, and ultimately all of quantum computation and quantum information, is *what makes quantum information processing tick?* What is it that separates the quantum and the classical world? What resources, unavailable in a classical world, are being utilized in a quantum computation? Existing answers to these questions are foggy and incomplete; it is our hope that the fog may yet lift in the years to come, and we will obtain a clear appreciation for the possibilities and limitations of quantum information processing.

Problem 1.1: (Feynman-Gates conversation) Construct a friendly imaginary discussion of about 2000 words between Bill Gates and Richard Feynman, set in the present, on the future of computation. (*Comment*: You might like to try waiting until you've read the rest of the book before attempting this question. See the 'History and further reading' below for pointers to one possible answer for this question.)

Problem 1.2: What is the most significant discovery yet made in quantum computation and quantum information? Write an essay of about 2000 words for an educated lay audience about the discovery. (*Comment*: As for the previous problem, you might like to try waiting until you've read the rest of the book before attempting this question.)

History and further reading

Most of the material in this chapter is revisited in more depth in later chapters. Therefore the historical references and further reading below are limited to material which does not recur in later chapters.

Piecing together the historical context in which quantum computation and quantum information have developed requires a broad overview of the history of many fields. We have tried to tie this history together in this chapter, but inevitably much background material was omitted due to limited space and expertise. The following recommendations attempt to redress this omission.

The history of quantum mechanics has been told in many places. We recommend especially the outstanding works of Pais[Pai82, Pai86, Pai91]. Of these three, [Pai86] is most directly concerned with the development of quantum mechanics; however, Pais' biographies of Einstein[Pai82] and of Bohr[Pai91] also contain much material of interest, at a less intense level. The rise of technologies based upon quantum mechanics has been described by Milburn[Mil97, Mil98]. Turing's marvelous paper on the foundations of computer science[Tur36] is well worth reading. It can be found in the valuable historical collection of Davis[Dav65]. Hofstadter[Hof79] and Penrose[Pen89] contain entertaining and informative discussions of the foundations of computer science. Shasha and Lazere's biography of fifteen leading computer scientists[SL98] gives considerable insight into many different facets of the history of computer science. Finally, Knuth's awesome series of books[Knu97, Knu98a, Knu98b] contain an amazing amount of historical information. Shannon's brilliant papers founding information theory make excellent reading[Sha48] (also reprinted in [SW49]). MacWilliams and Sloane[MS77] is not only an excellent text on error-correcting codes, but also contains an enormous amount of useful historical information. Similarly, Cover and Thomas[CT91] is an excellent text on information theory, with extensive historical information. Shannon's collected works, together with many useful historical items have been collected in a large volume[SW93] edited by Sloane and Wyner. Slepian has also collected a useful set of reprints on information theory[Sle74]. Cryptography is an ancient art with an intricate and often interesting history. Kahn[Kah96] is a huge history of cryptography containing a wealth of information. For more recent developments we recommend the books by Menezes, van Oorschot, and Vanstone[MvOV96], Schneier[Sch96a], and by Diffie and Landau[DL98].

Quantum teleportation was discovered by Bennett, Brassard, Crépeau, Jozsa, Peres, and Wootters[BBC+93], and later experimentally realized in various different forms by Boschi, Branca, De Martini, Hardy and Popescu[BBM+98] using optical techniques, by Bouwmeester, Pan, Mattle, Eibl, Weinfurter, and Zeilinger[BPM+97] using photon polarization, by Furusawa, Sørensen, Braunstein, Fuchs, Kimble, and Polzik using 'squeezed' states of light[FSB+98], and by Nielsen, Knill, and Laflamme using NMR[NKL98].

Deutsch's problem was posed by Deutsch[Deu85], and a one-bit solution was given in the same paper. The extension to the general n-bit case was given by Deutsch and Jozsa[DJ92]. The algorithms in these early papers have been substantially improved subsequently by Cleve, Ekert, Macchiavello, and Mosca[CEMM98], and independently in unpublished work by Tapp. In this chapter we have given the improved version of the algorithm, which fits very nicely into the hidden subgroup problem framework that will later be discussed in Chapter 5. The original algorithm of Deutsch only worked probabilistically; Deutsch and Jozsa improved this to obtain a deterministic algorithm, but their method required two function evaluations, in contrast to the improved algorithms presented in this chapter. Nevertheless, it is still conventional to refer to these algorithms as Deutsch's algorithm and the Deutsch–Jozsa algorithm in honor of two huge leaps forward: the concrete demonstration by Deutsch that a quantum computer could do something faster than a classical computer; and the extension by Deutsch and Jozsa which demonstrated for the first time a similar gap for the scaling of the time required to solve a problem.

Excellent discussions of the Stern–Gerlach experiment can be found in standard quantum mechanics textbooks such as the texts by Sakurai[Sak95], Volume III of Feynman, Leighton and Sands[FLS65a], and Cohen-Tannoudji, Diu and Laloë[CTDL77a, CTDL77b].

Problem 1.1 was suggested by the lovely article of Rahim[Rah99].