Introduction to Quantum Information

Stephen M. Barnett

School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, UK

OXFORD
UNIVERSITY PRESS

Contents

1
Classical Information Theory

In these five lectures I shall give a short introduction to the field of quantum information. The lectures are drawn from my book *Quantum Information* (Barnett 2009). My aim is to make these notes self-contained, but cannot hope to cover the totality of what is now a large and active field of research. In these notes I aim, rather, to give a taster of the subject to whet the appetite of the reader. A more complete introduction may be found in (Barnett 2009) or in any of a now large collection of books and review papers devoted to the topic (far too may for me to attempt to make a list and to risk offence by innocent omission). One text that needs to be mentioned, however, is that by Nielsen and Chuang (2000), which did so much both to popularise the field and to lay out its founding principles.

I am grateful to Oxford University Press for their permission and, indeed, encouragement to reproduce some of the material from (Barnett 2009). I wish to express also my gratitude to Allison Yao for her help in preparing these notes.

1.1 A very short history

Our field starts with the work of the Reverend Thomas Bayes (1702–1761) and the celebrated theorem that bears his name (of which more below) (Bayes 1763). His key idea was that probabilities depend on what you know; if we acquire additional *information* then this modifies the probabilities. Today such reasoning is uncontentious and forms part of the prevailing paradigm in much of probability theory (Jeffreys, 1939; Box and Tiao 1973; Bretthorst 1988; Lee 1989; Jaynes 2003). This was not the case, however, for most of the 350 years of its history. An entertaining and informative presentation of its troubled history may be found in (Mcgrayne 2011).

The picture is completed by identifying, or formulating the quantity of information. It was Claude Shannon (1916–2001) who solved this problem and, by using it to devise his two coding theorems, founded information theory (Shannon 1948). Perhaps I can give an indication of the magnitude of Shannon's achievement by relating that the title of his paper was *A Mathematical Theory of Communication*, but a year later the paper was republished as a book (Shannon and Weaver 1949); apart from correcting a few typographical errors, there are only two changes, the inclusion of a short introductory article by Weaver and a change of title to *The Mathematical Theory of Communication.* The theory was born complete, the numerous textbooks on the topic have greatly broadened the scope and application of Shannon's ideas but have not departed from the fundamentals as explained by Shannon in his first paper (Brillouin 1956; Khinchin 1957; Kullback 1959; Hamming 1980; Cover and Thomas 1991; Goldie and Pinch 1991).

Shannon's information has a simple and familiar form. For a given set of probabilities $\{p_i\}$, the information is

$$H = -\sum_i p_i \log p_i \,. \tag{1.1}$$

Remarkably, this is simply Boltzmann's formula for the entropy (of which more later).

We can sum up the fundamental lessons learned from Bayes and from Shannon as follows: Bayes taught us that probabilities are not absolute but rather depend on available information. Shannon showed that information itself is a precisely defined function of the probabilities.

Shannon's work was aimed, initially, at the problem of providing a quantitative theory of communications, but any set of probabilities can be associated with a quantity of information and it follows that any probabilistic phenomenon has an associated information theory. This idea has been applied, for example, to statistical mechanics (Jaynes 1957a, 1957b). Quantum theory is a probabilistic theory, of course, and so it was inevitable that a quantum information theory would be developed.

1.2 Probabilities and conditional probabilities

Consider an event A with possible outcomes $\{a_i\}$. Everything we know is specified by the probabilities for the possible outcomes: $\{P(a_i)\}$. For tossing a fair coin, for example, the outcomes are 'heads' and 'tails' with $P(\text{heads}) = \frac{1}{2} = P(\text{tails})$. In general the probabilities satisfy the two conditions

$$\begin{aligned} 0 \le P(a_i) \le 1 \\ \sum_i P(a_i) = 1 \,. \end{aligned} \tag{1.2}$$

If we have two events A and B with outcomes $\{a_i\}$ and $\{b_j\}$ then the complete description is given by the *joint* probabilities, $P(a_i, b_j)$. Here the comma is read as 'and' so that $P(a_i, b_j)$ is the probability that both $A = a_i$ *and* $B = b_j$. If the two events are independent then $P(a_i, b_j) = P(a_i)P(b_j)$ but this is *not* true in general. More generally we have

$$\begin{aligned} P(a_i) &= \sum_j P(a_i, b_j) \\ P(b_j) &= \sum_i P(a_i, b_j) \,. \end{aligned} \tag{1.3}$$

If the events are created then what does learning the value of A tell us about the value of B? If we learn, for example, that $A = a_0$ then $P(b_j)$ is *replaced* by

$$P(b_j|a_0) \propto P(a_0, b_j) \,. \tag{1.4}$$

Here the vertical line is read as 'given that' so that $P(b_j|a_0)$ is the probability that $B = b_j$ *given that* $A = a_0$. We can find the constant of proportionality by noting that the sum over j of $P(b_j|a_0)$ must be unity and this leads to Bayes' rule:

$$P(a_i, b_j) = P(b_j|a_i)P(a_i)$$

$$= P(a_i|b_j)P(b_j)\,. \tag{1.5}$$

Bayes' theorem utilises this rule to relate the two conditional probabilities:

$$P(a_i|b_j) = \frac{P(b_j|a_i)P(a_i)}{P(b_j)}\,. \tag{1.6}$$

1.3 Entropy and information

The link between information and entropy is a subtle and beautiful one. Shannon himself gave a simple derivation of this as a mathematical theorem using only a few simple axioms he argued that information must satisfy (Shannon and Weaver 1949). We shall not reproduce this here bust rather present a simple argument to make the association plausible.

Suppose that we have a single event A with possible outcomes a_i. If one, say a_0, is certain to occur then we acquire no information by observing A. By simple extension, if $A = a_0$ is very *likely* to occur then we might confidently expect it and so when it happens we learn very little. If, however, $A = a_0$ is very *unlikely* then when it occurs we might need to drastically modify our actions. A very simple example is a two-state communication system that you may not have considered: a fire alarm. A fire alarm is either ringing or not ringing. Its most common state (hopefully not ringing) is so innocuous that we give it no thought. When it does ring, however, we stop what we are doing and leave the building (if we are sensible).

It seems reasonable to conclude that learning the value of A provides a quantity of information, h, that *increases* as the corresponding probability *decreases*:

$$h[P(a_i)] \Uparrow \quad \text{as} \quad P(a_i) \Downarrow\,.$$

We think of learning something new as *adding* to the available information. Therefore for independent probabilities for a pair of events, $P(a_i, b_j) = P(a_i)P(b_j)$, it is natural to require that

$$\begin{aligned} h[P(a_i, b_j)] &= h[P(a_i)P(b_j)] \\ &= h[P(a_i)] + h[P(b_j)]\,, \end{aligned} \tag{1.7}$$

which immediately suggests logarithms. Hence we set

$$h[P(a_i)] = -K \log P(a_i)\,. \tag{1.8}$$

Here K is a positive constant, yet to be determined, and the minus sign ensures both that h is positive and that it increases as the probability decreases (recall that the logarithm of a number less than unity is negative).

It is convenient to define the information as the *average* of h, which means weighting h for each outcome by its probability of occurring and then summing. This procedure leads us, of course, to the entropy:

$$H = -K \sum_i P(a_i) \log P(a_i)\,. \tag{1.9}$$

We can absorb the prefactor K into the choice of base for the logarithm. A convenient choice is to use base 2 so that

$$H = -\sum_i P(a_i)\log_2 P(a_i) \quad \text{bits}\,, \tag{1.10}$$

which is Shannon's formula for the information. Henceforth we shall drop the subscript 2. It is sometimes convenient, especially in analytical calculations, to use the natural base of logarithms, base e, which gives the entropy in nats:

$$H_e = -\sum_i P(a_i)\ln P(a_i) \quad \text{nats}\,. \tag{1.11}$$

It is straightforward to show that $H_e = H \ln 2$, so that 1 nat = $\ln 2$ bits.

For two events A and B we can write informations for the joint events or for the single events:

$$\begin{aligned} H(A,B) &= -\sum_{i,j} P(a_i,b_j)\log P(a_i,b_j) \\ H(A) &= -\sum_{i,j} P(a_i,b_j)\log \sum_k P(a_i,b_k) \\ H(B) &= -\sum_{i,j} P(a_i,b_j)\log \sum_l P(a_l,b_j)\,. \end{aligned} \tag{1.12}$$

We can also define information based in the conditional probabilities. An especially useful measure of correlation between the two events is the *mutual information*:

$$H(A:B) = H(A) + H(B) - H(A,B)\,. \tag{1.13}$$

This has the important properties that it is positive or zero and that it takes the value zero if and only if the events are independent:

$$\begin{aligned} &H(A:B) \geq 0 \\ &H(A:B) = 0 \qquad \text{iff} \quad P(a_i,b_j) = P(a_i)P(b_j) \quad \forall i,j\,. \end{aligned} \tag{1.14}$$

It is the mutual information that provides an upper bound on the rate at which we can communicate. A more detailed (but gentle) discussion of these entropies and exploration of their properties may be found in (Barnett 2009).

1.4 Information and thermodynamics

The fact that the mathematical form of the information is also the entropy begs the question as to whether information entropy is the *same* quantity that appears in statistical mechanics. It is!

An important and simple example is the way in which we can obtain the Boltzmann distribution by maximising the information (what we have yet to discover) subject only to a constraint on the average energy. This is such a nice calculation that I cannot resist the temptation to include it here, even though there was not time to present it in my lectures. Let a quantum system have a set of discrete energy levels E_i and be in thermodynamic equilibrium with its environment. Our task is to determine the probability, p_i, that the system is to be found in any one of its given energy levels, E_i. Of course we know how to do this by maximising the entropy, but information theory gives us a rather different take on the problem; we maximise the information of the system subject to the constraint only that its mean energy is fixed. In this way we maximise, in an unbiased manner, our uncertainty about the state. The derivation is readily performed as a variational calculation using Lagrange's method of undetermined multipliers (Boas 1983). It is convenient to work with the natural base of logarithms in this case and so we define the information to be

$$H_e = -\sum_i p_i \ln p_i \,. \tag{1.15}$$

Our task is to maximise this quantity subject to a fixed mean energy, $\sum_i p_i E_i = \bar{E}$, and the probabilities summing to unity, $\sum_i p_i = 1$. We can achieve this by varying, independently, the probabilities in the function

$$\tilde{H} = H_e + \lambda \left(1 - \sum_i p_i\right) + \beta \left(\bar{E} - \sum_i p_i E_i\right) \,. \tag{1.16}$$

We find

$$d\tilde{H} = \sum_i (-\ln p_i - 1 - \lambda - \beta E_i)\, dp_i \,. \tag{1.17}$$

We require this to be stationary (zero) for all dp_i, which leads to the solution

$$p_i = e^{-1-\lambda} e^{-\beta E_i} \,. \tag{1.18}$$

We can fix the value of the undetermined multipliers by enforcing the normalisation of the probabilities and the value of the average energy to arrive at the familiar Boltzmann distribution:

$$p_i = \frac{\exp\left(-E_i/(k_B T)\right)}{\sum_j \exp\left(-E_j/(k_B T)\right)} \,. \tag{1.19}$$

A more dramatic example was provided by Szilard in his paper *On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings* (Szilard 1929). This paper is all the more remarkable in that it precedes Shannon's work by nearly 20 years. Here is Szilard's argument. Consider a box, of volume V_0 with a movable partition dividing the box into two equal halves. There is a single molecule in one of the two sides, as depicted in Fig. 1.1. An intelligent being can look in the box and determine which side the molecule is in. By attaching a small weight to a pulley we can extract work from heat in an apparent violation of the second law of thermodynamics, an interesting paradox in the style of Maxwell's famous demon (Maxwell 1871).

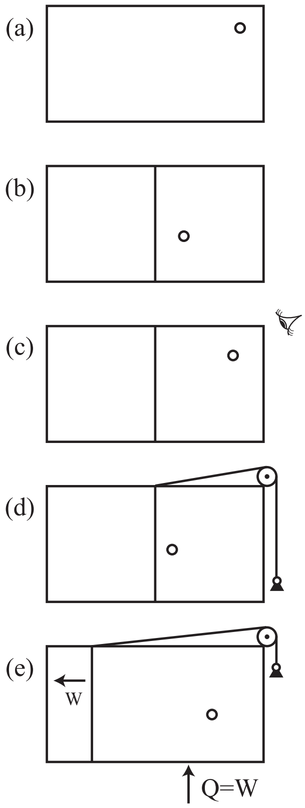


Fig. 1.1 Szilard's illustration of the connection between information and (thermodynamic) entropy. Reproduced, with permission, from Barnett(2009).

Applying the ideal gas law,

$$PV = k_B T\,, \tag{1.20}$$

to our single-molecule gas allows us to calculate the work extracted from the expanding gas in an isothermal expansion:

$$W = \int_{V_0/2}^{V_0} PdV = k_B T \ln 2\,. \tag{1.21}$$

We have extracted useful work from the reservoir in apparent conflict with the second law of thermodynamics. The second law can be saved, however, if the process of measuring and recording the position of the molecule produces an entropy change of *not less than*

$$\Delta S = k_B \ln 2\,. \tag{1.22}$$

Szilard concludes, in this way, a direct link between information and thermodynamic entropy.

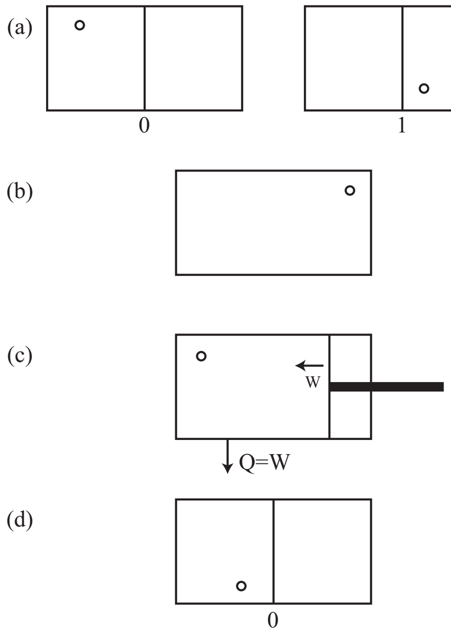


Fig. 1.2 Illustration of Landauer's derivation of the thermodynamic cost of information erasure. Reproduced, with permission, from Barnett(2009).

A more precise statement is that to complete the thermodynamic cycle, we need to include the process of the observer forgetting in which side the molecule was found. The process of forgetting was studied by Landauer (1961, Leff and Rex 1994, Plenio and Vitteli 2001). He considered a single-molecule gas, like Szilard's, and proposed using the position of the molecule to represent a logical bit: the molecule being to the left of the partion corresponding to a logical 0 and to the right corresponding to a logical 1. Landauer showed that erasing and resetting the bit to 0 requires the dissipation of at least $k_B T \ln 2$ worth of energy as heat. To see this we can simply remove the membrane (a reversible process) and then push in, slowly, a fresh partition from the right, as depicted in Fig. 1.2. When the partition reached halfway the 'memory' has been reset to the bit value 0 and, of course, all trace of the original bit value has been lost. The work that needs to be done to achieve this is

$$W = -\int_{V_0}^{V_0/2} PdV = k_B T \ln 2 \,, \tag{1.23}$$

and this is dissipated, of course, as heat. We can view this as a resolution of the paradox implicit in Szilard's model. This is not the end of the story, however. It has recently been shown that Landauer's limit can be beaten if we pay a cost, not in energy, but in some other quantity such as angular momentum (Vaccaro and Barnett 2011). The *key idea* to take from these models, however, is unaffected: information is physical.

1.5 Communications Theory

Shannon first introduced his information theory as a/the mathematical theory of communication (Shannon 1948, Shannon and Weaver 1949). In doing so he introduced the very general model of a communications system depicted in Fig. 1.3. Let us analyse this by introducing two characters, Alice and Bob, who have become very popular in quantum communications. Alice wishes to send a message to Bob. Alice's event, A, is the selection of the message from a set of possible messages $\{a_i\}$. Bob's event, B, is detecting the message from the possible set $\{b_j\}$. On Alice's side there is an information source, which produces the choice of message to be sent (this may be Alice herself, of course) and a transmitter which produces the signal (electrical, optical, acoustical or even a piece of paper) ready for transmission. The signal propagates from Alice to Bob and, whilst en route, is subject to noise. Bob receives the noisy signal and his receiving device turns the signal into a message.

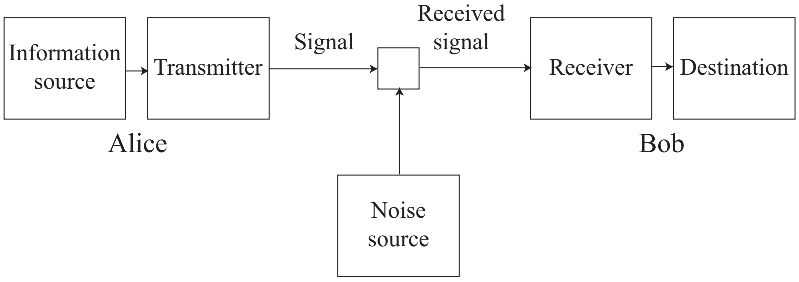


Fig. 1.3 Shannon's model of a communications channel. Reproduced, with permission, from Barnett(2009).

The limiting performance of such a communications channel is governed by two theorems derived by Shannon (Shannon 1948, Shannon and Weaver 1949): his noiseless coding theorem, which tells us how much a message can be compressed and still be read, and his noisy channel coding theorem, which tells us how much redundancy is needed to correct errors.

1.5.1 Noiseless coding theorem

Most messages have an element of redundancy and can be compressed but still be readable. As an example, you might like to try this one[1]:

THS LS HCHS SCHL HS NTRSTNG LCTRS

I compressed the message by removing the vowels. Shannon's noiseless coding theorem quantifies the redundancy in a message. Two simple examples are that in English we mostly don't need to put a 'u' after 'q' and a three-letter word beginning with 'th' will almost certainly be 'the'.

[1] If you are having difficulty, the message is THiS LeS HouCHeS SCHooL HaS iNTeReSTiNG LeCTuReS.

Shannon's theorem shows us the typical number of bits we need to encode a message efficiently. It is the *entropy* associated with the probabilities for the messages that is the key quantity. If we have one of a set of messages that is N bits long, for example

$$\underbrace{010\cdots111}_{N\text{ bits}},$$

and each of the possible possible messages $\{a_i\}$ is selected with the probability associated probability $P(a_i)$, then Shannon's noiseless coding theorem tells us that we need only an *average* of $H(A)$ bits, if we can find an optimum coding scheme.

We shall not prove Shannon's noiseless coding theorem, but rather present a simple example from Shannon's original paper (Shannon 1948, Shannon and Weaver 1949). A simple derivation, both of the noiseless and noisy coding theorems may be found in (Stenholm and Suominen 2005, Barnett 2009). Consider a message formed from an alphabet of four letters, A, B, C and D and let each entry in the message be one of these four with the probabilities

$$\begin{aligned}
P(\mathrm{A}) &= \frac{1}{2} \\
P(\mathrm{B}) &= \frac{1}{4} \\
P(\mathrm{C}) &= \frac{1}{8} = P(\mathrm{D})\,.
\end{aligned} \tag{1.24}$$

The information associated with this set of probabilities is

$$\begin{aligned}
H &= -\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{4}\log\frac{1}{4} + 2\times\frac{1}{8}\log\frac{1}{8}\right) \\
&= \frac{7}{4}\ \text{bits}\,.
\end{aligned} \tag{1.25}$$

Hence Shannon's noiseless coding theorem says that we should be able (on average) to reduce a message of N characters, or $2N$ bits, to 1.75 bits. The key to this reduction is to use short sequences for common elements of the message (here the letter A) and longer ones for the less likely ones (like C and D in our example). Here is a coding scheme that achieves this:

$$\begin{aligned}
A &= 0 \\
B &= 10 \\
C &= 110 \\
D &= 111\,.
\end{aligned} \tag{1.26}$$

A bit of thought will confirm that any message so encoded can be decoded uniquely to recover the original sequence of letters. The average number of bits used to encode a sequence of N letters is then

$$N\left(\frac{1}{2}\times 1 + \frac{1}{4}\times 2 + 2\times\frac{1}{8}\times 3\right) = \frac{7}{4}N = HN\,, \tag{1.27}$$

which is the bound provided by Shannon's theorem. The fact that Shannon's value is reached in this case tells us, moreover, that no better coding is possible: this is the shortest possible length of the message.

1.5.2 Noisy coding theorem

The presence of noise on the communications channel will introduce errors in the received signal. We can combat these errors by introducing some redundancy, indeed this is undoubtedly the reason why language evolved already containing redundancy. As a simple example, let us suppose that any given bit in the message is 'flipped' with probability q and so produces an error. How much redundancy do we need to be able to detect and correct these errors?

Shannon's noisy coding theorem tells us that, on average, we require at least

$$\frac{N_0}{1 - H(q)} \text{ bits} \tag{1.28}$$

to encode, faithfully, one of 2^{N_0} equiprobable messages. Here

$$H(q) = -\left[q \log q + (1 - q) \log(1 - q)\right] \tag{1.29}$$

is the entropy associated with the single-bit error probability. In other words if we first remove all the redundancy to get 2^{N_0} possible optimally compressed messages, we need to *put back* this much redundancy to combat errors.

The general statement is based on the mutual information. It says that the greatest number of messages that can be sent from Alice to Bob on a noisy channel, using N bits, *and be reconstructed by Bob* is

$$2^{NH(A:B)}\,. \tag{1.30}$$

Any more is impossible in that an attempt to do so will inevitably produce ambiguities in Bob's decoding process.

We conclude with a simple illustration of the principle of using redundancy to combat errors. Try to read the following message:

WNTM NARMQN THRS S FN

You probably didn't manage to do so. The reason for this is that I first compressed the message by removing the vowels and then added in errors. Because much of the redundancy was removed, the message has become unreadable. If I had left the full message (complete with vowels) and then added the errors, we might have:

WUANTFM INAORMAQION THEORS US FUN

Hopefully, after a bit of thought, you should be able to read the message[2]. Note that decoding the message is possible even though the errors affect both the characters in the compressed message and in the redundant characters added to combat the errors.

[2] If you are struggling, the message is qUANTuM INfORMAtION THEORy iS FUN.

2

Quantum Communications and Quantum Key Distribution

Quantum communications differ from their classical counterpart in that the transmitted signal is carried by a quantum system. At its simplest, classical information is expressed in *bits*; and is carried by a physical system with two distinct states (for example a high or low voltage or the presence or absence of an optical pulse). One state is used to represent the logical 0 and the other a logical 1.

We can take the same approach in quantum communications and use two orthogonal quantum states to represent 0 and 1. We label these, naturally enough, as $|0\rangle$ and $|1\rangle$. The additional element brought in by using a two-state *quantum* system is that we can also prepare any superposition state of the form

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

where α and β are complex probability amplitudes. A quantum bit, or *qubit*, is such a two-state quantum system.

2.1 Qubits

Before getting into the theory of quantum communications, we pause to elaborate on the idea of a qubit and the mathematical tools used to describe it. A qubit can be any quantum system with two orthogonal states. We choose a basis, often called the computational basis in quantum information, and label the two states in this basis as $|0\rangle$ and $|1\rangle$. It is convenient to think of this as a spin-half particle; this is not literally true in most cases, but it has the benefit that we can use the Pauli-operators to describe the properties of the qubit. There are four Pauli operators, which are $\hat{\sigma}_z$, $\hat{\sigma}_x$, $\hat{\sigma}_y$ and the identity operator $\hat{\mathrm{I}}$. We can define each of these by their action on the states in the computational basis:

$$\begin{aligned}
\hat{\sigma}_z|0\rangle &= |0\rangle & \qquad \hat{\sigma}_z|1\rangle &= -|1\rangle \\
\hat{\sigma}_x|0\rangle &= |1\rangle & \qquad \hat{\sigma}_x|1\rangle &= |0\rangle \\
\hat{\sigma}_y|0\rangle &= i|1\rangle & \qquad \hat{\sigma}_y|1\rangle &= -i|0\rangle \\
\hat{\mathrm{I}}|0\rangle &= |0\rangle & \qquad \hat{\mathrm{I}}|1\rangle &= |1\rangle\,.
\end{aligned} \tag{2.2}$$

The first three Pauli operators do not mutually commute. We can write the product rule for these operators in an appealingly simple form if we introduce, by means of a scalar product, the Pauli operator for an arbitrary direction associated with the unit vector **a**:

$$\mathbf{a}\cdot\hat{\boldsymbol{\sigma}} = a_x\hat{\sigma}_x + a_y\hat{\sigma}_y + a_z\hat{\sigma}_z\,. \tag{2.3}$$

The product rule for two such operators is then

$$(\mathbf{a}\cdot\hat{\boldsymbol{\sigma}})(\mathbf{b}\cdot\hat{\boldsymbol{\sigma}}) = (\mathbf{a}\cdot\mathbf{b})\hat{\mathrm{I}} + i(\mathbf{a}\times\mathbf{b})\cdot\hat{\boldsymbol{\sigma}}\,. \tag{2.4}$$

as may readily be verified.

2.2 Information security

The use of a quantum communications channel makes three essential modifications or additions to Shannon's model. The first is that the signal sent by Alice is encoded on the quantum state of a physical system sent to Bob. This means that each signal a_i is associated with a quantum state with corresponding density operator $\hat{\rho}_i$. In addition to any intrinsic noise in the channel, we have a special role for any eavesdropper who might be listening in. This is because in order to extract any information, the eavesdropper needs to perform a measurement and, as we know, a measurement will, in general, modify the state of the observed system. Finally, Bob cannot determine the state of the system but rather has to settle for measuring one observable. Hence Bob must choose what to measure. All of these elements are essential in quantum communications and it is the combination that, in particular, makes quantum key distribution possible. Before I get to that, let us set the scene by looking at secure communications is general.

In the information age we are all aware of the importance and difficulty of keeping information secure. The science that today underpins this effort is cryptography (Piper and Murphy 2002). The history of secure communications and keeping information secure, however, is a long and interesting one (Singh 1999, 2000). For important communications we might try enciphering a message to keep it safe. The simplest such scheme, at least conceptually, is the single-key cryptosystem. The principal idea is that Alice and Bob share a secret key (a number) which Alice can use to generate the cipher-text and Bob can use to decipher it. The general scheme is depicted in Fig.2.1.

We can write the transformation, formally, involving the plaintext $\mathcal{P}$, the key $\mathcal{K}$ and the ciphertext $\mathcal{C}$. The ciphertext is a function of the plaintext and the key, to be calculated by Alice, and the plaintext may be recovered by Bob as a function of the ciphertext and the key:

$$\begin{aligned}\mathcal{C} &= \mathcal{C}(\mathcal{P},\mathcal{K})\\ \mathcal{P} &= \mathcal{P}(\mathcal{C},\mathcal{K})\,.\end{aligned} \tag{2.5}$$

For example in a substitution cipher we replace each letter with a symbol. In principle there are

$$26! \approx 4\times10^{26}$$

possible substitution ciphers, so an exhaustive search is completely impractical. A substitution cipher is easily cracked, however, using the known letter frequencies. For

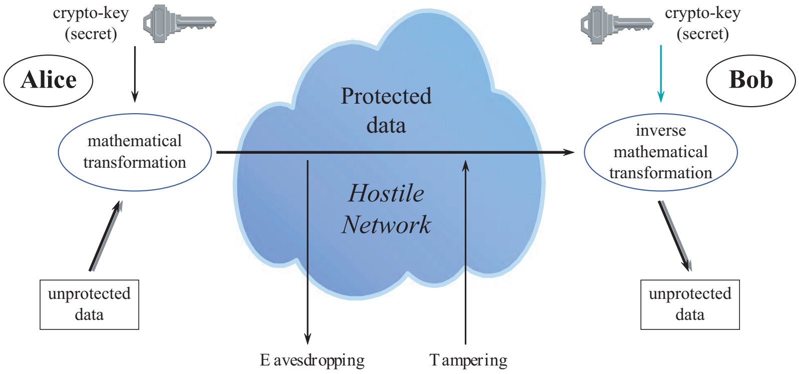


Fig. 2.1 Elements of a secret communications channel. Reproduced, with permission, from Barnett(2009).

example the most common symbol will be E which occurs 12.7% of the time, while Q makes up only about 0.1% of the symbols (or perhaps a bit more in these notes!). Sherlock Holmes makes use of precisely this technique in one of his cases (Conan Doyle 1903).

It was Shannon who gave us an objective criterion for *perfect secrecy* (Shannon 1949). Let $\{p_i\}$ be the set of possible plaintexts (messages) and $\{c_j\}$ be the set of possible cipher texts. Shannon's criterion for perfect secrecy is

$$P(p_i|c_j) = P(p_i) \qquad \forall\ i, j\,. \tag{2.6}$$

A straightforward application of Bayes' theorem shows that this is equivalent to

$$P(c_j|p_i) = P(c_j) \qquad \forall\ i, j\,. \tag{2.7}$$

This means that discovering or intercepting the ciphertext does not provide any information on the plaintext. The second condition states that any given ciphertext is equally likely to have been generated by any plaintext. A question that you might like to ponder is how many possible keys does this require?

The simplest perfectly secure cipher is the Vernam cipher, or one-time pad. It uses a key in the form of a *random* sequence of bits, $\cdots 101110 \cdots$, that is at least as long as the plaintext (which is also a sequence of bits, of course). The ciphertext is produces by bit-wise modulo addition of the plaintext and the key:

$$\begin{aligned} 0 \oplus 0 &= 0 \qquad & 0 \oplus 1 &= 1 \\ 1 \oplus 0 &= 1 \qquad & 1 \oplus 1 &= 0\,. \end{aligned} \tag{2.8}$$

A simple example is

$$\begin{array}{ll} \mathcal{P} & \cdots 0011010100 \cdots \\ \mathcal{K} & \cdots 1011101000 \cdots \end{array}$$

$$\mathcal{C} = \mathcal{P} \oplus \mathcal{K} \qquad \cdots 1000111100 \cdots .$$

The cipertext $\mathcal{C}$ is *random* because the key $\mathcal{K}$ is random. Deciphering is performed by Bob by a second modulo addition of the key. This works because $\mathcal{K} \oplus \mathcal{K} = \cdots 0000000 \cdots$:

$$\begin{aligned} \mathcal{C} &\qquad \cdots 1000111100 \cdots \\ \mathcal{K} &\qquad \cdots 1011101000 \cdots \\ \mathcal{P} = \mathcal{C} \oplus \mathcal{K} &\qquad \cdots 0011010100 \cdots . \end{aligned}$$

Clearly the secrecy of the key is crucial, as anyone in possession of it can decipher the message.

The remaining problem, of course, is how can Alice and Bob exchange the key $\mathcal{K}$? To see how this and other secure communications are realised in practice we need to introduce *public-key cryptography*, which is based on the fact that some mathematical operations are *easy* but the inverse inverse operation is *very difficult* and, hopefully, effectively impossible in practice (Buchmann 2001, Loepp and Wootters 2006, Barnett 2009). The classic example is multiplying and factoring.

In public-key cryptography Bob generates *two keys*, an enciphering key e and a deciphering key d. He publishes e but keeps d secret. Alice can use e to encode her message which should be all but impossible (she hopes!) for anyone other than Bob to decipher. This is the RSA cryptosystem (Buchmann 2001, Loepp and Wootters 2006, Barnett 2009).

RSA scheme

1. Bob finds two large prime numbers, p and q, and calculates $m = p \times q$. This is easy.

2. He then finds two numbers e and d such that $de = 1$ modulo $(p-1)(q-1)$. There are efficient algorithms for doing this if you know p and q.

3. The *public* key is m and e. The *private* key is d.

4. Alice takes her message, which is a number x, and turns it into a ciphertext by the transformation

$$x \rightarrow x^e \text{ modulo } m$$

5. By the wonders of pure mathematics (actually it is not so hard to prove this)

$$(x^e \text{modulo } m)^d \text{ modulo } m = x \text{ moludo } m\,,$$

which is the original message. So Bob can recover the message using his secret key d.

The security of the RSA scheme is believed to be equivalent to the difficulty of factoring m into its constituent primes p and q. It is certainly true that if p and q are known then finding d from e and m is straightforward.

How big does m have to be? Numbers of order 10^{90} can be factored in less than a day, so much larger numbers are needed. There is an RSA factoring challenge, which you might like to try. The number RSA-212, which is $\approx 7 \times 10^{212}$ was factored in July 2012 to claim a \$30,000 prize. There is currently a \$75,000 prize to anyone who can produce the two prime factors of the 270 decimal digit number (http://en.wikipedia.org/wiki/RSA accessed October 21 (2013)):

> RSA-896 = 412023436986659543855531365332575948179811699844327982845456
> 2643387644556524842619809887042316184187926142024718886949256093177637503
> 3342113098239748515094490910691026986103186270411488086697056490290365365
> 88674337317208131041051908642547932826013912576240339463732693 91

Public key cryptography is routinely used to distribute keys and for proof of identity via so-called digital signatures.

2.3 Quantum copying?

Quantum key distribution is a radically different approach to secure communications. It relies on the difficulty for anyone eavesdropping in determining the quantum signal generated by Alice. Before describing this in detail, we show that it is impossible to copy an *unknown* quantum state. This is the famous *no-cloning theorem* of Wootters and Zurek (1982) and of Dieks (1982). It works by exposing a contradiction.

Suppose that you are given a qubit in some unknown (to you) quantum state:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle\,, \tag{2.9}$$

that is, you do not know the amplitudes α and β. Clearly you could do a measurement but that will not tell you $|\psi\rangle$. What you would like to achieve is a transformation on this quit and a 'blank' prepared in the state $|B\rangle$ such that

$$|\psi\rangle \otimes |B\rangle \to |\psi\rangle \otimes |\psi\rangle \qquad \forall\ |\psi\rangle\,. \tag{2.10}$$

Let us suppose that this works if $|\psi\rangle = |0\rangle$ or $|1\rangle$, so that

$$\begin{aligned} |0\rangle \otimes |B\rangle &\to |0\rangle \otimes |0\rangle \\ |1\rangle \otimes |B\rangle &\to |1\rangle \otimes |1\rangle\,. \end{aligned} \tag{2.11}$$

The superposition principle then tells us that for a general state $|\psi\rangle$ the corresponding transformation is

$$\begin{aligned} |\psi\rangle \otimes |B\rangle &= (\alpha|0\rangle + \beta|1\rangle) \otimes |B\rangle \\ &\to \alpha|0\rangle \otimes |0\rangle + \beta|1\rangle \otimes |1\rangle \\ &\neq (\alpha|0\rangle + \beta|1\rangle) \otimes (\alpha|0\rangle + \beta|1\rangle)\,. \end{aligned} \tag{2.12}$$

So it necessarily follows from the superposition principle that a quantum cloner cannot work perfectly for all quantum states. Having established that strictly exact cloning of an unknown state is not possible, we might very reasonably ask what is the best that is possible. There has been a great deal of work on this topic, but probably the most important is the universal quantum-copying machine the operation of which is the best possible with the same performance for all possible input qubit states (Bužek and Hillery 1996, Scarani *et al* 2005).

2.4 Optical polarization

Recall that for a plane wave, the electric field $\vec{E}$, the magnetic field $\vec{H}$ and the wavevector $\vec{k}$ are all mutually orthogonal and are oriented such that $\vec{E} \times \vec{H}$ points in the direction of $\vec{k}$ see Fig. 2.2 . This fixes $\vec{E}$ and $\vec{H}$ to be in the plane perpendicular to $\vec{k}$.

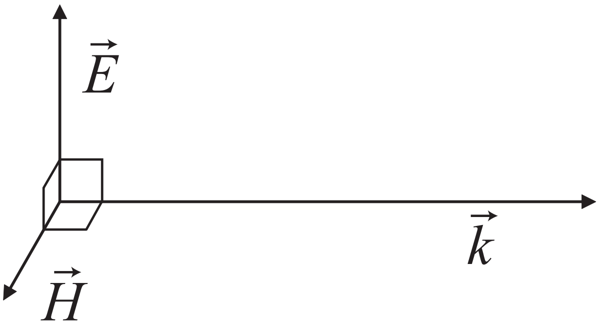


Fig. 2.2 Relative orientations of the electric and magnetic fields and the wavevector. Reproduced, with permission, from Barnett(2009).

Consider a complex electric field for a plane wave propagating in the z-direction:

$$\vec{E} = \vec{E}_0 e^{i(kz-\omega t)} \,. \tag{2.13}$$

We can characterise the polarization by the direction of the complex vector $\vec{E}_0$ in the $x-y$ plane

$$\vec{E}_0 = E_{0x}\vec{\imath} + E_{0y}\vec{\jmath} \tag{2.14}$$

or as a column vector, the Jones vector,

$$\vec{E}_0 = \begin{bmatrix} E_{0x} \\ E_{0y} \end{bmatrix} \,. \tag{2.15}$$

Only the relative sizes of E_{0x} and E_{0y} matter, so we can use a normalized Jones vector. The Jones vectors corresponding to horizontal/vertical, diagonal and circular polarizations are depicted in Fig. 2.3.

For a single photon we can associate each of the Jones vectors with a qubit state according to the mapping

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \to \alpha|0\rangle + \beta|1\rangle \,, \tag{2.16}$$

which leads to the identifications listed in Fig. 2.4. This simple idea has been used widely in experimental demonstrations of a variety of quantum information and communications protocols, most especially in quantum cryptography.

2.5 Quantum cryptography

Quantum cryptography, or perhaps more precisely quantum key distribution, has become an advanced experimental technique and is on the verge, perhaps, of becoming a

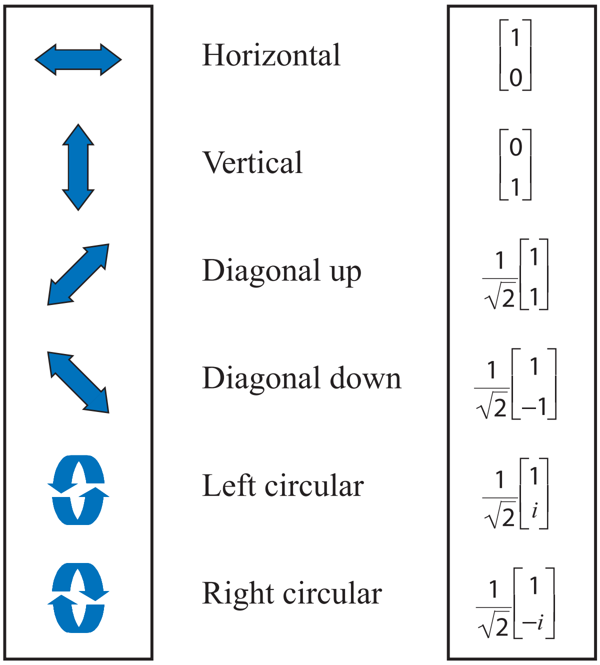

Horizontal
$\begin{bmatrix}1\\0\end{bmatrix}$
Vertical
$\begin{bmatrix}0\\1\end{bmatrix}$
Diagonal up
$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$
Diagonal down
$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}$
Left circular
$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\i\end{bmatrix}$
Right circular
$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-i\end{bmatrix}$

Fig. 2.3 Linear and circular polarizations, together with their Jones vectors. Reproduced, with permission, from Barnett(2009).

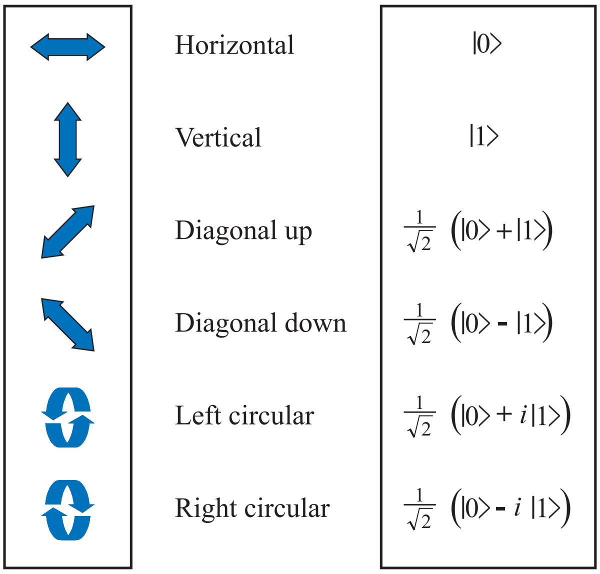

Horizontal
$|0\rangle$
Vertical
$|1\rangle$
Diagonal up
$\frac{1}{\sqrt{2}}\left(|0\rangle+|1\rangle\right)$
Diagonal down
$\frac{1}{\sqrt{2}}\left(|0\rangle-|1\rangle\right)$
Left circular
$\frac{1}{\sqrt{2}}\left(|0\rangle+i|1\rangle\right)$
Right circular
$\frac{1}{\sqrt{2}}\left(|0\rangle-i|1\rangle\right)$

Fig. 2.4 Linear and circular polarizations for single photons, together with their associated qubit states. Reproduced, with permission, from Barnett(2009).

practical technology. In these lectures I can give only a very brief overview of the topic and, for a more complete description, refer the reader to some of the introductory literature on the subject (Phoenix and Townsend 1995, Bouwmeester *et al* 2000, Gisin *et al* 2002, Loepp and Wootters 2006, Van Assche 2006, Barnett 2009).

The idea of using quantum effects for security is due to Stephen Wiesner (1983) who suggested (in a paper written in about 1970 but published only much later) the idea of unforgeable money. He supposed a high-value banknote worth, say, $\pounds 10M$. Each 'note' would have traps for 20 single photons, each of which was prepared in one of the polarization states $|0\rangle$, $|1\rangle$, $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$. The note also has a serial number which identifies to the back the states of these photon polarizations. The problem face by a would-be forger is how to copy these unknown quantum states? The no-cloning theorem tells us, of course that this is impossible. The situation is illustrated in Fig. 2.5. The counterfeiter needs to work out the polarization of each of the trapped photons and do so only by performing measurements, but he/she has no idea which measurement to perform. Let us focus our attention on one of the photons that happens to be prepared in a state of vertical polarization. If the counterfeiter measures this in the horizontal/vertical basis then the result will be a bank note with the correct horizontally polarised photon. If, however, the counterfeiter measures in the diagonal basis then either of the two possible results will occur with equal probability and, importantly, there is no indication possible that the measurement has been performed in the incorrect basis. If the polarization is checked in the bank then an error will be produced with a probability $\frac{1}{2}$. It follows that for a bank note produced in this way, anyone checking in the bank will see an error (indicating a forgery) with a probability of $\frac{1}{4}$ for *each* of the twenty photons. The probability that all 20 photons in the forged note pass a test in the bank is $\left(\frac{3}{4}\right)^{20} \approx 0.003$. If this isn't considered low enough, then one simply has to add a few more light traps.

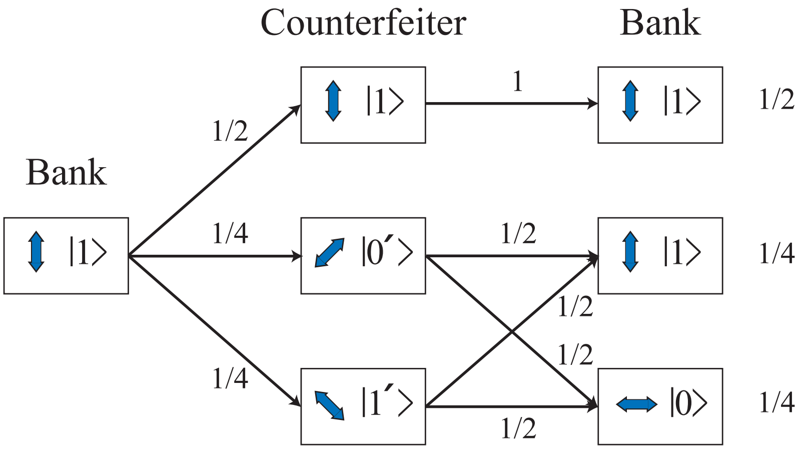


Fig. 2.5 Possible outcomes due to the intervention of a counterfeiter, together with their associated probabilities of occurrence. Reproduced, with permission, from Barnett(2009).

Quantum money is impractical but Bennett and Brassard (1984) proposed a variant on this idea: quantum key distribution. The idea is that Alice associates the bit values 0 and 1 with a linear polarization state (either horizontal/vertical or diagonal/anti-diagonal):

$$\begin{aligned} H \text{ or } D &\longleftrightarrow 0 \\ V \text{ or } A &\longleftrightarrow 1\,. \end{aligned} \tag{2.17}$$

Suppose that Alice prepares a vertically polarised photon and that Bob measures the state in the horizontal/vertical basis. An eavesdropper, Eve, does not know that the photon was prepared in this basis and so can only make a choice of what to measure. In this way she will introduce, for $\frac{1}{4}$ of the photons intercepted, a disagreement between the values obtained by Alice and by Bob. This 25% error rate indicates to Alice and to Bob the activity of the eavesdropper.

All that is needed is a sequence of instructions (a protocol) to ensure that Eve is trapped in this way. One such protocol is depicted in Fig. 2.6. Alice prepares a random sequence of bits and randomly encodes each onto the polarization as either a horizontally/vertically or diagonally/anti-diagonally polarised photon, which she sends to Bob. For each photon, Bob randomly (and independently of Alice, of course) chooses to make a polarization measurement in one of the two bases. Roughly half the time he will guess correctly and half incorrectly. Bob then uses a classical channel to tell Alice which basis he used in each time slot for which he detected a photon (but not, of course, the measured value). Alice tells Bob the slots in which he made the correct choice (time-slots 2, 4, 6, 8, 9, 11, 14, 15, 16, and 19, in the case depicted in the Fig. 2.6). These should comprise a secret shared random bit string (0111010100 in this case) which can be used as a secret key.

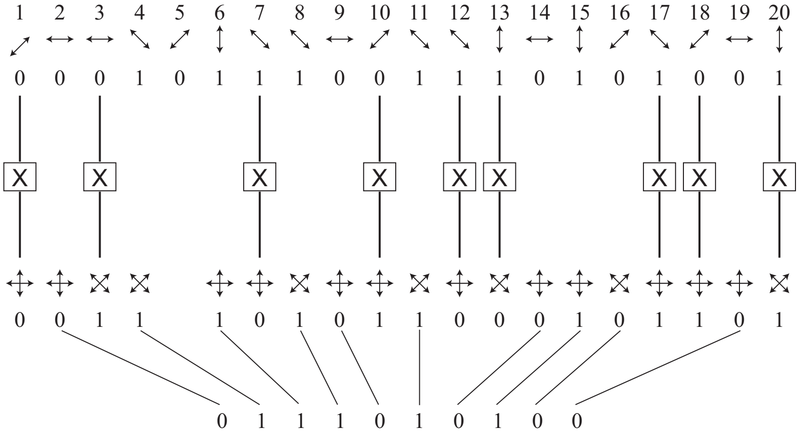
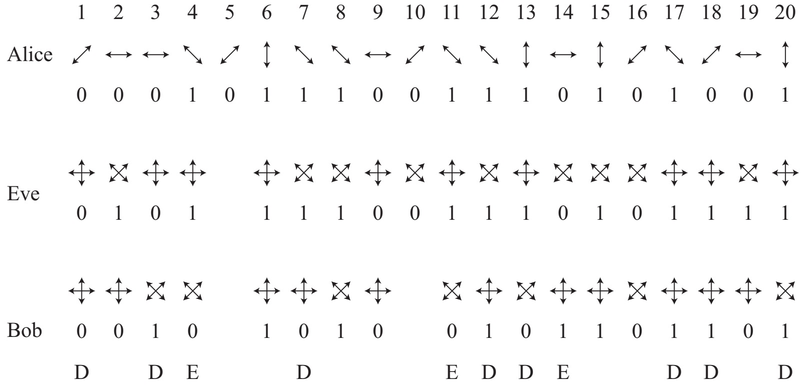


Fig. 2.6 An example of a quantum cryptographic protocol. Alice prepares single photons in one of two polarization bases and these are measured by Bob. Reproduced, with permission, from Barnett(2009).

It remains only to determine whether an eavesdropper has been listening in. To determine whether this has happened or not, Alice and Bob can publicly reveal a subset of their shared data and look for errors. An example of what might be expected to happen is depicted in Fig. 2.7. Alice and Bob delete those bits associated with time-bins 1, 3, 7, 12, 13, 17, 18 and 20, in which the preparation and measurement were performed using different bases. In time slots 2, 4, 11, 14, 15, and 19, Eve used a different basis to Alice and Bob and of these, this led to a difference in the recorded bit values for Alice and Bob in time-bins 4, 11 and 14. The probability that Eve has been active and does not cause an error is then $\left(\frac{3}{4}\right)^N$, where N is the number of bits tested. Naturally, those bits used in the open discussion between Alice and Bob need to be discarded.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Alice	⤢	↔	↔	⤡	⤢	↕	⤡	⤡	↔	⤢	⤡	⤡	↕	↔	↕	⤢	⤡	⤢	↔	↕
	0	0	0	1	0	1	1	1	0	0	1	1	1	0	1	0	1	0	0	1
	✛	✕	✛	✛		✛	✕	✕	✛	✕	✛	✕	✛	✕	✕	✕	✛	✛	✕	✛
Eve	0	1	0	1		1	1	1	0	0	1	1	1	0	1	0	1	1	1	1
	✛	✛	✕	✕		✛	✛	✕	✛		✕	✛	✕	✛	✛	✕	✛	✛	✛	✕
Bob	0	0	1	0		1	0	1	0		0	1	0	1	1	0	1	1	0	1
	D		D	E			D				E	D	D	E			D	D		D

Fig. 2.7 An example of a quantum cryptographic protocol in which an eavesdropper has been active. Reproduced, with permission, from Barnett(2009).

Naturally, there are many subtleties that need to be accounted for and which complicate, to a greater or lesser extent, the above simple picture. There will always be errors and, to be safe, we need to assume that these are due to eavesdropper activity. If we simply discard the communication when we find an error then communication of a key of any useful size becomes impossible. We need to detect and correct errors, which is achieved by parity checks which shorten the key. The error rate is then used to place a bound on the information an eavesdropper might have and this can then be reduced by taking as key bits the parity of a number of bits (this is called privacy amplification). Practical systems, moreover, may not have access to single-photon sources and there will usually be other departures from the ideal. Proving and assessing the security of real-world systems has become a research topic in its own right (Scarani *et al* 2009).

3 Generalized Measurements

The extraction of information from a quantum system requires us to perform measurements. Our task in this lecture is to set up a general description of this process[1]. We seek two things: (i) the probability for any given measurement result to occur and (ii) the state of the system after the measurement has been made, that is the post-measurement state conditioned on the measurement outcome.

3.1 Ideal von Neumann measurements

Let us start with the measurement process as it is usually encountered in quantum mechanics courses. This formulation is essentially that given by von Neumann in his famous and early book on quantum mechanics (von Neumann 1955). We start by representing each observable A by a Hermitian operator[2], $\hat{A}$. This operator will have a complete set of eigenvectors $|\lambda_n\rangle$ and associated eigenvalues λ_n:

$$\hat{A}|\lambda_n\rangle = \lambda_n|\lambda_n\rangle\,, \tag{3.1}$$

which means that we can write $\hat{A}$ in the form

$$\hat{A} = \sum_n \lambda_n|\lambda_n\rangle\langle\lambda_n|\,. \tag{3.2}$$

Let us assume, for the moment, that each of the eigenvalues is distinct from the others. The von Neumann description then states that if we perform a measurement of $\hat{A}$ then we will find the measurement result to be one of the eigenvalues and, moreover, the probability for finding any one of these is

$$P(\lambda_n) = |\langle\lambda_n|\psi\rangle|^2\,, \tag{3.3}$$

where $|\psi\rangle$ is the pre-measurement state. More generally, for a mixed state with density operator $\hat{\rho}$, we have

$$\begin{aligned} P(\lambda_n) &= \langle\lambda_n|\hat{\rho}|\lambda_n\rangle \\ &= \mathrm{Tr}\left(\hat{\rho}|\lambda_n\rangle\langle\lambda_n|\right)\,. \end{aligned} \tag{3.4}$$

Immediately after the measurement, the von Neumann description has the system left in the eigenstate corresponding to the measurement outcome. Hence if we make a

[1] We should note that we do not include in this, measurements involving post-selection and so will not cover the topic of 'weak measurements' (Aharonov *et al* 1988).

[2] The distinction between Hermitian and self-adjoint operators will not concern us.

measurement of $\hat{A}$ and find the value λ_n, then we know that the post-measurement state is $|\lambda_n\rangle$ and repeating the measurement, if it is done quickly enough, should give the same result.

There is one further point that we need to consider, which is that the eigenvalues of $\hat{A}$ might be degenerate, which means that a set of orthonormal eigenvectors, $|\lambda_n^j\rangle$, will correspond to the same measurement outcome, λ_n. To incorporate this case, it is useful to introduce a projector onto the set of states with a common eigenvalue:

$$\hat{P}_n = \sum_j |\lambda_n^j\rangle\langle\lambda_n^j| \,. \tag{3.5}$$

The probability that the measurement will give the result λ_n is then

$$\begin{aligned} P(\lambda_n) &= \sum_j \langle\lambda_n^j|\hat{\rho}|\lambda_n^j\rangle \\ &= \mathrm{Tr}\left(\hat{\rho}\hat{P}_n\right) \,. \end{aligned} \tag{3.6}$$

The post-measurement state will simply be that part of the pre-measurement state that was in the subspace spanned by the corresponding eigenstates or, in other words, that part of the state selected by the projector. Hence if our measurement gives the result λ_n then the density matrix changes as

$$\hat{\rho} \to \frac{\hat{P}_n\hat{\rho}\hat{P}_n}{\mathrm{Tr}(\hat{\rho}\hat{P}_n)} \,, \tag{3.7}$$

where the denominator, which is the prior probability for the observed measurement result, ensures correct normalisation of the post-measurement state.

Let us conclude this brief review of von Neumann measurements with a summary of the properties of projectors.

Properties of projectors

1. They are Hermitian, $\hat{P}_n = \hat{P}_n^\dagger$.
2. They are positive operators, $\hat{P}_n \geq \hat{\mathrm{I}}$.
3. They are complete, $\sum_n \hat{P}_n = \hat{\mathrm{I}}$.
4. They are orthonormal, $\hat{P}_n\hat{P}_m = \hat{P}_n\delta_{nm}$.

Here $\hat{\mathrm{I}}$ is the identity operator. We note that the first three of these properties have physical significance in that they are required in order that the probability rule, $P(\lambda_n) = \mathrm{Tr}\left(\hat{\rho}\hat{P}_n\right)$, be true. They correspond, respectively, to the requirements that the projectors are observables, that they give positive probabilities and that the probabilities for the complete set of possible outcomes must sum to unity. The final property, however, does not seem to have a similar physical significance and, indeed, we shall see that generalised measurements correspond to dropping this requirement.

3.2 Non-ideal measurements

Most measurements are not of the form described in the preceding section. Consider, for example, the operation of a photodetector. Real photodetectors have a finite efficiency and so do note resolve perfectly the number of photons. In detecting the photons, moreover, the detector absorbs the light and so leaves the field in its zero-photon (or vacuum) state. Hence neither the von Neumann forms for the detection probabilities nor for the post-measurement state are appropriate and something more general is required.

As a simple example, both of the problem and of how we might proceed, let us consider a measurement of a spin-component for a single quit. The ideal measurement would correspond to the pair of projectors

$$\begin{aligned}\hat{P}_0 &= |0\rangle\langle 0| \\ \hat{P}_1 &= |1\rangle\langle 1| \,.\end{aligned} \tag{3.8}$$

Suppose that our detector gives the *wrong* answer with probability p so that the two possible measurement results occur with probabilities

$$\begin{aligned}P(0) &= (1-p)\mathrm{Tr}\left(\hat{\rho}\hat{P}_0\right) + p\mathrm{Tr}\left(\hat{\rho}\hat{P}_1\right) \\ P(1) &= (1-p)\mathrm{Tr}\left(\hat{\rho}\hat{P}_1\right) + p\mathrm{Tr}\left(\hat{\rho}\hat{P}_0\right) \,,\end{aligned} \tag{3.9}$$

the sum of which is clearly unity, as it should be. We can write these in a form reminiscent of the von Neumann expressions,

$$\begin{aligned}P(0) &= \mathrm{Tr}\left(\hat{\rho}\hat{\pi}_0\right) \\ P(1) &= \mathrm{Tr}\left(\hat{\rho}\hat{\pi}_0\right) \,,\end{aligned} \tag{3.10}$$

by introducing the probability operators

$$\begin{aligned}\hat{\pi}_0 &= (1-p)\hat{P}_0 + p\hat{P}_1 \\ \hat{\pi}_1 &= (1-p)\hat{P}_1 + p\hat{P}_0 \,.\end{aligned} \tag{3.11}$$

These operators satisfy the first three properties that we listed for the projectors, they are Hermitian, positive and complete. They are *not* orthonormal, however:

$$\hat{\pi}_0\hat{\pi}_1 = p(1-p)\hat{\mathrm{I}} \,. \tag{3.12}$$

This is our first indication of a more general description of the measurement process, which we develop further below.

3.3 Probability operator measures

To calculate the portability for any given result of a generalized measurement we need a probability rule and, in particular, a set of operators that characterise the measurement, one for each of the possible measurement outcomes. The required operators are the probability operators, the set of which is a probability operator measure or a positive operator-valued measure (Helstrom 1976, Holevo 1982, 2001, Peres 1993, Barnett 2009).

That the first three of the properties we listed for the projectors have physical significance, but the third does not, tells us in what way we are allowed to generalise the von Neumann description; we simply drop the fourth and final property from our list. Hence we describe any measurement by a set of probability operators $\{\hat{\pi}_m\}$ such that the probability for getting measurement outcome m is

$$P_m = \mathrm{Tr}\left(\hat{\rho}\hat{\pi}_m\right) , \tag{3.13}$$

where the probability operators have the following *three* properties

Properties of probability operators

1. They are Hermitian, $\hat{\pi}_n = \hat{\pi}_n^\dagger$.
2. They are positive operators, $\hat{\pi}_n \geq \hat{\mathrm{I}}$.
3. They are complete, $\sum_n \hat{\pi}_n = \hat{\mathrm{I}}$.

We should emphasise that there is no restriction on the number of probability operators; the number can be greater or less than the dimension of the state space. For example, we shall analyse an example in which a generalised measurement on a qubit which has three outcomes, even though the dimension of the state space is only 2. There is no need, moreover, for the probability operators to commute with each other.

The set of probability operators characterising a measurement is called a probability operator measure (or POM) which you will also find called a positive operator-valued measure (POVM). The latter has become the commonly used expression, but I prefer the former[3]. The description of a measurement in terms of a POM is very useful because of two points:

1. All measurements can be described in this way. (This is at least reasonable given the physical significance of the three properties used to define a POM.)

2. Any set of the probability operators (that is any POM) is realisable, at least in principle, in the laboratory.

This combination is very useful because if we seek to find the best possible measurement in any situation, we can separate out the purely mathematical optimisation of the POM from the experimental question of how to achieve it. That any POM is realisable as a measurement is a consequence of Naimark's theorem. We shall not prove this here but rather give an indication of how it works. A more complete presentation may be found in (Barnett 2009). The key idea is to map the generalized measurement onto a projective, or von Neumann, measurement in an extended state space. To see

[3] POM or POVM? The term 'probability operator measure' tells us that the elements forming the measure are the probability operators. The more popular expression 'positive operator-valued measure' expresses the fact that the elements of the measure, the probability operators, are positive operators. Calling the set of operators a POM reminds us of their physical significance, while the term POVM recalls their mathematical properties.

how this might be done, consider as quantum system s, which we wish to measure, and let us prepare also an ancillary quantum system a, so that we know its initial state. The state of the combined system and ancilla is then $|\psi\rangle_s \otimes |A\rangle_a$. Next we engineer an interaction between the system and the ancilla, so that a unitary transformation of our choosing occurs:

$$|\psi\rangle_s \otimes |A\rangle_a \to \hat{U}|\psi\rangle_s \otimes |A\rangle_a \,. \tag{3.14}$$

Finally, we perform a von Neumann measurement on both the system and the ancilla, so that the probability for a given outcome will be

$$\begin{aligned} P(m,l) &= \left| {}_s\langle m| \otimes {}_a\langle l|\hat{U}|\psi\rangle_s \otimes |A\rangle_a \right|^2 \\ &= {}_s\langle\psi|\hat{\pi}_{ml}|\psi\rangle_s \,, \end{aligned} \tag{3.15}$$

where

$$\hat{\pi}_{ml} = {}_a\langle A|\hat{U}^\dagger|m\rangle_s \otimes |l\rangle_{aa}\langle l| \otimes {}_s\langle m|\hat{U}|A\rangle_a \,, \tag{3.16}$$

which is an operator acting *only* on the system state space. It is clearly Hermitian and positive. That it sums to the identity follows, moreover, from the completeness of the orthonomal states $\{|m\rangle_s\}$ and $\{|l\rangle_a\}$. It follows that the operators $\hat{\pi}_{ml}$ are probability operators.

Let us give a simple but important example of a generalised measurement; the simultaneous measurement of the position and momentum of a particle. This is something that we do all the time in our predominantly classical world but, because position and momentum are incompatible observables, *cannot* be described as a von Neumann measurement. Let us write the probability density for a joint measurement to give a position between x_m and $x_m + dx_m$ and also a momentum between p_m and $p_m + dp_m$ as

$$\mathcal{P}(x_m, p_m) = \mathrm{Tr}\left[\hat{\rho}\hat{\pi}(x_m, p_m)\right] \,, \tag{3.17}$$

where the $\hat{\pi}(x_m, p_m)$ are positive operators satisfying the continuum condition

$$\begin{aligned} &\int_{-\infty}^{\infty} dx_m \int_{-\infty}^{\infty} dp_m \hat{\pi}(x_m, p_m) = \hat{\mathrm{I}} \\ \Rightarrow \quad &\int_{-\infty}^{\infty} dx_m \int_{-\infty}^{\infty} dp_m \mathcal{P}(x_m, p_m) = 1 \,. \end{aligned} \tag{3.18}$$

A good simultaneous measurement of the position and momentum of a body will localise rather well both quantities. In order to produce a plausible measurement operator, let us introduce a Gaussian state:

$$|x_m, p_m\rangle = (2\pi\sigma^2)^{-1/4} \int_{-\infty}^{\infty} dx \exp\left[-\frac{(x - x_m)^2}{4\sigma^2} + i\frac{p_m x}{\hbar}\right] |x\rangle \,, \tag{3.19}$$

where $|x\rangle$ is a position eigenstate. We note that these states form an over-complete set in that

$$\frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} dx_m \int_{-\infty}^{\infty} dp_m |x_m, p_m\rangle\langle x_m, p_m| = \hat{\mathrm{I}} \,. \tag{3.20}$$

Comparing this with our normalisation condition leads us to identify our probability operators as

$$\hat{\pi}(x_m, p_m) = \frac{1}{2\pi\hbar}|x_m, p_m\rangle\langle x_m, p_m| \,. \tag{3.21}$$

It follows, for example, that the probability distribution for the measured position is

$$\mathcal{P}(x_m) = \int_{-\infty}^{\infty} dx\, \langle x|\hat{\rho}|x\rangle \exp\left[-\frac{(x-x_m)^2}{2\sigma^2}\right] \,, \tag{3.22}$$

which is a convolution of the true momentum distribution, $\langle x|\hat{\rho}|x\rangle$, with a Gaussian of width associated with the state $|x_m, p_m\rangle$. A similar expression applies also for the measured momentum distribution. The variances found for the measurement results are

$$\begin{aligned} \mathrm{Var}(x_m) &= \Delta x^2 + \sigma^2 \\ \mathrm{Var}(p_m) &= \Delta p^2 + \frac{\hbar^2}{4\sigma^2} \,. \end{aligned} \tag{3.23}$$

The additional contributions, over and above Δx^2 and Δp^2, are a consequence of the fact that we have performed a simultaneous measurement of two incompatible observables.

3.4 Optimized measurements

We can seek the best possible measurement in any given situation in two stages: first we have the mathematical optimization by finding the best POM and secondly design an experiment that gives the corresponding probabilities. The mathematical optimization is simply a search over all possible sets of probability operators forming a POM. The optimal measurements to perform have been determined in a number of cases and some of these have also been realised in the laboratory (Chefles 2000, Paris and Řeháček 2004, Bergou 2007, Barnett 2009, Barnett and Croke 2009b).

Let us begin by considering a simple example motivated by our treatment of quantum communications. Suppose we have a qubit which we know to have been prepared in one of two non-orthogonal states $|\psi_1\rangle$ or $|\psi_2\rangle$:

$$\langle\psi_1|\psi_2\rangle \neq 0 \,. \tag{3.24}$$

We can use our probability operators to ask whether a measurement exists that will determine which state has been prepared *with certainty.* Not surprisingly, the answer is no, but we can prove this. To proceed, we seek two probability operators, $\hat{\pi}_1$ and $\hat{\pi}_2$, such that

$$\begin{aligned} \langle\psi_1|\hat{\pi}_1|\psi_1\rangle &= 1 = \langle\psi_2|\hat{\pi}_2|\psi_2\rangle \\ \langle\psi_2|\hat{\pi}_1|\psi_2\rangle &= 0 = \langle\psi_1|\hat{\pi}_2|\psi_1\rangle \end{aligned}$$

$$\hat{\pi}_1 + \hat{\pi}_2 = \hat{\mathrm{I}} \,. \tag{3.25}$$

From our first condition, $\langle\psi_1|\hat{\pi}_1|\psi_1\rangle = 1$, we can write

$$\hat{\pi}_1 = |\psi_1\rangle\langle\psi_1| + \hat{A} \,, \tag{3.26}$$

where $\hat{A}$ is a Hermitian, positive operator such that $\langle\psi_1|\hat{A}|\psi_1\rangle = 0$, which implies that $\hat{A}|\psi_1\rangle = 0$. From this it follows that

$$\langle\psi_2|\hat{\pi}_1|\psi_2\rangle = |\langle\psi_2|\psi_1\rangle|^2 + \langle\psi_2|\hat{A}|\psi_2\rangle \tag{3.27}$$

which, being the sum of a positive quantity and a quantity greater than or equal to zero, must be greater than zero. Hence we cannot discriminate between these states perfectly. A natural question then is “what is the best we can do?”. The answer depends, of course, on what we mean by ‘best’. Here we consider only two such strategies: measurement with the minimum probability of error and unambiguous state discrimination.

3.4.1 Maximum probability of being correct

Consider the following task, motivated by our discussion of quantum communications. We are given a quantum system which we know to have been prepared in one of a set of possible states $\{\hat{\rho}_j\}$ and also the probabilities, $\{p_j\}$, that each state was prepared. We seek to identify the state with the maximum probability of being correct or, equivalently, with the minimum value of the error probability. Hence we construct a POM with probability operators $\{\hat{\pi}_j\}$ and associate each measurement outcome, $\hat{\pi}_j$, with the corresponding state $\hat{\rho}_j$ and then minimise the quantity

$$P_{\mathrm{error}} = 1 - \sum_j p_j \mathrm{Tr}\left(\hat{\rho}_j\hat{\pi}_j\right) \,. \tag{3.28}$$

Remarkably, the necessary and sufficient conditions for reaching this minimum are known (Holevo 1973, Helstrom 1976, Barnett 2009):

$$\begin{aligned} \hat{\pi}_j\left(p_j\hat{\rho}_j - p_k\hat{\rho}_k\right)\hat{\pi}_k &= 0 \\ \sum_i p_i\hat{\rho}_i\hat{\pi}_i - p_j\hat{\rho}_j &\geq 0 \,. \end{aligned} \tag{3.29}$$

A simple derivation of these is given in (Barnett and Croke 2009a). These conditions do not provide a direct means to construct the minimum error measurement and, indeed, there are many cases in which the strategy for measuring with minimum probability of error is not unique. Our strategy, therefore, is to try a measurement strategy and if it satisfies these conditions then we know it is optimal. The problem is usually a tractable one where there is some symmetry amongst the set of possible states and the minimum-error strategy has been derived in a variety of cases (Chefles 2000, Paris and Řeháček 2004, Bergou 2007, Barnett 2009, Barnett and Croke 2009b).

If we have just two possible states, $\hat{\rho}_1$ and $\hat{\rho}_1$, then the minimum-error measurement is known: we need only perform a projective (or von Neumann) measurement in which the two measurement outcomes correspond to projectors onto the positive and negative eigenvalues of $p_1\hat{\rho}_1 - p_2\hat{\rho}_2$, corresponding, respectively, to identifying the state as $\hat{\rho}_1$ and $\hat{\rho}_1$. This gives as the minimum error probability the Helstrom bound:

$$P_{\text{error}}^{\min} = \frac{1}{2}\left[1 - \text{Tr}\,|p_1\hat{\rho}_1 - p_2\hat{\rho}_2|\right] . \tag{3.30}$$

It is interesting to note that sometimes a measurement does not help and the minimum-error strategy is simply to guess that the most likely state was prepared (Hunter 2003). As an example, let us suppose that we have a single qubit which we know to have been prepared with equal probability ($p_1 = \frac{1}{2} = p_2$) in one of the two non-orthogonal states

$$\begin{aligned}|\psi_1\rangle &= \cos\theta|0\rangle + \sin\theta|1\rangle\\ |\psi_2\rangle &= \cos\theta|0\rangle - \sin\theta|1\rangle\,,\end{aligned} \tag{3.31}$$

so that the overlap of the states is

$$\langle\psi_1|\psi_2\rangle = \cos 2\theta\,. \tag{3.32}$$

It is straightforward to confirm that the minimum error strategy corresponds to the two probability operators

$$\begin{aligned}\hat{\pi}_1 &= \frac{1}{2}(|0\rangle + |1\rangle)(\langle 0| + \langle 1|) = |\hat{\pi}_1\rangle\langle\hat{\pi}_1|\\ \hat{\pi}_2 &= \frac{1}{2}(|0\rangle - |1\rangle)(\langle 0| - \langle 1|) = |\hat{\pi}_2\rangle\langle\hat{\pi}_2|\,.\end{aligned} \tag{3.33}$$

The arrangement of these states is depicted in Fig. 3.1. We see that the optimal measurement in this case corresponds to projection onto one of a pair of states $|\hat{\pi}_1\rangle$ or $|\hat{\pi}_2\rangle$ which are 'close' to the states $|\psi_1\rangle$ and $|\psi_2\rangle$. The error probability arises from the fact that $|\hat{\pi}_{1(2)}\rangle$ is not perpendicular to $|\psi_{2(1)}\rangle$. A simple experiment has been performed at this limit to discriminate between two states of photon linear polarization (Barnett and Riis 1997).

By no means all optimal measurements are von Neumann measurements. Indeed a generalised measurement is normally required. As a simple illustration, consider the trine ensemble in which a quit is prepared (with equal probability) in one of the three states

$$\begin{aligned}|\psi_1\rangle &= |0\rangle\\ |\psi_2\rangle &= \frac{1}{2}\left(-|0\rangle + \sqrt{3}|1\rangle\right)\\ |\psi_3\rangle &= \frac{1}{2}\left(|0\rangle + \sqrt{3}|1\rangle\right)\,.\end{aligned} \tag{3.34}$$

These correspond to states separated by 120° on a great circle of the Bloch sphere, as depicted in Fig. 3.2. The minimum-error strategy in this case corresponds to the three probability operators

$$\hat{\pi}_i = \frac{2}{3}|\psi_i\rangle\langle\psi_i| \tag{3.35}$$

and gives an error probability of $\frac{1}{3}$ so that this measurement strategy determines the correct state with probability $\frac{2}{3}$. The required generalized measurement has been

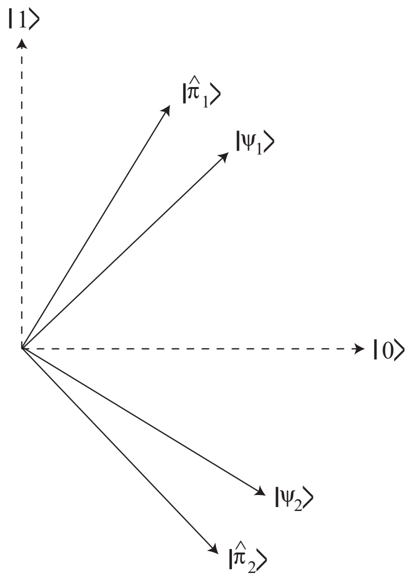


Fig. 3.1 Orientation in state space of the two non-orthogonal states to be discriminated and the directions along which it is best to measure. Reproduced, with permission, from Barnett(2009).

demonstrated using photon polarization qubits; the device is an interferometer in which the path taken through the device provides the ancillary degree of freedom (Clarke *et al* 2001b).

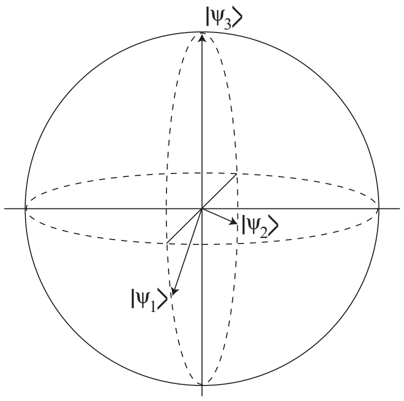


Fig. 3.2 The trine states depicted on the Bloch sphere. Reproduced, with permission, from Barnett(2009).

3.4.2 Unambiguous discrimination

We have found the minimum-error strategy for discriminating between a pair of equiprobable qubit states, $|\psi_1\rangle$ and $|\psi_2\rangle$, but can we determine the state *for certain*? This might sound like a contradiction but it is not if we allow for the possibility of an ambiguous measurement outcome. To see this, consider the von Neumann measurement in which the projectors are

$$\begin{aligned}\hat{P}_1 &= |\psi_1\rangle\langle\psi_1| \\ \hat{P}_{\bar{1}} &= |\psi_1^\perp\rangle\langle\psi_1^\perp| \,,\end{aligned} \tag{3.36}$$

where $|\psi_1^\perp\rangle$ is the state orthogonal to $|\psi_1\rangle$. If we get the result $\bar{1}$ then we know for certain that the state was not $|\psi_1\rangle$ and hence that it *must* have been $|\psi_2\rangle$. If we get the result 1, however, then the state could have been either $|\psi_1\rangle$ or $|\psi_2\rangle$ and the outcome is ambiguous. Hence this simple measurement strategy determines the state unambiguously but only some of the time.

It is natural to ask if we can find an optimal unambiguous discrimination strategy (Ivanovic 1987, Dieks 1988, Peres 1988, Chefles 2000, Barnett 2009, Barnett and Croke 2009b). By this we mean a measurement strategy that identifies the state unambiguously with the highest probability or, equivalently, produces the smallest probability for an ambiguous outcome. The optimal strategy in this case is the three-element POM with probability operators

$$\begin{aligned}\hat{\pi}_1 &= \frac{1}{1+|\langle\psi_1|\psi_2\rangle|}|\psi_2^\perp\rangle\langle\psi_2^\perp| \\ \hat{\pi}_2 &= \frac{1}{1+|\langle\psi_1|\psi_2\rangle|}|\psi_1^\perp\rangle\langle\psi_1^\perp| \\ \hat{\pi}_3 &= \hat{\mathrm{I}} - \hat{\pi}_1 - \hat{\pi}_2 \,.\end{aligned} \tag{3.37}$$

Note that the first two probability operators, $\hat{\pi}_{1(2)}$, are proportional to projectors onto the states *perpendicular* to $|\psi_{2(1)}\rangle$. this means that the measurement outcomes corresponding to these two operators *unambiguously* identify the state as $|\psi_{1(2)}\rangle$. The state will be correctly identified with probability $1-|\langle\psi_1|\psi_2\rangle|$ and an ambiguous result occurs with probability $|\langle\psi_1|\psi_2\rangle|$. It is noteworthy that it is the modulus of the overlap of the states and not the modulus *squared* that appears in these probabilities.

To realise this unambiguous detection we require a three-element POM and unambiguous discrimination between two non-orthogonal photon polarisation states has been demonstrated experimentally, using optical fibre with polarisation-dependent losses (Huttner *et al* 1996) and in an interferometer similar to that used for minimum error discrimination between the trine states (Clarke *et al* 2001a).

We have presented here only two of a variety of optimal measurements. Others that have been investigated include maximising the mutual information and determining the state with maximum confidence. These, the relationships between them and the experiments that have been performed are discussed further in (Barnett and Croke 2009b).

3.5 Operations and the post-measurement state

We have not as yet addressed the question of how the measurement process modifies the quantum state in a generalized measurement[4]. There are two pressing reasons for proceeding beyond the von Neumann ideal in which the quantum system is left in an eigenstate corresponding to the measurement outcome. The first is that most real measurements are more destructive than this, and the second is that it gives us no idea how to describe the post measurement state for a generalised, that is non-projective, measurement.

A rigorous treatment takes us into the mathematical world of effects and operations (Kraus K 1983). Rather than this, we present only an indication of what is required. For a more complete treatment we refer the reader to (Croke *et al* 2008, Barnett 2009). We start by noting that quantum theory is linear in the density operator and this suggests that an allowed transformation of the density operator should be of the form

$$\hat{\rho} \to \hat{\rho}' = \sum_i \hat{A}_i \hat{\rho} \hat{B}_i \, . \tag{3.38}$$

Not every set of operators $\{\hat{A}_i, \hat{B}_i\}$ will be allowed, however, as the transformed density operator must, itself, be a density operator. This means that it must be Hermitian, $\hat{\rho}' = \hat{\rho}'^\dagger$, it must be positive, $\langle\psi|\hat{\rho}'|\psi\rangle \geq 0$, and it must have unit trace, $\mathrm{Tr}\hat{\rho}' = 1$. The first of these conditions suggests that we should set $\hat{B}_i = \hat{A}_i^\dagger$ and doing so automatically ensures that the second is fulfilled. The final one, the preservation of the unit trace, is satisfied if we set $\sum_i \hat{A}_i^\dagger \hat{A}_i = \hat{\mathrm{I}}$.

The operator $\hat{A}_i^\dagger \hat{A}_i$ is positive and also Hermitian. The fact that we require the sum of these products to equal the identity operator, moreover, suggests the natural identification

$$\hat{\pi}_i = \hat{A}_i^\dagger \hat{A}_i \, , \tag{3.39}$$

and this allows us to complete the required description of the change of state after a measurement. If we know that a measurement has been performed but do not know the measurement outcome then the density operator is transformed as

$$\hat{\rho} \to \sum_i \hat{A}_i \hat{\rho} \hat{A}_i^\dagger \, . \tag{3.40}$$

If we know that measurement result i was recorded, however, then the state is transformed as

$$\hat{\rho} \to \frac{\hat{A}_i \hat{\rho} \hat{A}_i^\dagger}{\mathrm{Tr}(\hat{A}_i \hat{\rho} \hat{A}_i^\dagger)} = \frac{\hat{A}_i \hat{\rho} \hat{A}_i^\dagger}{\mathrm{Tr}(\hat{\pi}_i \hat{\rho})} \, . \tag{3.41}$$

In order to arrive at a unit-trace density operator in this case, we have divided by the a priori probability for the measurement result. This is directly analogous to the procedure in the von Neumann scheme, in which the transformation is

[4] Time did not permit me to address this question at the School, but these notes would be incomplete without at least a brief account of it.

$$\hat{\rho} \to \frac{\hat{P}_n \hat{\rho} \hat{P}_n}{\mathrm{Tr}(\hat{P}_n \hat{\rho})} \,. \tag{3.42}$$

It should be noted that a set of Kraus operators $\{\hat{A}_i\}$ determine uniquely a corresponding set of probability operators $\{\hat{\pi}_i\}$, but the converse is *not* true; assigning the probability operators does not determine a unique transformation on the state.

4
Entanglement and its Applications

More than any other feature, it is entanglement that gives quantum theory its distinctive character. It is the source of paradoxes, the most famous of which is the much discussed EPR paradox (Einstein *et al* 1935, Bohr 1935, Wheeler and Zurek 1983, Whitaker 2006), of tests of distinctive 'quantumness', such as the violation of Bell's famous inequality (Bell 1964, 1987), and it underlies some of the more unexpected (and headline-grabbing) features of quantum information such as teleportation (Bennett *et al* 1993).

4.1 Entangled states and non-locality

We should start by stating what entanglement is. The simplest answer is that it is the distinctive property of entangled states, which leads us to ask "what are entangled states?". This is, in fact, a surprisingly difficult question to answer, at least in full generality. For the purposes of these all too brief lectures we shall avoid completely the tricky issue of entanglement for mixed states and discuss only pure states. For pure states there is a clear definition but it is the definition of states that are *not* entangled. A state that is not entangled is a state of two or more quantum systems that can be factorized into a product of single-system states. States that do not have this property are entangled. For two quantum systems, A and B, the combined state, $|\Psi\rangle_{AB}$, is entangled if[1]

$$|\Psi\rangle_{\mathrm{AB}} \neq |\psi\rangle_{\mathrm{A}} \otimes |\psi\rangle_{\mathrm{B}} \, . \tag{4.1}$$

Consider, for example, the two-qubit state

$$|\Psi\rangle = \cos\theta|0\rangle \otimes |1\rangle - \sin\theta|1\rangle \otimes |0\rangle \, . \tag{4.2}$$

This state will not be entangled if $\theta = n\pi/2$, but will be entangled for all other values of θ. For entangled states, the properties of the A and B systems are quantum correlated. We can see this in the above example in that determining whether the first qubit is in the state $|0\rangle$ or the state $|1\rangle$ also determines the, previously undetermined, state of the second qubit.

The implications of the existence of entanglement are profound and many. In these lectures we shall address only five:

1. EPR and related non-locality paradoxes.

[1] For those not familiar with the notation, it is often convenient to use the symbol $\otimes$ to separate quantum states and especially operators. Thus $\hat{\sigma}_x\hat{\sigma}_y$ denotes acting first with $\hat{\sigma}_y$ then with $\hat{\sigma}_x$ on a single qubit, but $\hat{\sigma}_x \otimes \hat{\sigma}_y$ means act with $\hat{\sigma}_x$ on qubit 1 and with $\hat{\sigma}_y$ on qubit 2 (Barnett 2009).

2. The abiliy to perform “magic”.

3. Quantum dense coding.

4. Teleportation.

5. Dramatic speed up in quantum computing.

We shall treat the first four of these in this lecture and leave the final one for the next and final lecture.

Let us begin with the EPR, in the form given by Bohm (1951). To this end, consider two qubits in the entangled state

$$|\Psi^-\rangle_{\mathrm{AB}} = \frac{1}{\sqrt{2}}(|0\rangle_{\mathrm{A}} \otimes |1\rangle_{\mathrm{B}} - |1\rangle_{\mathrm{A}} \otimes |0\rangle_{\mathrm{B}})\,. \tag{4.3}$$

Let us suppose that qubit A is held by Alice and qubit B by Bob and that they are at a considerable distance from each other. If either party measures their qubit in any basis then they find either of two possible results, each occurring with probability 1/2. If they measure the same observable as each other then they find opposite or anti-correlated results. This is a simple consequence of the eigenvalue equation:

$$\mathbf{a} \cdot \hat{\boldsymbol{\sigma}} \otimes \mathbf{a} \cdot \hat{\boldsymbol{\sigma}} |\Psi^-\rangle = -|\Psi^-\rangle\,, \tag{4.4}$$

which holds for all unit vectors **a**. The paradox is that a measurement by Alice of her particle will *instantaneously* project the state of its partner into an eigenstate of the observable selected by Alice. If we assume that Alice has a free choice of what to measure and that any influence of her measurement cannot travel arbitrarily fast, then Bob’s qubit must have carried values for all possible observables, something that is clearly at odds with complementarity. We give, here, only a partial resolution of this in the form of the no-signalling theorem.

4.1.1 Bell’s theorem

Correlations are also common in the classical world and, indeed, underlie classical communications. Are not the quantum correlations associated with entanglement simply the same thing? It was Bell’s theorem, in the form of the violation of his celebrated inequality, that gave the definitive answer “no”!

We present here a derivation of Bell’s inequality, in its most common form (Clauser *et al* 1969, Bell 1987). Let us start by thinking of our qubit as a spin-half particle. A measurement of the spin along a direction in space, given by the unit vector **a**, will reveal the spin to be aligned or anti-aligned with this direction. We write the first of these as +1 and the second as −1, so that in any given measurement our measurement result A with be ±1. We can do the same thing for the second qubit (entangled with the first) and write the measurement result as $B = \pm 1$. If we take the product of these two values and average over many experimental realisations then we obtain the correlation function

$$E(\mathbf{a}, \mathbf{b}) = \langle AB \rangle\,, \tag{4.5}$$

which is clearly bounded in magnitude:

$$-1 \leq E(\mathbf{a}, \mathbf{b}) \leq 1\,. \tag{4.6}$$

In order to arrive at Bell's inequality we need to suspend disbelief and wonder what a theory would look like if the indeterminacy of quantum theory arose from hidden variables or, more precisely, local hidden variables which we denote by λ and take to be governed by a probability distribution $P(\lambda)$. To this end we write the correlation function in the form

$$E(\mathbf{a}, \mathbf{b}) = \int d\lambda P(\lambda) A(\mathbf{a}, \lambda) B(\mathbf{b}, \lambda)\,. \tag{4.7}$$

Some words of explanation are in order. We say that this is a form based on a local hidden-variable theory because each measured result A and B depends only on (i) the choice of observable made at the observation site ($\mathbf{a}$ and $\mathbf{b}$ respectively) and (ii) the hidden variable λ, presumed to have been determined in the source of the particles. The measurement results do *not* depend on the choice of measurement made at the distant site. These considerations lead us to Bell's inequality as follows. First we write

$$\begin{aligned} E(\mathbf{a}, \mathbf{b}) - E(\mathbf{a}, \mathbf{b}') &= \int d\lambda P(\lambda) [A(\mathbf{a}, \lambda) B(\mathbf{b}, \lambda) - A(\mathbf{a}, \lambda) B(\mathbf{b}', \lambda)] \\ &= \int d\lambda P(\lambda) A(\mathbf{a}, \lambda) B(\mathbf{b}, \lambda) [1 \pm A(\mathbf{a}', \lambda) B(\mathbf{b}'] \\ &\quad - \int d\lambda P(\lambda) A(\mathbf{a}, \lambda) B(\mathbf{b}', \lambda) [1 \pm A(\mathbf{a}', \lambda) B(\mathbf{b}]\,. \end{aligned} \tag{4.8}$$

In the second line we have added and subtracted the same expression. In doing so, we have assumed it is meaningful for quantities such as $A(\mathbf{a}, \lambda)$ and $A(\mathbf{a}', \lambda)$ to coexist. This is the realism part of local-realism; it states that properties exist even if we do not measure them. We recall that the product AB lies between -1 and $+1$ and hence we can obtain from this equality an inequality of the form

$$\begin{aligned} |E(\mathbf{a}, \mathbf{b}) - E(\mathbf{a}, \mathbf{b}')| &\leq \int d\lambda P(\lambda) [1 \pm A(\mathbf{a}', \lambda) B(\mathbf{b}', \lambda)] \\ &\quad + \int d\lambda P(\lambda) [1 \pm A(\mathbf{a}', \lambda) B(\mathbf{b}, \lambda)] \\ &= 2 \pm [E(\mathbf{a}', \mathbf{b}') + E(\mathbf{a}', \mathbf{b})]\,, \end{aligned} \tag{4.9}$$

which implies that

$$|E(\mathbf{a}, \mathbf{b}) - E(\mathbf{a}, \mathbf{b}')| + |E(\mathbf{a}', \mathbf{b}') + E(\mathbf{a}', \mathbf{b})| \leq 2\,. \tag{4.10}$$

This is Bell's inequality; it states that any correlations of a local-realistic nature must satisfy this inequality. What do we find for entangled quantum states? Consider again the EPR spin state:

$$|\Psi^-\rangle_{\mathrm{AB}} = \frac{1}{\sqrt{2}} (|0\rangle_{\mathrm{A}} \otimes |1\rangle_{\mathrm{B}} - |1\rangle_{\mathrm{A}} \otimes |0\rangle_{\mathrm{B}})\,. \tag{4.11}$$

It is straightforward to confirm that for spin measurements on the two qubits prepared in this state we find

$$E(\mathbf{a},\mathbf{b}) = \langle \mathbf{a}\cdot\hat{\boldsymbol{\sigma}} \otimes \mathbf{a}\cdot\hat{\boldsymbol{\sigma}} \rangle = -\mathbf{a}\cdot\mathbf{b}\,. \tag{4.12}$$

If we put this into our Bell inequality we find

$$|E(\mathbf{a},\mathbf{b}) - E(\mathbf{a},\mathbf{b}')| + |E(\mathbf{a}',\mathbf{b}') + E(\mathbf{a}',\mathbf{b})| = |-\mathbf{a}\cdot(\mathbf{b}-\mathbf{b}')| + |-\mathbf{a}'\cdot(\mathbf{b}+\mathbf{b}')|\,. \tag{4.13}$$

To find the maximum value we simply choose $\mathbf{a}$ and $\mathbf{a}'$ to be parallel, respectively, to $\mathbf{b}-\mathbf{b}'$ and $\mathbf{b}+\mathbf{b}'$ and $\mathbf{b}$ and $\mathbf{b}'$ to be mutually perpendicular, as depicted in Fig. 4.1. With this choice we find

$$|E(\mathbf{a},\mathbf{b}) - E(\mathbf{a},\mathbf{b}')| + |E(\mathbf{a}',\mathbf{b}') + E(\mathbf{a}',\mathbf{b})| = 2\sqrt{2}\,, \tag{4.14}$$

which clearly exceeds 2 and hence violates the inequality. It is this, perhaps more than anything else, that signified the exceptional nature of entanglement. In quantum information it is often by violating a Bell inequality that we demonstrate most clearly the presence of entanglement: all entangled pure states violate a Bell inequality (Gisin 1991).

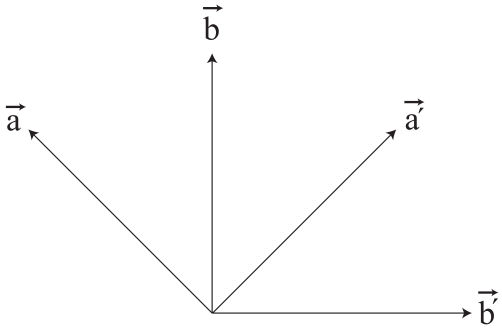


Fig. 4.1 Orientation of the optimal measurement directions for maximal violation of Bell's inequality. Reproduced, with permission, from Barnett(2009).

4.1.2 No-signalling theorem

A resolution, of sorts, to the EPR paradox is that no operation carried out on one of an entangled pair of quantum systems is detectable by an observation on its entangled partner. This means, of course, that no signal can be transmitted in this way. The no signalling theorem was stated rigorously by Ghirardi *et al* (1980), but rather than follow their analysis, it is instructive to make use of what we have learned already about generalised measurements (Barnett 2009).

Let us suppose that Alice and Bob each have one of a pair of entangled systems and that Alice attempts to communicate with Bob by making a measurement on hers. She cannot, of course, control the measurement result but can decide what to measure. We shall allow Alice to choose between any two measurements, which we allow to be generalised. Let Alice's measurements have the probability operators $\{\hat{\pi}_j^{\mathrm{A1}}\}$ for measurement choice 1 and $\{\hat{\pi}_j^{\mathrm{A2}}\}$ for choice 2. In an attempt to detect the effect of Alice's choice, Bob makes a measurement characterised by the probability operators $\{\hat{\pi}_k^{\mathrm{B}}\}$. The joint probabilities for Alice's and Bob's measurement results are

$$P(j,k) = \langle\Psi|\hat{\pi}_j^{\mathrm{A1,A2}} \otimes \hat{\pi}_k^{\mathrm{B}}|\Psi\rangle\,, \tag{4.15}$$

where $|\Psi\rangle$ is the joint state of Alice's and Bob's system. If Alice's choice is to affect the Bob's observation, then we need the probabilities for Bob's measurement results to depend on Alice's choice of measurement. This is clearly not the case, however, as

$$P(k) = \sum_j P(j,k) = \langle\Psi|\hat{\pi}_k^{\mathrm{B}}|\Psi\rangle\,, \tag{4.16}$$

where we have used the fact that our probability operators sum to the identity operator:

$$\sum_j \hat{\pi}_j^{\mathrm{A1,A2}} = \hat{\mathrm{I}}\,. \tag{4.17}$$

The probabilities of any of Bob's possible measurement results is *independent* of Alice's choice of measurement and, indeed, are the same whether Alice performs a measurement at all. Nothing Alice does to her quantum system is detectable by Bob and hence no signalling is possible by quantum measurement of a distant system.

The no-signalling theorem is a rigorous consequence of quantum theory and can be used, in quantum information theory, to place bounds on what is and is not possible. As a simple example, we can derive a lower bound on the probability for an inconclusive outcome in unambiguous state-discrimination using the no-signalling condition (Barnett and Andersson 2002).

4.2 Quantum "magic tricks"

The Bell inequality is only one of an impressive array of startling and testable consequences of entanglement (Redhead 1987, Greenberger *et al* 1990, Mermin 1990). Were these lectures about the mysteries of quantum theory, then we might enjoy exploring these, but our purpose is not to test entanglement and its consequences, but rather to exploit it. We shall present one example, however, of a testable consequence of entanglement as it points to a necessary distinction between logical reasoning in quantum theory and the classical world.

Hardy (1993, Goldstein 1994) presented a simple demonstration of non-locality without an inequality. This is instructive in that it provides a warning against reasoning on the basis of what might have been measured rather than what actually was observed. The following short presentation is reproduced, essentially verbatim, from Barnett (2009).

Consider a pair of qubits, one held by Alice and the other by Bob, prepared in the pure state

$$|\mathrm{Hardy}\rangle = \frac{1}{\sqrt{3}}(|0\rangle_{\mathrm{A}} \otimes |0\rangle_{\mathrm{B}} + |1\rangle_{\mathrm{A}} \otimes |0\rangle_{\mathrm{B}} + |0\rangle_{\mathrm{A}} \otimes |1\rangle_{\mathrm{B}})\,. \tag{4.18}$$

We proceed by noting that this state can also be written in the form

$$\begin{aligned}|\mathrm{Hardy}\rangle &= \sqrt{\frac{2}{3}}|0'\rangle_{\mathrm{A}} \otimes |0\rangle_{\mathrm{B}} + \frac{1}{\sqrt{3}}|0\rangle_{\mathrm{A}} \otimes |1\rangle_{\mathrm{B}} \\ &= \frac{1}{\sqrt{3}}|1\rangle_{\mathrm{A}} \otimes |0\rangle_{\mathrm{B}} + \sqrt{\frac{2}{3}}|0\rangle_{\mathrm{A}} \otimes |0'\rangle_{\mathrm{B}}\,, \end{aligned} \tag{4.19}$$

where $|0'\rangle = 2^{-1/2}(|0\rangle + |1\rangle)$, is the eigenstate of $\hat{\sigma}_x$ with eigenvalue +1. The following statements follow directly from the form of $|\mathrm{Hardy}\rangle$:

(i) If both Alice and Bob measure the observable correspondding to $\hat{\sigma}_z$, with eigenstates $|0\rangle$ and $|1\rangle$, then at least one of them will get the result +1, corresponding to the state $|0\rangle$.

(ii) If Alice measures $\hat{\sigma}_z$ and gets the value +1 then a measurement by Bob of $\hat{\sigma}_x$ will, with certainty, find the value +1, corresponding to the state $|0'\rangle$.

(iii) If Bob measures $\hat{\sigma}_z$ and gets the value +1 then a measurement by Alice of $\hat{\sigma}_x$ will, with certainty, find the value +1, corresponding to the state $|0'\rangle$.

Local realistic ideas lead us to treat as simultaneously real the values ±1 of the observables corresponding to the operators $\hat{\sigma}_z$ and $\hat{\sigma}_x$. The values of these, which we denote σ_z and σ_x, respectively, should be independent of any choice of an observation carried out on the other qubit. This leads us to express the above three properties as the following probabilities:

$$\begin{aligned}P(\sigma_z^{\mathrm{A}} = -1, \sigma_z^{\mathrm{B}} = -1) &= 0 \\ P(\sigma_x^{\mathrm{B}} = +1|\sigma_z^{\mathrm{A}} = +1) &= 1 \\ P(\sigma_x^{\mathrm{A}} = +1|\sigma_z^{\mathrm{B}} = +1) &= 1\,. \end{aligned} \tag{4.20}$$

The first of these tells us that at least one of the properties σ_z^{A} and σ_z^{B} must take the value +1 and the following two then tell us that at least one of the properties σ_x^{A} and σ_z^{B} must take the value +1. It is a prediction of local realism, therefore, that σ_x^{A} and σ_z^{B} cannot *both* take the value −1:

$$P(\sigma_x^{\mathrm{A}} = -1, \sigma_x^{\mathrm{B}} = -1) = 0\,. \tag{4.21}$$

A quantum mechanical treatment, however, shows that measurement by both Alice and Bob of $\hat{\sigma}_x$ can *both* give the result +1 and that this occurs with probability $\frac{1}{12}$. This is clearly at odds with the local-realistic reasoning presented above.

As with all the best magic tricks, it is what the audience assumes to have happened that makes the trick appear to be impossible.

4.3 Ebits and shared entanglement

In quantum information we think of entanglement not as a mystery, but rather as a *resource* to be exploited. The first question to be addressed, as with any resource, is to decide how to quantify it. This is a subtle question but we can give, at least, a simple preliminary answer in terms of "ebits". If two spatially separated individuals share a pair of maximally entangled qubits then we say that they share 1 ebit. By maximally entangled we mean pure state such that the reduced density operator for each of the component qubits is of the form $\hat{\mathrm{I}}/2$. This means that we can identify the state only by bringing together the component qubits and that a measurement of just a single qubit is maximally uncertain.

It is important to ask how we can get to a situation in which two parties share an ebit. One way, of course, is to prepare an entangled state at one site and then transport (carefully) one of the entangled qubits to the other site. We might very reasonably ask, however, if it is not possible to achieve the same result by means of classical communications, by exchanging instructions between the distant sites. That this is *not* possible follows directly from what we already know. Let us suppose that Alice prepares the entangled state

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|0\rangle \otimes |1\rangle - |1\rangle \otimes |0\rangle)\,. \tag{4.22}$$

To share this state with Bob she can send the second qubit to Bob by means of a suitable quantum communications channel. The resulting quantum correlations between the qubits (at least for a sufficient number of entangled pairs) could be used to demonstrate a violation a Bell inequality. Such a violation is not possible, of course, for any classical or classically-mediated correlations. Hence we can conclude, without the need for further analysis, that classical and quantum communications are radically different in nature.

The antisymmetric state, $|\Psi^-\rangle$ is not the only maximally entangled state of two qubits. Indeed, we can create a complete basis of four orthonomal states for two qubits, each of which is similarly maximally entangled. One simple choice is the set of states, now referred to as the Bell states or the Bell basis. These four states are

$$\begin{aligned}|\Psi^\pm\rangle &= \frac{1}{\sqrt{2}}(|0\rangle \otimes |1\rangle \pm |1\rangle \otimes |0\rangle)\\ |\Phi^\pm\rangle &= \frac{1}{\sqrt{2}}(|0\rangle \otimes |0\rangle \pm |1\rangle \otimes |1\rangle)\,.\end{aligned} \tag{4.23}$$

As these states form an orthonormal basis, we can perform a von Neumann measurement on the two qubits with each of these four corresponding to a different outcome. This "Bell measurement" cannot be performed, however, simply combining a von Neumann measurement on each individual qubit.

4.4 Quantum dense coding

A very simple application of the Bell states, perhaps the simplest, is in quantum dense coding (Bennett and Wiesner 1992). The starting point is the observation that a (von Neumann) measurement of a qubit can provide at most one bit of information. To see how this works, we need only consider a qubit that has been prepared with equal probability in either of the two orthogonal states, $|0\rangle$ and $|1\rangle$. A simple projective measurement in this basis will reveal the state that has been prepared and provide a single bit.

By using an ebit we can, in a sense, double this rate by communicating two bits of information for every qubit sent from Alice to Bob. Let us start with a single ebit shared by Alice and Bob, prepared in the state

$$|\Psi^-\rangle_{\mathrm{AB}} = \frac{1}{\sqrt{2}}(|0\rangle \otimes |1\rangle - |1\rangle \otimes |0\rangle)\,. \tag{4.24}$$

Bob selects one of four unitary transformations to perform on his qubit. These transformations, and the effect of the corresponding entangled state, are listed below:

$$\begin{array}{cccccc}
\hat{U}_1 & \begin{array}{c}|0\rangle_{\mathrm{B}}\\|1\rangle_{\mathrm{B}}\end{array} & \Rightarrow & \begin{array}{c}|0\rangle_{\mathrm{B}}\\|1\rangle_{\mathrm{B}}\end{array} & & |\Psi^-\rangle_{\mathrm{AB}} \Rightarrow |\Psi^-\rangle_{\mathrm{AB}} \\
\hat{U}_2 & \begin{array}{c}|0\rangle_{\mathrm{B}}\\|1\rangle_{\mathrm{B}}\end{array} & \Rightarrow & \begin{array}{c}|1\rangle_{\mathrm{B}}\\|0\rangle_{\mathrm{B}}\end{array} & & |\Psi^-\rangle_{\mathrm{AB}} \Rightarrow |\Phi^-\rangle_{\mathrm{AB}} \\
\hat{U}_3 & \begin{array}{c}|0\rangle_{\mathrm{B}}\\|1\rangle_{\mathrm{B}}\end{array} & \Rightarrow & \begin{array}{c}-|0\rangle_{\mathrm{B}}\\|1\rangle_{\mathrm{B}}\end{array} & & |\Psi^-\rangle_{\mathrm{AB}} \Rightarrow |\Psi^+\rangle_{\mathrm{AB}} \\
\hat{U}_4 & \begin{array}{c}|0\rangle_{\mathrm{B}}\\|1\rangle_{\mathrm{B}}\end{array} & \Rightarrow & \begin{array}{c}-|1\rangle_{\mathrm{B}}\\|0\rangle_{\mathrm{B}}\end{array} & & |\Psi^-\rangle_{\mathrm{AB}} \Rightarrow |\Phi^+\rangle_{\mathrm{AB}}\,.
\end{array} \tag{4.25}$$

We see that the effect of Bob's choice is to transform the state of the entangled pair of qubits into one of the four orthogonal Bell states. Hence by sending just his single qubit to Alice he can send *two* bits of information (recall that $4 = 2^2$ states corresponds to 2 bits). In order to retrieve the information Alice simply performs a Bell measurement on the two qubits comprising the ebit and extracts the two bits[2].

It is interesting to reflect on the nature of a Bell measurement and, in particular, to try to understand it in terms of spin measurements. We know, of course, that the Pauli operators $\hat{\sigma}_x$, $\hat{\sigma}_y$ and $\hat{\sigma}_z$ do not mutually commute. It may come as something of a surprise, therefore, that the products of these cartesian components of the spin, $\hat{\sigma}_x \otimes \hat{\sigma}_x$, $\hat{\sigma}_y \otimes \hat{\sigma}_y$ and $\hat{\sigma}_z \otimes \hat{\sigma}_z$, commute for *any* pair of qubits. The Bell states are the simultaneous eigenstates of these three product operators:

$$\begin{array}{lll}
\hat{\sigma}_x \otimes \hat{\sigma}_x|\Psi^-\rangle = -|\Psi^-\rangle & \hat{\sigma}_y \otimes \hat{\sigma}_y|\Psi^-\rangle = -|\Psi^-\rangle & \hat{\sigma}_z \otimes \hat{\sigma}_z|\Psi^-\rangle = -|\Psi^-\rangle \\
\hat{\sigma}_x \otimes \hat{\sigma}_x|\Psi^+\rangle = |\Psi^+\rangle & \hat{\sigma}_y \otimes \hat{\sigma}_y|\Psi^+\rangle = |\Psi^+\rangle & \hat{\sigma}_z \otimes \hat{\sigma}_z|\Psi^+\rangle = -|\Psi^+\rangle \\
\hat{\sigma}_x \otimes \hat{\sigma}_x|\Phi^-\rangle = -|\Phi^-\rangle & \hat{\sigma}_y \otimes \hat{\sigma}_y|\Phi^-\rangle = |\Phi^-\rangle & \hat{\sigma}_z \otimes \hat{\sigma}_z|\Phi^-\rangle = |\Phi^-\rangle \\
\hat{\sigma}_x \otimes \hat{\sigma}_x|\Phi^+\rangle = |\Phi^+\rangle & \hat{\sigma}_y \otimes \hat{\sigma}_y|\Phi^+\rangle = -|\Phi^+\rangle & \hat{\sigma}_z \otimes \hat{\sigma}_z|\Phi^+\rangle = |\Phi^+\rangle\,.
\end{array} \tag{4.26}$$

Thus the Bell states are those in which the products of the cartesian components of the spin are defined and we can view a Bell measurement as a *comparison* of the three components of spin for the two qubits. You may have noticed, in fact, that it suffices to compare only two components of the spin as the final one is then defined. The reason for this has its origins in the fact that the product of any two Pauli spin components is proportional to the third and, in particular $\hat{\sigma}_x\hat{\sigma}_y = i\hat{\sigma}_z$. It follows that

$$(\hat{\sigma}_x \otimes \hat{\sigma}_x)\,(\hat{\sigma}_y \otimes \hat{\sigma}_y) = -\hat{\sigma}_z \otimes \hat{\sigma}_z \tag{4.27}$$

[2] Although it sounds easy to perform a Bell measurement, achieving this in practice is, perhaps not surprisingly rather more of a challenge (Mattle *et al* 1996).

and hence that the product of the eigenvalues of $\hat{\sigma}_x \otimes \hat{\sigma}_x$, $\hat{\sigma}_y \otimes \hat{\sigma}_y$ and $\hat{\sigma}_z \otimes \hat{\sigma}_z$ must be -1. This means that there can be no state of two qubits, for example, in which the x-, y- and z- components of the spin are all the same.

4.5 Teleportation

Our final application is quantum teleportation. This is an important application of entanglement and came as something of a surprise when it was first proposed (Bennett *et al* 1993). We should be very clear before proceeding however, that quantum teleportation is *not* matter transportation like something out of the popular science fiction series Star Trek. A better understanding may be reached, I feel, by adopting the description proposed by Haroche, which is quantum teleportation is a "fax machine for quantum information". It is the quantum information, or rather the quantum state that is transferred and not the physical system on which the qubit state is stored.

Let us suppose that Alice wishes to send a qubit to Bob. She may know the state of the qubit or she may not. We have seen that quantum communications and classical communications are very different and so should not be surprised that there is, in general, no classical way to achieve this. Alice has two ways to achieve the desired result (i) she can send the physical qubit carrying the quantum information to Bob (we might call this "qmail"), or (ii) she can teleport the quantum information to Bob using an ebit of shared entanglement (which we might call a "qfax").

Alice can teleport the qubit state

$$|\psi\rangle_1 = \alpha|0\rangle_1 + \beta|1\rangle_1 \tag{4.28}$$

by making use of an ebit of the form

$$|\Psi^-\rangle_{\mathrm{AB}} = \frac{1}{\sqrt{2}}(|0\rangle_{\mathrm{A}} \otimes |1\rangle_{\mathrm{B}} - |1\rangle_{\mathrm{A}} \otimes |0\rangle_{\mathrm{B}})\,, \tag{4.29}$$

shared with Bob. We can write the combined state of the three qubits as

$$\begin{aligned}|\Psi\rangle_{1\mathrm{AB}} &= |\psi\rangle_1 \otimes |\Psi^-\rangle_{\mathrm{AB}} \\ &= \frac{1}{2}\left[|\Psi^-\rangle_{1\mathrm{A}}\,(-\alpha|0\rangle_{\mathrm{B}} - \beta|1\rangle_{\mathrm{B}})\right. \\ &\qquad +|\Psi^+\rangle_{1\mathrm{A}}\,(-\alpha|0\rangle_{\mathrm{B}} + \beta|1\rangle_{\mathrm{B}}) \\ &\qquad +|\Phi^-\rangle_{1\mathrm{A}}\,(\alpha|1\rangle_{\mathrm{B}} + \beta|0\rangle_{\mathrm{B}}) \\ &\qquad \left.+|\Phi^+\rangle_{1\mathrm{A}}\,(\alpha|1\rangle_{\mathrm{B}} - \beta|0\rangle_{\mathrm{B}})\right]\,. \end{aligned} \tag{4.30}$$

In arriving at the final expression, we have only rewritten the state, but the quantum information, in the form of the coefficients α and β, have appeared connected with Bob's qubit. We know, however, that Bob cannot yet access this information, for to do so would violate the no-signalling theorem.

If Alice performs a Bell measurement on her two qubits (qubits 1 and A) she can then tell Bob which of four possible transformations to perform on qubit B in order to change its state into $|\psi\rangle_{\mathrm{B}}$. The idea is best understood in terms of an example. Let us suppose that Alice's Bell measurement gives a result corresponding to the state

$|\Psi^+\rangle_{\text{1A}}$, then she knows that Bob's qubit is in the state $-\alpha|0\rangle_{\text{B}}+\beta|1\rangle_{\text{B}}$. She then has only to use a two-bit classical channel to tell Bob to perform unitary transformation number 3 so that:

$$\hat{U}_3 \quad \begin{matrix}|0\rangle_{\text{B}}\\|1\rangle_{\text{B}}\end{matrix} \Rightarrow \begin{matrix}-|0\rangle_{\text{B}}\\|1\rangle_{\text{B}}\end{matrix} \qquad\qquad -\alpha|0\rangle_{\text{B}}+\beta|1\rangle_{\text{B}} \Rightarrow \alpha|0\rangle_{\text{B}}+\beta|1\rangle_{\text{B}} \tag{4.31}$$

and the teleportation is complete in that Bob's qubit is in the same state as that in which the original qubit started, $|\psi\rangle_1$. Each of the other three possible Bell-measurement outcomes corresponds to a different unitary transformation that Bob is required to perform.

We can understand why teleportation works by recalling that a Bell measurement is one in which we compare the cartesian components of the two spins. In particular the four possible measurement results tell us that

$$\begin{aligned}|\Psi^-\rangle &\Rightarrow \sigma_x \text{ different}, \ \sigma_y \text{ different}, \ \sigma_z \text{ different}\\ |\Psi^+\rangle &\Rightarrow \sigma_x \text{ same}, \ \sigma_y \text{ same}, \ \sigma_z \text{ different}\\ |\Phi^-\rangle &\Rightarrow \sigma_x \text{ different}, \ \sigma_y \text{ same}, \ \sigma_z \text{ same}\\ |\Phi^+\rangle &\Rightarrow \sigma_x \text{ same}, \ \sigma_y \text{ different}, \ \sigma_z \text{ same}\,.\end{aligned} \tag{4.32}$$

We know, also, that in the shared entangled state, $|\Psi^-\rangle_{\text{AB}}$, has the three components of the spins all opposite to each other. Hence we can reason as follows: (i) If the Bell measurement gives the state $|\Psi^-\rangle_{1A}$ then the spin components for the qubits 1 and A are anti-aligned *but* we know also that the qubits A and B are anti-aligned and hence we are left with Bob's qubit in initial state of qubit 1. (ii) If the Bell measurement gives the state $|\Psi^+\rangle_{1A}$ then the x- and y-components of the spins for the qubits 1 and A are the same but the z-component is different. It follows that qubit B is left in a state in which the x- and y-components of the spin are opposite to that for the initial state of qubit 1, but the z-component is the same. Instructing Bob to perform a rotation, through π about the z-axis will leave his qubit in the initial state of qubit 1. Similarly, (iii) if the Bell measurement gives the state $|\Phi^-\rangle$ then Bob needs to perform a qubit-rotation about the x-axis and (iv) a Bell measurement giving the state $|\Phi^+\rangle$ means that Bob needs to perform a rotation about the y-axis.

There is much more that could be said about teleportation, but we conclude this all too brief introduction by describing a few features worthy of further thought:

1. We note that there is a sense in which Bob already has the information after Alice's measurement. This is a direct consequence of the projective nature of Alice's Bell measurement. What saves us from violating the no-signalling theorem is simply that Bob does not know where to look for it.

2. Alice's copy of the original qubit is destroyed in the teleportation process. This is inevitable and not simply a consequence of the scheme we have adopted. Were it otherwise, then we would be violating the no-cloning theorem.

3. Alice does not have to know the state that is to be teleported. We have chosen a general qubit state parametrized by the amplitudes α and β, but no process undertaken

by either Alice or Bob is dependent on these amplitudes and it follows that Alice does not need to know the state to be teleported.

4. The qubit to be teleported may itself be part of an entangled state and in this way we can teleport entanglement. In this way, if Alice shares an ebit with Bob and also one with Claire then she can teleport the entanglement to leave Bob and Claire sharing an ebit:

$$|\Psi^-\rangle_{\mathrm{C1}} \otimes |\Psi^-\rangle_{\mathrm{AB}} \Rightarrow |\Psi^-\rangle_{\mathrm{CB}} \, . \tag{4.33}$$

This process of transferring entanglement by teleportation is commonly referred to as entanglement-swapping.

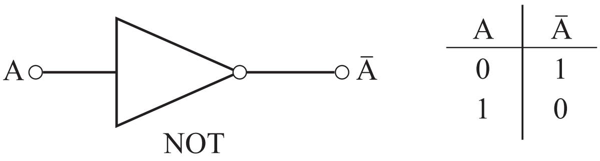
5
Quantum Computation

No course on quantum information would be complete without a treatment of quantum computation. Yet covering such a diverse topic in a single lecture presents, if anything, an even greater challenge than the areas addressed in the preceding lectures. The material selected for this lecture can only provide a modest foretaste of the topic that spans a number of disciplines, including physics and computer science, and for a more satiating introduction I can only recommend some texts for further reading (Nielsen and Chuang 2000, Stenholm and Suominen 2005, Vedral 2006, Mermin 2007, Kaye *et al* 2007, Barnett 2009, Gay and Mackie 2010, Pachos 2012).

We should start by asking why the idea of quantum computers seems to have become so prominent. There are, I think, two very good reasons. The first is in response to a very pressing need. The speed of development in computers has been truly remarkable and follows an exponential increase in performance first described by Moore in 1965 (Moore 1965). There are many versions of this law, but perhaps the simplest is that the number of transistors on chip doubles roughly every two years. The way this is achieved is by making the individual transistors ever smaller. As we make components ever smaller, however, we must inevitably run into quantum effects. Rather than fight against quantum mechanics (a battle we must lose at some size scale) perhaps it would be better to embrace the new possibilities provided by quantum information processing. The second reason is the distinctively new possibilities offered by quantum algorithms (which can only run on a quantum computer). These address problems that will always remain intractable for a computer based on classical logic and include simulations of complex quantum systems (like large molecules) and the headline-grabbing Shor's algorithm for factoring, of which more later.

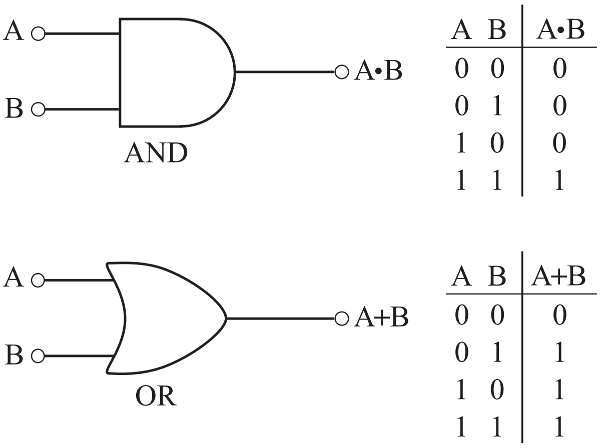
5.1 Digital electronics

We begin our discussion by describing the way in which logical bits and logic operations are implemented classically. In the simplest form, the logical bits 1 and 0 are encoded as voltages - a high voltage for 1 and ground or zero volts for 0. These values are manipulated by transistor-based devices called gates (Smith 1983). The most common of these act either on a single bit value or couple two together. There is only one single-bit gate, the NOT gate, which simply changes a 0 to 1 and a 1 to 0. The symbol for the NOT gate and its truth table are given in Fig 5.1. The simplest two-bit gates are the AND and the OR gate, depicted in Fig 5.2. Combinations of these together with a NOT gate to form NAND and NOR gates are also common. These two bit gates, together with the NOT gate allow us to perform any logical operation (strictly a Boolean operation) on a string of bits. Such an operation is, at a fundamental level, what we mean by a computation.



A	Ā
0	1
1	0

Fig. 5.1 The NOT gate and its truth table. Reproduced, with permission, from Barnett(2009).



A	B	A•B
0	0	0
0	1	0
1	0	0
1	1	1

A	B	A+B
0	0	0
0	1	1
1	0	1
1	1	1

Fig. 5.2 The AND and OR gates and their truth tables. Reproduced, with permission, from Barnett(2009).

5.2 Quantum gates

The simplest way in which to introduce quantum elements into information processing is to replace each classical bit with a qubit and to design devices that operate upon them. We are not completely free in the operations we can devise, however, as we must respect the laws of quantum mechanics. This means, in particular, that the transformations we can perform are necessarily limited to the unitary transformations. Even at the single qubit level, however, there is already a great deal more we can do that for a classical bit, as we are allowed to generate superpositions of the two qubit states $|0\rangle$ and $|1\rangle$. A single qubit-gate may perform any single-qubit unitary transformation, that is any rotation on the Bloch sphere. Six of the more commonly occurring one-qubit gates, together with the unitary transformations they enact are presented in Fig 5.3.

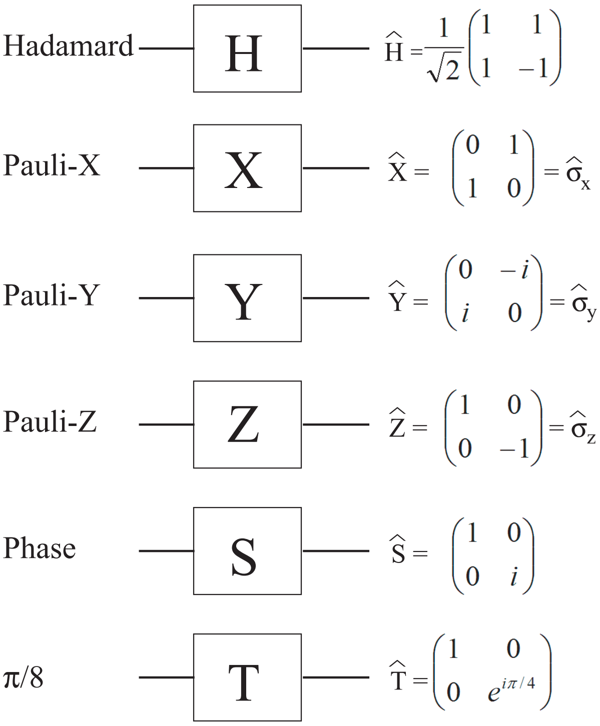
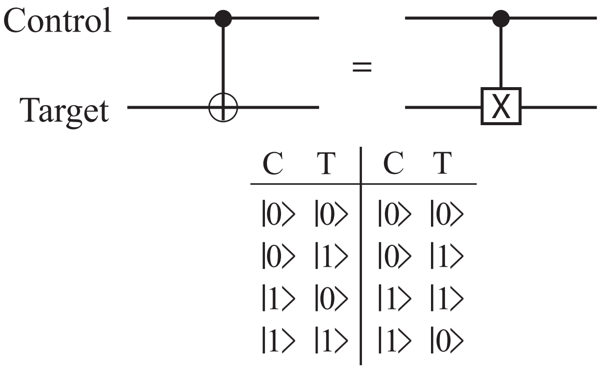

Hadamard
H
$\hat{H} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$
Pauli-X
X
$\hat{X} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \hat{\sigma}_x$
Pauli-Y
Y
$\hat{Y} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \hat{\sigma}_y$
Pauli-Z
Z
$\hat{Z} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \hat{\sigma}_z$
Phase
S
$\hat{S} = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$
π/8
T
$\hat{T} = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}$

Fig. 5.3 Some of the more common one-qubit gates together with their associated unitary transformations. Reproduced, with permission, from Barnett(2009).

We require in addition gates that couple qubits. Unlike their classical counterparts, these gates have two output qubits as well as two input ones; where this not the case then we would not respect unitarity. Principal among the vast array of possible two-bit gates is the controlled NOT gate, or CNOT gate, depicted in Fig 5.4. The two qubits entering the gate are designated the control qubit, C and the target qubit, T. The gate enacts the transformation $|1\rangle \rightarrow |0\rangle$, $|0\rangle \rightarrow |1\rangle$ if the control qubit is in the state $|1\rangle$ but leaves it unchanged if the if the control qubit is in the state $|0\rangle$. Despite its innocuous looking truth table, it is straightforward to show that this gate is intrinsically quantum-mechanical in nature. To see this, let the control bit entering the gate be in a superposition of computational basis states. The transformation is then

$$\frac{1}{\sqrt{2}}\left(|0\rangle_C + |1\rangle_C\right)|0\rangle_T \rightarrow \frac{1}{\sqrt{2}}\left(|0\rangle_C|0\rangle_T + |1\rangle_C|1\rangle_T\right) , \tag{5.1}$$

which we recognise as one of the entangled Bell states. The fact that the CNOT gate has generated an entangled state from an unentangled one (and therefore introduced the possibility of non-locality) suffices to establish the intrinsically quantum nature of the CNOT gate.



C	T	C	T
$\|0\rangle$	$\|0\rangle$	$\|0\rangle$	$\|0\rangle$
$\|0\rangle$	$\|1\rangle$	$\|0\rangle$	$\|1\rangle$
$\|1\rangle$	$\|0\rangle$	$\|1\rangle$	$\|1\rangle$
$\|1\rangle$	$\|1\rangle$	$\|1\rangle$	$\|0\rangle$

Fig. 5.4 The CNOT gate and its effect on the computational basis states. Reproduced, with permission, from Barnett(2009).

In digital electronics, a small number of gates suffices to allow all possible Boolean operations and, remarkably, something very similar holds for quantum gates. We can perform any unitary evolution of a set of qubits by acting on them only with single-qubit gates and a suitable two-qubit gate; the CNOT gate suffices for this purpose. We do not prove this assertion here as the required demonstration would take too long and occupy too much space. The proof is not particularly difficult, however, and may be find in the (Nielsen and Chuang 2000, Barnett 2009). As an example we can construct the three-qubit Toffoli gate, or controlled, controlled NOT gate using single-qubit and two-bit gates. The operation of the Toffoli gate is most readily understood in the computational basis, $|0\rangle$, $|1\rangle$ for the three qubits:

$$|\mathrm{A}\rangle \otimes |\mathrm{B}\rangle \otimes |\mathrm{C}\rangle \to |\mathrm{A}\rangle \otimes |\mathrm{B}\rangle \otimes |\mathrm{C} \oplus (\mathrm{A} \cdot \mathrm{B})\rangle \,, \tag{5.2}$$

where A, B, and C take the values 0 or 1. In words, the gate leaves the computational-basis of the first and second qubits unchanged but flips the state, $|0\rangle \leftrightarrow |1\rangle$, of the third qubit if *both* the first and second qubits are in the state $|1\rangle$. The Toffoli gate and its implementation in terms of two-qubit gates is given in Fig 5.5. This involves two CNOT gates and three controlled unitary gates, in which the designated unitary transformation,

$$\begin{aligned} \hat{W} &= \frac{1-i}{2}\left(\hat{\mathrm{I}} + i\hat{\sigma}_x\right) \\ \hat{W}^\dagger &= \frac{1+i}{2}\left(\hat{\mathrm{I}} - i\hat{\sigma}_x\right) \,, \end{aligned} \tag{5.3}$$

is performed on the target qubit if the control qubit is in the state $|1\rangle$ and the identity is applied if it is in the state $|0\rangle$. The controlled unitary operation may also be implemented using CNOT gates and single-qubit gates if desired (Barnett 2009).

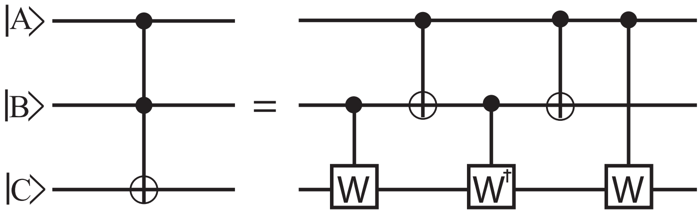


Fig. 5.5 The three-qubit Toffoli gate constructed from two-qubit gates. Reproduced, with permission, from Barnett(2009).

If we can combine enough qubits and produce enough gates then we can produce any desired unitary transformation. This would then constitute a quantum processor, designed to take input data in the form of an input multi-qubit state and generate output in the form of another state. The information could then be extracted by means of a measurement on the qubits. What is necessary in order to achieve this goal? The answer was provided by DiVincenzo (1996) in the form of five critteria for implementing a quantum computer. These have been modified somewhat over the intervening period, but the original five serve to convey the key requirements:

1) We need well-defined extendible qubit array that is stable.

2) The qubit array should be preparable in a suitable starting state, such as that in which all the qubits are in the state $|0\rangle$.

3) We need good isolation from the environment, i.e. long coherence times.

4) It must be possible to perform a universal set of gate operations, such as single-qubit rotations and CNOT operations for any chosen pair of qubits.

5) Finally, we need to be able to perform something close to ideal von Neumann measurements on each of the qubits.

These demands present a significant technical challenge and, although great advances have been made, I think it safe to say that, at present, not proposed implementation of a quantum processor has managed to achieve them all.

We should note that the gate model of a quantum processor, described above, is not the only possible one. An alternative is provided by the use of cluster states (Raussendorf *et al* 2003). In this approach we first prepare a highly entangled state of our qubits and then proceed by performing a sequence of single-qubit measurements followed by single-qubit unitary transformations the form of which depends on the preceding measurement results.

5.3 Principles of quantum computation

The basic idea of a quantum computation is quite simple. We start with an input string of bits, which we encode onto the initial states of our by preparing each in one of the states $|0\rangle$ or $|1\rangle$. We then use our collection of gates to perform on this state a unitary transformation:

$$\begin{aligned}101101001 &\to |1\rangle \otimes |0\rangle \otimes |1\rangle \otimes |1\rangle \otimes |0\rangle \otimes |1\rangle \otimes |0\rangle \otimes |0\rangle \otimes |1\rangle \otimes |1\rangle \\ &\to \hat{U}|1\rangle \otimes |0\rangle \otimes |1\rangle \otimes |1\rangle \otimes |0\rangle \otimes |1\rangle \otimes |0\rangle \otimes |0\rangle \otimes |1\rangle \otimes |1\rangle\,. \qquad (5.4)\end{aligned}$$

The computation is completed by measuring each qubit to give back a (classical) bit string which, hopefully, is the desired output:

$$\hat{U}|1\rangle \otimes |0\rangle \otimes |1\rangle \otimes |1\rangle \otimes |0\rangle \otimes |1\rangle \otimes |0\rangle \otimes |0\rangle \otimes |1\rangle \otimes |1\rangle \to 000111010\,. \qquad (5.5)$$

This is not quite all there is to it. There is a problem if the function we wish to calculate gives the same value for two distinct inputs. Consider, for example, the two-bit transformation in two bits, A and B, are transformed into A and (A AND B):

$$00 \to 00 \quad 01 \to 00 \quad 10 \to 10 \quad 11 \to 11\,. \qquad (5.6)$$

The difficulty in realising this computation on a quantum processor in the manner indicated above is that the transformation must be unitary and that unitary transformation maintain the overlap of the initial states and our computation requires

$$|0\rangle \otimes |0\rangle \to |0\rangle \otimes |0\rangle \qquad |0\rangle \otimes |1\rangle \to |0\rangle \otimes |0\rangle$$

$$\langle 0,0|0,1\rangle = 0 \quad \longrightarrow \quad \langle 0,0|0,0\rangle = 1\,, \qquad (5.7)$$

which clearly violates this requirement.

In order to be able to compute any function, we input into our quantum processor not one string of qubits but two. Let the input states of the two strings be $|a\rangle$ and $|b\rangle$. The first of these is the input data, with the sequence of bits encoded onto the qubits as in Eq (5.4). The second string is to act as our output and is often prepared in the state $|b\rangle = |0\rangle^{\otimes N}$, so that every qubit is in the state $|0\rangle$. If the processor is to calculate the Boolean function $f(a)$ then we require it to realise the transformation

$$|a\rangle \otimes |b\rangle \to |a\rangle \otimes |b \oplus f(a)\rangle\,, \qquad (5.8)$$

where $\oplus$ denotes the bit-wise modulo addition. (See in Fig 5.6.) It is straightforward to

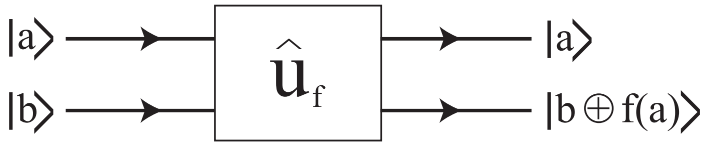


Fig. 5.6 A quantum processor designed to compute the function $f(a)$. Reproduced, with permission, from Barnett(2009).

confirm, by explicit construction, that such a unitary transformation is always possible:

$$\hat{U}_f = \sum_a |a\rangle\langle a| \otimes \left(|f(a)\rangle\langle a| + |a\rangle\langle f(a)| + \sum_{b \neq a, f(a)} |b\rangle\langle b| \right) . \tag{5.9}$$

Of course this is not the only unitary operator that performs this task, but this simple operator suffices to demonstrate that a suitable unitary operator exists.

The remarkable properties of a quantum computer derive largely from the fact that we can put into the processor not just a single number but rather a superposition of many. If each qubit in the first qubit string is prepared in the state $2^{-1/2}(|0\rangle + |1\rangle)$ then the string of N qubits is in a superposition of every number between 0 and $2^N - 1$:

$$2^{-N/2}\left(|0\rangle + |1\rangle\right)^{\otimes N} = 2^{-N/2} \sum_{a=0}^{2^N-1} |a\rangle . \tag{5.10}$$

The linearity of quantum mechanics means that the output state is an entangled one in which the function $f(a)$ has been calculated for *every* possible input number, as in Fig 5.7.

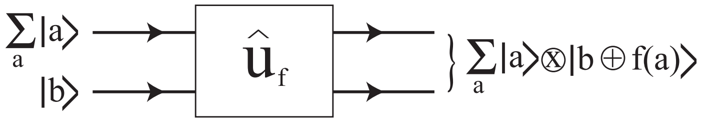


Fig. 5.7 If a superposition of possible input numbers is prepared then the output state is an entangled one in which values of the function for each of the inputs appear. Reproduced, with permission, from Barnett(2009).

A simple example may serve to illustrate the potential of a quantum processor. With this aim in mind we present Deutsch's algorithm and its many-bit extension, the Deutsch-Jozsa algorithm (Deutsch 1985, Deutsch and Jozsa 1992, Cleve *et al* 1998). Although somewhat contrived in nature, these serve to illustrate in a very simple way the potential for speed-up offered by a quantum computer. These are examples of a class of challenges called oracle-problems.

We start with Deutsch's algorithm (Deutsch 1985). Let us suppose that we have a 'black box' (an oracle) and that this device evaluates a one-bit function of a one-bit input. This means that we input a single bit, with value either 0 or 1, and that this generates an output value of either 0 or 1. There are four possible such functions, two constant functions:

$$\begin{aligned} & f(0) = 0 \qquad f(1) = 0 \,, \\ \text{and} \quad & f(0) = 1 \qquad f(1) = 1 \,, \end{aligned} \tag{5.11}$$

and two balanced functions (balanced in the sense that both possible bit values occur in the output, one for each input):

$$f(0) = 0 \qquad f(1) = 1 \,,$$

$$\text{and} \quad f(0) = 1 \qquad f(1) = 0\,. \tag{5.12}$$

If we want to know which function the oracle calculates then we need to input both bits values and examine the corresponding outputs. Let us suppose, however, that our task is simply to determine whether the function calculated is a constant function or a balanced one. Classically, of course, this requires us to input both possible bit values and so also requires two computations. A suitable quantum processor, however, can do this in a single step. Consider the general transformation given in Eq. (5.8), but with the two subit strings each replaced by a single qubit so that the transformation enacted by the processor is

$$|\mathrm{A}\rangle \otimes |\mathrm{B}\rangle \rightarrow |\mathrm{A}\rangle \otimes |\mathrm{B} \oplus f(\mathrm{A})\rangle\,. \tag{5.13}$$

We can make use of the superposition principle to prepare *both* of our qubits in a superposition state and so produce the transformation

$$\begin{aligned}\frac{1}{2}(|0\rangle + |1\rangle) \otimes (|0\rangle - |1\rangle) \rightarrow &\ \frac{1}{2}\left(|0\rangle \otimes |f(0)\rangle - |0\rangle \otimes |\bar{f}(0)\rangle + |1\rangle \otimes |f(1)\rangle - |1\rangle \otimes |\bar{f}(1)\rangle\right) \\ =&\ \frac{1}{2}\left[|0\rangle \otimes \left(|f(0)\rangle - |\bar{f}(0)\rangle\right) + |1\rangle \otimes \left(|f(1)\rangle - |\bar{f}(1)\rangle\right)\right] \\ =&\ \frac{1}{2}\Big((-1)^{f(0)}|0\rangle + (-1)^{f(1)}|1\rangle\Big) \otimes (|0\rangle - |1\rangle)\,. \end{aligned} \tag{5.14}$$

A simple measurement carried out on the first qubit then tells us what we need to know: if it is found to be in the state $2^{-1/2}(|0\rangle + |1\rangle)$ then the function is constant and if we find it in the orthogonal state $2^{-1/2}(|0\rangle - |1\rangle)$ then the function is balanced. We have found out what we need to know by addressing the oracle only once, rather than twice as would be required classically.

The Deutsch algorithm may seem, at least at first sight, not to be an especially convincing illustration, after all we do input two qubits into the processor. The real power appears when we consider an extension of it to the Deutsch-Jozsa algorithm. In this algorithm, our input is an n-bit number a and our output is again a single bit, 0 or 1. The function is either constant, giving either 0 or 1 for all inputs, or balanced, giving 0 for half of the inputs and 1 for the other half. Our task is to determine an algorithm that determines *with certainty* whether the function is constant or balanced. We can start putting in different strings and has soon as we get two different output values then we know for certain that the function compute is balanced. To know its nature for certain, however, we may have to input over half of the possible inputs which means addressing the oracle $2^{n-1} + 1$ times. A quantum processor can achieve this task in a single shot, however. To see how this works, we let the first string consist of n qubits, each prepared in the superposition state $2^{-1/2}(|0\rangle + |1\rangle)$ with the second string being just a single qubit prepared in the state $2^{-1/2}(|0\rangle - |1\rangle)$. The oracle then performs the transformation

$$2^{-(n+1)/2}(|0\rangle + |1\rangle^{\otimes n}) \otimes (|0\rangle - |1\rangle) \rightarrow 2^{-(n+1)/2} \sum_{a=0}^{2^n-1} (-1)^{f(a)}|a\rangle \otimes (|0\rangle - |1\rangle)\,. \tag{5.15}$$

If the function is constant then the first string of n qubits remain in their input state, as equally-weighted superposition of the states $|a\rangle$, but if the function is balanced then the qubit-string will be in a state orthogonal to this. Hence only a single computation is required to determine whether the function is constant or balanced. This represents an improvement over the classical requirement of $2^{(n-1)}+1$ that is *exponential* in n, the number of bits. Such dramatic improvements are characteristic of a number of quantum algorithms, but not all.

5.4 Quantum algorithms

Let us suppose that we have a suitable quantum processor, what might we do with it? One important thing that we might do is to use it to simulate a complicated quantum process (Feynman 1982). As quantum systems get larger it becomes ever more difficult to simulate them; were this not the case, then we could emulate a large quantum computer on a classical one. We might also use a quantum computer to speed-up searching in an unstructured database using Grover's algorithm, which provides a dramatic improvement (albeit not an exponential one) over classical methods (Grover 1998). The most dramatic possibility to date, however, is Shor's algorithm for determining the two prime factors of a large number. You will recall that it is the difficulty in performing this task that underlies to security of the RSA public-key cryptosystem and with it, much of the world's secure communications (Shor 1997). It was Shor's proposal more than any other single factor that truly changed quantum computation, and with it quantum information, from a small-scale research field into a major international endeavour. There is space, in these lectures, to present briefly only one quantum algorithm and, because of its significance, we choose Shor's algorithm. Several others, including Grover's algorithm, and a more complete presentation of Shor's algorithm may be found in (Barnett 2009).

At the heart of Shor's algorithm is the remarkable ability of a quantum computer to perform, highly efficiently, a quantum Fourier transform and hence to find the period of a function. To see why this helps in factoring, we need to consider an idea from number theory. We start with three bits of input data: N, the number to be factorized, m, a small integer chosen at random and the non-negative integers, $n = 0, 1, 2, \cdots$. We first make the series $F_N(n) = m^n \mathrm{mod} N$. If we can then find the period of this function, r, such that $F_N(n+r) = F_N(n)$, then the greatest common divisor of $m^{r/2} \pm 1$ and N divides N. In other words, one of the factors of each of $m^{r/2} \pm 1$ is also a prime factor of N. There are, of course, some additional subtleties, but the number theory to prove this is by no means difficult, and may be found in Barnett (2009). Let us consider two examples to demonstrate the idea. Let us consider the problem of factoring 15 by means of this process. Let us first select $m = 2$, so that our series is

$$\begin{aligned} 2^0 \mathrm{mod} 15 &= 1 \\ 2^1 \mathrm{mod} 15 &= 2 \\ 2^2 \mathrm{mod} 15 &= 4 \\ 2^3 \mathrm{mod} 15 &= 8 \\ 2^4 \mathrm{mod} 15 &= 1 \\ \vdots \quad &\quad \vdots . \end{aligned} \tag{5.16}$$

We see that the period is 4 and hence

$$\begin{aligned} m^{r/2}+1&=5\\ m^{r/2}-1&=3 \end{aligned} \tag{5.17}$$

and these are the required prime factors. Let us try another example, $m = 11$, for which our series is

$$\begin{aligned} 11^0\mathrm{mod}15&=1\\ 11^1\mathrm{mod}15&=11\\ 11^2\mathrm{mod}15&=1\\ \vdots\quad&\quad\vdots, \end{aligned} \tag{5.18}$$

so the period is 2. This gives

$$\begin{aligned} m^{r/2}+1&=12\\ m^{r/2}-1&=10\,. \end{aligned} \tag{5.19}$$

The greatest common divisor of 12 and 15 is 3, which is one of the prime factors we seek and the creates common divisor of 10 and 15 is 5, which is the other factor.

So why do we need a quantum computer to run this seemingly simple algorithm? The answer is that each step can be run efficiently, even for very large numbers, on a classical computer with the single exception of finding the period of the function $F_N(n)$. This is a very difficult task and it is no exaggeration to state that it is this period finding problem that underlies the security of the RSA cryptosysem.

Central to Shor's algorithm for factoring on a quantum computer is the quantum Fourier transform (Nielsen and Chuang 2000, Mermin 2007, Barnett 2009). We do not attempt a detailed account of the quantum Fourier transform, but present instead only an outline of the steps involved in Shor's algorithm. It is, perhaps, clearest to give a list of the five major steps:

Shor's algorithm to find the two prime factors of N

1. We start by finding two integers, q and M, such that

$$q = 2^M > N^2 \tag{5.20}$$

and prepare two registers, each containing M qubits.

2. We set each on the qubits in the first register to the state $2^{-1/2}(|0\rangle+|1\rangle)$ and each in the second register in the state $|0\rangle$, so that the state input into our quantum processor is

$$|\psi\rangle = \frac{1}{\sqrt{q}}\sum_{n=0}^{q-1}|n\rangle_1\otimes|0\rangle_2\,, \tag{5.21}$$

where we have used the short-hand notation

$$|0\rangle = |00\cdots000\rangle$$

$$\begin{aligned}|1\rangle &= |00\cdots001\rangle\\ |2\rangle &= |00\cdots010\rangle\\ |3\rangle &= |00\cdots011\rangle\\ \vdots \qquad & \qquad \vdots\,. \end{aligned} \tag{5.22}$$

3. Next we chose an integer m at random and use the quantum processor to entangle the two registers so that

$$|\psi\rangle \to \frac{1}{\sqrt{q}} \sum_{n=0}^{q-1} |n\rangle_1 \otimes |m^n \mathrm{mod} N\rangle_2\,. \tag{5.23}$$

This can be achieved efficiently by means of a unitary transformation on a suitably programmed quantum computer.

4. Next comes the crucial quantum Fourier transform, which again can be performed efficiently as a unitary transform on a quantum computer. We use our quantum computer to perform a quantum Fourier transform on the first register to generate the transformation

$$\frac{1}{\sqrt{q}} \sum_{n=0}^{q-1} |n\rangle_1 \otimes |m^n \mathrm{mod} N\rangle_2 \to \frac{1}{q} \sum_{n=0}^{q-1} \sum_{k=0}^{q-1} |k\rangle_1 \otimes |m^n \mathrm{mod} N\rangle_2 \exp\left(i2\pi \frac{kn}{q}\right)\,. \tag{5.24}$$

Let us pause to think and see what has been achieved by all of this. The number $m^n \mathrm{mod} N$ has period r, which means, of course, that

$$m^{n+sr} \mathrm{mod} N = m^n \mathrm{mod} N\,, \tag{5.25}$$

so that we have some identical states in the second register:

$$|m^n \mathrm{mod} N\rangle = |m^{n+r} \mathrm{mod} N\rangle = |m^{n+2r} \mathrm{mod} N\rangle = \cdots\,. \tag{5.26}$$

The coefficients of these states in our state will have the phases are

$$\exp\left(i2\pi \frac{k(n+sr)}{q}\right) = \exp\left(i2\pi \frac{kn}{q}\right) \times \exp\left(i2\pi \frac{ksr}{q}\right)\,, \tag{5.27}$$

which will be in phase (or nearly in phase) for $k \approx q/r$.

5. Finally, we need only measure the first register in the computational basis and we will find, because of constructive interference, with high probability a value k for which the amplitude is large, one for which $k \approx q/r$, or $r \approx q/k$. This allows us to greatly narrow the range of allowed values of r and hence greatly restrict the range of possible factors. All possible candidate factors can be checked efficiently and simply by dividing them into N.

This is just one example, albeit the most prominent and newsworthy, of an efficient quantum algorithm. There are others and the field of quantum algorithm design promises to gain in importance as the first quantum processors become available.

5.5 Errors and decoherence

So what stops us building a quantum processor, revolutionising computation and breaking into RSA? The short answer is that we don't yet have any suitably scalable implementation of quantum logic elements such as qubits and gates.

Many technical challenges remain to be solved before we can build a quantum computer to challenge current devices based on classical logic. I mention here only one problem, that of decoherence. In a quantum computer our information is encoded onto two-state quantum systems, our qubits and these are made to interact with each other by the implementation of desired Hamiltonians. Yet the quantum natures, both of the qubits and of the Hamiltonian, makes them extremely sensitive to their environment and even weak interactions can have a disastrous effect. As an illustration, let us consider the effect of a single-qubit error in an implementation of Deutsch's algorithm. To keep things simple, let us consider only a phase error:

$$|0\rangle \to |0\rangle \qquad |1\rangle \to -|1\rangle\,, \tag{5.28}$$

and a bit-flip error

$$|0\rangle \to |1\rangle \qquad |1\rangle \to |0\rangle\,. \tag{5.29}$$

Recall that we obtain the desired information, as to whether the computed function is constant or balanced, by measuring the first qubit to be in the state $2^{-1/2}(|0\rangle + |1\rangle)$ or $2^{-1/2}(|0\rangle - |1\rangle)$, respectively. If a bit-flip error occurs for this qubit then we are safe, but if a phase-error occurs then we get the wrong answer!

Clearly we need to work hard to suppress all sources of errors and decoherence, but this is not the end of the problem. For large-scale computations we need a large number of qubits and the scaling is not at all favourable. To see this let us suppose that the probability that a single qubit has no error in a time t is

$$\exp(-\Gamma t)\ .$$

If we have n qubits then the probability that *none* of these experiences an error in this time is

$$\exp(-n\Gamma t)\ ,$$

which is exponential in n. But worse is to come. Let us suppose that t is the typical amount of time it takes to perform a single gate operation. If we need a sequence of m of these then we find

$$\exp(-nm\Gamma t)\ .$$

Typically we might need each qubit to interact with every other qubit so m may be of the same rode as n. We do not need to perform the gate operations one at a time, of course, and by suitably optimising the order of operations, we might have $m \approx \log n$ to give a final zero-error probability of

$$\exp(-n \log n \Gamma t)\ . \tag{5.30}$$

For 300 qubits, the *exponent* is about 2,000 times smaller than the single-qubit and single-gate error rate, $e^{-\Gamma t}$, that we started with. The zero-error probability is the $2,000^{th}$ *power* of the no error probability for a single qubit and a single-gate operation.

Clearly we need to control sources of decoherence to an extraordinary degree but ultimately this will always be a losing battle. Is it all hopeless? By no means; the solution is essentially the same as we encountered in the first lecture. We can combat errors in a quantum computer by redundancy, just as we do to combat classical errors. This is a more subtle problem than in the classical regime, however. Firstly there is the no-cloning theorem which tells us that we cannot literally make multiple copies of our qubits and there is also the problem that quantum measurements, if not performed carefully, will modify the very qubit states that we are trying to protect. Nevertheless, there are protocols detecting and correcting qubit errors. Perhaps the most important of these is the Steane code, in which each logical qubit is ended onto seven logical qubits (Steane 1996, Barnett 2009), but a description of how this works will have to wait for next time.

References

Aharonov, Y., Albert, D. Z. and Vaidman, L. (1988) How the result of a measurement of a component of spin of a spin$-\frac{1}{2}$ particle can turn out to be 100. *Phys. Rev. Lett.* **60**, 1351–1354.

Barnett, S. M. and Riis, E. (1997) Experimental demonstration of polarization discrimination at the Helstrom bound. *J. Mod. Opt.* **44**, 1061–1064.

Barnett, S. M. and Andersson, E. (2002) Bound on measurement based on the no-signalling condition. *Phys. Rev. A* **65**, 044307.

Barnett, S. M. (2009) *Quantum information.* Oxford University Press, Oxford.

Barnett, S. M. and Croke, S. (2009a) On the conditions for discrimination between quantum states with minimum error. *J. Phys. A: Math. Theor.* **42**, 062001.

Barnett, S. M. and Croke, S. (2009b) Quantum state discrimination. *Adv. Opt. Photon.* **1**, 238–278.

Bayes, T. (1763) An essay towards solving a problem in the doctrine of chances. *Phil. Trans. R. Soc.* **53**, 370–418.

Bell, J. S. (1964) On the Einstein-Podolsky-Rosen paradox. *Physics* **1**, 195–200. Reprinted in (Wheeler and Zurek 1983; Bell 1987).

Bell, J. S. (1987) *Speakable and unspeakable in quantum mecahnics.* Cambridge University Press, Cambridge.

Bennett, C. H. and Brassard, G. (1984) Quantum cryptography: public key distribution and coin tossing. *Proceedings of IEEE international conference on computers systems and signal processing, Bangalore India* 175–179.

Bennett, C. H. and Wiesner, S. J. (1992) Communication via one- and two-particle operators on Einsetin-Podolsky-Rosen states. *Phys. Rev. Lett.* **69**, 2881–2884.

Bennett, C. H., Brassard, G., Crépeau, C., Jozsa, R., Peres, A. and Wootters, W. K. (1993) Teleporting an unknown quantum state via dual Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **70**, 1895–1899.

Bergou, J. (2007) Quantum state discrimination and selected applications. *J. Phys. Conf. Ser.* **84**, 012002.

Boas, M. L. (1983) *Mathematical methods in the physical sciences.* Wiley, New York.

Bohm, D. (1951) *Quantum theory*. Prentice-Hall, New Jersey. Reprinted by Dover, New York (1989).

Bohr, N. (1935) Can quantum-mechanical description of reality be considered complete?. *Phys. Rev.* **47**, 696–702.

Bouwmeester, D., Ekert, A. and Zeilinger, A. (eds) (2000) *The physics of quantum information.* Springer-Verglag, Berlin.

Box, G. E. P. and Tiao, G. C. (1973) *Bayesian inference in statistical analysis.* Wiley, New York.

Bretthorst, G. L. (1988) *Bayesian spectrum analysis and parameter estimation.* Springer-Verlag, New York.

Brillouin, L. (1956) *Science and information theory*. Academic Press, New York.

Buchmann, J. A. (2001) *Introduction to cryptography*. Spring, New York.

Bužek, V. and Hillery, M. (1996) Quantum copying: beyond the no-cloning theorem. *Phys. Rev. A* **54**, 1844–1852.

Chefles, A. (2000) Quantum state discrimination. *Contemp. Phys.* **41**, 401–424.

Clarke, R. B. M., Chefles, A., Barnett, S. M. and Riis, E. (2001a) Experimental demonstration of optimal unambiguous state discrimination. *Phys. Rev. A* **63**, 040305(R).

Clarke, R. B. M., Kendon, V. M., Chefles, A., Barnett, S. M. and Riis, E. (2001b) Experimental realization of optimal detection strategies for overcomplete states. *Phys. Rev. A* **64**, 012303.

Clauser, J. F., Horne, M. A., Shimony, A. and Holt, R. A. (1969) Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **23**, 880–884.

Cleve, R., Ekert, A., Macchiavello, C. and Mosca, M. (1998) Quantum algorithms revisited *Proc. R. Soc. Lond. A* **454**, 339–354.

Conan Doyle, A. (1903) The adventure of the dancing men. *Strand Magazine* **26** December issue. Reprinted in *The original illustrated Sherlock Holmes*, Castle, New Jersey.

Cover, T. M. and Thomas, J. A. (1991) *Elements of information theory.* Wiley, New York.

Croke, S., Barnett, S. M. and Stenholm, S. (2008) Linear transformation of quantum states. *Ann. Phys.* **323**, 893–906.

Deutsch, D. (1985) Quantum theory, the Church-Turing principle and the universal quantum computer. *Proc. R. Soc. Lond. A* **400**, 97–117.

Deutsch, D. and Jozsa, R. (1992) Rapid solution of problems by quantum computation. *Proc. Roy. Soc. Lond. A* **439**, 553–558.

Dieks, D. (1982) Communication by EPR devices. *Phys. Lett. A* **92**, 271–272.

Dieks, D. (1988) Overlap and distinguishability of quantum states. *Phys. Lett. A* **126**, 303–306.

DiVincenzo, D. (1996) Topics in quantum computers. http://arxiv.org/abs/cond-mat/9612126.

Einstein, A., Podolsky, B. and Rosen, N. (1935) Can quantum-mechanical description of reality be considered complete?. *Phys. Rev.* **47**, 777–780.

Feynman, R. P. (1982) Simulating physics with computers. *Int. J. Theo. Phys.* **B2**, 467–488.

Gay, S. and Mackie, I. (eds) (2010) *Semantic techniques in quantum computation* Cambridge University Press, Cambridge.

Ghirardi, G. C., Rimini, A. and Weber, T. (1980) A general argument against superluminal transmissions through the quantum mechanical measurement process. *Lett. al. Nuovo Cim.* **27**, 293–298.

Gisin, N. (1991) Bell's inequality holds for all non-product states. *Phys. Lett. A* **154**, 201–202. It should be noted that the paper contains some typographical errors, the most serious of which is the title; the paper proves the exact opposite of what is stated in the title!

Gisin, N., Ribordy, G., Tittel, W. and Zbinden, H. (2002) Quantum cryptography. *Rev. Mod. Phys.* **74**, 145–195.

Goldie, C. M. and Pinch, R. G. E. (1991) *Communications theory*. Cambridge University Press, Cambridge.

Goldstein, S. (1994) Nonlocality without inequalities for almost all entangled states of two particles. *Pys. Rev. Lett.* **72**, 1951–1951.

Greenberger, D. M., Horne, M. A., Shimony, A. and Zeilinger, A. (1990) Bell's theorem without inequalities. *Am. J. Phys.* **58**, 1131–1143.

Grover, L. (1998) Quantum computers can search rapidly by using almost any transformation. *Phys. Rev. Lett.* **80**, 4329–4332.

Hamming, R. W., (1980) *Coding and information theory*. Prentice-Hall, London.

Hardy, L. (1993) Nonlocality for two particles without inequalities for almost all entangled states. *Phys. Rev. Lett.* **71**, 1665–1668.

Helstrom, C. W. (1976) *Quantum detection and estimation theory*. Academic Press, New York.

Holevo, A. S. (1973) Statistical decision theory for quantum systems. *J. Multivariate Anal.* **3**, 337–394.

Holevo, A. S. (1982) *Probabilistic and statistical aspects of quantum theory*. North Holland, Amsterdam.

Holevo, A. S. (2001) *Statistical structure of quantum theory*. Springer-Verlag, Berlin.

Hunter, K. (2003) Measurement does not always aid state discrimination. *Phys. Rev. A* **68**, 012306.

Huttner, B., Muller, A. Gautier, G., Zbinden, H. and Gisin, N. (1996) Unambiguous quantum measurement of nonorthogonal quantum states. *Phys. Rev. A* **54**, 3783–3789.

Ivanovic, I. D. (1987) How to differentiate between non-orthogonal states. *Phys. Lett. A* **123**. 257–259.

Jaynes, E. T. (1957a) Information theory and statistical mechanics I. *Phys. Rev.* **106**, 620–630.

Jaynes, E. T. (1957b) Information theory and statistical mechanics II. *Phys. Rev.* **108**, 171–190.

Jaynes, E. T. (2003) *Probability theory the logic of science*. Cambridge University Press, Cambridge.

Jeffreys, H. (1939) *The theory of probability*. Clarendon Press, Oxford.

Kaye, P., Laflamme, R. and Mosca, M. (2007) *An Introduction to quantum computing* Oxford University Press, Oxford.

Khinchin, A. I. (1957) *Mathematical foundations of information theory*. Dover, New York.

Kullback, S. (1959) *Information theory and statistics*. Dover, New York.

Kraus, K. (1983) *States, effects and operations*. Springer-Verlag, Berlin.

Landauer, R. (1961) Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.* **5**, 183–191. Reprinted in (Leff and Rex 1990, 2003).

Lee, P. M. (1989) *Bayesian statistics: an introduction*. Edward Arnold, London.

Leff, H. S. and Rex, A. F. eds. (1990) *Maxwell's demon: entropy, information, computing*. Adam Hilger, Bristol.

Leff, H. S. and Rex, A. F. (1994) Entropy of measurement and erasure: Szilard's membrane model revisited. *Am. J. Phys.* **52**, 3495–3499. Reprinted in (Leff and Rex 2003).

Leff, H. S. and Rex, A. F. (2003) *Maxwell's demon 2: entropy, classical and quantum information, computing*. Institute of Physics Publishing, Bristol.

Loepp, S and Wootters, W. K. (2006) *Protecting information: from classical error correction to quantum cryptography*. (Cambridge University Press, Cambridge).

Mattle, K., Weinfurter, H., Kwiat, P. G. and Zeilinger, A. (1996) Dense coding in experimental quantum communication. *Phys. Rev. Lett.* **76**, 4656–4659.

Maxwell, J. C. (1871) *Theory of heat*. Longmans, Green, and Co., London.

Mcgrayne, S. B. (2011) *The theory that would not die*. Yale University Press, New Haven.

Mermin, D. (1990) What's wrong with these elements of reality? *Physics Today* June issue 9–11

Mermin, D. (2007) *Quantum computer science* Cambridge University Press, Cambridge.

Moore, G. E. (1965) Cramming More Components onto Integrated Circuits *Proc. IEEE* **86** 82–85.

Nielsen, M. A. and Chuang, I. L. (2000) *Quantum computation and quantum information* Cambridge University Press, Cambridge.

Pachos, J. K. (2012) *Introduction to topological quantum computation* Cambridge University Press, Cambridge.

Peres, A. (1988) How to differentiate between two non-orthogonal states. *Phys. Lett. A* **128**, 19–19.

Peres, A. (1993) *Quantum theory: concepts and methods*. Academic, Dordrecht.

Plenio, M. and Vitelli, V. (2001) The physics of forgetting: Landauer's erasure principle and information theory. *Contemp. Phys.* **42**, 25–60.

Phoenix, S. J. D. and Townsend, P. D. (1995) Quantum cryptography: how to beat the code breakers using quantum mechanics. *Contemp. Phys.* **36**, 165–195.

Piper, F. and Murphy, S. (2002) *Cryptography: a very short introduction*. Oxford University Press, Oxford.

Raussendorf, R., Browne, D. E. and Briegel, H. J. (2003) Measurement-based quantum computation on cluster states. *Phys. Rev. A* **68**, 022312.

Redhead, M. (1987) *Incompleteness, Nonlocality and Realism* Clarendon Press, Oxford.

Scarani, V., Iblisdir, S., Gisin, N. and Acín, A. (2005) Quantum cloning. *Rev. Mod. Phys.* **77**, 1225–1256.

Scarani, V., Bechmann-Pasquinucci, H., Cerf, N. J., Dušek, M., Lütknehaus, N. and Peev, M. (2009) The security of practical quantum key distribution. *Rev. Mod. Phys.* **81**, 1301–1350.

Shannon, C. E. (1948) A mathematical theory of communication. *Bell Sys. Tech. J.* **27**, 379–423 and 623–656.

Shannon, C. E. (1949) Communication theory of secrecy systems. *Bell Sys. Tech. J.* **28**, 656–715.

Shannon, C. E. and Weaver, W. (1949) *The mathematical theory of communication.* University of Illinois Press, Urbana.

Shor, P. W. (1997), Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.* **26**, 1484–1509, arXiv:quant-ph/9508027v2

Singh, S. (1999) *The code book*. Fourth Estate, London.

Singh, S. (2000) *The science of secrecy*. Fourth Estate, London.

Smith, R. J. (1983) *Circuits, devices and systems* (4th edn). Wiley, New York.

Steane, A. M. Error correcting codes in quantum theory. *Phys. Rev. Lett.* **77** 793–797.

Stenholm, S. and Suominen, K.-A. (2005) *Quantum approach to informatics.* Wiley, Hoboken.

Szillard, L. (1929) Über die Entropieveminderung in einem thermodynamikschen System bei eingriffen intelligenter Wesen. *Zeits. Phys.* **53**, 840–856. Reprinted in translation in (Wheeler and Zurek 1983; Leff and Rex 1990, 2003).

Vaccaro, J. A. and Barnett, S. M. (2011) Information erasure without an energy cost. *Proc. R. Soc. A* **467**, 1770–1778.

Van Assche, G. (2006) *Quantum cryptography and secret-key distillation.* Cambridge University Press, Cambridge.

Vedral, V. (2006) *Introduction to quantum information science* Oxford University Press, Oxford.

von Neumann, J. (1955) *Mathematical foundations of quantum mechanics.* Princeton University Press, New Jersey.

Wheeler, J. A. and Zurek, W. H. (eds) (1983) *Quantum theory and measurement.* Princeton University Press, New Jersey.

Whitaker, A. (2006) *Einstein, Bohr and the quantum dilemma* (2nd edn). Cambridge University Press, Cambridge.

Wiesner, S. (1983) Conjugate coding. *SIGACT News* **15**, 78–88.

Wootters, W. K. and Zurek, W. H. (1982) A single quantum cannot be cloned. *Nature* **299**, 802–803.