7

Quantum information

In this chapter we address the problems of the storage and transmission of quantum information. Quantum cryptography should in principle be included here, but we preferred to describe it in Chapter 2, where it was used to illustrate the basic principles of quantum mechanics. In Section 7.1 we explain quantum teleportation and in Section 7.2 we give a short and schematic review of classical information theory. Section 7.3 is devoted to an introduction to the storage and communication of quantum information, and, finally, in Section 7.4 we take a quick look at the important but difficult topic of quantum error correction.

7.1 Teleportation

Teleportation is an interesting application of entangled states which may have applications to quantum information transfer (Fig. 7.1). Let us suppose that Alice wishes to transfer to Bob the information about the spin state $|\varphi_A\rangle$ of a particle A of spin 1/2,

$$|\varphi_A\rangle = \lambda|0_A\rangle + \mu|1_A\rangle, \tag{7.1}$$

which is *a priori* unknown, without sending Bob this particle directly. She cannot measure its spin, because she does not know the basis in which the spin of particle A was prepared, and any measurement would in general project $|\varphi_A\rangle$ onto another state. The principle of information transfer consists of using an auxiliary pair of entangled particles B and C of spin 1/2 shared between Alice and Bob. Particle B is used by Alice and particle C is sent to Bob (Fig. 7.1). These particles B and C may be, for example, in an entangled spin state $|\Psi_{BC}\rangle$:

$$|\Psi_{BC}\rangle = \frac{1}{\sqrt{2}}\left(|0_B 0_C\rangle + |1_B 1_C\rangle\right). \tag{7.2}$$

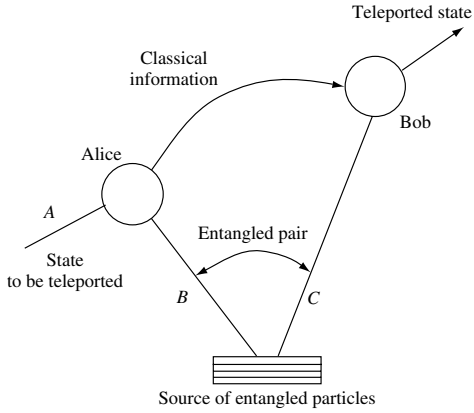

Figure 7.1 Teleportation. Alice makes a Bell measurement on the qubits A and B and informs Bob of the result by a classical path.

The initial state of the three particles $|\Phi_{ABC}\rangle$ then is

$$\begin{aligned} |\Phi_{ABC}\rangle &= (\lambda|0_A\rangle + \mu|1_A\rangle)\,\frac{1}{\sqrt{2}}\,(|0_B0_C\rangle + |1_B1_C\rangle) \\ &= \frac{\lambda}{\sqrt{2}}|0_A\rangle\,(|0_B0_C\rangle + |1_B1_C\rangle) + \frac{\mu}{\sqrt{2}}|1_A\rangle\,(|0_B0_C\rangle + |1_B1_C\rangle)\,. \end{aligned} \tag{7.3}$$

Alice first applies a cNOT gate to the qubits A and B, with the qubit A acting as the control qubit and the qubit B acting as the target qubit (Fig. 7.2). This operation transforms the initial state (7.3) of three qubits into

$$|\Phi'_{ABC}\rangle = \frac{\lambda}{\sqrt{2}}\,(|0_A\rangle(|0_B0_C\rangle + |1_B1_C\rangle) + \frac{\mu}{\sqrt{2}}\,(|1_A\rangle(|1_B0_C\rangle + |0_B1_C\rangle)\,. \tag{7.4}$$

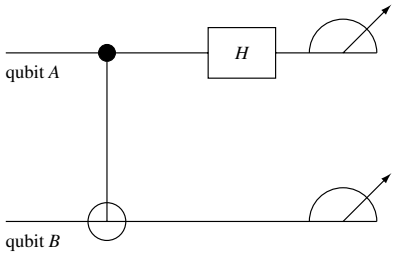

Figure 7.2 Alice applies a cNOT gate to the qubits A and B and then a Hadamard gate to the qubit A.

Alice then applies a Hadamard gate to the qubit A, which transforms (7.4) into

$$|\Phi''_{ABC}\rangle = \frac{1}{2}\Big[\lambda|0_A0_B0_C\rangle + \lambda|0_A1_B1_C\rangle + \lambda|1_A0_B0_C\rangle + \lambda|1_A1_B1_C\rangle + \mu|0_A1_B0_C\rangle + \mu|0_A0_B1_C\rangle - \mu|1_A1_B0_C\rangle - \mu|1_A0_B1_C\rangle\Big]. \tag{7.5}$$

This equation can be rewritten as

$$\begin{aligned}|\Phi''_{ABC}\rangle = &\frac{1}{2}|0_A0_B\rangle\,(\lambda|0_C\rangle + \mu|1_C\rangle)\\ &+\frac{1}{2}|0_A1_B\rangle\,(\mu|0_C\rangle + \lambda|1_C\rangle)\\ &+\frac{1}{2}|1_A0_B\rangle\,(\lambda|0_C\rangle - \mu|1_C\rangle)\\ &+\frac{1}{2}|1_A1_B\rangle\,(-\mu|0_C\rangle + \lambda|1_C\rangle)\,.\end{aligned} \tag{7.6}$$

The last operation that Alice performs is measurement of the two qubits in the basis $\{|0\rangle, |1\rangle\}$. The joint measurement by Alice of qubits A and B is called a *Bell measurement*. This measurement projects the pair (AB) onto one of the four states $|i_Aj_B\rangle$, $i, j = 0, 1$, and the state vector of the qubit C is then read on each of the lines of (7.6).

The simplest case occurs when the measurement result is $|0_A0_B\rangle$. The qubit C then reaches Bob in the state

$$\lambda|0_C\rangle + \mu|1_C\rangle,$$

that is, in the initial state of the qubit A, with the *same* coefficients λ and μ. So, Alice informs Bob by a classical channel (for example, a telephone) that the qubit will reach him in the same state as the qubit A. If on the contrary she measures $|0_A1_B\rangle$, the qubit C is in the state

$$\mu|0_C\rangle + \lambda|1_C\rangle,$$

she then informs Bob that he must apply to qubit C a rotation of π about Ox, or, equivalently, the matrix σ_x:

$$\exp\left(-\mathrm{i}\frac{\pi\sigma_x}{2}\right) = -\mathrm{i}\sigma_x.$$

In the third case ($|1_A0_B\rangle$) it is necessary to apply a rotation of π about Oz, and in the final case ($|1_A1_B\rangle$) a rotation of π about Oy. We note that in these four cases Alice does not know the coefficients λ and μ, and she sends Bob only the information about which rotation he should apply.

It is useful to add a few final remarks.

- The coefficients λ and μ are never measured, and the state $|\varphi_A\rangle$ is destroyed during Alice's measurement. There is therefore no contradiction with the no-cloning theorem.
- Bob "knows" the state of particle C only once he has received the result of Alice's measurement. This information must be sent by a classical channel, at a speed at most equal to that of light. There is therefore no instantaneous transmission of information at a distance.
- Teleportation never involves the transport of matter.

7.2 Shannon entropy

The two fundamental theorems of information theory were stated by Shannon in 1948. Before discussing their quantum generalization, let us give a very schematic review of these theorems without going into detailed proofs. These theorems answer the following questions.

(i) What is the maximal compression that can be applied to a message? In other words, how can redundant information be quantified?
(ii) At what rate can one communicate via a noisy channel, that is, what redundancy must be incorporated in a message to protect against errors?

It can easily be seen from the following example that a message can be compressed when compared to its naive encoding. Let us suppose that we are using four different letters, (a_0, a_1, a_2, a_3), which we can encode in the usual manner using two bits: $a_0 = 00$, $a_1 = 01$, $a_2 = 10$, and $a_3 = 11$. A message n letters long will then be encoded by $2n$ bits. However, suppose that the letters occur with different probabilities: a_0 with probability 1/2, a_1 with probability 1/4, and a_2 and a_3 with probability 1/8. We can then use the following encoding: $a_0 = 0$, $a_1 = 10$, $a_2 = 110$, and $a_3 = 111$. This can easily be verified to be unambiguous: a letter stops after every 0, or after a sequence of three 1. The average length of a message n letters long will then be

$$n\left(\frac{1}{2}1+\frac{1}{4}2+\frac{1}{4}3\right)=\frac{7}{4}n<2n.$$

Shannon's first theorem shows that this is in fact the best possible compression. Let us take a set of letters a_x, $0 \leq x \leq k$, and a sequence $\{a_1, \ldots, a_n\}$ of n letters forming a message. Each letter occurs *a priori* with a probability $\mathsf{p}(a_x)$, $\sum_x \mathsf{p}(a_x) = 1$. We consider a message n letters long, $n \gg 1$. Is it possible to compress the message into a shorter sequence containing essentially the same

information? The simplest case is that of two letters, $\mathsf{p}(a_0) = \mathsf{p}$, $\mathsf{p}(a_1) = 1 - \mathsf{p}$. The probability $\mathsf{p}(q)$ that an n-letter message contains q letters a_0 is [1]

$$\mathsf{p}(q) = \binom{n}{q} \mathsf{p}^q (1-\mathsf{p})^{n-q}.$$

Let us find the maximum of $\mathsf{p}(q)$ by using Stirling's formula $\ln n! \simeq n \ln n/\mathrm{e}$, from which $\mathrm{d} \ln n!/\mathrm{d}n \simeq \ln n$. We compute the q-derivative of $\mathsf{p}(q)$

$$\frac{\mathrm{d} \ln \mathsf{p}(q)}{\mathrm{d}q} = -\ln q + \ln(n-q) + \ln \mathsf{p} - \ln(1-\mathsf{p})$$

which vanishes for $q = \overline{q} = n\mathsf{p}$. As was to be expected, the most probable value [2] of q is $\overline{q} = n\mathsf{p}$. The dispersion around $\overline{q}$ is found from the second derivative of $\ln \mathsf{p}(q)$

$$\left. \frac{\mathrm{d}^2 \ln \mathsf{p}(q)}{\mathrm{d}q^2} \right|_{q=\overline{q}} = -\frac{1}{n\mathsf{p}(1-\mathsf{p})}$$

so that

$$\langle \Delta q^2 \rangle = \langle (q - \overline{q})^2 \rangle = n\mathsf{p}(1-\mathsf{p}).$$

With a negligible error when $n \to \infty$, the variable q lies in the range

$$n\mathsf{p} - \mathcal{O}(\sqrt{n}) \lesssim q \lesssim n\mathsf{p} + \mathcal{O}(\sqrt{n}).$$

The number of occurrences of the letter a_0 in an n-letter message will lie in this range, and the number of typical messages (or sequences) will be of order $\binom{n}{n\mathsf{p}}$. Instead of coding 2^n sequences, it is sufficient to code the $\simeq \binom{n}{n\mathsf{p}}$ *typical sequences*. Stirling's formula allows us to compute $\ln \binom{n}{n\mathsf{p}}$

$$\ln \binom{n}{n\mathsf{p}} \simeq -n[\mathsf{p} \ln \mathsf{p} + (1-\mathsf{p}) \ln(1-\mathsf{p})] = n\overline{H}_{\mathrm{Sh}}(\mathsf{p}),$$

or

$$\binom{n}{n\mathsf{p}} \simeq \mathrm{e}^{n\overline{H}_{\mathrm{Sh}}(\mathsf{p})} = 2^{nH_{\mathrm{Sh}}(\mathsf{p})}.$$

In information theory it is usual to work with base-2 logarithms, and the *Shannon entropy* is then defined by the second expression in the preceding equation, or

$$H_{\mathrm{Sh}}(\mathsf{p}) = -\mathsf{p} \log \mathsf{p} - (1-\mathsf{p}) \log(1-\mathsf{p}), \tag{7.7}$$

[1] We assume that the correlations between letters can be neglected.
[2] The function $\mathsf{p}(q)$ is approximately Gaussian around $q = \overline{q}$, so that the most probable value is also the mean value $\langle q \rangle$.

where log is a base-2 logarithm. The number of *typical* sequences is of order $2^{nH_{\mathrm{Sh}}(\mathsf{p})}$. Let us illustrate this by two limiting cases.

(i) $\mathsf{p} = 1$. In this case the 2^n messages are identical and it is sufficient to send only one: $H_{\mathrm{Sh}}(\mathsf{p}) = 0$.
(ii) $\mathsf{p} = 1/2$. All messages are equally probable and $H_{\mathrm{Sh}} = 1$. In this case it is necessary to send the 2^n messages and no compression is possible.

In an intermediate case, for example, $\mathsf{p} = 1/4$, it is sufficient to encode typical sequences, and one never has to encode sequences of letters containing very few a_0 or very few a_1, which are highly unlikely.

In the case of k letters a_x with probabilities $\mathsf{p}(x)$, the number of typical sequences is

$$\frac{n!}{\prod_x (n\mathsf{p}(x))!} \simeq 2^{nH_{\mathrm{Sh}}(X)}$$

with

$$\boxed{H_{\mathrm{Sh}}(X) = -\sum_{x=0}^{k} \mathsf{p}(x) \log \mathsf{p}(x)} \tag{7.8}$$

where X denotes the probability distribution of the a_x. It can be rigorously shown that if $n \to \infty$, an optimal encoding compresses each letter into $H_{\mathrm{Sh}}(X)$ bits. This is the content of the first Shannon theorem, which also states that no further data compression is possible without introducing errors. In the example given above

$$-\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{4}\log\frac{1}{4} + \frac{1}{4}\log\frac{1}{8}\right) = \frac{7}{4},$$

which shows that the proposed encoding is optimal.

Let us now turn to the problem of a noisy channel. Let $\mathsf{p}(y|x)$ be the *conditional* probability for y to be read when the letter x is sent, the letter[3] x being sent with probability $\mathsf{p}(x)$. The entropy H_{Sh} quantifies our *a priori* ignorance per letter before receiving the message. Once y is known, we have at our disposal supplementary information, and our ignorance is not as great. We shall make use of Bayes' law

$$\mathsf{p}(x|y) = \frac{\mathsf{p}(x,y)}{\mathsf{p}(y)} \Longrightarrow \mathsf{p}(x|y) = \frac{\mathsf{p}(y|x)\mathsf{p}(x)}{\mathsf{p}(y)} \tag{7.9}$$

and of

$$\mathsf{p}(y) = \sum_x \mathsf{p}(y|x)\mathsf{p}(x).$$

[3] To simplify the notation we write $a_x = x$.

The number of bits needed to send a message knowing that y is read is then

$$\begin{aligned} H_{\mathrm{Sh}}(X|Y) &= \langle -\log \mathsf{p}(x|y)\rangle = \sum_y \mathsf{p}(y) \sum_x \mathsf{p}(x|y) \ln \mathsf{p}(x|y) \\ &= H_{\mathrm{Sh}}(X,Y) - H_{\mathrm{Sh}}(Y). \end{aligned} \tag{7.10}$$

The *information gain* or *mutual information* $I(X:Y)$ quantifies the information that is acquired about x when y is read:

$$\boxed{I(X:Y) = I(Y:X) = H_{\mathrm{Sh}}(X) - H_{\mathrm{Sh}}(X|Y) = H_{\mathrm{Sh}}(X) + H_{\mathrm{Sh}}(Y) - H_{\mathrm{Sh}}(X,Y)} \tag{7.11}$$

In other words, $I(X:Y) = I(Y:X)$ is the number of bits per letter of X which can be acquired by reading Y (or vice versa). If $\mathsf{p}(y|x)$ characterizes a noisy channel, $I(X:Y)$ is the information per letter which can be sent via the channel given the probability distribution X, and the *transmission capacity* C of the channel is the maximum of $I(X:Y)$ over the ensemble of these probability distributions:

$$\boxed{C = \max_{\{\mathsf{p}(x)\}} I(X:Y)} \tag{7.12}$$

The second Shannon theorem states that error-free transmission by a noisy channel is possible if the transmission rate of the channel is less than C.

Let us give an example for a symmetric binary channel, defined as

$$\begin{aligned} \mathsf{p}(x=0|y=0) &= \mathsf{p}(x=1|y=1) = 1-\mathsf{p}, \\ \mathsf{p}(x=0|y=1) &= \mathsf{p}(x=1|y=0) = \mathsf{p}, \end{aligned}$$

where the mutual information is

$$I(X:Y) = H_{\mathrm{Sh}}(X) - H_{\mathrm{Sh}}(\mathsf{p}),$$

$H_{\mathrm{Sh}}(X)$ being given by (7.8). The maximal value of $H_{\mathrm{Sh}}(X)$ is 1, and so

$$C(\mathsf{p}) = 1 - H_{\mathrm{Sh}}(\mathsf{p}).$$

Another illustration of the concept of information gain is given in Exercise 7.5.3, where it is applied to quantum cryptography.

7.3 von Neumann entropy

In the quantum case, the letters are replaced by quantum states $|\alpha\rangle$ whose frequency is p_α. The state operator is

$$\rho = \sum_\alpha \mathsf{p}_\alpha |\alpha\rangle\langle\alpha|, \quad \sum_\alpha \mathsf{p}_\alpha = 1. \tag{7.13}$$

The state operator ρ represents a statistical mixture of states $|\alpha\rangle$, each state $|\alpha\rangle$ having probability p_α. The states $|\alpha\rangle$ are normalized ($\langle\alpha|\alpha\rangle = 1$) but *not necessarily orthogonal* ($\langle\alpha|\beta\rangle \neq \delta_{\alpha\beta}$), and in general there exist an infinite number

of decompositions of ρ of the type (7.13). It can also be said that there are an infinite number of ways of preparing ρ. The way it is prepared determines ρ, but not the reverse. However, ρ is Hermitian and can always be diagonalized:

$$\rho = \sum_i \mathsf{p}_i |i\rangle\langle i|, \quad \langle i|j\rangle = \delta_{ij}. \tag{7.14}$$

This leads to a generalization of the Shannon entropy, the *von Neumann entropy*, which is independent of the preparation:

$$\boxed{H_{\text{vN}} = -\sum_i \mathsf{p}_i \log \mathsf{p}_i = -\text{Tr}\rho\log\rho} \tag{7.15}$$

We note that the entropy of a pure case is zero, because all the p_i are zero except for one, which is unity. As in the classical case, we define the Shannon entropy of the preparation (7.13) as

$$H_{\text{Sh}} = -\sum_\alpha \mathsf{p}_\alpha \log \mathsf{p}_\alpha. \tag{7.16}$$

As already mentioned, there are in general an infinite number of different statistical mixtures $\{\mathsf{p}_\alpha, |\alpha\rangle\}$ which give the same state operator, and it can be shown that the Shannon entropy is always greater than the von Neumann entropy (see Exercise 7.5.2):

$$-\sum_\alpha \mathsf{p}_\alpha \log \mathsf{p}_\alpha \geq -\text{Tr}\,\rho\log\rho, \quad H_{\text{Sh}} \geq H_{\text{vN}}. \tag{7.17}$$

As we shall see later on, the entropy H_{vN} quantifies the incompressible information contained in the source described by the state operator ρ. The difference between the Shannon entropy and the von Neumann entropy is particularly clear for a product state AB represented by a state operator ρ_{AB}. The operator ρ_{AB} is used to construct the state operators of A and B, ρ_A and ρ_B, by taking the trace of ρ_{AB} over the spaces $\mathcal{H}_B$ and $\mathcal{H}_A$, respectively [*cf.* (4.12)]:

$$\rho_A = \text{Tr}_B \rho_{AB}, \quad \rho_B = \text{Tr}_A \rho_{AB},$$

or in matrix form[4]

$$\rho^A_{ij} = \sum_\mu \rho^{AB}_{i\mu,j\mu}, \quad \rho^B_{\mu\nu} = \sum_i \rho^{AB}_{i\mu,i\nu}.$$

The operators ρ_A and ρ_B are the reduced state operators of A and B. The following inequalities can be derived for the von Neumann entropy:

$$|H_{\text{vN}}(\rho_A) - H_{\text{vN}}(\rho_B)| \leq H_{\text{vN}}(\rho_{AB}) \leq H_{\text{vN}}(\rho_A) + H_{\text{vN}}(\rho_B). \tag{7.18}$$

[4] Here AB is written as a superscript to make room for the subscripts labeling the matrix elements.

On the contrary, the Shannon entropy of a joint probability distribution $H_{\rm Sh}(\mathsf{p}_{AB})$ satisfies

$$\max\left[H_{\rm Sh}(\mathsf{p}_A), H_{\rm Sh}(\mathsf{p}_B)\right] \le H_{\rm Sh}(\mathsf{p}_{AB}) \le H_{\rm Sh}(\mathsf{p}_A) + H_{\rm Sh}(\mathsf{p}_B), \tag{7.19}$$

where p_A and p_B are the probability distributions of x_A and x_B

$$\mathsf{p}_A(x_A) = \sum_{x_B} \mathsf{p}_{AB}(x_A, x_B), \quad \mathsf{p}_B(x_B) = \sum_{x_A} \mathsf{p}_{AB}(x_A, x_B).$$

The inequality on the right-hand side is the same for the two entropies, but that on the left (called the Araki–Lieb inequality) is different. For example, if ρ_{AB} is the state operator describing the pure state (4.4) of two qubits we have $H_{\rm vN}(\rho_{AB}) = 0$, whereas

$$H_{\rm vN}(\rho_A) = H_{\rm vN}(\rho_B) = 1.$$

The von Neumann entropy provides the key to the quantum generalization of the two Shannon theorems on data compression and on the maximum transmission capacity of a noisy channel. To explain this, let us consider an ensemble of n letters, where each letter is drawn from an ensemble $\{\mathsf{p}_\alpha, |\alpha\rangle\}$ such that the state operator of a single letter is given by (7.13). Successive letters are assumed to be independent, and the state operator of the ensemble of letters is

$$\rho^{\otimes n} = \rho \otimes \rho \otimes \cdots \otimes \rho := \sigma, \quad n \gg 1.$$

Let us suppose that we wish to send (or store) a message of n letters, by trying to encode the quantum system in a smaller system. This smaller system is sent to one end of a channel and decoded at the other end. The state operator of the transmitted system is σ', and the *fidelity* $\mathcal{F}$ of the transmission is defined as[5]

$$\mathcal{F}(\sigma, \sigma') := \left(\mathrm{Tr}\sqrt{\sigma^{1/2}\sigma'\sigma^{1/2}}\right)^2. \tag{7.20}$$

This expression is not very intuitive and does not even look symmetric in σ and σ', although one may prove that *it is symmetric* (see Exercise 7.5.4)

$$\mathcal{F}(\sigma, \sigma') = \mathcal{F}(\sigma', \sigma).$$

If σ is a pure state, $\sigma = \sigma^{1/2} = |\psi\rangle\langle\psi|$ and σ' a state matrix of the form (7.13)

$$\sigma' = \sum_\tau \mathsf{p}'_\tau |\tau\rangle\langle\tau|,$$

[5] Many authors, including Nielsen and Chuang (2000), define the fidelity as the square root of $\mathcal{F}$ in (7.20).

then

$$\sigma^{1/2}\sigma'\sigma^{1/2} = \left(\sum_\tau \mathsf{p}'_\tau |\langle\psi|\tau\rangle|^2\right)|\psi\rangle\langle\psi|$$

and one finds

$$\mathcal{F}(|\psi\rangle\langle\psi|, \sigma') = \sum_\tau \mathsf{p}'_\tau |\langle\psi|\tau\rangle|^2 = \langle\psi|\sigma'|\psi\rangle. \tag{7.21}$$

When σ and σ' represent pure states $|\psi\rangle$ and $|\psi'\rangle$, the fidelity reduces to

$$\mathcal{F}\left(|\psi\rangle\langle\psi|, |\psi'\rangle\langle\psi'|\right) = |\langle\psi|\psi'\rangle|^2 = \mathsf{p}(\psi' \to \psi)$$

according to (2.18), which in this case is a natural definition because $\mathcal{F}$ is simply the overlap of the two states.

We wish to find the smallest possible system such that $\mathcal{F} \geq 1 - \varepsilon$, for ε arbitrarily small and when letters are qubits. The Hilbert space $\mathcal{H}^{\otimes n}$ of n qubits has dimension 2^n. However, if $H_{\text{vN}}(\rho) < 1$, we are going to show that the state operator can be restricted to a typical Hilbert subspace of $\mathcal{H}^{\otimes n}$, and this *typical subspace* will have dimension smaller than 2^n. The fundamental result of Shumacher (and Josza) is that the dimension of this subspace is $2^{nH_{\text{vN}}(\rho)}$ for $n \gg 1$. It is therefore sufficient to use $nH_{\text{vN}}(\rho)$ qubits to represent faithfully the quantum information. This result transposes the classical result of Shannon, with the idea of a typical sequence of letters replaced by that of a typical subspace, and the Shannon entropy replaced by the von Neumann entropy.

Before giving the proof of Schumacher's theorem, let us explain intuitively why such a compression of qubits is possible. Suppose Alice has drawn her qubits from the ensemble

$$\begin{aligned} |0\rangle &= \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \mathsf{p} = \frac{1}{2}, \\ |+\rangle &= \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathsf{p} = \frac{1}{2}, \end{aligned} \tag{7.22}$$

so that the state matrix (7.13) is

$$\rho = \frac{1}{4}\begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}. \tag{7.23}$$

Observing that $|0\rangle$ is an eigenstate of σ_z and $|+\rangle$, an eigenstate of σ_x, both with eigenvalue $+1$, it is obvious from symmetry considerations that the eigenstates

of ρ are the vectors $|0'\rangle = |0, \hat{n}\rangle$ and $|1'\rangle = |1, \hat{n}\rangle$, where $\hat{n}$ is the unit vector $(\hat{x}+\hat{z})/\sqrt{2}$:

$$\begin{aligned} |0'\rangle = |0, \hat{n}\rangle &= \begin{pmatrix} \cos \pi/8 \\ \sin \pi/8 \end{pmatrix} = \begin{pmatrix} \beta \\ \gamma \end{pmatrix}, \\ |1'\rangle = |1, \hat{n}\rangle &= \begin{pmatrix} -\sin \pi/8 \\ \cos \pi/8 \end{pmatrix} = \begin{pmatrix} -\gamma \\ \beta \end{pmatrix}. \end{aligned} \tag{7.24}$$

The vectors $|0, \hat{n}\rangle$ and $|1, \hat{n}\rangle$ are eigenvectors of $\vec{\sigma} \cdot \hat{n}$ with eigenvalues $+1$ and -1, respectively. The eigenvalues of ρ are

$$\lambda(0') = \cos^2 \pi/8 = \beta^2 \simeq 0.8536$$

and

$$\lambda(1') = \sin^2 \pi/8 = \gamma^2 \simeq 0.1464.$$

By construction, the state $|0'\rangle$ has the same (large) overlap with $|0\rangle$ and $|+\rangle$:

$$|\langle 0'|0\rangle|^2 = |\langle 0'|+\rangle|^2 = \beta^2 \simeq 0.8536,$$

while $|1'\rangle$ has the same (small) overlap with $|0\rangle$ and $|+\rangle$:

$$|\langle 1'|0\rangle|^2 = |\langle 1'|+\rangle|^2 = \gamma^2 \simeq 0.1564.$$

If we do not know which of the states (7.22) was sent, our best guess is $|\psi\rangle = |0'\rangle$, and from (7.21) the probability of a successful guess is just the fidelity:

$$\frac{1}{2}\left(|\langle 0'|0\rangle|^2 + |\langle 0'|+\rangle|^2\right) = \mathcal{F}(\rho, |0'\rangle\langle 0'|) \simeq 0.8536.$$

Another way of obtaining the preceding result is to start from a trial vector

$$|\varphi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\phi} \sin\frac{\theta}{2}|1\rangle,$$

compute the fidelity $\mathcal{F}(\rho, |\varphi\rangle\langle\varphi|)$, and then check that it has a maximum for $\theta = \pi/4$ (see Exercise 2.6.4).

Now, suppose that Alice wants to send Bob a three-qubit message compressed into only two qubits with maximum fidelity. Let the message be

$$|\Psi\rangle = |\psi_A \otimes \psi_B \otimes \psi_C\rangle,$$

where $|\psi_i\rangle = |0_i\rangle$ or $|+_i\rangle$, and let us examine the three-qubit Hilbert space $\mathcal{H}^{\otimes 3}$. A possible basis of $\mathcal{H}^{\otimes 3}$ is

$$\begin{aligned} &|b_1\rangle = |0'_A 0'_B 0'_C\rangle, \quad |b_2\rangle = |0'_A 0'_B 1'_C\rangle, \quad |b_3\rangle = |0'_A 1'_B 0'_C\rangle, \quad |b_4\rangle = |1'_A 0'_B 0'_C\rangle, \\ &|b_5\rangle = |0'_A 1'_B 1'_C\rangle, \quad |b_6\rangle = |1'_A 0'_B 1'_C\rangle, \quad |b_7\rangle = |1'_A 1'_B 0'_C\rangle, \quad |b_8\rangle = |1'_A 1'_B 1'_C\rangle. \end{aligned} \tag{7.25}$$

The vectors $|b_1\rangle$ to $|b_4\rangle$ span a subspace $\mathcal{G}$ in $\mathcal{H}^{\otimes 3}$, and the probability that $|\Psi\rangle$ belongs to $\mathcal{G}$ is

$$\mathsf{p}_{\mathcal{G}} = \beta^6 + 3\beta^4\gamma^2 \simeq 0.9419.$$

Therefore, it is much more likely (roughly 20 times more likely) that the message belongs to $\mathcal{G}$ rather than to its orthogonal complement $\mathcal{G}_\perp$. Since $\mathcal{G}$ is a four-dimensional space, two qubits should be enough to send the message with a very good fidelity.

In practice, Alice can use a unitary transformation U which rotates the four basis vectors $|b_1\rangle \cdots |b_4\rangle$ of $\mathcal{G}$ into basis states of the form $|\varphi_A \varphi_B 0_C\rangle$ and the four basis vectors $|b_5\rangle \cdots |b_8\rangle$ of $\mathcal{G}_\perp$ into basis states $|\overline{\varphi}_A \overline{\varphi}_B 1_C\rangle$. She measures qubit C, and if the outcome is 0 she sends the remaining two qubits to Bob. The compressed message Ψ_{comp} is given by

$$|\Psi_{\text{comp}} \otimes 0_C\rangle = U|\Psi\rangle, \qquad |\Psi\rangle \in \mathcal{G}.$$

Bob receives the two-qubit message, takes its tensor product with $|0_C\rangle$ and he reads it by applying U^{-1}:

$$|\Psi'\rangle = U^{-1}|\Psi_{\text{comp}} \otimes 0_C\rangle \equiv |\Psi\rangle.$$

If Alice's measurement of qubit C gives $|1\rangle$, then the best she can do is send Bob the state that he will decompress to the most likely state $|0'_A 0'_B 0'_C\rangle$, that is, she sends the state $|\Psi_{\text{comp}}\rangle$ such that

$$|\Psi'\rangle = U^{-1}|\Psi_{\text{comp}} \otimes 1_C\rangle = |0'_A 0'_B 0'_C\rangle.$$

The outcome of the procedure is that Bob obtains the state matrix

$$\sigma' = \mathcal{P}|\Psi\rangle\langle\Psi|\mathcal{P} + |b_1\rangle\langle\Psi|(I - \mathcal{P})|\Psi\rangle\langle b_1|, \tag{7.26}$$

where $\mathcal{P}$ is the projector onto $\mathcal{G}$.

Let us now proceed to the general case. The key to Schumacher's theorem is that it is sufficient to encode typical subspaces of $\mathcal{H}^{\otimes n}$ if we wish to send n qubits drawn from the ensemble $\{\mathsf{p}_\alpha, |\alpha\rangle\}$ which defines the state matrix (7.13). Since the letters are drawn independently, the state matrix of the n qubits is

$$\rho^{\otimes n} = \rho_1 \otimes \cdots \otimes \rho_n = \sigma. \tag{7.27}$$

Now, each ρ_i can be written in diagonal form (7.14). The eigenvalues of ρ_i are $\lambda_1 = \mathsf{p}$ and $\lambda_2 = 1 - \mathsf{p}$. An eigenvalue of $\rho^{\otimes n}$ will be of the form

$$\lambda_1^q \lambda_2^{n-q} = \mathsf{p}^q (1 - \mathsf{p})^{n-q}$$

and it will appear $\binom{n}{q}$ times. From our preceding discussion of Shannon's theorem, we see that almost all the eigenvalues of $\rho^{\otimes n}$ lie in a domain defined by the following range of q:

$$n\mathsf{p} - \mathcal{O}(\sqrt{n}) \lesssim q \lesssim n\mathsf{p} + \mathcal{O}(\sqrt{n}).$$

For these values of q the eigenvalues of $\rho^{\otimes n}$ will be

$$\lambda \simeq 2^{-nH_{\mathrm{Sh}}(\mathsf{p})} = 2^{-nH_{\mathrm{vN}}(\rho)},$$

with $H_{\mathrm{vN}}(\rho) = -\sum_i \mathsf{p}_i \ln \mathsf{p}_i$. In other words, the typical eigenvalue of $\rho^{\otimes n}$ is $\lambda = 2^{-nH_{\mathrm{vN}}(\rho)}$. Let $\mathcal{G}$ be the subspace of $\mathcal{H}^{\otimes n}$ spanned by the eigenvectors corresponding to these eigenvalues, and let $\mathcal{P}$ be the projector onto this subspace. Then for any $\varepsilon > 0$ we can choose n large enough that

$$\mathrm{Tr}\left(\rho^{\otimes n}\mathcal{P}\right) \geq 1 - \varepsilon. \tag{7.28}$$

Suppose that Alice wants to send an n-letter message drawn from the ensemble $\{\alpha\}$:

$$|\Psi\rangle = |\alpha_1 \cdots \alpha_n\rangle.$$

As above, she uses a unitary transformation U such that its action on a typical message belonging to $\mathcal{G}$ is

$$U|\Psi_{\mathrm{typ}}\rangle = |\Psi_{\mathrm{comp}} \otimes 0 \otimes \cdots \otimes 0\rangle.$$

Then she sends Bob the nH_{vN} qubits corresponding to the space $\mathcal{G}$ and Bob decodes it using U^{-1}. The state received by Bob is

$$\sigma' = \mathcal{P}|\Psi\rangle\langle\Psi|\mathcal{P} + \overline{\sigma}', \tag{7.29}$$

where $\overline{\sigma}'$ is what Alice sends if $|\Psi\rangle \notin \mathcal{G}$. The fidelity of σ' obeys the inequality

$$\mathcal{F}(|\Psi\rangle\langle\Psi|, \sigma') \geq |\langle\Psi|\mathcal{P}|\Psi\rangle|^2 \tag{7.30}$$

because $\overline{\sigma}'$ is a positive operator. The fidelity depends on the message $|\alpha_1 \cdots \alpha_n\rangle$, and to obtain the final result we must sum over the p_α:

$$\begin{aligned}
\mathcal{F}(\sigma, \sigma') &\geq \sum_\alpha \mathsf{p}_\alpha |\langle\Psi|\mathcal{P}|\Psi\rangle|^2 \\
&\geq \sum_\alpha \mathsf{p}_\alpha (2\langle\Psi|\mathcal{P}|\Psi\rangle - 1) \\
&= 2\mathrm{Tr}\left(\rho^{\otimes n}\mathcal{P}\right) - 1 \geq 1 - \varepsilon,
\end{aligned} \tag{7.31}$$

where we have used $x^2 \geq 2x - 1$ and (7.28). In order to obtain a fidelity arbitrarily close to one, it is enough to send nH_{vN} qubits when $n \to \infty$.

https://doi.org/10.1017/CBO9780511755361.009

7.4 Quantum error correction

Noise is omnipresent in all classical data processing, communication, and storage and it introduces errors. For example, noise can flip the initial value 0 of a bit to the value 1, something we wish to avoid so that all our operations do not fall apart. The task of error correction is to detect the erroneous bits and correct them. Modern (classical) error-correcting codes are extremely sophisticated in their details, but they are all based on redundancy. A very simple example is the following. Instead of encoding information in a single bit, we encode it in three bits:

$$0 \to 000, \qquad 1 \to 111.$$

Suppose that the effect of noise is to flip a bit from 0 to 1 or vice versa with probability p. Then if we start with the bits in the state 000, for example, a single-bit flip will occur with probability $3\mathsf{p}(1-\mathsf{p})^2$ and a two-bit flip with probability $3\mathsf{p}^2(1-\mathsf{p})$. If we read

$$000, \qquad 100, \qquad 010, \qquad \text{or} \qquad 001$$

we can decide with probability $(1-\mathsf{p})^2(1+2\mathsf{p})$ that the original bit had the value 0. Therefore, a majority rule gives the correct result with probability $(1-\mathsf{p})^2(1+2\mathsf{p})$. If $\mathsf{p} = 10^{-2}$, the probability of error is $\mathsf{p}^2(3-2\mathsf{p}) \simeq 3\times 10^{-4}$, while it would be 10^{-2} without corrrection.

Classical error correction cannot be transposed directly to qubits for four reasons.

1. The no-cloning theorem forbids duplicating qubits in an unknown state.
2. There is no classical analog to the superposition of qubits, so that a phase flip such as (4.25) has no classical equivalent.
3. Errors may be continuous. For example, in the general qubit state

$$|\varphi\rangle = \cos\frac{\theta}{2}|0\rangle + \mathrm{e}^{\mathrm{i}\phi}\sin\frac{\theta}{2}|1\rangle$$

   noise could lead to continuous variations of the angles θ and ϕ.
4. Finally, measurement destroys quantum information: one should not affect the information encoded in a qubit by a projective measurement.

In spite of these difficulties, it has been possible to devise quantum error-correcting codes. These codes are rather involved, and we shall limit ourselves to a simple but illustrative example for the case of the phase flip σ_z introduced in (4.25):

$$\lambda|0\rangle + \mu|1\rangle \to \sigma_z(\lambda|0\rangle + \mu|1\rangle). \tag{7.32}$$

https://doi.org/10.1017/CBO9780511755361.009

This phase flip is a typical quantum error, because a classical bit cannot be in a linear superposition. It will be convenient to rephrase (7.32) by going to the $|\pm\rangle$ basis:

$$|\pm\rangle = \frac{1}{\sqrt{2}}\Big(|0\rangle \pm |1\rangle\Big). \tag{7.33}$$

We recall that the $|\pm\rangle$ vectors are obtained from $|0\rangle$ and $|1\rangle$ by application of the H gate:

$$H|0\rangle = |+\rangle, \qquad H|1\rangle = |-\rangle.$$

Then a phase flip in the $\{|0\rangle, |1\rangle\}$ basis corresponds to a bit flip in the $|\pm\rangle$ basis:

$$\begin{aligned} H\Big(\lambda|0\rangle + \mu|1\rangle\Big) &= \lambda|+\rangle + \mu|-\rangle, \\ \sigma_z\, H\Big(\lambda|0\rangle + \mu|1\rangle\Big) &= \lambda|-\rangle + \mu|+\rangle. \end{aligned} \tag{7.34}$$

The redundancy is introduced by using, in addition to the original qubit A, two auxiliary qubits B and C in the state $|0\rangle$:

$$(\lambda|0\rangle + \mu|1\rangle) \otimes |00\rangle = \lambda|000\rangle + \mu|100\rangle,$$

to which we apply two cNOT gates controlled by qubit A (left-hand side of the circuit drawn in Fig. 7.3)

$$\mathrm{cNOT}_B\, \mathrm{cNOT}_C\, (\lambda|000\rangle + \mu|100\rangle) = \lambda|000\rangle + \mu|111\rangle, \tag{7.35}$$

followed by three Hadamard gates

$$H^{\otimes 3}\, (\lambda|000\rangle + \mu|111\rangle) = \lambda|{+}{+}{+}\rangle + \mu|{-}{-}{-}\rangle = |\Psi_0\rangle. \tag{7.36}$$

In the absence of any phase flip, the final state of the three qubits is $|\Psi_0\rangle$ (7.36). If the phase of one of the three qubits is flipped, we get

$$|\Psi_A\rangle = \lambda|{-}{+}{+}\rangle + \mu|{+}{-}{-}\rangle \qquad \text{qubit } A \text{ flipped}, \tag{7.37}$$

$$|\Psi_B\rangle = \lambda|{+}{-}{+}\rangle + \mu|{-}{+}{-}\rangle \qquad \text{qubit } B \text{ flipped}, \tag{7.38}$$

$$|\Psi_C\rangle = \lambda|{+}{+}{-}\rangle + \mu|{-}{-}{+}\rangle \qquad \text{qubit } C \text{ flipped}. \tag{7.39}$$

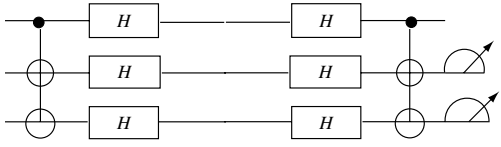

Figure 7.3 Circuit for error-correcting code.

https://doi.org/10.1017/CBO9780511755361.009

Let us introduce the two operators $X_A X_B$ and $X_A X_C$, where $X \equiv \sigma_x$. The vectors $|\Psi_0\rangle \cdots |\Psi_C\rangle$ are eigenvectors of these two operators with eigenvalues $+1$ or -1:

$$\begin{aligned} &X_A X_B|\Psi_0\rangle = +|\Psi_0\rangle, \quad X_A X_C|\Psi_0\rangle = +|\Psi_0\rangle, \quad X_A X_B|\Psi_A\rangle = -|\Psi_A\rangle, \\ &X_A X_C|\Psi_A\rangle = -|\Psi_A\rangle, \quad X_A X_B|\Psi_B\rangle = -|\Psi_B\rangle, \quad X_A X_C|\Psi_B\rangle = +|\Psi_B\rangle, \\ &X_A X_B|\Psi_C\rangle = +|\Psi_C\rangle, \quad X_A X_C|\Psi_C\rangle = -|\Psi_C\rangle. \end{aligned} \tag{7.40}$$

The measurement of $X_A X_B$ and $X_A X_C$ allows us to determine the type of error which has occurred:

$$\begin{aligned} &X_A X_B = +1, \quad X_A X_C = +1 \quad \text{no error}, \\ &X_A X_B = -1, \quad X_A X_C = -1 \quad \text{bit } A \text{ flipped}, \\ &X_A X_B = -1, \quad X_A X_C = +1 \quad \text{bit } B \text{ flipped}, \\ &X_A X_B = +1, \quad X_A X_C = -1 \quad \text{bit } C \text{ flipped}. \end{aligned} \tag{7.41}$$

However, the qubits should not be measured individually, as this would lead to the destruction of information on qubit A. The measurement is performed according to the right-hand side of the circuit in Fig. 7.3. For example,

$$\begin{aligned} \text{cNOT}_B\,\text{cNOT}_C(H_A \otimes H_B \otimes H_C)|\Psi_B\rangle &= \text{cNOT}_B\,\text{cNOT}_C\,(\lambda|010\rangle + \mu|110\rangle) \\ &= \lambda|010\rangle + \mu|110\rangle \\ &= (\lambda|0\rangle + \mu|1\rangle) \otimes |10\rangle. \end{aligned} \tag{7.42}$$

If the qubits B and C are found in the states $|1\rangle$ and $|0\rangle$, respectively, this implies that qubit B was flipped. The reader will easily check (Exercise 7.5.5) that the final states of the qubits B and C are

$$|\Psi_0\rangle \to |00\rangle, \quad |\Psi_A\rangle \to |11\rangle, \quad |\Psi_B\rangle \to |10\rangle, \quad |\Psi_C\rangle \to |01\rangle.$$

If the measured values of qubits B and C give the state $|11\rangle$, then we apply X_A to qubit A. The correct quantum state is therefore recovered without ever measuring this qubit: the measurement does not give any information on the values of λ and μ.

There are other types of error in addition to phase flip. In order to deal with all the errors it is necessary to use at least four auxiliary qubits, but this so-called five-qubit correcting code is extremely cumbersome. At present the most favored code is the seven-qubit correcting code devised by Sheane, although the first code devised by Shor, which is a nine-qubit code, also has interesting properties.

7.5 Exercises

7.5.1 Superdense coding

Alice and Bob share a pair of entangled qubits A and B in state $|\Psi\rangle$ (see Fig. 7.4)

$$|\Psi\rangle = \frac{1}{\sqrt{2}}\left(|0_A \otimes 0_B\rangle + |1_A \otimes 1_B\rangle\right).$$

Alice wishes to send Bob two *classical* bits of information i and j, $i, j = 0, 1$, while using a single qubit. She transforms the state of her qubit by applying on it the operator A_{ij} acting on qubit A

$$A_{ij} = (\sigma_{xA})^i (\sigma_{zA})^j$$

where i and j are exponents. She then sends her qubit to Bob, who gets the pair in the state $A_{ij}|\Psi\rangle$.

1. Give the explicit expression of $A_{00}|\Psi\rangle$, $A_{01}|\Psi\rangle$, $A_{10}|\Psi\rangle$, $A_{11}|\Psi\rangle$ in terms of the states $|0_A \otimes 0_B\rangle$, $|0_A \otimes 1_B\rangle$, $|1_A \otimes 0_B\rangle$, $|1_A \otimes 1_B\rangle$.
2. Bob uses the logic circuit of Fig. 7.4 with a cNOT gate and a Hadamard gate H. Examining the four possibilities for $A_{ij}|\Psi\rangle$, show that the cNOT gate transforms $A_{ij}|\Psi\rangle$ into a tensor product and that measurement of qubit B gives the value of i. Show finally that measurement of qubit A gives the value of j. Thus Alice transmits two bits of information while sending only one qubit.

7.5.2 Shannon entropy versus von Neumann entropy

Let us consider a two-dimensional space and define the state $|\theta\rangle$ as

$$|\theta\rangle = \cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2}|1\rangle.$$

Let a state matrix ρ be given by

$$\rho = \mathsf{p}|0\rangle\langle 0| + (1-\mathsf{p})|\theta\rangle\langle\theta|.$$

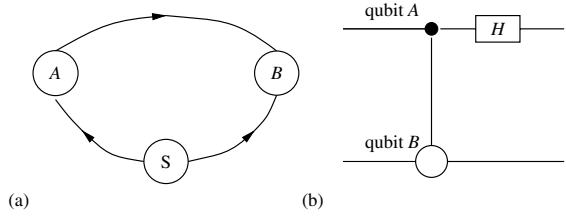

Figure 7.4 (a) General depiction; S, source of entangled particles. (b) Gates applied by Bob.

Compute the Shannon and von Neumann entropies. Show that

$$H_{\mathrm{Sh}} \geq H_{\mathrm{vN}}.$$

7.5.3 Information gain of Eve

Let us compute the information gain $I(\alpha : \varepsilon)$ (see (7.11)) of Eve, where α stands for Alice and ε for Eve, in two different situations.

1. Let i characterize the bit sent by Alice in the $\{|x\rangle, |y\rangle\}$ basis or in the $\{|\pi/4\rangle, |-\pi/4\rangle\}$ basis. Thus i can take four different values with equal probabilities $\mathsf{p}(i) = 1/4$. Let Eve use the $\{|x\rangle, |y\rangle\}$ basis in which she measures a result r, where r takes two different values. Establish a table of the conditional probabilities $\mathsf{p}(r|i)$ and deduce from it $\mathsf{p}(i|r)$. Show that Eve's information gain is 1/2.
2. Now Eve uses a symmetric $\{|\pi/8\rangle, |-\pi/8\rangle\}$ basis (see Exercise 2.6.4). Show that in this case the information gain is only $I(\alpha : \varepsilon) \simeq 0.4$. The information gain is smaller when Eve uses the symmetric basis.

7.5.4 Symmetry of the fidelity

Show that

$$\mathcal{F}(\rho, |\Psi\rangle\langle\Psi|) = \mathcal{F}(|\Psi\rangle\langle\Psi|, \rho)\ .$$

Hint: to evaluate the first expression for $\mathcal{F}$, use a basis (7.14) where ρ is diagonal, and observe that a product of matrices is of rank one if one of the matrices in the product is of rank one. What is then the nonzero eigenvalue of the product

$$\rho^{1/2}|\Psi\rangle\langle\Psi|\rho^{1/2}?$$

It may be instructive to examine first the case of two-dimensional matrices.

7.5.5 Quantum error correcting code

Work out the details of the calculations leading to (7.41) and the action of the transformation on the right-hand side of the circuit in Fig. 7.3 in the four different cases.

7.6 Further reading

Zeilinger (2000) gives an elementary account of teleportation. Recent experiments demonstrating teleportation using atoms are described by Barret *et al.* (2004) and

Riebe *et al.* (2004). Shannon entropy, von Neumann entropy and Schumacher theorem are explained by Preskill (1999), Chapter 5, Nielsen and Chuang (2000), Chapters 11 and 12 or Stolze and Suter (2004), Chapter 13. For quantum error correction, see Nielsen and Chuang (2000), Chapter 10 or Stolze and Suter (2004), Chapter 7.