Exercise 2.68: Prove that $|\psi\rangle \neq |a\rangle|b\rangle$ for all single qubit states $|a\rangle$ and $|b\rangle$.

We say that a state of a composite system having this property (that it can't be written as a product of states of its component systems) is an *entangled* state. For reasons which nobody fully understands, entangled states play a crucial role in quantum computation and quantum information, and arise repeatedly through the remainder of this book. We have already seen entanglement play a crucial role in quantum teleportation, as described in Section 1.3.7. In this chapter we give two examples of the strange effects enabled by entangled quantum states, superdense coding (Section 2.3), and the violation of Bell's inequality (Section 2.6).

2.2.9 Quantum mechanics: a global view

We have now explained *all* the fundamental postulates of quantum mechanics. Most of the rest of the book is taken up with deriving consequences of these postulates. Let's quickly review the postulates and try to place them in some kind of global perspective.

Postulate 1 sets the arena for quantum mechanics, by specifying how the state of an isolated quantum system is to be described. Postulate 2 tells us that the dynamics of *closed* quantum systems are described by the Schrödinger equation, and thus by unitary evolution. Postulate 3 tells us how to extract information from our quantum systems by giving a prescription for the description of measurement. Postulate 4 tells us how the state spaces of different quantum systems may be combined to give a description of the composite system.

What's odd about quantum mechanics, at least by our classical lights, is that we can't directly observe the state vector. It's a little bit like a game of chess where you can never find out exactly where each piece is, but only know the rank of the board they are on. Classical physics – and our intuition – tells us that the fundamental properties of an object, like energy, position, and velocity, are directly accessible to observation. In quantum mechanics these quantities no longer appear as fundamental, being replaced by the state vector, which can't be directly observed. It is as though there is a *hidden world* in quantum mechanics, which we can only indirectly and imperfectly access. Moreover, merely observing a classical system does not necessarily change the state of the system. Imagine how difficult it would be to play tennis if each time you looked at the ball its position changed! But according to Postulate 3, observation in quantum mechanics is an invasive procedure that typically changes the state of the system.

What conclusions should we draw from these strange features of quantum mechanics? Might it be possible to reformulate quantum mechanics in a mathematically equivalent way so that it had a structure more like classical physics? In Section 2.6 we'll prove *Bell's inequality*, a surprising result that shows any attempt at such a reformulation is doomed to failure. We're stuck with the counter-intuitive nature of quantum mechanics. Of course, the proper reaction to this is glee, not sorrow! It gives us an opportunity to develop tools of thought that make quantum mechanics intuitive. Moreover, we can exploit the hidden nature of the state vector to do information processing tasks beyond what is possible in the classical world. Without this counter-intuitive behavior, quantum computation and quantum information would be a lot less interesting.

We can also turn this discussion about, and ask ourselves: 'If quantum mechanics is so different from classical physics, then how come the everyday world looks so classical?' Why do we see no evidence of a hidden state vector in our everyday lives? It turns out that the classical world we see can be *derived* from quantum mechanics as an approximate description of the world that will be valid on the sort of time, length and mass scales we commonly encounter in our everyday lives. Explaining the details of how quantum mechanics gives rise to classical physics is beyond the scope of this book, but the interested reader should check out the discussion of this topic in 'History and further reading'at the end of Chapter 8.

2.3 Application: superdense coding

Superdense coding is a simple yet surprising application of elementary quantum mechanics. It combines in a concrete, non-trivial way all the basic ideas of elementary quantum mechanics, as covered in the previous sections, and is therefore an ideal example of the information processing tasks that can be accomplished using quantum mechanics.

Superdense coding involves two parties, conventionally known as 'Alice' and 'Bob', who are a long way away from one another. Their goal is to transmit some classical information from Alice to Bob. Suppose Alice is in possession of two classical bits of information which she wishes to send Bob, but is only allowed to send a single qubit to Bob. Can she achieve her goal?

Superdense coding tells us that the answer to this question is yes. Suppose Alice and Bob initially share a pair of qubits in the entangled state

$$|\psi\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}. \tag{2.133}$$

Alice is initially in possession of the first qubit, while Bob has possession of the second qubit, as illustrated in Figure 2.3. Note that $|\psi\rangle$ is a fixed state; there is no need for Alice to have sent Bob any qubits in order to prepare this state. Instead, some third party may prepare the entangled state ahead of time, sending one of the qubits to Alice, and the other to Bob.

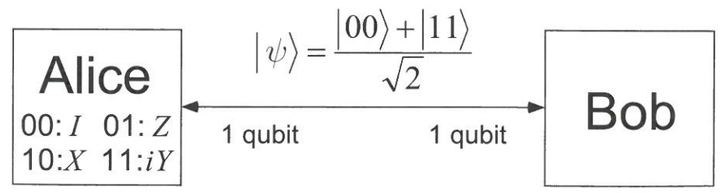

Figure 2.3. The initial setup for superdense coding, with Alice and Bob each in possession of one half of an entangled pair of qubits. Alice can use superdense coding to transmit two classical bits of information to Bob, using only a single qubit of communication and this preshared entanglement.

By sending the single qubit in her possession to Bob, it turns out that Alice can communicate two bits of classical information to Bob. Here is the procedure she uses. If she wishes to send the bit string '00' to Bob then she does nothing at all to her qubit. If she wishes to send '01' then she applies the phase flip Z to her qubit. If she wishes to send '10' then she applies the quantum NOT gate, X, to her qubit. If she wishes to send '11' then she applies the iY gate to her qubit. The four resulting states are easily seen

to be:

$$00 : |\psi\rangle \rightarrow \frac{|00\rangle + |11\rangle}{\sqrt{2}} \tag{2.134}$$

$$01 : |\psi\rangle \rightarrow \frac{|00\rangle - |11\rangle}{\sqrt{2}} \tag{2.135}$$

$$10 : |\psi\rangle \rightarrow \frac{|10\rangle + |01\rangle}{\sqrt{2}} \tag{2.136}$$

$$11 : |\psi\rangle \rightarrow \frac{|01\rangle - |10\rangle}{\sqrt{2}}. \tag{2.137}$$

As we noted in Section 1.3.6, these four states are known as the *Bell basis*, *Bell states*, or *EPR pairs*, in honor of several of the pioneers who first appreciated the novelty of entanglement. Notice that the Bell states form an orthonormal basis, and can therefore be distinguished by an appropriate quantum measurement. If Alice sends her qubit to Bob, giving Bob possession of both qubits, then by doing a measurement in the Bell basis Bob can determine which of the four possible bit strings Alice sent.

Summarizing, Alice, interacting with only a single qubit, is able to transmit two bits of information to Bob. Of course, two qubits are involved in the protocol, but Alice never need interact with the second qubit. Classically, the task Alice accomplishes would have been impossible had she only transmitted a single classical bit, as we will show in Chapter 12. Furthermore, this remarkable superdense coding protocol has received partial verification in the laboratory. (See 'History and further reading' for references to the experimental verification.) In later chapters we will see many other examples, some of them much more spectacular than superdense coding, of quantum mechanics being harnessed to perform information processing tasks. However, a key point can already be seen in this beautiful example: information is physical, and surprising physical theories such as quantum mechanics may predict surprising information processing abilities.

Exercise 2.69: Verify that the Bell basis forms an orthonormal basis for the two qubit state space.

Exercise 2.70: Suppose E is any positive operator acting on Alice's qubit. Show that $\langle\psi|E \otimes I|\psi\rangle$ *takes the same value* when $|\psi\rangle$ is any of the four Bell states. Suppose some malevolent third party ('Eve') intercepts Alice's qubit on the way to Bob in the superdense coding protocol. Can Eve infer anything about which of the four possible bit strings $00, 01, 10, 11$ Alice is trying to send? If so, how, or if not, why not?

2.4 The density operator

We have formulated quantum mechanics using the language of state vectors. An alternate formulation is possible using a tool known as the *density operator* or *density matrix*. This alternate formulation is mathematically equivalent to the state vector approach, but it provides a much more convenient language for thinking about some commonly encountered scenarios in quantum mechanics. The next three sections describe the density operator formulation of quantum mechanics. Section 2.4.1 introduces the density operator using the concept of an ensemble of quantum states. Section 2.4.2 develops some general

properties of the density operator. Finally, Section 2.4.3 describes an application where the density operator really shines – as a tool for the description of *individual subsystems* of a composite quantum system.

2.4.1 Ensembles of quantum states

The density operator language provides a convenient means for describing quantum systems whose state is not completely known. More precisely, suppose a quantum system is in one of a number of states $|\psi_i\rangle$, where i is an index, with respective probabilities p_i. We shall call $\{p_i, |\psi_i\rangle\}$ an *ensemble of pure states*. The density operator for the system is defined by the equation

$$\rho \equiv \sum_i p_i |\psi_i\rangle\langle\psi_i|. \tag{2.138}$$

The density operator is often known as the *density matrix*; we will use the two terms interchangeably. It turns out that all the postulates of quantum mechanics can be reformulated in terms of the density operator language. The purpose of this section and the next is to explain how to perform this reformulation, and explain when it is useful. Whether one uses the density operator language or the state vector language is a matter of taste, since both give the same results; however it is sometimes much easier to approach problems from one point of view rather than the other.

Suppose, for example, that the evolution of a closed quantum system is described by the unitary operator U. If the system was initially in the state $|\psi_i\rangle$ with probability p_i then after the evolution has occurred the system will be in the state $U|\psi_i\rangle$ with probability p_i. Thus, the evolution of the density operator is described by the equation

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| \xrightarrow{U} \sum_i p_i U|\psi_i\rangle\langle\psi_i|U^\dagger = U\rho U^\dagger. \tag{2.139}$$

Measurements are also easily described in the density operator language. Suppose we perform a measurement described by measurement operators M_m. If the initial state was $|\psi_i\rangle$, then the probability of getting result m is

$$p(m|i) = \langle\psi_i|M_m^\dagger M_m|\psi_i\rangle = \mathrm{tr}(M_m^\dagger M_m|\psi_i\rangle\langle\psi_i|), \tag{2.140}$$

where we have used Equation (2.61) to obtain the last equality. By the law of total probability (see Appendix 1 for an explanation of this and other elementary notions of probability theory) the probability of obtaining result m is

$$p(m) = \sum_i p(m|i)p_i \tag{2.141}$$

$$= \sum_i p_i \mathrm{tr}(M_m^\dagger M_m|\psi_i\rangle\langle\psi_i|) \tag{2.142}$$

$$= \mathrm{tr}(M_m^\dagger M_m \rho). \tag{2.143}$$

What is the density operator of the system after obtaining the measurement result m? If the initial state was $|\psi_i\rangle$ then the state after obtaining the result m is

$$|\psi_i^m\rangle = \frac{M_m|\psi_i\rangle}{\sqrt{\langle\psi_i|M_m^\dagger M_m|\psi_i\rangle}}. \tag{2.144}$$

Thus, after a measurement which yields the result m we have an ensemble of states $|\psi_i^m\rangle$ with respective probabilities $p(i|m)$. The corresponding density operator ρ_m is therefore

$$\rho_m = \sum_i p(i|m)|\psi_i^m\rangle\langle\psi_i^m| = \sum_i p(i|m)\frac{M_m|\psi_i\rangle\langle\psi_i|M_m^\dagger}{\langle\psi_i|M_m^\dagger M_m|\psi_i\rangle}. \tag{2.145}$$

But by elementary probability theory, $p(i|m) = p(m,i)/p(m) = p(m|i)p_i/p(m)$. Substituting from (2.143) and (2.140) we obtain

$$\rho_m = \sum_i p_i \frac{M_m|\psi_i\rangle\langle\psi_i|M_m^\dagger}{\mathrm{tr}(M_m^\dagger M_m \rho)} \tag{2.146}$$

$$= \frac{M_m \rho M_m^\dagger}{\mathrm{tr}(M_m^\dagger M_m \rho)}. \tag{2.147}$$

What we have shown is that the basic postulates of quantum mechanics related to unitary evolution and measurement can be rephrased in the language of density operators. In the next section we complete this rephrasing by giving an intrinsic characterization of the density operator that does not rely on the idea of a state vector.

Before doing so, however, it is useful to introduce some more language, and one more fact about the density operator. First, the language. A quantum system whose state $|\psi\rangle$ is known exactly is said to be in a *pure state*. In this case the density operator is simply $\rho = |\psi\rangle\langle\psi|$. Otherwise, ρ is in a *mixed state*; it is said to be a *mixture* of the different pure states in the ensemble for ρ. In the exercises you will be asked to demonstrate a simple criterion for determining whether a state is pure or mixed: a pure state satisfies $\mathrm{tr}(\rho^2) = 1$, while a mixed state satisfies $\mathrm{tr}(\rho^2) < 1$. A few words of warning about the nomenclature: sometimes people use the term 'mixed state' as a catch-all to include both pure and mixed quantum states. The origin for this usage seems to be that it implies that the writer is not necessarily *assuming* that a state is pure. Second, the term 'pure state' is often used in reference to a state vector $|\psi\rangle$, to distinguish it from a density operator ρ.

Finally, imagine a quantum system is prepared in the state ρ_i with probability p_i. It is not difficult to convince yourself that the system may be described by the density matrix $\sum_i p_i\rho_i$. A proof of this is to suppose that ρ_i arises from some ensemble $\{p_{ij}, |\psi_{ij}\rangle\}$ (note that i is fixed) of pure states, so the probability for being in the state $|\psi_{ij}\rangle$ is $p_i p_{ij}$. The density matrix for the system is thus

$$\rho = \sum_{ij} p_i p_{ij} |\psi_{ij}\rangle\langle\psi_{ij}| \tag{2.148}$$

$$= \sum_i p_i \rho_i, \tag{2.149}$$

where we have used the definition $\rho_i = \sum_j p_{ij}|\psi_{ij}\rangle\langle\psi_{ij}|$. We say that ρ is a *mixture* of the states ρ_i with probabilities p_i. This concept of a mixture comes up repeatedly in the analysis of problems like quantum noise, where the effect of the noise is to introduce ignorance into our knowledge of the quantum state. A simple example is provided by the measurement scenario described above. Imagine that, for some reason, our record of the result m of the measurement was lost. We would have a quantum system in the state ρ_m with probability $p(m)$, but would no longer know the actual value of m. The state of

such a quantum system would therefore be described by the density operator

$$\rho = \sum_m p(m)\rho_m \tag{2.150}$$

$$= \sum_m \mathrm{tr}(M_m^\dagger M_m \rho)\frac{M_m \rho M_m^\dagger}{\mathrm{tr}(M_m^\dagger M_m \rho)} \tag{2.151}$$

$$= \sum_m M_m \rho M_m^\dagger, \tag{2.152}$$

a nice compact formula which may be used as the starting point for analysis of further operations on the system.

2.4.2 General properties of the density operator

The density operator was introduced as a means of describing ensembles of quantum states. In this section we move away from this description to develop an intrinsic characterization of density operators that does not rely on an ensemble interpretation. This allows us to complete the program of giving a description of quantum mechanics that does not take as its foundation the state vector. We also take the opportunity to develop numerous other elementary properties of the density operator.

The class of operators that are density operators are characterized by the following useful theorem:

Theorem 2.5: **(Characterization of density operators)** An operator ρ is the density operator associated to some ensemble $\{p_i, |\psi_i\rangle\}$ if and only if it satisfies the conditions:

(1) **(Trace condition)** ρ has trace equal to one.
(2) **(Positivity condition)** ρ is a positive operator.

Proof
Suppose $\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|$ is a density operator. Then

$$\mathrm{tr}(\rho) = \sum_i p_i \mathrm{tr}(|\psi_i\rangle\langle\psi_i|) = \sum_i p_i = 1, \tag{2.153}$$

so the trace condition $\mathrm{tr}(\rho) = 1$ is satisfied. Suppose $|\varphi\rangle$ is an arbitrary vector in state space. Then

$$\langle\varphi|\rho|\varphi\rangle = \sum_i p_i \langle\varphi|\psi_i\rangle\langle\psi_i|\varphi\rangle \tag{2.154}$$

$$= \sum_i p_i |\langle\varphi|\psi_i\rangle|^2 \tag{2.155}$$

$$\geq 0, \tag{2.156}$$

so the positivity condition is satisfied.

Conversely, suppose ρ is any operator satisfying the trace and positivity conditions. Since ρ is positive, it must have a spectral decomposition

$$\rho = \sum_j \lambda_j |j\rangle\langle j|, \tag{2.157}$$

where the vectors $|j\rangle$ are orthogonal, and λ_j are real, non-negative eigenvalues of ρ.

From the trace condition we see that $\sum_j \lambda_j = 1$. Therefore, a system in state $|j\rangle$ with probability λ_j will have density operator ρ. That is, the ensemble $\{\lambda_j, |j\rangle\}$ is an ensemble of states giving rise to the density operator ρ. □

This theorem provides a characterization of density operators that is intrinsic to the operator itself: we can *define* a density operator to be a positive operator ρ which has trace equal to one. Making this definition allows us to reformulate the postulates of quantum mechanics in the density operator picture. For ease of reference we state all the reformulated postulates here:

Postulate 1: Associated to any isolated physical system is a complex vector space with inner product (that is, a Hilbert space) known as the *state space* of the system. The system is completely described by its *density operator*, which is a positive operator ρ with trace one, acting on the state space of the system. If a quantum system is in the state ρ_i with probability p_i, then the density operator for the system is $\sum_i p_i \rho_i$.

Postulate 2: The evolution of a *closed* quantum system is described by a *unitary transformation*. That is, the state ρ of the system at time t_1 is related to the state ρ' of the system at time t_2 by a unitary operator U which depends only on the times t_1 and t_2,

$$\rho' = U \rho U^\dagger. \tag{2.158}$$

Postulate 3: Quantum measurements are described by a collection $\{M_m\}$ of *measurement operators*. These are operators acting on the state space of the system being measured. The index m refers to the measurement outcomes that may occur in the experiment. If the state of the quantum system is ρ immediately before the measurement then the probability that result m occurs is given by

$$p(m) = \mathrm{tr}(M_m^\dagger M_m \rho), \tag{2.159}$$

and the state of the system after the measurement is

$$\frac{M_m \rho M_m^\dagger}{\mathrm{tr}(M_m^\dagger M_m \rho)}. \tag{2.160}$$

The measurement operators satisfy the *completeness equation*,

$$\sum_m M_m^\dagger M_m = I. \tag{2.161}$$

Postulate 4: The state space of a composite physical system is the tensor product of the state spaces of the component physical systems. Moreover, if we have systems numbered 1 through n, and system number i is prepared in the state ρ_i, then the joint state of the total system is $\rho_1 \otimes \rho_2 \otimes \ldots \rho_n$.

These reformulations of the fundamental postulates of quantum mechanics in terms of the density operator are, of course, mathematically equivalent to the description in terms of the state vector. Nevertheless, as a way of thinking about quantum mechanics, the density operator approach really shines for two applications: the description of quantum systems whose state is not known, and the description of subsystems of a composite

quantum system, as will be described in the next section. For the remainder of this section we flesh out the properties of the density matrix in more detail.

Exercise 2.71: (Criterion to decide if a state is mixed or pure) Let ρ be a density operator. Show that $\mathrm{tr}(\rho^2) \leq 1$, with equality if and only if ρ is a pure state.

It is a tempting (and surprisingly common) fallacy to suppose that the eigenvalues and eigenvectors of a density matrix have some special significance with regard to the ensemble of quantum states represented by that density matrix. For example, one might suppose that a quantum system with density matrix

$$\rho = \frac{3}{4}|0\rangle\langle 0| + \frac{1}{4}|1\rangle\langle 1|. \tag{2.162}$$

must be in the state $|0\rangle$ with probability 3/4 and in the state $|1\rangle$ with probability 1/4. However, this is not necessarily the case. Suppose we define

$$|a\rangle \equiv \sqrt{\frac{3}{4}}|0\rangle + \sqrt{\frac{1}{4}}|1\rangle \tag{2.163}$$

$$|b\rangle \equiv \sqrt{\frac{3}{4}}|0\rangle - \sqrt{\frac{1}{4}}|1\rangle, \tag{2.164}$$

and the quantum system is prepared in the state $|a\rangle$ with probability 1/2 and in the state $|b\rangle$ with probability 1/2. Then it is easily checked that the corresponding density matrix is

$$\rho = \frac{1}{2}|a\rangle\langle a| + \frac{1}{2}|b\rangle\langle b| = \frac{3}{4}|0\rangle\langle 0| + \frac{1}{4}|1\rangle\langle 1|. \tag{2.165}$$

That is, these two *different* ensembles of quantum states give rise to the *same* density matrix. In general, the eigenvectors and eigenvalues of a density matrix just indicate *one* of many possible ensembles that may give rise to a specific density matrix, and there is no reason to suppose it is an especially privileged ensemble.

A natural question to ask in the light of this discussion is what class of ensembles does give rise to a particular density matrix? The solution to this problem, which we now give, has surprisingly many applications in quantum computation and quantum information, notably in the understanding of quantum noise and quantum error-correction (Chapters 8 and 10). For the solution it is convenient to make use of vectors $|\tilde{\psi}_i\rangle$ which may not be normalized to unit length. We say the set $|\tilde{\psi}_i\rangle$ *generates* the operator $\rho \equiv \sum_i |\tilde{\psi}_i\rangle\langle\tilde{\psi}_i|$, and thus the connection to the usual ensemble picture of density operators is expressed by the equation $|\tilde{\psi}_i\rangle = \sqrt{p_i}|\psi_i\rangle$. When do two sets of vectors, $|\tilde{\psi}_i\rangle$ and $|\tilde{\varphi}_j\rangle$ generate the same operator ρ? The solution to this problem will enable us to answer the question of what ensembles give rise to a given density matrix.

Theorem 2.6: **(Unitary freedom in the ensemble for density matrices)** The sets $|\tilde{\psi}_i\rangle$ and $|\tilde{\varphi}_j\rangle$ generate the same density matrix if and only if

$$|\tilde{\psi}_i\rangle = \sum_j u_{ij} |\tilde{\varphi}_j\rangle, \tag{2.166}$$

where u_{ij} is a unitary matrix of complex numbers, with indices i and j, and we

'pad' whichever set of vectors $|\tilde{\psi}_i\rangle$ or $|\tilde{\varphi}_j\rangle$ is smaller with additional vectors 0 so that the two sets have the same number of elements.

As a consequence of the theorem, note that $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| = \sum_j q_j |\varphi_j\rangle\langle\varphi_j|$ for *normalized* states $|\psi_i\rangle, |\varphi_j\rangle$ and probability distributions p_i and q_j if and only if

$$\sqrt{p_i}|\psi_i\rangle = \sum_j u_{ij}\sqrt{q_j}|\varphi_j\rangle, \tag{2.167}$$

for some unitary matrix u_{ij}, and we may pad the smaller ensemble with entries having probability zero in order to make the two ensembles the same size. Thus, Theorem 2.6 characterizes the freedom in ensembles $\{p_i, |\psi_i\rangle\}$ giving rise to a given density matrix ρ. Indeed, it is easily checked that our earlier example of a density matrix with two different decompositions, (2.162), arises as a special case of this general result. Let's turn now to the proof of the theorem.

Proof

Suppose $|\tilde{\psi}_i\rangle = \sum_j u_{ij}|\tilde{\varphi}_j\rangle$ for some unitary u_{ij}. Then

$$\sum_i |\tilde{\psi}_i\rangle\langle\tilde{\psi}_i| = \sum_{ijk} u_{ij}u_{ik}^*|\tilde{\varphi}_j\rangle\langle\tilde{\varphi}_k| \tag{2.168}$$

$$= \sum_{jk}\left(\sum_i u_{ki}^\dagger u_{ij}\right)|\tilde{\varphi}_j\rangle\langle\tilde{\varphi}_k| \tag{2.169}$$

$$= \sum_{jk}\delta_{kj}|\tilde{\varphi}_j\rangle\langle\tilde{\varphi}_k| \tag{2.170}$$

$$= \sum_j |\tilde{\varphi}_j\rangle\langle\tilde{\varphi}_j|, \tag{2.171}$$

which shows that $|\tilde{\psi}_i\rangle$ and $|\tilde{\varphi}_j\rangle$ generate the same operator.

Conversely, suppose

$$A = \sum_i |\tilde{\psi}_i\rangle\langle\tilde{\psi}_i| = \sum_j |\tilde{\varphi}_j\rangle\langle\tilde{\varphi}_j|\,. \tag{2.172}$$

Let $A = \sum_k \lambda_k |k\rangle\langle k|$ be a decomposition for A such that the states $|k\rangle$ are orthonormal, and the λ_k are strictly positive. Our strategy is to relate the states $|\tilde{\psi}_i\rangle$ to the states $|\tilde{k}\rangle \equiv \sqrt{\lambda_k}|k\rangle$, and similarly relate the states $|\tilde{\varphi}_j\rangle$ to the states $|\tilde{k}\rangle$. Combining the two relations will give the result. Let $|\psi\rangle$ be any vector orthonormal to the space spanned by the $|\tilde{k}\rangle$, so $\langle\psi|\tilde{k}\rangle\langle\tilde{k}|\psi\rangle = 0$ for all k, and thus we see that

$$0 = \langle\psi|A|\psi\rangle = \sum_i \langle\psi|\tilde{\psi}_i\rangle\langle\tilde{\psi}_i|\psi\rangle = \sum_i |\langle\psi|\tilde{\psi}_i\rangle|^2. \tag{2.173}$$

Thus $\langle\psi|\tilde{\psi}_i\rangle = 0$ for all i and all $|\psi\rangle$ orthonormal to the space spanned by the $|\tilde{k}\rangle$. It follows that each $|\tilde{\psi}_i\rangle$ can be expressed as a linear combination of the $|\tilde{k}\rangle$, $|\tilde{\psi}_i\rangle = \sum_k c_{ik}|\tilde{k}\rangle$. Since $A = \sum_k |\tilde{k}\rangle\langle\tilde{k}| = \sum_i |\tilde{\psi}_i\rangle\langle\tilde{\psi}_i|$ we see that

$$\sum_k |\tilde{k}\rangle\langle\tilde{k}| = \sum_{kl}\left(\sum_i c_{ik}c_{il}^*\right)|\tilde{k}\rangle\langle\tilde{l}|. \tag{2.174}$$

The operators $|\tilde{k}\rangle\langle\tilde{l}|$ are easily seen to be linearly independent, and thus it must be that

$\sum_i c_{ik}c_{il}^* = \delta_{kl}$. This ensures that we may append extra columns to c to obtain a unitary matrix v such that $|\tilde{\psi}_i\rangle = \sum_k v_{ik}|\tilde{k}\rangle$, where we have appended zero vectors to the list of $|\tilde{k}\rangle$. Similarly, we can find a unitary matrix w such that $|\tilde{\varphi}_j\rangle = \sum_k w_{jk}|\tilde{k}\rangle$. Thus $|\tilde{\psi}_i\rangle = \sum_j u_{ij}|\tilde{\varphi}_j\rangle$, where $u = vw^\dagger$ is unitary. □

Exercise 2.72: (Bloch sphere for mixed states) The Bloch sphere picture for pure states of a single qubit was introduced in Section 1.2. This description has an important generalization to mixed states as follows.

(1) Show that an arbitrary density matrix for a mixed state qubit may be written as

$$\rho = \frac{I + \vec{r}\cdot\vec{\sigma}}{2}, \tag{2.175}$$

where $\vec{r}$ is a real three-dimensional vector such that $\|\vec{r}\| \leq 1$. This vector is known as the *Bloch vector* for the state ρ.

(2) What is the Bloch vector representation for the state $\rho = I/2$?

(3) Show that a state ρ is pure if and only if $\|\vec{r}\| = 1$.

(4) Show that for pure states the description of the Bloch vector we have given coincides with that in Section 1.2.

Exercise 2.73: Let ρ be a density operator. A *minimal ensemble* for ρ is an ensemble $\{p_i, |\psi_i\rangle\}$ containing a number of elements equal to the rank of ρ. Let $|\psi\rangle$ be any state in the support of ρ. (The *support* of a Hermitian operator A is the vector space spanned by the eigenvectors of A with non-zero eigenvalues.) Show that there is a minimal ensemble for ρ that contains $|\psi\rangle$, and moreover that in any such ensemble $|\psi\rangle$ must appear with probability

$$p_i = \frac{1}{\langle\psi_i|\rho^{-1}|\psi_i\rangle}, \tag{2.176}$$

where ρ^{-1} is defined to be the inverse of ρ, when ρ is considered as an operator acting only on the support of ρ. (This definition removes the problem that ρ may not have an inverse.)

2.4.3 The reduced density operator

Perhaps the deepest application of the density operator is as a descriptive tool for *subsystems* of a composite quantum system. Such a description is provided by the *reduced density operator*, which is the subject of this section. The reduced density operator is so useful as to be virtually indispensable in the analysis of composite quantum systems.

Suppose we have physical systems A and B, whose state is described by a density operator ρ^{AB}. The reduced density operator for system A is defined by

$$\rho^A \equiv \mathrm{tr}_B(\rho^{AB}), \tag{2.177}$$

where tr_B is a map of operators known as the *partial trace* over system B. The partial trace is defined by

$$\mathrm{tr}_B\left(|a_1\rangle\langle a_2| \otimes |b_1\rangle\langle b_2|\right) \equiv |a_1\rangle\langle a_2|\,\mathrm{tr}(|b_1\rangle\langle b_2|), \tag{2.178}$$

where $|a_1\rangle$ and $|a_2\rangle$ are any two vectors in the state space of A, and $|b_1\rangle$ and $|b_2\rangle$ are any two vectors in the state space of B. The trace operation appearing on the right hand side

is the usual trace operation for system B, so $\mathrm{tr}(|b_1\rangle\langle b_2|) = \langle b_2|b_1\rangle$. We have defined the partial trace operation only on a special subclass of operators on AB; the specification is completed by requiring in addition to Equation (2.178) that the partial trace be linear in its input.

It is not obvious that the reduced density operator for system A is in any sense a description for the state of system A. The physical justification for making this identification is that the reduced density operator provides the correct measurement statistics for measurements made on system A. This is explained in more detail in Box 2.6 on page 107. The following simple example calculations may also help understand the reduced density operator. First, suppose a quantum system is in the product state $\rho^{AB} = \rho \otimes \sigma$, where ρ is a density operator for system A, and σ is a density operator for system B. Then

$$\rho^A = \mathrm{tr}_B(\rho \otimes \sigma) = \rho\,\mathrm{tr}(\sigma) = \rho, \tag{2.184}$$

which is the result we intuitively expect. Similarly, $\rho^B = \sigma$ for this state. A less trivial example is the Bell state $(|00\rangle + |11\rangle)/\sqrt{2}$. This has density operator

$$\rho = \left(\frac{|00\rangle + |11\rangle}{\sqrt{2}}\right)\left(\frac{\langle 00| + \langle 11|}{\sqrt{2}}\right) \tag{2.185}$$

$$= \frac{|00\rangle\langle 00| + |11\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 11|}{2}. \tag{2.186}$$

Tracing out the second qubit, we find the reduced density operator of the first qubit,

$$\rho^1 = \mathrm{tr}_2(\rho) \tag{2.187}$$

$$= \frac{\mathrm{tr}_2(|00\rangle\langle 00|) + \mathrm{tr}_2(|11\rangle\langle 00|) + \mathrm{tr}_2(|00\rangle\langle 11|) + \mathrm{tr}_2(|11\rangle\langle 11|)}{2} \tag{2.188}$$

$$= \frac{|0\rangle\langle 0|\langle 0|0\rangle + |1\rangle\langle 0|\langle 0|1\rangle + |0\rangle\langle 1|\langle 1|0\rangle + |1\rangle\langle 1|\langle 1|1\rangle}{2} \tag{2.189}$$

$$= \frac{|0\rangle\langle 0| + |1\rangle\langle 1|}{2} \tag{2.190}$$

$$= \frac{I}{2}. \tag{2.191}$$

Notice that this state is a *mixed state*, since $\mathrm{tr}((I/2)^2) = 1/2 < 1$. This is quite a remarkable result. The state of the joint system of two qubits is a pure state, that is, it is known *exactly*; however, the first qubit is in a mixed state, that is, a state about which we apparently do not have maximal knowledge. This strange property, that the joint state of a system can be completely known, yet a subsystem be in mixed states, is another hallmark of quantum entanglement.

Exercise 2.74: Suppose a composite of systems A and B is in the state $|a\rangle|b\rangle$, where $|a\rangle$ is a pure state of system A, and $|b\rangle$ is a pure state of system B. Show that the reduced density operator of system A alone is a pure state.

Exercise 2.75: For each of the four Bell states, find the reduced density operator for each qubit.

Quantum teleportation and the reduced density operator

A useful application of the reduced density operator is to the analysis of quantum teleportation. Recall from Section 1.3.7 that quantum teleportation is a procedure for sending quantum information from Alice to Bob, given that Alice and Bob share an EPR pair, and have a classical communications channel.

Box 2.6: Why the partial trace?

Why is the partial trace used to describe part of a larger quantum system? The reason for doing this is because the partial trace operation is the *unique* operation which gives rise to the correct description of *observable* quantities for subsystems of a composite system, in the following sense.

Suppose M is any observable on system A, and we have some measuring device which is capable of realizing measurements of M. Let $\tilde{M}$ denote the corresponding observable for the same measurement, performed on the composite system AB. Our immediate goal is to argue that $\tilde{M}$ is necessarily equal to $M \otimes I_B$. Note that if the system AB is prepared in the state $|m\rangle|\psi\rangle$, where $|m\rangle$ is an eigenstate of M with eigenvalue m, and $|\psi\rangle$ is any state of B, then the measuring device must yield the result m for the measurement, with probability one. Thus, if P_m is the projector onto the m eigenspace of the observable M, then the corresponding projector for $\tilde{M}$ is $P_m \otimes I_B$. We therefore have

$$\tilde{M} = \sum_m m P_m \otimes I_B = M \otimes I_B. \tag{2.179}$$

The next step is to show that the partial trace procedure gives the correct measurement statistics for observations on part of a system. Suppose we perform a measurement on system A described by the observable M. Physical consistency requires that any prescription for associating a 'state', ρ^A, to system A, must have the property that measurement averages be the same whether computed via ρ^A or ρ^{AB},

$$\mathrm{tr}(M\rho^A) = \mathrm{tr}(\tilde{M}\rho^{AB}) = \mathrm{tr}((M \otimes I_B)\rho^{AB}). \tag{2.180}$$

This equation is certainly satisfied if we choose $\rho^A \equiv \mathrm{tr}_B(\rho^{AB})$. In fact, the partial trace turns out to be the *unique* function having this property. To see this uniqueness property, let $f(\cdot)$ be any map of density operators on AB to density operators on A such that

$$\mathrm{tr}(M f(\rho^{AB})) = \mathrm{tr}((M \otimes I_B)\rho^{AB}), \tag{2.181}$$

for all observables M. Let M_i be an orthonormal basis of operators for the space of Hermitian operators with respect to the Hilbert–Schmidt inner product $(X, Y) \equiv \mathrm{tr}(XY)$ (compare Exercise 2.39 on page 76). Then expanding $f(\rho^{AB})$ in this basis gives

$$f(\rho^{AB}) = \sum_i M_i \mathrm{tr}(M_i f(\rho^{AB})) \tag{2.182}$$

$$= \sum_i M_i \mathrm{tr}((M_i \otimes I_B)\rho^{AB}). \tag{2.183}$$

It follows that f is uniquely determined by Equation (2.180). Moreover, the partial trace satisfies (2.180), so it is the unique function having this property.

At first sight it appears as though teleportation can be used to do faster than light communication, a big no-no according to the theory of relativity. We surmised in Section 1.3.7 that what prevents faster than light communication is the need for Alice to communicate her measurement result to Bob. The reduced density operator allows us to make this rigorous.

Recall that immediately before Alice makes her measurement the quantum state of the three qubits is (Equation (1.32)):

$$|\psi_2\rangle = \frac{1}{2}\Big[|00\rangle\left(\alpha|0\rangle + \beta|1\rangle\right) + |01\rangle\left(\alpha|1\rangle + \beta|0\rangle\right) + |10\rangle\left(\alpha|0\rangle - \beta|1\rangle\right) + |11\rangle\left(\alpha|1\rangle - \beta|0\rangle\right)\Big]. \tag{2.192}$$

Measuring in Alice's computational basis, the state of the system after the measurement is:

$$|00\rangle\left[\alpha|0\rangle + \beta|1\rangle\right] \text{ with probability } \frac{1}{4} \tag{2.193}$$

$$|01\rangle\left[\alpha|1\rangle + \beta|0\rangle\right] \text{ with probability } \frac{1}{4} \tag{2.194}$$

$$|10\rangle\left[\alpha|0\rangle - \beta|1\rangle\right] \text{ with probability } \frac{1}{4} \tag{2.195}$$

$$|11\rangle\left[\alpha|1\rangle - \beta|0\rangle\right] \text{ with probability } \frac{1}{4}. \tag{2.196}$$

The density operator of the system is thus

$$\rho = \frac{1}{4}\Big[|00\rangle\langle 00|(\alpha|0\rangle + \beta|1\rangle)(\alpha^*\langle 0| + \beta^*\langle 1|) + |01\rangle\langle 01|(\alpha|1\rangle + \beta|0\rangle)(\alpha^*\langle 1| + \beta^*\langle 0|) + |10\rangle\langle 10|(\alpha|0\rangle - \beta|1\rangle)(\alpha^*\langle 0| - \beta^*\langle 1|) + |11\rangle\langle 11|(\alpha|1\rangle - \beta|0\rangle)(\alpha^*\langle 1| - \beta^*\langle 0|)\Big]. \tag{2.197}$$

Tracing out Alice's system, we see that the reduced density operator of Bob's system is

$$\rho^B = \frac{1}{4}\Big[(\alpha|0\rangle + \beta|1\rangle)(\alpha^*\langle 0| + \beta^*\langle 1|) + (\alpha|1\rangle + \beta|0\rangle)(\alpha^*\langle 1| + \beta^*\langle 0|) + (\alpha|0\rangle - \beta|1\rangle)(\alpha^*\langle 0| - \beta^*\langle 1|) + (\alpha|1\rangle - \beta|0\rangle)(\alpha^*\langle 1| - \beta^*\langle 0|)\Big] \tag{2.198}$$

$$= \frac{2(|\alpha|^2 + |\beta|^2)|0\rangle\langle 0| + 2(|\alpha|^2 + |\beta|^2)|1\rangle\langle 1|}{4} \tag{2.199}$$

$$= \frac{|0\rangle\langle 0| + |1\rangle\langle 1|}{2} \tag{2.200}$$

$$= \frac{I}{2}, \tag{2.201}$$

where we have used the completeness relation in the last line. Thus, the state of Bob's system *after* Alice has performed the measurement but *before* Bob has learned the measurement result is $I/2$. This state has no dependence upon the state $|\psi\rangle$ being teleported, and thus any measurements performed by Bob will contain no information about $|\psi\rangle$, thus preventing Alice from using teleportation to transmit information to Bob faster than light.

2.5 The Schmidt decomposition and purifications

Density operators and the partial trace are just the beginning of a wide array of tools useful for the study of composite quantum systems, which are at the heart of quantum computation and quantum information. Two additional tools of great value are the *Schmidt decomposition* and *purifications*. In this section we present both these tools, and try to give the flavor of their power.

Theorem 2.7: **(Schmidt decomposition)** Suppose $|\psi\rangle$ is a pure state of a composite system, AB. Then there exist orthonormal states $|i_A\rangle$ for system A, and orthonormal states $|i_B\rangle$ of system B such that

$$|\psi\rangle = \sum_i \lambda_i |i_A\rangle|i_B\rangle, \tag{2.202}$$

where λ_i are non-negative real numbers satisfying $\sum_i \lambda_i^2 = 1$ known as *Schmidt co-efficients*.

This result is very useful. As a taste of its power, consider the following consequence: let $|\psi\rangle$ be a pure state of a composite system, AB. Then by the Schmidt decomposition $\rho^A = \sum_i \lambda_i^2 |i_A\rangle\langle i_A|$ and $\rho^B = \sum_i \lambda_i^2 |i_B\rangle\langle i_B|$, so the eigenvalues of ρ^A and ρ^B are identical, namely λ_i^2 for both density operators. Many important properties of quantum systems are completely determined by the eigenvalues of the reduced density operator of the system, so for a pure state of a composite system such properties will be the same for both systems. As an example, consider the state of two qubits, $(|00\rangle + |01\rangle + |11\rangle)/\sqrt{3}$. This has no obvious symmetry property, yet if you calculate $\text{tr}\left((\rho^A)^2\right)$ and $\text{tr}\left((\rho^B)^2\right)$ you will discover that they have the same value, $7/9$ in each case. This is but one small consequence of the Schmidt decomposition.

Proof

We give the proof for the case where systems A and B have state spaces of the same dimension, and leave the general case to Exercise 2.76. Let $|j\rangle$ and $|k\rangle$ be any fixed orthonormal bases for systems A and B, respectively. Then $|\psi\rangle$ can be written

$$|\psi\rangle = \sum_{jk} a_{jk}|j\rangle|k\rangle, \tag{2.203}$$

for some matrix a of complex numbers a_{jk}. By the singular value decomposition, $a = udv$, where d is a diagonal matrix with non-negative elements, and u and v are unitary matrices. Thus

$$|\psi\rangle = \sum_{ijk} u_{ji} d_{ii} v_{ik} |j\rangle|k\rangle. \tag{2.204}$$

Defining $|i_A\rangle \equiv \sum_j u_{ji}|j\rangle$, $|i_B\rangle \equiv \sum_k v_{ik}|k\rangle$, and $\lambda_i \equiv d_{ii}$, we see that this gives

$$|\psi\rangle = \sum_i \lambda_i |i_A\rangle|i_B\rangle. \tag{2.205}$$

It is easy to check that $|i_A\rangle$ forms an orthonormal set, from the unitarity of u and the orthonormality of $|j\rangle$, and similarly that the $|i_B\rangle$ form an orthonormal set. □

Exercise 2.76: Extend the proof of the Schmidt decomposition to the case where A and B may have state spaces of different dimensionality.

Exercise 2.77: Suppose ABC is a three component quantum system. Show by example that there are quantum states $|\psi\rangle$ of such systems which can not be written in the form

$$|\psi\rangle = \sum_i \lambda_i |i_A\rangle |i_B\rangle |i_C\rangle, \tag{2.206}$$

where λ_i are real numbers, and $|i_A\rangle, |i_B\rangle, |i_C\rangle$ are orthonormal bases of the respective systems.

The bases $|i_A\rangle$ and $|i_B\rangle$ are called the *Schmidt bases* for A and B, respectively, and the number of non-zero values λ_i is called the *Schmidt number* for the state $|\psi\rangle$. The Schmidt number is an important property of a composite quantum system, which in some sense quantifies the 'amount' of entanglement between systems A and B. To get some idea of why this is the case, consider the following obvious but important property: the Schmidt number is preserved under unitary transformations on system A or system B alone. To see this, notice that if $\sum_i \lambda_i |i_A\rangle |i_B\rangle$ is the Schmidt decomposition for $|\psi\rangle$ then $\sum_i \lambda_i (U|i_A\rangle)|i_B\rangle$ is the Schmidt decomposition for $U|\psi\rangle$, where U is a unitary operator acting on system A alone. Algebraic invariance properties of this type make the Schmidt number a very useful tool.

Exercise 2.78: Prove that a state $|\psi\rangle$ of a composite system AB is a product state if and only if it has Schmidt number 1. Prove that $|\psi\rangle$ is a product state if and only if ρ^A (and thus ρ^B) are pure states.

A second, related technique for quantum computation and quantum information is *purification*. Suppose we are given a state ρ^A of a quantum system A. It is possible to introduce another system, which we denote R, and *define* a *pure state* $|AR\rangle$ for the joint system AR such that $\rho^A = \mathrm{tr}_R(|AR\rangle\langle AR|)$. That is, the pure state $|AR\rangle$ reduces to ρ^A when we look at system A alone. This is a purely mathematical procedure, known as *purification*, which allows us to associate pure states with mixed states. For this reason we call system R a *reference* system: it is a fictitious system, without a direct physical significance.

To prove that purification can be done for *any* state, we explain how to construct a system R and purification $|AR\rangle$ for ρ^A. Suppose ρ^A has orthonormal decomposition $\rho^A = \sum_i p_i |i^A\rangle\langle i^A|$. To purify ρ^A we introduce a system R which has the same state space as system A, with orthonormal basis states $|i^R\rangle$, and define a pure state for the combined system

$$|AR\rangle \equiv \sum_i \sqrt{p_i} |i^A\rangle |i^R\rangle. \tag{2.207}$$

We now calculate the reduced density operator for system A corresponding to the state $|AR\rangle$:

$$\mathrm{tr}_R(|AR\rangle\langle AR|) = \sum_{ij} \sqrt{p_i p_j} |i^A\rangle\langle j^A| \, \mathrm{tr}(|i^R\rangle\langle j^R|) \tag{2.208}$$

$$= \sum_{ij} \sqrt{p_i p_j} |i^A\rangle\langle j^A| \, \delta_{ij} \tag{2.209}$$

$$= \sum_i p_i |i^A\rangle\langle i^A| \tag{2.210}$$

$$= \rho^A. \tag{2.211}$$

Thus $|AR\rangle$ is a purification of ρ^A.

Notice the close relationship of the Schmidt decomposition to purification: the procedure used to purify a mixed state of system A is to define a pure state whose Schmidt basis for system A is just the basis in which the mixed state is diagonal, with the Schmidt coefficients being the square root of the eigenvalues of the density operator being purified.

In this section we've explained two tools for studying composite quantum systems, the Schmidt decomposition and purifications. These tools will be indispensable to the study of quantum computation and quantum information, especially quantum information, which is the subject of Part III of this book.

Exercise 2.79: Consider a composite system consisting of two qubits. Find the Schmidt decompositions of the states

$$\frac{|00\rangle + |11\rangle}{\sqrt{2}}; \quad \frac{|00\rangle + |01\rangle + |10\rangle + |11\rangle}{2}; \text{ and } \quad \frac{|00\rangle + |01\rangle + |10\rangle}{\sqrt{3}}. \tag{2.212}$$

Exercise 2.80: Suppose $|\psi\rangle$ and $|\varphi\rangle$ are two pure states of a composite quantum system with components A and B, with identical Schmidt coefficients. Show that there are unitary transformations U on system A and V on system B such that $|\psi\rangle = (U \otimes V)|\varphi\rangle$.

Exercise 2.81: (Freedom in purifications) Let $|AR_1\rangle$ and $|AR_2\rangle$ be two purifications of a state ρ^A to a composite system AR. Prove that there exists a unitary transformation U_R acting on system R such that $|AR_1\rangle = (I_A \otimes U_R)|AR_2\rangle$.

Exercise 2.82: Suppose $\{p_i, |\psi_i\rangle\}$ is an ensemble of states generating a density matrix $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ for a quantum system A. Introduce a system R with orthonormal basis $|i\rangle$.

(1) Show that $\sum_i \sqrt{p_i}|\psi_i\rangle|i\rangle$ is a purification of ρ.
(2) Suppose we measure R in the basis $|i\rangle$, obtaining outcome i. With what probability do we obtain the result i, and what is the corresponding state of system A?
(3) Let $|AR\rangle$ be *any* purification of ρ to the system AR. Show that there exists an orthonormal basis $|i\rangle$ in which R can be measured such that the corresponding post-measurement state for system A is $|\psi_i\rangle$ with probability p_i.

2.6 EPR and the Bell inequality

Anybody who is not shocked by quantum theory has not understood it.
– Niels Bohr

I recall that during one walk Einstein suddenly stopped, turned to me and asked whether I really believed that the moon exists only when I look at it. The rest of this walk was devoted to a discussion of what a physicist should mean by the term 'to exist'.
– Abraham Pais

...quantum phenomena do not occur in a Hilbert space, they occur in a laboratory.
– Asher Peres

...what is proved by impossibility proofs is lack of imagination.
– John Bell

This chapter has focused on introducing the tools and mathematics of quantum mechanics. As these techniques are applied in the following chapters of this book, an important recurring theme is the unusual, *non-classical* properties of quantum mechanics. But what exactly is the difference between quantum mechanics and the classical world? Understanding this difference is vital in learning how to perform information processing tasks that are difficult or impossible with classical physics. This section concludes the chapter with a discussion of the Bell inequality, a compelling example of an essential difference between quantum and classical physics.

When we speak of an object such as a person or a book, we assume that the physical properties of that object have an existence independent of observation. That is, measurements merely act to *reveal* such physical properties. For example, a tennis ball has as one of its physical properties its *position*, which we typically measure using light scattered from the surface of the ball. As quantum mechanics was being developed in the 1920s and 1930s a strange point of view arose that differs markedly from the classical view. As described earlier in the chapter, according to quantum mechanics, an unobserved particle does not possess physical properties that exist independent of observation. Rather, such physical properties arise as a consequence of measurements performed upon the system. For example, according to quantum mechanics a qubit does not possess definite properties of 'spin in the z direction, σ_z', and 'spin in the x direction, σ_x', each of which can be revealed by performing the appropriate measurement. Rather, quantum mechanics gives a set of rules which specify, given the state vector, the probabilities for the possible measurement outcomes when the observable σ_z is measured, or when the observable σ_x is measured.

Many physicists rejected this new view of Nature. The most prominent objector was Albert Einstein. In the famous 'EPR paper', co-authored with Nathan Rosen and Boris Podolsky, Einstein proposed a thought experiment which, he believed, demonstrated that quantum mechanics is not a complete theory of Nature.

The essence of the EPR argument is as follows. EPR were interested in what they termed 'elements of reality'. Their belief was that any such element of reality *must* be represented in any complete physical theory. The goal of the argument was to show that quantum mechanics is not a complete physical theory, by identifying elements of reality that were not included in quantum mechanics. The way they attempted to do this was by introducing what they claimed was a *sufficient condition* for a physical property to

be an element of reality, namely, that it be possible to predict with certainty the value that property will have, immediately before measurement.

Box 2.7: Anti-correlations in the EPR experiment

Suppose we prepare the two qubit state

$$|\psi\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}, \tag{2.213}$$

a state sometimes known as the *spin singlet* for historical reasons. It is not difficult to show that this state is an entangled state of the two qubit system. Suppose we perform a measurement of spin along the $\vec{v}$ axis on both qubits, that is, we measure the observable $\vec{v} \cdot \vec{\sigma}$ (defined in Equation (2.116) on page 90) on each qubit, getting a result of $+1$ or -1 for each qubit. It turns out that no matter what choice of $\vec{v}$ we make, the results of the two measurements are always opposite to one another. That is, if the measurement on the first qubit yields $+1$, then the measurement on the second qubit will yield -1, and vice versa. It is as though the second qubit knows the result of the measurement on the first, no matter how the first qubit is measured. To see why this is true, suppose $|a\rangle$ and $|b\rangle$ are the eigenstates of $\vec{v} \cdot \vec{\sigma}$. Then there exist complex numbers $\alpha, \beta, \gamma, \delta$ such that

$$|0\rangle = \alpha|a\rangle + \beta|b\rangle \tag{2.214}$$
$$|1\rangle = \gamma|a\rangle + \delta|b\rangle. \tag{2.215}$$

Substituting we obtain

$$\frac{|01\rangle - |10\rangle}{\sqrt{2}} = (\alpha\delta - \beta\gamma)\frac{|ab\rangle - |ba\rangle}{\sqrt{2}}. \tag{2.216}$$

But $\alpha\delta - \beta\gamma$ is the determinant of the unitary matrix $\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}$, and thus is equal to a phase factor $e^{i\theta}$ for some real θ. Thus

$$\frac{|01\rangle - |10\rangle}{\sqrt{2}} = \frac{|ab\rangle - |ba\rangle}{\sqrt{2}}, \tag{2.217}$$

up to an unobservable global phase factor. As a result, if a measurement of $\vec{v} \cdot \vec{\sigma}$ is performed on both qubits, then we can see that a result of $+1$ (-1) on the first qubit implies a result of -1 ($+1$) on the second qubit.

Consider, for example, an entangled pair of qubits belonging to Alice and Bob, respectively:

$$\frac{|01\rangle - |10\rangle}{\sqrt{2}}. \tag{2.218}$$

Suppose Alice and Bob are a long way away from one another. Alice performs a measurement of spin along the $\vec{v}$ axis, that is, she measures the observable $\vec{v} \cdot \vec{\sigma}$ (defined in Equation (2.116) on page 90). Suppose Alice receives the result $+1$. Then a simple quantum mechanical calculation, given in Box 2.7, shows that she can predict with certainty

that Bob will measure -1 on his qubit if he also measures spin along the $\vec{v}$ axis. Similarly, if Alice measured -1, then she can predict with certainty that Bob will measure $+1$ on his qubit. Because it is always possible for Alice to predict the value of the measurement result recorded when Bob's qubit is measured in the $\vec{v}$ direction, that physical property must correspond to an element of reality, by the EPR criterion, and should be represented in any complete physical theory. However, standard quantum mechanics, as we have presented it, merely tells one how to calculate the probabilities of the respective measurement outcomes if $\vec{v} \cdot \vec{\sigma}$ is measured. Standard quantum mechanics certainly does not include any fundamental element intended to represent the value of $\vec{v} \cdot \vec{\sigma}$, for all unit vectors $\vec{v}$.

The goal of EPR was to show that quantum mechanics is incomplete, by demonstrating that quantum mechanics lacked some essential 'element of reality', by their criterion. They hoped to force a return to a more classical view of the world, one in which systems could be ascribed properties which existed independently of measurements performed on those systems. Unfortunately for EPR, most physicists did not accept the above reasoning as convincing. The attempt to impose on Nature *by fiat* properties which she must obey seems a most peculiar way of studying her laws.

Indeed, Nature has had the last laugh on EPR. Nearly thirty years after the EPR paper was published, an *experimental test* was proposed that could be used to check whether or not the picture of the world which EPR were hoping to force a return to is valid or not. It turns out that Nature *experimentally invalidates* that point of view, while agreeing with quantum mechanics.

The key to this experimental invalidation is a result known as *Bell's inequality*. Bell's inequality is *not* a result about quantum mechanics, so the first thing we need to do is momentarily *forget* all our knowledge of quantum mechanics. To obtain Bell's inequality, we're going to do a thought experiment, which we will analyze using our common sense notions of how the world works – the sort of notions Einstein and his collaborators thought Nature ought to obey. After we have done the common sense analysis, we will perform a quantum mechanical analysis which we can show *is not consistent with the common sense analysis*. Nature can then be asked, by means of a real experiment, to decide between our common sense notions of how the world works, and quantum mechanics.

Imagine we perform the following experiment, illustrated in Figure 2.4. Charlie prepares two particles. It doesn't matter how he prepares the particles, just that he is capable of repeating the experimental procedure which he uses. Once he has performed the preparation, he sends one particle to Alice, and the second particle to Bob.

Once Alice receives her particle, she performs a measurement on it. Imagine that she has available two different measurement apparatuses, so she could choose to do one of two different measurements. These measurements are of physical properties which we shall label P_Q and P_R, respectively. Alice doesn't know in advance which measurement she will choose to perform. Rather, when she receives the particle she flips a coin or uses some other random method to decide which measurement to perform. We suppose for simplicity that the measurements can each have one of two outcomes, $+1$ or -1. Suppose Alice's particle has a value Q for the property P_Q. Q is assumed to be an *objective property* of Alice's particle, which is merely revealed by the measurement, much as we imagine the position of a tennis ball to be revealed by the particles of light being scattered off it. Similarly, let R denote the value revealed by a measurement of the property P_R.

Similarly, suppose that Bob is capable of measuring one of two properties, P_S or P_T, once again revealing an objectively existing value S or T for the property, each taking value $+1$ or -1. Bob does not decide beforehand which property he will measure, but waits until he has received the particle and then chooses randomly. The timing of the experiment is arranged so that Alice and Bob do their measurements *at the same time* (or, to use the more precise language of relativity, in a causally disconnected manner). Therefore, the measurement which Alice performs cannot disturb the result of Bob's measurement (or vice versa), since physical influences cannot propagate faster than light.

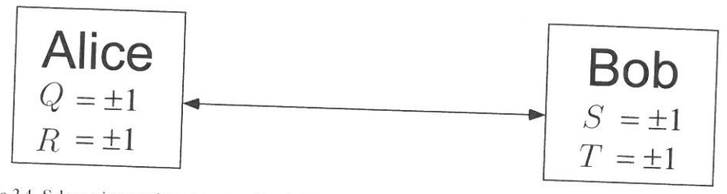

Figure 2.4. Schematic experimental setup for the Bell inequalities. Alice can choose to measure either Q or R, and Bob chooses to measure either S or T. They perform their measurements simultaneously. Alice and Bob are assumed to be far enough apart that performing a measurement on one system can not have any effect on the result of measurements on the other.

We are going to do some simple algebra with the quantity $QS + RS + RT - QT$. Notice that

$$QS + RS + RT - QT = (Q + R)S + (R - Q)T. \tag{2.219}$$

Because $R, Q = \pm 1$ it follows that either $(Q + R)S = 0$ or $(R - Q)T = 0$. In either case, it is easy to see from (2.219) that $QS + RS + RT - QT = \pm 2$. Suppose next that $p(q, r, s, t)$ is the probability that, before the measurements are performed, the system is in a state where $Q = q, R = r, S = s$, and $T = t$. These probabilities may depend on how Charlie performs his preparation, and on experimental noise. Letting $\mathbf{E}(\cdot)$ denote the mean value of a quantity, we have

$$\mathbf{E}(QS + RS + RT - QT) = \sum_{qrst} p(q, r, s, t)(qs + rs + rt - qt) \tag{2.220}$$

$$\leq \sum_{qrst} p(q, r, s, t) \times 2 \tag{2.221}$$

$$= 2. \tag{2.222}$$

Also,

$$\mathbf{E}(QS + RS + RT - QT) = \sum_{qrst} p(q, r, s, t)qs + \sum_{qrst} p(q, r, s, t)rs + \sum_{qrst} p(q, r, s, t)rt - \sum_{qrst} p(q, r, s, t)qt \tag{2.223}$$

$$= \mathbf{E}(QS) + \mathbf{E}(RS) + \mathbf{E}(RT) - \mathbf{E}(QT). \tag{2.224}$$

Comparing (2.222) and (2.224) we obtain the *Bell inequality*,

$$\mathbf{E}(QS) + \mathbf{E}(RS) + \mathbf{E}(RT) - \mathbf{E}(QT) \leq 2. \tag{2.225}$$

This result is also often known as the *CHSH inequality* after the initials of its four discoverers. It is part of a larger set of inequalities known generically as Bell inequalities, since the first was found by John Bell.

By repeating the experiment many times, Alice and Bob can determine each quantity on the left hand side of the Bell inequality. For example, after finishing a set of experiments, Alice and Bob get together to analyze their data. They look at all the experiments where Alice measured P_Q and Bob measured P_S. By multiplying the results of their experiments together, they get a sample of values for QS. By averaging over this sample, they can estimate $\mathbf{E}(QS)$ to an accuracy only limited by the number of experiments which they perform. Similarly, they can estimate all the other quantities on the left hand side of the Bell inequality, and thus check to see whether it is obeyed in a real experiment.

It's time to put some quantum mechanics back in the picture. Imagine we perform the following quantum mechanical experiment. Charlie prepares a quantum system of two qubits in the state

$$|\psi\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}. \tag{2.226}$$

He passes the first qubit to Alice, and the second qubit to Bob. They perform measurements of the following observables:

$$Q = Z_1 \qquad S = \frac{-Z_2 - X_2}{\sqrt{2}} \tag{2.227}$$

$$R = X_1 \qquad T = \frac{Z_2 - X_2}{\sqrt{2}}. \tag{2.228}$$

Simple calculations show that the average values for these observables, written in the quantum mechanical $\langle \cdot \rangle$ notation, are:

$$\langle QS \rangle = \frac{1}{\sqrt{2}}; \ \langle RS \rangle = \frac{1}{\sqrt{2}}; \ \langle RT \rangle = \frac{1}{\sqrt{2}}; \ \langle QT \rangle = -\frac{1}{\sqrt{2}}. \tag{2.229}$$

Thus,

$$\langle QS \rangle + \langle RS \rangle + \langle RT \rangle - \langle QT \rangle = 2\sqrt{2}. \tag{2.230}$$

Hold on! We learned back in (2.225) that the average value of QS plus the average value of RS plus the average value of RT minus the average value of QT can never exceed two. Yet here, quantum mechanics predicts that this sum of averages yields $2\sqrt{2}$!

Fortunately, we can ask Nature to resolve the apparent paradox for us. Clever experiments using photons – particles of light – have been done to check the prediction (2.230) of quantum mechanics versus the Bell inequality (2.225) which we were led to by our common sense reasoning. The details of the experiments are outside the scope of the book, but the results were resoundingly in favor of the quantum mechanical prediction. The Bell inequality (2.225) is *not* obeyed by Nature.

What does this mean? It means that one or more of the assumptions that went into the derivation of the Bell inequality must be incorrect. Vast tomes have been written analyzing the various forms in which this type of argument can be made, and analyzing the subtly different assumptions which must be made to reach Bell-like inequalities. Here we merely summarize the main points.

There are two assumptions made in the proof of (2.225) which are questionable:

(1) The assumption that the physical properties P_Q, P_R, P_S, P_T have definite values Q, R, S, T which exist independent of observation. This is sometimes known as the assumption of *realism*.
(2) The assumption that Alice performing her measurement does not influence the result of Bob's measurement. This is sometimes known as the assumption of *locality*.

These two assumptions together are known as the assumptions of *local realism*. They are certainly intuitively plausible assumptions about how the world works, and they fit our everyday experience. Yet the Bell inequalities show that at least one of these assumptions is not correct.

What can we learn from Bell's inequality? For physicists, the most important lesson is that their deeply held commonsense intuitions about how the world works are wrong. The world is *not* locally realistic. Most physicists take the point of view that it is the assumption of realism which needs to be dropped from our worldview in quantum mechanics, although others have argued that the assumption of locality should be dropped instead. Regardless, Bell's inequality together with substantial experimental evidence now points to the conclusion that either or both of locality and realism must be dropped from our view of the world if we are to develop a good intuitive understanding of quantum mechanics.

What lessons can the fields of quantum computation and quantum information learn from Bell's inequality? Historically the most useful lesson has perhaps also been the most vague: there is something profoundly 'up' with entangled states like the EPR state. A lot of mileage in quantum computation and, especially, quantum information, has come from asking the simple question: 'what would some entanglement buy me in this problem?' As we saw in teleportation and superdense coding, and as we will see repeatedly later in the book, by throwing some entanglement into a problem we open up a new world of possibilities unimaginable with classical information. The bigger picture is that Bell's inequality teaches us that entanglement is a fundamentally new resource in the world that goes essentially *beyond* classical resources; iron to the classical world's bronze age. A major task of quantum computation and quantum information is to exploit this new resource to do information processing tasks impossible or much more difficult with classical resources.

Problem 2.1: (Functions of the Pauli matrices) Let $f(\cdot)$ be any function from complex numbers to complex numbers. Let $\vec{n}$ be a normalized vector in three dimensions, and let θ be real. Show that

$$f(\theta \vec{n} \cdot \vec{\sigma}) = \frac{f(\theta) + f(-\theta)}{2} I + \frac{f(\theta) - f(-\theta)}{2} \vec{n} \cdot \vec{\sigma}. \tag{2.231}$$

Problem 2.2: (Properties of the Schmidt number) Suppose $|\psi\rangle$ is a pure state of a composite system with components A and B.

(1) Prove that the Schmidt number of $|\psi\rangle$ is equal to the rank of the reduced density matrix $\rho_A \equiv \mathrm{tr}_B(|\psi\rangle\langle\psi|)$. (Note that the rank of a Hermitian operator is equal to the dimension of its support.)
(2) Suppose $|\psi\rangle = \sum_j |\alpha_j\rangle|\beta_j\rangle$ is a representation for $|\psi\rangle$, where $|\alpha_j\rangle$ and $|\beta_j\rangle$ are (un-normalized) states for systems A and B, respectively. Prove that the

number of terms in such a decomposition is greater than or equal to the Schmidt number of $|\psi\rangle$, $\mathrm{Sch}(\psi)$.

(3) Suppose $|\psi\rangle = \alpha|\varphi\rangle + \beta|\gamma\rangle$. Prove that

$$\mathrm{Sch}(\psi) \geq |\mathrm{Sch}(\varphi) - \mathrm{Sch}(\gamma)| . \tag{2.232}$$

Problem 2.3: (Tsirelson's inequality) Suppose $Q = \vec{q}\cdot\vec{\sigma}, R = \vec{r}\cdot\vec{\sigma}, S = \vec{s}\cdot\vec{\sigma}, T = \vec{t}\cdot\vec{\sigma}$, where $\vec{q}, \vec{r}, \vec{s}$ and $\vec{t}$ are real unit vectors in three dimensions. Show that

$$(Q\otimes S + R\otimes S + R\otimes T - Q\otimes T)^2 = 4I + [Q,R]\otimes[S,T]. \tag{2.233}$$

Use this result to prove that

$$\langle Q\otimes S\rangle + \langle R\otimes S\rangle + \langle R\otimes T\rangle - \langle Q\otimes T\rangle \leq 2\sqrt{2}, \tag{2.234}$$

so the violation of the Bell inequality found in Equation (2.230) is the maximum possible in quantum mechanics.

History and further reading

There are an enormous number of books on linear algebra at levels ranging from High School through to Graduate School. Perhaps our favorites are the two volume set by Horn and Johnson[HJ85, HJ91], which cover an extensive range of topics in an accessible manner. Other useful references include Marcus and Minc[MM92], and Bhatia[Bha97]. Good introductions to linear algebra include Halmos[Hal58], Perlis[Per52], and Strang[Str76].

There are many excellent books on quantum mechanics. Unfortunately, most of these books focus on topics of tangential interest to quantum information and computation. Perhaps the most relevant in the existing literature is Peres' superb book[Per93]. Beside an extremely clear exposition of elementary quantum mechanics, Peres gives an extensive discussion of the Bell inequalities and related results. Good introductory level texts include Sakurai's book[Sak95], Volume III of the superb series by Feynman, Leighton, and Sands[FLS65a], and the two volume work by Cohen-Tannoudji, Diu and Laloë[CTDL77a, CTDL77b]. All three of these works are somewhat closer in spirit to quantum computation and quantum information than are most other quantum mechanics texts, although the great bulk of each is still taken up by applications far removed from quantum computation and quantum information. As a result, none of these texts need be read in detail by someone interested in learning about quantum computation and quantum information. However, any one of these texts may prove handy as a reference, especially when reading articles by physicists. References for the history of quantum mechanics may be found at the end of Chapter 1.

Many texts on quantum mechanics deal only with projective measurements. For applications to quantum computing and quantum information it is more convenient – and, we believe, easier for novices – to start with the general description of measurements, of which projective measurements can be regarded as a special case. Of course, ultimately, as we have shown, the two approaches are equivalent. The theory of generalized measurements which we have employed was developed between the 1940s and 1970s. Much of the history can be distilled from the book of Kraus[Kra83]. Interesting discussion of quantum measurements may be found in Section 2.2 of Gardiner[Gar91], and in the book by Braginsky and Khahili[BK92]. The POVM measurement for distinguishing

non-orthogonal states described in Section 2.2.6 is due to Peres[Per88]. The extension described in Exercise 2.64 appeared in Duan and Guo[DG98].

Superdense coding was invented by Bennett and Wiesner[BW92]. An experiment implementing a variant of superdense coding using entangled photon pairs was performed by Mattle, Weinfurter, Kwiat, and Zeilinger[MWKZ96].

The density operator formalism was introduced independently by Landau[Lan27] and by von Neumann[von27]. The unitary freedom in the ensemble for density matrices, Theorem 2.6, was first pointed out by Schrödinger[Sch36], and was later rediscovered and extended by Jaynes[Jay57] and by Hughston, Jozsa and Wootters[HJW93]. The result of Exercise 2.73 is from the paper by Jaynes, and the results of Exercises 2.81 and 2.82 appear in the paper by Hughston, Jozsa and Wootters. The class of probability distributions which may appear in a density matrix decomposition for a given density matrix has been studied by Uhlmann[Uhl70] and by Nielsen[Nie99b]. Schmidt's eponymous decomposition appeared in[Sch06]. The result of Exercise 2.77 was noted by Peres[Per95].

The EPR thought experiment is due to Einstein, Podolsky and Rosen[EPR35], and was recast in essentially the form we have given here by Bohm[Boh51]. It is sometimes misleadingly referred to as the EPR 'paradox'. The Bell inequality is named in honour of Bell[Bel64], who first derived inequalities of this type. The form we have presented is due to Clauser, Horne, Shimony, and Holt[CHSH69], and is often known as the CHSH inequality. This inequality was derived independently by Bell, who did not publish the result.

Part 3 of Problem 2.2 is due to Thapliyal (private communication). Tsirelson's inequality is due to Tsirelson[Tsi80].