Chapter 11
Quantum Information

"*One could caricature quantum information processing as the science of turning quantum conundrums into potentially useful applications.*"
– Nicolas Gisin[1]

Classical information theory, invented by Claude Shannon in 1948, addresses two main issues: the degree to which a classical message (i.e., a sequence of symbols) can be compressed, and the maximum rate at which reliable communications can be sustained over a noisy communications channel. The quantitative statement regarding the maximum compressibility of a symbol sequence is enshrined in Shannon's "Noiseless Source Coding Theorem", and the quantitative statement regarding the maximum rate of reliable communications, for a given noise level in the channel, is enshrined in Shannon's "Noisy Channel Coding Theorem". Together, these theorems laid the foundations for several multi-billion dollar industries such as telecommunications, cellular phone networks, internet, and disk drives. In fact, we make use of information theory everyday but barely give it any thought whatsoever.

Since information theory was invented, engineers have refined communications and data storage devices constantly so that they use fewer physical resources to encode more information. This has enabled dramatic increases in the storage capacity of computer memories, significant reductions in the power consumption of communications devices, and large increases in the rate at which information can be exchanged. Indeed, codes are now known that operate surprisingly close to the limits implied by Shannon's theorems.

In this chapter we consider how information theory needs to be modified once we use the quantum states of simple systems (such as photons) to encode symbols. We might expect that some modification is necessary because, e.g., whereas symbols encoded in the states of classical physical systems are guaranteed to be distinguishable, the same cannot be said for symbols encoded in the states of quantum systems (e.g., if they are non-orthogonal). But, in fact, the reasons for modification runs much deeper than this: Some elementary information processing operations, such

[1] Source: in "Quantum Cryptography" Reviews of Modern Physics, Volume **74**, January (2002).

C.P. Williams, *Explorations in Quantum Computing,*
Texts in Computer Science,
DOI 10.1007/978-1-84628-887-6_11, © Springer-Verlag London Limited 2011

as copying data, which are permitted on classical information are impossible when attempted on quantum information. Conversely, other operations, such as teleportation, which are impossible when using classical information, can be achieved using quantum information.

As in the case of computer science, this shift in the foundations of the field turns out to have profound consequences. In particular, it leads to new (quantum) versions of both the noiseless coding theorem and the noisy channel coding theorem. As you shall see, quantum information theory forces us to revise our most cherished assumptions regarding how information should behave.

11.1 What is Classical Information?

> "*It might even be fair to observe that the concept that information is fundamental is very old knowledge of humanity, witness for example the beginning of the gospel according to John*: "*In the beginning was the Word*" "
> – Anton Zeilinger[2]

Most people have an intuitive understanding of what they mean by "information". It's the stuff they read in newspapers, copy off blackboards, or absorb while watching CNN etc. However, when pressed to give a more precise definition, I find that most people equate "information" with the knowledge imparted during some communicative act, i.e., what they know now that they didn't know before. This implicitly connects "information" with the meaning of a communication, i.e., its qualitative aspects.

A problem with this position, is that it makes the "information" contained within a message highly subjective, hard to quantify, and context dependent. For example, the "information" two people may attach to a CNN report would then depend on what they knew beforehand. It is tricky to make any mathematical headway with such a subjective basis for a notion of "information". So the commonsense view of "information" as the knowledge imparted during some communicative act is not very useful in a practical sense.

In 1948 Claude Shannon hit upon an alternative view of what we should mean by "information". He suggested the information within a message was simply the minimum number of 0s and 1s needed to transmit it. In Shannon's own words [14]:

"*The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.*"

[2] Source: [560].

Shannon's insight was as ingenious as it was dehumanizing! By equating "information" with the minimal resources needed to *represent* a message, rather than its *knowledge content* per se, it became possible to derive laws describing how the amount of information would change under various operations, such as compressing messages or sending them through noisy communications channels. In turn, such understanding led to breakthroughs in data compression, encryption, and telecommunications.

Yet the cost is severe. Shannon's perspective strips all humanity from the notion of information. In Shannon's theory a love letter might have the same information content as a bus schedule, since his notion of information only addresses its quantitative aspects not its qualitative aspects. "Information" became something sterile, lifeless, and devoid of passion or creativity. Nevertheless, the operational utility of regarding information as the minimum number of 0s and 1s needed to encode some message is currently the best handle we have on quantifying the elusive and abstract notion of "information".

11.1.1 Classical Sources: The Shannon Entropy

We can think of a source of classical information as a device that produces a stream of classical symbols, such as lowercase letters, uppercase letters, numbers, and punctuation marks. After large numbers of such symbols have been produced we can determine their probability distribution. In principle, all sorts of subtle correlations amongst the symbols are possible. For example, in English the symbol "*q*" is followed, invariably, by the symbol "*u*" as in the words such "*quantum*", "*quest*", "*quibble*", and "*quoff*". Nevertheless, as each distinct symbol can be encoded as a corresponding binary string, we can equally think of a source of classical information as a device that produces sequences of *bits*, i.e., 0s and 1s. Consequently, correlations amongst the symbols would then appear as correlations amongst *subsequences* of bits. However, correlations at the level of *individual* bits would tend to be diluted out.

How one sets up the mapping between symbols and bit strings makes a difference. For example, the frequencies with which different letters arise in written English are different (see Fig. 11.1) with "*e*" being the most common letter. Similarly, one could treat whole words as "symbols" and plot *their* frequency of occurrence too. Such statistical insights into the structure of natural languages have permitted modern marvels such as smarter internet search engines (which exploit word correlations to infer context and relevance) and statistical machine translation tools (which can teach themselves to translate documents by being "trained" to infer the mathematical correlations between the words and phrases found in matching pairs of human-translations of large corpora of documents). When one makes the sensible choice of using shorter bit strings to encode more frequent symbols in a language, one finds that although we can model a source of the language as a stream of independent, identically distributed, bits in which 0 occurs with probability $p_0 = p$ and

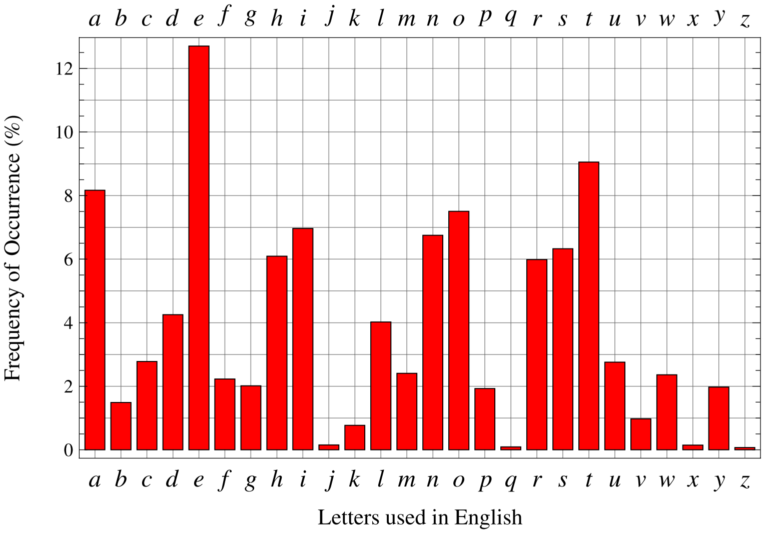

Fig. 11.1 Letter frequency distribution in English

1 occurs with probability $p_1 = 1 - p$, that for real languages there is an asymmetry between p_0 and p_1. Ultimately, this asymmetry is what allows us to compress messages.

Specifically, if 0 occurs with probability $p_0 = p$ and 1 occurs with probability $p_1 = 1 - p$, a "typical" n-bit message will have roughly np 0s and $n(1-p)$ 1s. Hence, the number of "typical" bit strings is therefore:

$$\binom{n}{np} = \frac{n!}{(np)!(n-np)!} \tag{11.1}$$

Using Stirling's formula $N! \approx N^N e^{-N}\sqrt{2\pi N}$ for $N \gg 1$ we have $\log_e N! \approx N \log_e N - N$ and so:

$$\begin{aligned}\log_e \binom{n}{np} &\approx n\log_e n - n - (np\log_e np - np \\ &\quad + (n-np)\log_e(n-np) - (n-np)) \\ &= n(-p\log_e p - (1-p)\log_e(1-p)) \\ &= \frac{1}{\log_2 e}\, n(-p\log_2 p - (1-p)\log_2(1-p)) \\ &\approx nH(\{p, 1-p\}) \end{aligned} \tag{11.2}$$

where $\{p_i\}$ is the set of positive real numbers defining the *probability* with which each possible symbol appears in the message. In the case of bit string messages there are only two symbols, $i = 0$ and $i = 1$, and so the probabilities are simply p_0 and

$p_1 = 1 - p_0$. The function $H(\{p_0, p_1\}) = -\sum_{i=0}^{1} p_i \log_2 p_i$ is called the *Shannon entropy*. For symbols that are just single bits we have $H(\{p_i\}) \equiv H(\{p_0, p_1\}) \equiv H(\{p, 1-p\})$.

The choice of which base to use for the logarithm is somewhat arbitrary as different choices only serve to re-scale the measure of information (or entropy) by a constant factor. If we choose to use base 2, our scale has a certain natural feel to it. Using base 2 logarithms, if $p_0 = p_1 = \frac{1}{2}$, an n-bit classical message would be completely random (and hence incompressible) and would convey exactly $nH(\{\frac{1}{2}, \frac{1}{2}\}) = n$ bits of information. At the other extreme, a string of n identical bits, such as n 0s (and hence devoid of any useful information), would convey $nH(\{1, 0\}) = 0$ bits of information. So by choosing base 2, we arrive at a fairly intuitive scale for information.

11.1.2 Maximal Compression (Source Coding Theorem)

"Source coding" refers to the data compression problem. That is, given a source producing a sequence of symbols in accordance with some a priori probability distribution, by what factor can we compress a typical message from this source without corrupting it? If no information whatsoever is lost, the compression is said to be "lossless". But in many cases we are content with a "lossy" compression provided the losses do not rise to a level we perceive as significant.

We can approach this question with the help of Shannon information theory. Suppose we model the source as emitting a sequence of independent, identically distributed, bits in which 0 occurs with probability p_0 and 1 occurs with probability $p_1 = 1 - p_0$. Then most n-bit messages generated by such a source will be close to the "typical" messages. That is, they will have close to np_0 0's and $n(1 - p_0)$ 1's. Therefore, we need only worry about how sending "typical" messages. So rather than there being $\mathcal{O}(2^n)$ messages to worry about, we only really need to figure out how to handle $\mathcal{O}(2^{nH(\{p_0,p_1\})})$ typical messages. All we need to do is to assign a unique positive integer to each typical message, and send that integer, which requires only $nH(\{p_0, p_1\})$ bits, rather than the message, which requires n bits. As $n \to$ inf almost all message will be close to typical. For example, if $p_0 = 0.3$ and $p_1 = 0.7$, then a "typical" 20-bit message would have six 0's and fourteen 1's, and instead of there being $2^{20} \approx 1{,}000{,}000$ *possible* messages to send there would be only $2^{nH(\{0.3,0.7\})} \approx 200{,}000)$ *typical* messages to send.

The notion of the entropy of a source that emits one of two possible symbols, i.e., a binary source, can be generalized readily to one that emits one of d possible symbols, $x_1, x_2, \ldots, x_d$. Assuming symbol x_i appears with probability p_i, a typical message of length $n \gg 1$ symbols from such a source will have roughly np_1 occurrences of x_1, np_2 occurrences of x_2, etc. Hence the number of such typical messages is given by the number of ways np_1 x_1's, np_2 x_2's, etc. can be placed within a string of length n symbols, which is just the multinomial formula:

$$\frac{n!}{\prod_{i=1}^{d}(np_i)!} \tag{11.3}$$

such that $0 \leq p_i \leq 1$ and $\sum_{i=1}^{d} p_i = 1$. We can write this approximately as an exponential function of a modified entropy function

$$\frac{n!}{\prod_i (np_i)!} \approx 2^{nH(\{p_1, p_2, \ldots, p_d\})} \tag{11.4}$$

if we define

$$H(\{p_1, p_2, \ldots, p_d\}) = -\sum_{i=1}^{d} p_i \log_2 p_i \tag{11.5}$$

Such a generalization to the case of alphabets having d-symbols gives the Source Coding Theorem:

Source Coding Theorem *If n independent, identically distributed, random variables taken from a finite d-symbol alphabet each with entropy $H(\{p_1, p_2, \ldots, p_d\})$ are compressed into no fewer than $nH(\{p_1, p_2, \ldots, p_d\})$ bits then there is negligible risk of information loss, but compression beyond this limit makes some loss almost certain.*

For natural languages this notion of source coding is appropriate. But in other fields, e.g., mathematics and computer science, strings of letters and symbols arise that although outwardly complex if viewed as a symbol sequence, are actually much simpler if one understands the underlying generator. In such cases algorithmic information theory is a better tool for understanding their compressibility. In particular, Kolmogorov complexity is the shortest program needed to reproduce some sequence. So the Kolmogorov complexity of a truly random sequence is the sequence itself as a random sequence is, be definition, incompressible. In contrast, the sequence of (say) Fibonacci numbers, in which each successive number is the sum of the last two numbers, i.e., $1, 1, 2, 3, 5, 8, 13, 21, 34, \ldots$ can be describe more compactly via the recursive formula $f(n) : f(n) = f(n-1) + f(n-2)$ for $n \geq 3 \wedge f(1) = f(2) = 1$. This is dramatically shorter than writing out the sequence itself.

11.1.3 Reliable Transmission (Channel Coding Theorem)

Besides compression, another aspect of information theory is to ask how *reliably* information may be *conveyed* over a noisy communications channel. A typical communications channel adds noise to any signal sent through it causing errors in the data received. Attempts to correct such errors are prone to errors themselves. It is not obvious a priori, therefore, that a noisy communications channel *can* be used to transmit messages without error. Remarkably, in 1948 Claude Shannon proved a theorem that showed, regardless of how noisy a given channel may be, that it is always possible to communicate information over such a channel almost error free up to a certain maximum rate set by the Channel Coding theorem. The method for doing so relies upon the use of error correcting codes, but the Channel Coding theorem does not tell us how to find these good codes, only that they exist. Nevertheless, since the advent of the Channel Coding theorem many excellent codes have been discovered, driven in large part by the needs of deep Space communications for supporting reliable communications during NASA Space missions. In particular, Turbo Codes, and Low Density Parity-Check Codes now come close to saturating the limit set by Shannon's Channel Coding theorem.

To state the theorem quantitatively we need a few key ideas. First the notion of a discrete channel is one consisting of an input alphabet $\mathcal{X}$ and an output alphabet $\mathcal{Y}$ and a probability transition matrix $p(Y|X)$, which specifies the probability of receiving symbol $Y \in \mathcal{Y}$ given that symbol $X \in \mathcal{X}$ was sent. When this probability distribution only depends on the last input to the channel, the channel is said to be "memoryless". We can also define the marginal probabilities of seeing the different symbols as $p(x = X) = \sum_y p(x, y)$ and $p(y = Y) = \sum_x p(x, y)$, where $p(x, y)$ is the joint probability of seeing $x = X$ and $y = Y$. From these we construct the mutual information $I(X : Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x,y)}{p(x)p(y)}$, which is a measure of how much the two variables depend on each another. Then the channel capacity, C, of a discrete memoryless channel, can be defined to be the mutual information maximized over all probability distributions, i.e.,

$$C = \max_{p(X)} I(X : Y) \tag{11.6}$$

The relationship between entropy, conditional entropy, joint entropy, and mutual information is shown in Fig. 11.2. Formally, the Channel Coding theorem then establishes the maximum rate at which reliable communications can be supported given the characteristics of the channel.

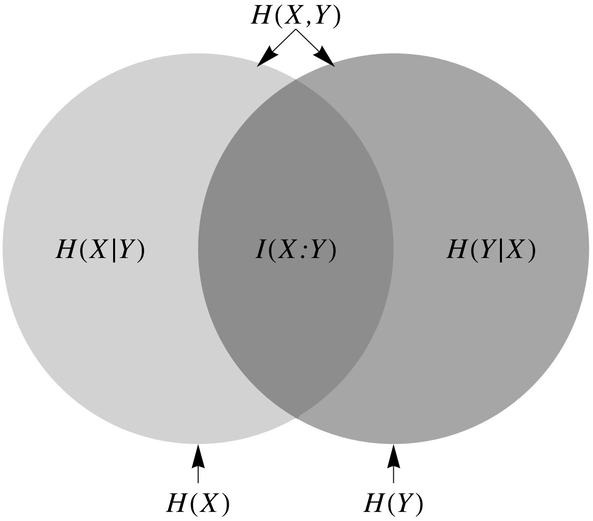

Fig. 11.2 Graphical illustration of the relationship between entropy ($H(X)$, $H(Y)$), conditional entropy ($H(X|Y)$ and $H(Y|X)$), joint entropy ($H(X, Y)$) and mutual information ($I(X : Y)$). Formally we have $I(X : Y) = H(X) + H(Y) - H(X|Y)$ or, equivalently, $I(X : Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$. Furthermore, $I(X : Y) = I(Y : X)$ and $I(X : X) = H(X)$. Mutual information is a way to quantify the degree to which two variables depend on each other

Channel Coding Theorem *For any upper bound on the acceptable block error rate,* $\epsilon > 0$, *and for any rate* $R < C$ (*where* $C = \max_{p(X)} I(X; Y)$ *is the channel capacity*), *there is an encoding/decoding protocol that can guarantee that the probability of block error is less than* ϵ *for a sufficiently long code. Moreover, for any rate* $R > C$, *i.e., if the communications rate attempted exceeds the channel capacity, errors are inevitable.*

A proof the Channel Coding theorem is given in Chapter 8 of Cover & Thomas's "Elements of Information Theory" [117].

11.1.4 Unstated Assumptions Regarding Classical Information

> "*Information is physical.*"
> – Rolf Landauer

Just as the inventors of classical computer science had attempted to construct a theory of computation that was independent of how computers were implemented, so too did Shannon attempt to construct a theory of information that was supposed to be independent of how symbols were implemented. By building information theory on such a mathematical ideal, Shannon was able to make heroic advances in modeling data compression and communications channels and hence designing superior telecommunications systems. However, accepting this mathematical ideal as reality, causes people to assume (implicitly perhaps) that information has certain eminently reasonable properties. Back in 1948 and for most of the time since then, these assumptions have in fact been so obvious that no-one has ever really questioned them—until now. For example, thinking of information as the mathematical ideal of a stream of symbols invites the following presumptions:

- Information consists of a stream of *distinguishable* symbols
- Information can be compressed to no more than the Shannon bound
- Information does not change upon being read
- Information can be read in part without it affecting the unread parts
- Information can be copied exactly deterministically
- Information can be negated trivially by flipping every bit value

Indeed, the remarkable advances in communications systems since 1948 bear witness to how effective Shannon's theory has been, and how solidly these assumptions have been upheld.

Yet when we reduce the scale of the systems encoding information to individual quantum systems, then the nature of information itself begins to change. Under the right circumstances *every one of the aforementioned plausible statements about information can be made false*. The fundamental reason for this, as Richard Feynman put it, is that "Nature isn't classical dammit!". Indeed it is not. Our preconceptions of the properties that information should possess are intimately tied to the (more implicit) assumptions for how such information is implemented. Just as computation should be seen as a *physical* process that depends in an essential way on the physical systems being used to enact computations, so it is for quantum information systems too.

11.2 What is Quantum Information?

The concept of quantum information is derived quite readily from that of classical information. Whereas classical information is a sequence of bits quantum information is a sequence of qubits. Entirely new types of phenomena are possible with quantum information that have no counterparts in classical information. For example, the successive qubits in a quantum message need not, and generally are not, orthogonal to one another, nor are they necessarily unentangled from one another. Thus a typical quantum memory register holds within it *quantum* information rather than classical information. As such it will typically hold information in an entangled superposition state, and the strengths of the correlations between bit values can exceed that which is possible classically.

11.2.1 Pure States cf. Mixed States

So far we have been mostly concerned with situations in which we have *complete* knowledge of the state of some n-qubit quantum memory register. That is, there is no uncertainty whatsoever regarding its state. It exists in some superposition of the possible bit string configurations of n bits, weighted by various amplitudes corresponding (via their modulus squared) to the probabilities of obtaining that particular bit string configuration if the memory register were to be read in the computational basis. In other words, the n-qubit register is in a state of the form:

$$|\psi\rangle = c_0|00\dots 0\rangle + c_1|00\dots 1\rangle + \cdots + c_{2^n-1}|11\dots 1\rangle \tag{11.7}$$

such that $\sum_{i=0}^{2^n-1}|c_i|^2 = 1$. Such a quantum state is said to be a *pure* state.

There are, however, situations in which we have only *incomplete* knowledge about some quantum state. Such states are called mixed states, as they correspond to weighted mixtures of different pure states.

11.2.2 Mixed States from Partial Knowledge: The Density Operator

One way mixed states can arise is when we only have probabilistic knowledge regarding the composition of a quantum state. Suppose, for example, that we only known that a quantum system is in one of the (not necessarily orthogonal) states $|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_N\rangle$ with probabilities $p_1, p_2, \ldots, p_N$ respectively such that $\sum_{i=1}^{N} p_i = 1$. We are therefore a little uncertain of what the state actually is. How are we to characterize the quantum state of such a system?

One way we might learn something about the state is to make some sort of measurement on it. If we performed a measurement, described by the observable $\mathcal{O}$, on this system, the result we would expect to obtain would be the weighted average of the results we would obtain if the system was in each of the states $|\psi_1\rangle, |\psi_2\rangle, \ldots,$ or $|\psi_N\rangle$, namely:

$$\langle \mathcal{O} \rangle = \sum_{i=1}^{N} p_i \langle \psi_i | \mathcal{O} | \psi_i \rangle \tag{11.8}$$

which after some manipulations (see problem 11.12) can be re-written as:

$$\sum_{i=1}^{N} p_i \langle \psi_i | \mathcal{O} | \psi_i \rangle = \text{tr}\left(\left(\sum_{i=1}^{N} p_i | \psi_i \rangle \langle \psi_i | \right) \cdot \mathcal{O} \right) = \text{tr}(\rho \cdot \mathcal{O}) \tag{11.9}$$

where "tr(·)" is the sum of the diagonal elements (i.e. the "trace") of its argument (which is a matrix), and $\rho = \sum_{i=1}^{N} p_i |\psi_i\rangle\langle\psi_i|$ (which is also a matrix). Notice that ρ contains information only about the statistical mixture of pure states that contribute to the state, and $\mathcal{O}$ contains information only about the observable being measured. Hence, ρ must be a complete characterization of the mixed state.

Density Operator If a quantum system exists in the state $|\psi_1\rangle$ with probability p_1, $|\psi_2\rangle$ with probability p_2, ..., $|\psi_N\rangle$ with probability p_N, where in general $\langle\psi_i|\psi_j\rangle \neq 1$ for $i \neq j$, then the best description of its state is given by the density operator:

$$\rho = \sum_{i=1}^{N} p_i |\psi_i\rangle\langle\psi_i| \tag{11.10}$$

11.2.2.1 Density Operator for a Mixed State

Although you can use density operators to describe pure states, the main motivation for introducing them is to be able to represent *mixed* states, i.e., statistical mixtures of pure states. This allows us to model circumstances in which we only have partial knowledge regarding the state. Specifically, if a quantum system exists in the state $|\psi_1\rangle$ with probability p_1, $|\psi_2\rangle$ with probability p_2, ..., $|\psi_N\rangle$ with probability p_N, where in general $\langle\psi_i|\psi_j\rangle \neq 1$ for $i \neq j$, then the best description of its state is given by the density operator:

$$\rho = \sum_{i=1}^{N} p_i |\psi_i\rangle\langle\psi_i| \tag{11.11}$$

where $\sum_{i=1} p_N = 1$. Here the component states need not be orthogonal with respect to one another, i.e., in general $\langle\psi_i|\psi_j\rangle \neq 0$ for $i \neq j$.

Many people are puzzled about the distinction between a mixed state and a superposition state, so it is worth stating this explicitly. A superposition state is a completely known pure state consisting of a weighted sum of eigenstates, $|\psi\rangle = \sum_i c_i|i\rangle$, which are all orthogonal with respect to one another, i.e., $\langle i|j\rangle = 0$ for all $i \neq j$. In principle, given knowledge of a superposition state, $|\psi\rangle$, one could build a measuring device that always yielded the same predictable result each time you used it to measure state $|\psi\rangle$. For example, if we had a single qubit in the superposition state $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ we could rotate a measuring device that measures in the $\{|0\rangle, |1\rangle\}$ basis by 45° and then it would be measuring in the $\{|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\}$ basis, and always yield the result $|+\rangle$.

In contrast, a mixed state $\rho = \sum_j p_j|\phi_j\rangle\langle\phi_j|$ is an incompletely known state in which the component pure states (described by density operators $|\phi_j\rangle\langle\phi_j|$) need not be, and generally are not, orthogonal to one another. The fact that the state is incompletely known means that you can never be sure whether you really are dealing with a $|\phi_1\rangle$, or a $|\phi_2\rangle$, etc. Consequently, even if you know ρ, you cannot pick a measurement basis for a mixed state that is always guaranteed to yield the same predictable outcome.

The following example illustrates how to calculate the density operator of a mixed state that is a combination of three non-orthogonal pure states, $|\psi_1\rangle$, $|\psi_2\rangle$, and $|\psi_3\rangle$, with probabilities $p_1 = \frac{1}{3}$, $p_2 = \frac{2}{5}$ and $p_3 = \frac{4}{15}$ respectively where:

$$|\psi_1\rangle = |0\rangle \tag{11.12}$$

$$|\psi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \tag{11.13}$$

$$|\psi_3\rangle = \frac{1}{2}|0\rangle + i\frac{\sqrt{3}}{2}|1\rangle) \tag{11.14}$$

The corresponding density operator is:

$$\begin{aligned}
\rho &= p_1|\psi_1\rangle\langle\psi_1| + p_2|\psi_2\rangle\langle\psi_2| + p_3|\psi_3\rangle\langle\psi_3| \\
&= \frac{1}{3}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{2}{5}\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} + \frac{4}{15}\begin{pmatrix} \frac{1}{4} & -i\frac{\sqrt{3}}{4} \\ i\frac{\sqrt{3}}{4} & \frac{3}{4} \end{pmatrix} \\
&= \begin{pmatrix} \frac{3}{5} & -\frac{1}{5} - i\frac{1}{5\sqrt{3}} \\ -\frac{1}{5} + i\frac{1}{5\sqrt{3}} & \frac{2}{5} \end{pmatrix}
\end{aligned} \tag{11.15}$$

Note that $\mathrm{tr}(\rho) = 1$ (as for a pure state), but since

$$\rho^2 = \rho \cdot \rho = \begin{pmatrix} \frac{31}{75} & -\frac{1}{5} - i\frac{1}{5\sqrt{3}} \\ -\frac{1}{5} + i\frac{1}{5\sqrt{3}} & \frac{16}{75} \end{pmatrix},$$

$\text{tr}(\rho^2) = \frac{47}{75} < 1$. Seeing $\text{tr}(\rho^2) < 1$ is sufficient to conclude that ρ is a mixed state. This criterion holds true whatever of the dimensions of ρ.

11.2.2.2 Density Operator for a Pure State

Although we don't have to, we can certainly express a pure state in terms of its density operator. As the state is pure we have complete knowledge about it. Hence the ensemble contains exactly one kind of state, namely $|\psi\rangle$, and so the probability of this state being in the ensemble is 1 and all others are 0. Hence the density operator corresponding to pure state $|\psi\rangle = a|0\rangle + b|1\rangle$ is:

$$\rho_{\text{pure}} = |\psi\rangle\langle\psi| \tag{11.16}$$

with no summation. By expanding out the implied bras and kets, we can compute the density matrix explicitly as:

$$\rho_{\text{pure}} = |\psi\rangle\langle\psi| = \begin{pmatrix} a \\ b \end{pmatrix} \cdot (a^* \; b^*) = \begin{pmatrix} |a|^2 & ab^* \\ ba^* & |b|^2 \end{pmatrix} \tag{11.17}$$

where $\langle\psi| \equiv (a^* \; b^*)$ is the bra vector associated with the ket $|\psi\rangle \equiv \binom{a}{b}$. It is obtained by computing the conjugate transpose of the column vector associated with $|\psi\rangle$.

Notice that the sum of the diagonal elements of the density operator is unity, i.e., $\text{tr}(\rho_{\text{pure}}) = 1$. However, as $\rho = |\psi\rangle\langle\psi|$ is actually a pure state (written in density operator formalism) it also happens to be true that $\text{tr}(\rho^2) = 1$ too. Specifically we have,

$$\begin{aligned} \rho^2 &= \begin{pmatrix} |a|^4 + |a|^2|b|^2 & ab^*(|a|^2 + |b|^2) \\ ba^*(|a|^2 + |b|^2) & |b|^4 + |a|^2|b|^2 \end{pmatrix} \\ &= \begin{pmatrix} |a|^2(|a|^2|b|^2) & ab^*(|a|^2 + |b|^2) \\ ba^*(|a|^2 + |b|^2) & |b|^2(|a|^2|b|^2) \end{pmatrix} \\ &= \begin{pmatrix} |a|^2 & ab^* \\ ba^* & |b|^2 \end{pmatrix} = \rho \end{aligned} \tag{11.18}$$

Hence, $\rho^2 = \rho$ and so $\text{tr}(\rho^2) = |a|^2 + |b|^2 = 1$ when ρ is a 1-qubit pure state.

The foregoing results carry over to multi-qubit pure states too. Thus, the density operator associated with a 2-qubit pure state $|\psi\rangle = a|00\rangle + b|01\rangle + c|10\rangle + d|11\rangle$ is:

$$\rho = |\psi\rangle\langle\psi| = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \cdot \left(a^* \; b^* \; c^* \; d^*\right) = \begin{pmatrix} |a|^2 & ab^* & ac^* & ac^* \\ ba^* & |b|^2 & bc^* & bd^* \\ ca^* & cb^* & |c|^2 & cd^* \\ da^* & db^* & dc^* & |d|^2 \end{pmatrix} \tag{11.19}$$

and so on. As you will show in problem ***** for this 2-qubit pure state it is also true that $\rho^2 = \rho$ and $\text{tr}(\rho^2) = (|a|^2 + |b|^2 + |c|^2 + |d|^2)^2 = 1$ (the latter factorization is a hint).

It turns out, whatever the dimensions of ρ, that $\rho^2 = \rho$ and $\text{tr}(\rho^2) = 1$ if and only if ρ is the density operator corresponding to a *pure* state. If the state described by ρ is not pure, but is instead mixed, then $\rho^2 \neq \rho$ and $\text{tr}(\rho^2) < 1$. These properties can be used to decide whether a given state is pure or mixed.

11.2.2.3 The Bloch Ball

In Chap. 1 we introduced the Bloch *sphere* as a way of visualizing single qubit *pure* states. In this picture, the pure states are always points on the *surface* of the Bloch sphere. Since all pure states that differ only by an overall phase factor are indistinguishable, this overall phase factor is not depicted in the Bloch sphere representation. One might wonder where single qubit *mixed* states would reside in this Bloch sphere picture?

The answer is that single qubit mixed states correspond to points *inside* the Bloch sphere, a region we shall henceforth call the Bloch *ball*. After a little algebra, we find that the (x, y, z) coordinates within the Bloch ball corresponding to the mixed state ρ are given by [164]:

$$\begin{aligned} x &= \langle 0|\rho|1\rangle + \langle 1|\rho|0\rangle \\ y &= i\langle 0|\rho|1\rangle - i\langle 1|\rho|0\rangle \\ z &= \langle 0|\rho|0\rangle - \langle 1|\rho|1\rangle \end{aligned} \tag{11.20}$$

with the North Pole corresponding to pure state $|0\rangle$, and the South Pole to pure state $|1\rangle$. Hence, a superposition state such as $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ will have coordinate $(x, y, z) = (1, 0, 0)$ etc. The maximally mixed state is a point, as shown in Fig. 11.3, at the center of the Bloch ball with coordinates $(x, y, z) = (0, 0, 0)$. Non-maximally mixed states lie between the center of the Bloch ball and its surface.

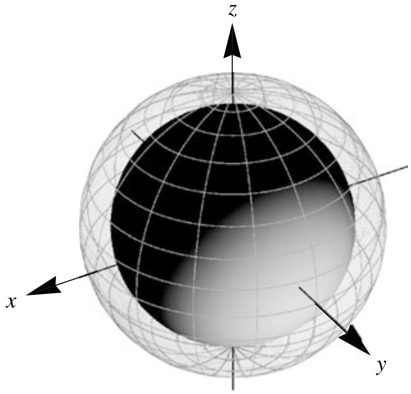

Fig. 11.3 The Bloch Ball can be used to visualize mixed states of a single qubit, which reside on the interior of the Bloch sphere

11.2.2.4 Properties of Density Operators

The quantum mechanical equations based on state vectors, which we have thus far used to describe the evolution and measurement of *pure* states can be re-expressed in the language of density operators. However, the density operator versions apply to the evolution and measurement of both *pure and mixed* states. Consequently, they are more useful, especially when we are dealing with quantum systems for which we have only incomplete knowledge.

- The sum of the diagonal elements of ρ is always 1, i.e., $\mathrm{tr}(\rho(t)) = 1$
- The expected value of an observable $\langle A \rangle = \mathrm{tr}(\rho A)$
- The time evolution of a density operation obeys $i\hbar \frac{d\rho}{dt} = [\mathcal{H}, \rho]$
- The density operator is Hermitian $\rho^\dagger = \rho$
- If ρ corresponds to a pure state $\rho^2 = \rho$
- If ρ corresponds to a pure state $\mathrm{tr}(\rho^2) = 1$
- If ρ corresponds to a pure state the eigenvalues of ρ are either 0 or 1 only
- If ρ corresponds to a mixed state $\frac{1}{d} \leq \mathrm{tr}(\rho^2) < 1$ where d is the dimension of ρ
- A measure of the similarity between two density matrices is given by the fidelity

$$\mathcal{F}(\rho_1, \rho_2) = \left[\mathrm{tr}\left(\sqrt{\rho_1^{1/2} \rho_2 \rho_1^{1/2}}\right)\right]^2$$

In Table 11.1 we compare and contrast formulae for performing similar operations on pure states and mixed states. Note that the formulae for mixed states encompass pure states too as a special case, namely, when the density operator takes the form $\rho = |\psi\rangle\langle\psi|$.

11.2.2.5 Non-unique Interpretation of Density Operators

The decomposition of a given density operator into a weighted sum of pure states is *non-unique*. Any decomposition that synthesizes the density operator is as legitimate as any other. This means that there is no unique mixed state to which each density operator corresponds. Moreover, as the expectation value of an observable, $\mathcal{O}$, is computed from $\mathrm{tr}(\rho\mathcal{O})$, then all these different mixed states (having the same ρ)

Table 11.1 Analogous quantum mechanical formulae for n-qubit pure and mixed states. Note that the mixed state formulae can also be used to describe pure states but not vice versa

Characteristic	Pure state description	Mixed state description
State	$\lvert\psi\rangle = \sum_{j=0}^{2^n-1} c_j \lvert j\rangle$	$\rho = \sum_k p_k \lvert\phi_k\rangle\langle\phi_k\rvert$ where $\lvert\phi_k\rangle$ is an arbitrary n-qubit pure state and $\sum_k p_k = 1$
State evolution	$i\hbar \frac{\partial \lvert\psi\rangle}{\partial t} = \mathcal{H}\lvert\psi\rangle$	$i\hbar \frac{\partial\rho}{\partial t} = [\mathcal{H}, \rho]$
Component evolution	$\frac{\partial c_j}{\partial t} = -\frac{i}{\hbar} \sum_{\ell=0}^{2^n-1} \mathcal{H}_{j\ell} c_\ell$	$\frac{\partial \rho_{jk}}{\partial t} = -\frac{i}{\hbar} \sum_{\ell=1}^{2^n} [\mathcal{H}_{j\ell}\rho_{\ell k} - \rho_{j\ell}\mathcal{H}_{\ell k}]$
Expected value of observable	$\langle\mathcal{O}\rangle = \langle\psi\rvert\mathcal{O}\lvert\psi\rangle$	$\langle\mathcal{O}\rangle = \mathrm{tr}(\rho\mathcal{O})$

would produce identical statistical distributions of measurement outcomes whatever observable is used! So there is no experiment we can do that will distinguish between these different mixed states. Operationally, they are all equivalent.

To illustrate the non-uniqueness of the mixed state associated with a given density operator consider the following. Let $|\psi_A\rangle = \frac{1}{2}|0\rangle + \frac{\sqrt{3}}{2}|1\rangle$ and $|\psi_B\rangle = \frac{2}{3}|0\rangle + \frac{\sqrt{5}}{3}|1\rangle$. Then ρ, the density operator corresponding to a mixed state that is $\frac{1}{3}$ $|\psi_A\rangle$ and $\frac{2}{3}$ $|\psi_B\rangle$ can be written as:

$$\begin{aligned}\rho &= \frac{1}{3}|\psi_A\rangle\langle\psi_A| + \frac{2}{3}|\psi_B\rangle\langle\psi_B| \\ &= \begin{pmatrix} 0.37962962962962965 & 0.47560689729737526 \\ 0.47560689729737526 & 0.6203703703703703 \end{pmatrix}\end{aligned} \tag{11.21}$$

However, ρ can be obtained equally well from states $|\phi_A\rangle = a|+\rangle + \sqrt{1-|a|^2}\,|-\rangle$ and $|\phi_B\rangle = b|+\rangle + \sqrt{1-|b|^2}\,|-\rangle$ where $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ as:

$$\begin{aligned}\rho &= c|\phi_A\rangle\langle\phi_A| + (1-c)|\phi_B\rangle\langle\phi_B| \\ &= \begin{pmatrix} 0.37962962962962965 & 0.47560689729737526 \\ 0.47560689729737526 & 0.6203703703703703 \end{pmatrix}\end{aligned} \tag{11.22}$$

provided $a = -0.875949$, $b = -0.994988$, and $c = 0.064635$. So the question whether ρ is "really" a mixture of the states $|\psi_A\rangle$ and $|\psi_B\rangle$ or a mixture of the states $|\phi_A\rangle$ and $|\phi_B\rangle$ is unanswerable. Each decomposition is as valid as the other.

11.2.3 Mixed States from Partial Ignorance: The Partial Trace

In Sect. 11.2.1 we introduced the concept of the *partial trace* operation. There we explained *what* it was (i.e., the act of ignoring or discarding a subset of the qubits of a multi-partite quantum state) but we did explain *how* to compute it. That is the subject of this section.

The basic idea is that we start off with the quantum mechanical description of a multi-qubit state, and we ask how our description must change if we ignore part of that state. The easiest way to think about this is to partition the set of qubits into two sets A and B and consider a multi-qubit system having density operator ρ_{AB}. In general, ρ_{AB}, will not be a product state, i.e., in general Here the subscript AB signifies that we can arbitrarily

The (i, i')-th element of the reduced density operator, ρ_A obtained by tracing over the second set of qubits B from the state ρ_{AB} is given by:

$$\langle i_A|\rho_A|i'_A\rangle = \text{tr}_B(\rho_{AB}) = \sum_{j_B=0}^{d_B-1} \langle i_A|\langle j_B|\rho_{AB}|i'_A\rangle|j_B\rangle \tag{11.23}$$

where $|i_A\rangle$ and $|i'_A\rangle$ are eigenstates of A subsystem, and $|j_B\rangle$ are eigenstates of the subsystem B (which is a d_B dimensional subspace). Notice that, in the summation, the same eigenstate index j_B is used either side of the ρ_{AB} and the summation is computed over all values for this index. Hence, the reduced density operator ρ_A is obtained by computing each of its possible matrix elements in accordance with (11.23).

Likewise, the (j, j')-th element of the reduced density operator, ρ_B is obtained by tracing over the *first* set of qubits A from the state ρ_{AB}. We have:

$$\langle j_B|\rho_B|j'_B\rangle = \text{tr}_A(\rho_{AB}) = \sum_{i_A=0}^{d_A-1} \langle i_A|\langle j_B|\rho_{AB}|i_A\rangle|j'_B\rangle \tag{11.24}$$

where $|j_B\rangle$ and $|j'_B\rangle$ are eigenstates of the B subsystem, and $|i_A\rangle$ are eigenstates of the subsystem A (which is a d_A dimensional subspace). Notice that, in the summation, the same eigenstate index i_A is used either side of the ρ_{AB} and the summation is computed over all values for this index. Hence, the reduced density operator ρ_B is obtained by computing each of its possible matrix elements in accordance with (11.24).

11.2.3.1 Example: Computing the Partial Trace

For example, consider a pair of non-orthogonal quantum states $|\psi_{ABC}\rangle$ and $|\varphi_{ABC}\rangle$ defined as follows:

$$|\psi_{ABC}\rangle = \frac{1}{2}|000\rangle + \frac{\sqrt{3}}{2}|111\rangle \tag{11.25}$$

$$|\varphi_{ABC}\rangle = \frac{1}{2}|000\rangle + \frac{1}{2}|010\rangle + \frac{1}{2}|011\rangle + \frac{1}{2}|111\rangle \tag{11.26}$$

and imagine these are the components of the density operator weighted to be one third $|\psi_{ABC}\rangle$ and two thirds $|\varphi_{ABC}\rangle$. Thus we have:

$$\begin{aligned}\rho_{ABC} &= \frac{1}{3}|\psi_{ABC}\rangle\langle\psi_{ABC}| + \frac{2}{3}|\varphi_{ABC}\rangle\langle\varphi_{ABC}| \\ &= \begin{pmatrix} \frac{1}{4} & 0 & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & \frac{1}{6}+\frac{1}{4\sqrt{3}} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & \frac{1}{6} \\ \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6}+\frac{1}{4\sqrt{3}} & 0 & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 & \frac{5}{12} \end{pmatrix}\end{aligned} \tag{11.27}$$

Tracing over any one of the qubits we obtain the three reduced density matrices ρ_{BC}, ρ_{AC}, and ρ_{AB}:

$$\rho_{BC} = \text{tr}_A(\rho_{ABC}) = \begin{pmatrix} \frac{1}{4} & 0 & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & 0 & \frac{1}{6} & \frac{7}{12} \end{pmatrix} \tag{11.28}$$

$$\rho_{AC} = \text{tr}_B(\rho_{ABC}) = \begin{pmatrix} \frac{5}{12} & \frac{1}{6} & 0 & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & 0 & \frac{1}{6} \\ 0 & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{6} & 0 & \frac{5}{12} \end{pmatrix} \tag{11.29}$$

$$\rho_{AB} = \text{tr}_C(\rho_{ABC}) = \begin{pmatrix} \frac{1}{4} & \frac{1}{6} & 0 & 0 \\ \frac{1}{6} & \frac{1}{3} & 0 & \frac{1}{6} \\ 0 & 0 & 0 & 0 \\ 0 & \frac{1}{6} & 0 & \frac{5}{12} \end{pmatrix} \tag{11.30}$$

Likewise, tracing over any two of the three qubits we obtain the three reduced density matrices ρ_A, ρ_B, and ρ_C:

$$\rho_A = \text{tr}_{BC}(\rho_{ABC}) = \begin{pmatrix} \frac{7}{12} & \frac{1}{6} \\ \frac{1}{6} & \frac{5}{12} \end{pmatrix} \tag{11.31}$$

$$\rho_B = \text{tr}_{AC}(\rho_{ABC}) = \begin{pmatrix} \frac{1}{4} & \frac{1}{6} \\ \frac{1}{6} & \frac{3}{4} \end{pmatrix} \tag{11.32}$$

$$\rho_C = \text{tr}_{AB}(\rho_{ABC}) = \begin{pmatrix} \frac{5}{12} & \frac{1}{6} \\ \frac{1}{6} & \frac{7}{12} \end{pmatrix} \tag{11.33}$$

Thus, the partial trace operation provides the procedure for calculating the quantum state of *part* of a composite quantum system. In general, if the starting state is entangled and pure (say) the restriction of this state to some subset of its qubits, i.e., its partial trace, will, in general, be a mixed state described mathematically by a reduced density operator.

11.2.4 Mixed States as Parts of Larger Pure States: "Purifications"

The foregoing interpretation of the partial trace operation invites the question of whether there is a procedure for going in the opposite direction? That is, starting with a mixed state, which we can think of as the reduced density operator of some larger *pure* state, is there a procedure for finding this larger pure state? The answer is that there is such a procedure. It is called a state *purification* operation, and it works as follows:

Purification of a Mixed State Let $\rho_A = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ be a n-qubit mixed state defined on a Hilbert space $\mathcal{H}_A$ of dimension $d = 2^n$. Our goal is to find a pure state $|\Psi\rangle_{AB}$, defined on a Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ such that $\text{tr}_B(|\Psi\rangle_{AB}\langle\Psi|_{AB}) = \rho_A$. Such a $|\Psi\rangle_{AB}$ is a purification of the mixed state ρ_A.

1. Rewrite the mixed state ρ_A as:

$$\rho_A = \sum_{i=1}^{N} p_i |\psi_i\rangle\langle\psi_i| = \sum_{j=1}^{d} \lambda_j |\phi_j\rangle\langle\phi_j| \tag{11.34}$$

   where $\{\lambda_j\}$ are the eigenvalues of ρ_A and $\{|\phi_j\rangle\}$ are the eigenvectors of ρ_A. Note that there are d eigenvalues and eigenvectors, whereas there are N states contributing to the original definition of ρ_A.
2. Pick out just the first N eigenvalues and eigenvectors from the basis $\{|\phi_j\rangle\}$. Then construct the pure state $|\Psi\rangle_{AB}$ defined as:

$$|\Psi_{AB}\rangle = \sum_{i=1}^{N} \sqrt{p_i} |\psi_i\rangle|\phi_i\rangle \tag{11.35}$$

3. The given $|\Psi\rangle_{AB}$ is a purification of ρ_A since $\text{tr}_B(|\Psi\rangle_{AB}\langle\Psi|_{AB}) = \rho_A$.

11.2.5 Quantifying Mixedness

How do we quantify the degree of mixedness in a state given its description in terms of a density operator? Clearly, our measure of "mixedness" must range from zero (for pure states) to some maximum value (for maximally mixed states). But what measure should we use? In this section we look at some ways to quantify the degree of mixedness of a quantum state.

11.2.5.1 Linear Entropy as a Measure of Mixedness

The first measure of mixedness is related to its deviation from a pure state. In particular, we saw in Sect. 11.2.2.2 that if a state with density matrix ρ is pure $\rho^2 = \rho$ and therefore $\text{tr}(\rho^2) = \text{tr}(\rho) = 1$, whereas if it is mixed, $\frac{1}{d} \le \text{tr}(\rho^2) < 1$ where d is the dimension of ρ. Hence the deviation of $\text{tr}(\rho^2)$ from 1 can be used as a measure for the mixedness of ρ. This gives us our first measure of mixedness called the *linear entropy* of ρ, which is especially easy to calculate:

$$S_L(\rho) = \frac{d}{d-1}(1 - \text{tr}(\rho^2)) \tag{11.36}$$

where d is the dimension of ρ. Hence, $0 \le S_L(\rho) \le 1$: the linear entropy $S_L(\rho) = 0$ whenever ρ is a pure state, and $S_L(\rho) = 1$ whenever ρ is a maximally mixed state.

11.2.5.2 von Neumann Entropy as a Measure of Mixedness

A second measure of mixedness is the von Neumann entropy, $S_V(\rho)$, which is the proper quantum generalization of the Shannon entropy.

To remind you, in classical information theory, the Shannon entropy of a classical source that outputs d distinguishable symbols with corresponding probabilities $p_1, p_2, \ldots, p_d$ is given by $H(\{p_i\}) = -\sum_{i=1}^{d} p_i \log_2 p_i$ where $\sum_{i=1}^{n} p_i = 1$. This ranges from 0 in the case when all the symbols are the same, to $\log_2 d$ in the case when all the d symbols are equiprobable, and therefore maximally random. One can therefore think of the Shannon entropy as quantifying the degree of randomness in the symbols streaming from a classical source.

What is the analog of Shannon entropy in the quantum context? We can think of a quantum source as outputting *d not necessarily orthogonal* quantum states $|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_d\rangle$ with corresponding probabilities $p_1, p_2, \ldots, p_d$. Such a source is characterized by the density operator ρ given by:

$$\rho = \sum_{i=1}^{d} p_i |\psi_i\rangle\langle\psi_i| \tag{11.37}$$

where $\sum_{i=1}^{d} p_i = 1$.

However, a given density operator can be decomposed into a sum of component states in many different ways, which are all equivalent to one another. In particular, even if the states $|\psi_i\rangle$ are non-orthogonal, we can always *diagonalize* ρ by finding a unitary matrix, U, such that $U\rho U^\dagger$ is a diagonal matrix. Thus, any density operator ρ can also be written as:

$$\rho = \sum_{i=1}^{d} p_i |\psi_i\rangle\langle\psi_i| = \sum_{j} \lambda_j |\lambda_j\rangle\langle\lambda_j| \tag{11.38}$$

When so diagonalized, the eigenvalues of ρ, i.e., the λ_j appearing along the main diagonal, are positive real numbers that sum to one, and correspond to the probabilities with which we will see the corresponding eigenvectors of ρ, i.e., $|\lambda_j\rangle$, if ρ were measured in its eigenbasis. As these eigenvectors $|\lambda_j\rangle$ *are* orthogonal to one another they are distinguishable and we can therefore regard them as classical symbols. Therefore, when viewed in the diagonal basis, we would expect the quantum entropy of the quantum source to coincide with the Shannon entropy of the analogous classical source, i.e., one emitting the "classical" (i.e., perfectly distinguishable) symbols, or equivalently orthonormal states $|\lambda_j\rangle$, with corresponding probabilities λ_j. This allows us to define the entropy of the quantum source (which may or may not output distinguishable symbols) in terms of the Shannon entropy of a corresponding fictitious classical source (which outputs only distinguishable symbols). In particular, we have:

$$S_V(\rho) = -\sum_{j} \lambda_j \log_2 \lambda_j = H(\{\lambda_j\}) \tag{11.39}$$

where we take $0\log_2 0 = 0$. Using purely mathematical arguments (i.e. no new physics insights), we can rewrite (11.39) as:

$$S_V(\rho) = -\mathrm{tr}(\rho \log_2 \rho) \tag{11.40}$$

This is the von Neumann entropy of the quantum source described by density operator ρ.

It is apparent from its definition that the von Neumann entropy is bounded as follows:

$$0 \leq S_V(\rho) \leq \log_2 d \tag{11.41}$$

with the von Neumann entropy being 0 for a pure state $\rho = |\psi\rangle\langle\psi|$, and $\log_2 d$ for a maximally mixed state $\rho = \frac{1}{d}\mathbb{1}$, where $\mathbb{1}$ is the identity matrix. Thus, the numerical value of the von Neumann entropy is a measure of the mixedness of the state.

11.3 Entanglement

> "*No self is of itself alone.*"
> – Erwin Schrödinger

"Entanglement" describes a correlation between different parts of a quantum system that exceeds anything that is possible classically. It will appear when subsystems interact in such a way that the resulting state of the whole system cannot be expressed as the direct product of states for its parts. When a quantum system is in such an entangled state, actions performed on one sub-system will have a side-effect on another sub-system even though that sub-system is not acted upon directly. Moreover, provided the sub-systems are separated in such a way that neither is measured, such entanglement will persist regardless of how far apart the sub-systems become. This leads to highly counterintuitive phenomena, which Einstein dubbed "spooky action at a distance", which we will have more to say about in Chap. 12.

All the known quantum algorithms that display an exponential speedup over their classical counterparts exploit such entanglement-induced side effects in one way or another. In addition, some tasks that are impossible by classical standards, such as teleporting a quantum state, depend upon entanglement in an essential way. Hence, entanglement deserves to be called a "quintessential" quantum phenomenon that plays a major role in making quantum computing more powerful than classical computing, and in enabling quantum information tasks that are impossible in the classical context.

11.3.1 Separable States Versus Entangled States

Formally, the distinction between whether a state is entangled or not entangled rests upon whether its quantum state is separable or not. Therefore, let us examine this question in more mathematical terms.

Suppose we have two independent quantum systems with Hilbert spaces H_A and H_B of dimensions d_A and d_B respectively. There is some complete orthonormal basis for H_A consisting of d_A eigenstates, called $\{|j_A\rangle\}$ say, and some complete orthonormal basis for H_B consisting of d_B eigenstates, $\{|k_B\rangle\}$ say. In other words, any pure state in H_A can be expressed as $|\psi_A\rangle = a_0|0\rangle_A + a_1|1\rangle_A + \cdots + a_{d_A-1}|d_A - 1\rangle_A$. Likewise, any pure state in H_B can be expressed as $|\psi_B\rangle = b_0|0\rangle_B + b_1|1\rangle_B + \cdots + b_{d_B-1}|d_B - 1\rangle_B$. And the Hilbert space of the composite system is just the tensor product of the constituent Hilbert spaces $H = H_A \otimes H_B$.

Separable State If a pure (mixed) state, $|\psi^{(AB)}\rangle$ $(\rho^{(AB)})$, of a composite quantum system defined on a Hilbert space $H_A \otimes H_B$ can be written as $|\psi^{(AB)}\rangle = |\psi^{(A)}\rangle \otimes |\psi^{(B)}\rangle$ $(\rho^{(AB)} = \sum_i p_i \rho_i^{(A)} \otimes \rho_i^{(B)}$), then $|\psi^{(AB)}\rangle$ $(\rho^{(AB)})$ is said to be a *separable state*.

The linear entropy $S_L(\rho)$, can also be useful in deciding whether a given density operator ρ corresponds to a separable or entangled state. Specifically, it has been proven that if the linear entropy exceeds a certain threshold, i.e., if $S_L(\rho) \geq d(d-2)/(d-1)^2$, then any such ρ is separable [567].

Entangled State If a state, $|\psi^{(AB)}\rangle$ $(\rho^{(AB)})$, of a composite quantum system defined on a Hilbert space $H_A \otimes H_B$ is not a separable state it is an *entangled state*. Note that a state can be entangled and pure, or entangled and mixed, simultaneously.

As an example, consider the state $\frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$. Is this state separable or entangled? Well, if it were separable it could be written in the form $|\psi_A\rangle \otimes |\psi_B\rangle$ where $|\psi_A\rangle = a_0|0\rangle + a_1|1\rangle$ and $|\psi_B\rangle = b_0|0\rangle + b_1|1\rangle$. Thus, equating amplitudes and solving we have:

$$\begin{aligned} |\psi_A\rangle \otimes |\psi_B\rangle &= a_0b_0|00\rangle + a_0b_1|01\rangle + a_1b_0|10\rangle + a_1b_1|11\rangle \\ &= \frac{1}{\sqrt{2}}|01\rangle - \frac{1}{\sqrt{2}}|10\rangle \end{aligned} \tag{11.42}$$

which implies we need to find a solution to the simultaneous equations $a_0b_0 = 0$, $a_0b_1 = \frac{1}{\sqrt{2}}$, $a_1b_0 = -\frac{1}{\sqrt{2}}$, $a_1b_1 = 0$. Unfortunately, these equations admit no such solution and hence the state $\frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$ is not separable. Hence it is entangled.

11.3.2 Signalling Entanglement via Entanglement Witnesses

Given a purported entangled state, ρ, how can we verify that ρ is, in fact, entangled?

One approach is to synthesize several instances of the state ρ via identical preparation procedures and then perform quantum state tomography to reconstruct the density operator for ρ. This is, in one sense, the preferred option since we would obtain *complete* information about ρ—at least to within experimental error.

However, in general, quantum state tomography is an extraordinarily costly procedure. An n-qubit state is described by a $2^n \times 2^n$ dimensional density operator. If we are to determine each element of this density operator empirically, we would need to perform $O(2^{2n})$ different experiments! Thus full quantum state tomography becomes quite impractical for quantum systems having more than a mere handful of qubits.

This difficulty spawned the invention of *entanglement witnesses* [166]. Entanglement witnesses are tools for detecting entanglement that avoid having to perform a complete quantum state tomographic reconstruction of ρ.

The basic idea is to construct an observable operator, W, whose expectation value serves as a "witness" to whether the given state is entangled. If the expectation value of the witness observable W when the system is in state ρ, i.e., $\text{tr}(W\rho) = \langle W \rangle$, is less than some threshold, this provides sufficient evidence that ρ is an entangled state.

Although the fully theory of entanglement witnesses requires an understanding of the superoperator formalism of quantum mechanics, entanglement witnesses need not be that exotic.

11.3.2.1 Example: Entanglement Witness

For example, consider the one-dimensional "Heisenberg chain". This consists of a one-dimensional loop of spins coupled together in accordance with the Hamiltonian:

$$\mathcal{H} = \sum_{i=1}^{N} (B\sigma_z^i + J\boldsymbol{\sigma}^i \cdot \boldsymbol{\sigma}^{i+1}) \tag{11.43}$$

where B is the external magnetic field, and a $J < 0$ or $J > 0$ are, respectively, a ferromagnetic or anti-ferromagnetic coupling between the spins. The symbol σ_z^i stands for an the Pauli-Z operator that acts on the i-th qubit, and the vectors $\boldsymbol{\sigma}^i \equiv (\sigma_x^i, \sigma_y^i, \sigma_z^i)$. The one-dimensional chain of spins is made periodic by choosing $\boldsymbol{\sigma}^{N+1} = \boldsymbol{\sigma}^1$.

In the absence of an external magnetic field, i.e., with $B = 0$, the expectation value for the energy, $\langle \mathcal{H} \rangle = \text{tr}(\rho \mathcal{H})$, can be an entanglement witness. Specifically, suppose we have chain consisting of two spins, i.e., $N = 2$. In this case, if the input state is separable, i.e., $\rho_{AB} = \rho_A \otimes \rho_B$, then it is possible to show that the expectation value of the energy is *guaranteed* to be bounded between $-2J \le \langle \mathcal{H} \rangle \le +2J$.

However, if the state ρ_{AB} is entangled, $\rho_{AB} \neq \rho_A \otimes \rho_B$, we find that there are entangled states for which $\langle \mathcal{H} \rangle < -2J$. Thus, by measuring expectation value of the energy, $\langle \mathcal{H} \rangle$, we can sometimes decide if the state is entangled. Hence, $\langle \mathcal{H} \rangle$ serves as an entanglement witness.

11.3.3 Signalling Entanglement via the Peres-Horodecki Criterion

An alternative to relying on entanglement witnesses to decide if a state is entangled, is to use the Peres-Horodecki criterion [239, 387]. This criterion uses an operation on a density matrix known as the *partial transpose*.

Definition: Partial Transpose Let ρ be a bi-partite density operator expressed in the form:

$$\rho = \sum_{i,j,k,\ell} \rho_{ij;k\ell} |e_i^A \otimes e_j^B\rangle\langle e_k^A \otimes e_\ell^B| \tag{11.44}$$

where $\{|e_i^A\rangle\}$ is an eigenbasis for sub-space A and $\{|e_j^B\rangle\}$ is an eigenbasis for sub-space B. Then the partial transpose ρ^{T_B} of the density operator ρ is:

$$\rho^{T_B} = \sum_{i,j,k,\ell} \rho_{i\ell;kj} |e_i^A \otimes e_j^B\rangle\langle e_k^A \otimes e_\ell^B| \tag{11.45}$$

Note that, as implied by the definition, the partial transpose depends on the basis chosen but the *eigenvalues* of the partial transposed matrix do not. However, most practical applications of the partial transpose only need to make use of the eigenvalues of the partial transpose matrix.

The partial transpose is important within a test for entanglement known as the Peres-Horodecki criterion [239, 387].

Peres-Horodecki Criterion: a Necessary and Sufficient Test for Entanglement If a bi-partite state is entangled, its partial transpose always has one or more negative eigenvalues, but if it is separable its partial transpose has no negative eigenvalues.

Thus, given a density operator ρ we can decide whether or not it is entangled by examining the signs of the eigenvalues of its partial transpose.

Note that we can define an analogous partial transpose over the "A" space as follows:

$$\rho^{T_A} = \sum_{i,j,k,\ell} \rho_{k\ j;i\ \ell} |e_i^A \otimes e_j^B\rangle\langle e_k^A \otimes e_\ell^B| \tag{11.46}$$

Even though the partial transpose ρ^{T_A} will usually be a different matrix from the partial transpose ρ^{T_B} their eigenvalues will be the same. In applications of the partial transpose it is usually the eigenvalues of the partial transpose that we need rather than the partial transpose itself. If this is the case, whether we use ρ^{T_A} or ρ^{T_B} is immaterial as their eigenvalues are the same.

For example, let us compute ρ^{T_A} and ρ^{T_B} for a general 2-qubit density matrix defined by:

$$\rho = \begin{pmatrix} \rho_{11} & \rho_{12} & \rho_{13} & \rho_{14} \\ \rho_{21} & \rho_{22} & \rho_{23} & \rho_{24} \\ \rho_{31} & \rho_{32} & \rho_{33} & \rho_{34} \\ \rho_{41} & \rho_{42} & \rho_{43} & \rho_{44} \end{pmatrix} \tag{11.47}$$

Computing the partial transposes we obtain:

$$\begin{aligned} \rho^{T_A} &= \begin{pmatrix} \rho_{11} & \rho_{12} & \rho_{31} & \rho_{32} \\ \rho_{21} & \rho_{22} & \rho_{41} & \rho_{42} \\ \rho_{13} & \rho_{14} & \rho_{33} & \rho_{34} \\ \rho_{23} & \rho_{24} & \rho_{43} & \rho_{44} \end{pmatrix} \\ \rho^{T_B} &= \begin{pmatrix} \rho_{11} & \rho_{21} & \rho_{13} & \rho_{23} \\ \rho_{12} & \rho_{22} & \rho_{14} & \rho_{24} \\ \rho_{31} & \rho_{41} & \rho_{33} & \rho_{43} \\ \rho_{32} & \rho_{42} & \rho_{34} & \rho_{44} \end{pmatrix} \end{aligned} \tag{11.48}$$

However, the characteristic polynomials of ρ^{T_A} and ρ^{T_B} are identical, and so the eigenvalues of these matrices must be the same.

Case 1: a Separable Pure State

Let us look at some simple examples. Consider first the case of an unentangled pure state. In this case we have:

$$|\psi_{AB}\rangle = \left(\frac{1}{2}|0\rangle + \sqrt{\frac{3}{4}}|1\rangle\right) \otimes \left(\frac{1}{3}|0\rangle + \sqrt{\frac{8}{9}}|1\rangle\right) \tag{11.49}$$

$$\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}| = \begin{pmatrix} \frac{1}{36} & \frac{1}{9\sqrt{2}} & \frac{1}{12\sqrt{3}} & \frac{1}{3\sqrt{6}} \\ \frac{1}{9\sqrt{2}} & \frac{2}{9} & \frac{1}{3\sqrt{6}} & \frac{2}{3\sqrt{3}} \\ \frac{1}{12\sqrt{3}} & \frac{1}{3\sqrt{6}} & \frac{1}{12} & \frac{1}{3\sqrt{2}} \\ \frac{1}{3\sqrt{6}} & \frac{2}{3\sqrt{3}} & \frac{1}{3\sqrt{2}} & \frac{2}{3} \end{pmatrix} \tag{11.50}$$

$$\rho_{AB}^{T_B} = \begin{pmatrix} \frac{1}{36} & \frac{1}{9\sqrt{2}} & \frac{1}{12\sqrt{3}} & \frac{1}{3\sqrt{6}} \\ \frac{1}{9\sqrt{2}} & \frac{2}{9} & \frac{1}{3\sqrt{6}} & \frac{2}{3\sqrt{3}} \\ \frac{1}{12\sqrt{3}} & \frac{1}{3\sqrt{6}} & \frac{1}{12} & \frac{1}{3\sqrt{2}} \\ \frac{1}{3\sqrt{6}} & \frac{2}{3\sqrt{3}} & \frac{1}{3\sqrt{2}} & \frac{2}{3} \end{pmatrix} \tag{11.51}$$

$$\text{Eigenvalues}(\rho_{AB}^{T_B}) = \{1, 0, 0, 0\} \tag{11.52}$$

As all of the eigenvalues of the partial transpose of $\rho_{AB}^{T_B}$ are positive this guarantees, by the Peres-Horodecki criterion, that ρ_{AB} is separable.

Case 2: an Entangled Pure State

Now let's look what happens when we have an entangled pure state such as $|\psi_{AB}\rangle = (\frac{1}{2}|01\rangle - \sqrt{\frac{3}{4}}|10\rangle)$:

$$|\psi_{AB}\rangle = \left(\frac{1}{2}|01\rangle - \sqrt{\frac{3}{4}}|10\rangle\right) \tag{11.53}$$

$$\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}| = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & -\frac{\sqrt{3}}{4} & 0 \\ 0 & -\frac{\sqrt{3}}{4} & \frac{3}{4} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{11.54}$$

$$\rho_{AB}^{T_B} = \begin{pmatrix} 0 & 0 & 0 & -\frac{\sqrt{3}}{4} \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{3}{4} & 0 \\ -\frac{\sqrt{3}}{4} & 0 & 0 & 0 \end{pmatrix} \tag{11.55}$$

$$\text{Eigenvalues}(\rho_{AB}^{T_B}) = \left\{\frac{3}{4}, -\frac{\sqrt{3}}{4}, \frac{\sqrt{3}}{4}, \frac{1}{4}\right\} \tag{11.56}$$

As one of the eigenvalues of the partial transpose of $\rho_{AB}^{T_B}$ is negative, this guarantees by the Peres-Horodecki criterion, that ρ_{AB} is entangled.

Case 3: a Separable Mixed State

The Peres-Horodecki criterion is not limited to deciding whether only pure states are entangled or separable. It also applies to mixed states. For example, the mixed state $\rho_{AB} = \frac{1}{3}\rho_A \otimes \rho_B + \frac{2}{3}\rho'_A \otimes \rho'_B$ is, by construction, separable. The Peres-Horodecki criterion gives us:

$$\rho_A = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \tag{11.57}$$

$$\rho_B = \begin{pmatrix} \frac{2}{3} & -\frac{i}{3} \\ \frac{i}{3} & \frac{1}{3} \end{pmatrix} \tag{11.58}$$

$$\rho'_A = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \tag{11.59}$$

$$\rho'_B = \begin{pmatrix} \frac{1}{8} & \frac{i\sqrt{3}}{8} \\ -\frac{i\sqrt{3}}{8} & \frac{7}{8} \end{pmatrix} \tag{11.60}$$

$$\begin{aligned} \rho_{AB} &= \frac{1}{3}\rho_A \otimes \rho_B + \frac{2}{3}\rho'_A \otimes \rho'_B \\ &= \begin{pmatrix} \frac{11}{72} & -\frac{i}{18} + \frac{i}{8\sqrt{3}} & \frac{1}{24} & \frac{i}{8\sqrt{3}} \\ \frac{i}{18} - \frac{i}{8\sqrt{3}} & \frac{25}{72} & -\frac{i}{8\sqrt{3}} & \frac{7}{24} \\ \frac{1}{24} & \frac{i}{8\sqrt{3}} & \frac{11}{72} & -\frac{i}{18} + \frac{i}{8\sqrt{3}} \\ -\frac{i}{8\sqrt{3}} & \frac{7}{24} & \frac{i}{18} - \frac{i}{8\sqrt{3}} & \frac{25}{72} \end{pmatrix} \end{aligned} \tag{11.61}$$

$$\rho_{AB}^{T_B} = \begin{pmatrix} \frac{11}{72} & \frac{i}{18} - \frac{i}{8\sqrt{3}} & \frac{1}{24} & -\frac{i}{8\sqrt{3}} \\ -\frac{i}{18} + \frac{i}{8\sqrt{3}} & \frac{25}{72} & \frac{i}{8\sqrt{3}} & \frac{7}{24} \\ \frac{1}{24} & -\frac{i}{8\sqrt{3}} & \frac{11}{72} & \frac{i}{18} - \frac{i}{8\sqrt{3}} \\ \frac{i}{8\sqrt{3}} & \frac{7}{24} & -\frac{i}{18} + \frac{i}{8\sqrt{3}} & \frac{25}{72} \end{pmatrix} \tag{11.62}$$

$$\begin{aligned} \text{Eigenvalues}(\rho_{AB}^{T_B}) = \Bigg\{ & \frac{1}{36}\left(15 + \sqrt{95 - 12\sqrt{3}}\right), \frac{1}{36}\left(15 - \sqrt{95 - 12\sqrt{3}}\right), \\ & \frac{1}{36}\left(3 + \sqrt{5}\right), \frac{1}{36}\left(3 - \sqrt{5}\right) \Bigg\} \end{aligned} \tag{11.63}$$

As all of the eigenvalues of the partial transpose of $\rho_{AB}^{T_B}$ are positive this guarantees, by the Peres-Horodecki criterion, that ρ_{AB} is separable.

Case 4: an Entangled Mixed State

Finally we consider what happens when the state is entangled and mixed.

$$|\beta_{00}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{11.64}$$

$$|\beta_{11}\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \tag{11.65}$$

$$\rho_{AB} = \frac{1}{3}|\beta_{00}\rangle\langle\beta_{00}| + \frac{2}{3}|\beta_{11}\rangle\langle\beta_{11}| = \begin{pmatrix} \frac{1}{6} & 0 & 0 & \frac{1}{6} \\ 0 & \frac{1}{3} & -\frac{1}{3} & 0 \\ 0 & -\frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{6} & 0 & 0 & \frac{1}{6} \end{pmatrix} \tag{11.66}$$

$$\rho_{AB}^{T_B} = \begin{pmatrix} \frac{1}{6} & 0 & 0 & -\frac{1}{3} \\ 0 & \frac{1}{3} & \frac{1}{6} & 0 \\ 0 & \frac{1}{6} & \frac{1}{3} & 0 \\ -\frac{1}{3} & 0 & 0 & \frac{1}{6} \end{pmatrix} \tag{11.67}$$

$$\text{Eigenvalues}(\rho_{AB}^{T_B}) = \left\{\frac{1}{2}, \frac{1}{2}, -\frac{1}{6}, \frac{1}{6}\right\} \tag{11.68}$$

As one of the eigenvalues of the partial transpose of $\rho_{AB}^{T_B}$ is negative, this guarantees by the Peres-Horodecki criterion, that ρ_{AB} is entangled.

11.3.4 Quantifying Entanglement

Rather than merely witnessing of detecting the presence or absence or entanglement, we would prefer to be able to *quantify* the degree of entanglement in a quantum state. Such quantitative methods are necessary if we to have any hope of understanding entanglement properly and how it changes under various physical operations.

Although there is only one effective measure of entanglement in 2-qubit systems—the "tangle", which we used in Sect. 2.8 to quantify the entangling power of a 2-qubit quantum gate—once we go to three or more qubits the situation becomes extraordinarily complicated. Even at three qubits we start to encounter counterintuitive results such as the possibility of having 3-qubit states that possess 3-way entanglement but for which there is no 2-way entanglement amongst every pair of constituent qubits when considered in isolation!

At present we are stuck with having to wrestle with the notion of entanglement, and having to live with several different and inequivalent ways of quantifying how much entanglement there is in a multi-qubit state. These measures are called entanglement monotones and all share certain desirable properties.

11.3.4.1 Entanglement Monotones

> "*All science is either physics or stamp collecting.*"
> – Ernest Rutherford

"Entanglement monotones" are quantitative measures of entanglement of a quantum state, ρ, that increase, monotonically, with the degree or entanglement in the state. Such measures, some of which are summarized in Table 11.2, allow us to compare and contrast the amount of entanglement in different states and hence, to begin to develop a classification for entangled states. We start by stipulating general properties any reasonable measure of entanglement, $E(\rho)$, must possess and then outline some functions that meet these criteria.

Table 11.2 Entanglement monotones

Measure	Explanation
Entanglement of formation $E_F(\rho_{AB}) = \min_{\{p_i, \lvert\psi_i\rangle\}} \sum_i p_i S(\rho_i^{(A)})$	Quantifies the amount of entanglement needed to synthesize ρ. In essence, it measures how many maximally entangled pairs are needed to synthesize ρ. The minimization is computed over all possible decompositions of ρ_{AB} into sums of pure states making E_F very costly to compute
Entanglement of distillation $E_D(\rho_{AB}) = \lim_{n\to\infty} m/n$	Quantifies the number of Bell states that can be distilled from ρ per copy of ρ using the optimal purification procedure. Here m is the maximum number of Bell states that can be distilled from n preparations of the state ρ. E_D is also difficult to calculate in practice
Relative entropy of entanglement $E_R(\rho) = \min_{\sigma\in D} \mathrm{tr}(\rho\log\rho - \rho\log\sigma)$	Quantifies the distance of the entangled state ρ from the nearest separable state in the set of all separable density operators D. E_R is relatively easy to compute and happens to exactly equal E_F for pure states of 2-qubit systems
Negativity $E_N(\rho) = 2\max(0, -\sum_i \lambda_i^{\text{negative}}(\rho^{T_B}))$	Quantifies the entanglement in a state as the degree to which the positive partial transpose separability criterion is violated. If a state is not entangled, the partial transpose of its density operator, ρ^{T_B}, is also a valid density operator, i.e., a positive semi-definite matrix. However, if a state is entangled, the partial transpose of its density operator is not positive semi-definite because it has at least one negative eigenvalue. Hence, negativity quantifies entanglement as the degree to which the positive partial transpose separability criterion is violated. For 2-qubit pure states the negativity equals the concurrence. In the formula for negativity where $\lambda_i^{\text{negative}}(\rho^{T_B})$ is the i-th *negative* eigenvalue of the partial transpose of ρ

1. For any entanglement measure $E(\rho)$ we require $0 \leq E(\rho) \leq 1$ with $E(\rho) = 0$ if and only if ρ is not entangled, and $E(\rho) = 1$ at least when ρ is the density operator of any maximally entangled state, such as one of the Bell states.
2. The entanglement measure should be immune to local operations, i.e. $E(\rho) = E((U_A \otimes U_B)\rho(U_A \otimes U_B)^\dagger)$.
3. The entanglement measure of the full density operator, i.e., $E(\rho) = E(\sum_i p_i \rho_i)$ cannot be greater than the weighted sum of the entanglement measures of its parts, i.e., $E(\rho) = E(\sum_i p_i \rho_i) \leq \sum_i p_i E(\rho_i)$.

Given the aforemention desiderata, the following candidates have been identified as acceptable measures of entanglement.

For the case of 2-qubits the different measures of entanglement turn out to be equivalent, and it is therefore simplest to work with the tangle (see Sect. 2.8.1). However, this equivalence does not hold for larger numbers of qubits.

11.3.5 Maximally Entangled Pure States

The most famous maximally entangled pure states are the 2-qubit Bell states:

$$\begin{aligned}
|\beta_{00}\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \\
|\beta_{01}\rangle &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \\
|\beta_{10}\rangle &= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \\
|\beta_{11}\rangle &= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)
\end{aligned} \tag{11.69}$$

The structure of the Bell states invite generalizations in two ways: Either we can extend the pattern we see in the $|\beta_{00}\rangle$ state, and conceive of a two-component superposition having one state with all 0's and the other with all 1's, or we can extend the pattern we see in the $|\beta_{01}\rangle$ state, and conceive of an n-component superposition having a single 1 each in each component at each possible qubit position. This leads to two fundamentally different kinds of maximally entangled states called GHZ and W states respectively. GHZ and W states are defined as follows:

$$\begin{aligned}
|\text{GHZ}\rangle &= \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle) \\
|\text{GHZ}\rangle &= \frac{1}{\sqrt{2}}(|0000\rangle + |1111\rangle) \\
&\vdots \\
|\text{GHZ}\rangle &= \frac{1}{\sqrt{2}}(|00\ldots0\rangle + |11\ldots1\rangle)
\end{aligned} \tag{11.70}$$

and

$$\begin{aligned}
|\text{W}\rangle &= \frac{1}{\sqrt{3}}(|001\rangle + |010\rangle + |100\rangle) \\
|\text{W}\rangle &= \frac{1}{\sqrt{4}}(|0001\rangle + |0010\rangle + |0100\rangle + |1000\rangle) \\
&\vdots \\
|\text{W}\rangle &= \frac{1}{\sqrt{n}}(|0\ldots01\rangle + |0\ldots10\rangle + \cdots + |1\ldots00\rangle)
\end{aligned} \tag{11.71}$$

These two kinds of states are maximally entangled and pure, but are nevertheless fundamentally inequivalent! That is, we cannot inter-convert from GHZ states to W states, or vice versa, using any unitary transformation [253].

Moreover, GHZ and W states behave quite differently under the partial trace operation. For example, tracing over the last qubit in a 3-qubit GHZ state, we obtain:

$$\mathrm{tr}_3(|\mathrm{GHZ}\rangle\langle\mathrm{GHZ}|) = \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{pmatrix} = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|) \tag{11.72}$$

which has *no* residual 2-qubit entanglement. However, tracing over the last qubit in a 3-qubit W state results in a state that does have residual 2-qubit entanglement:

$$\begin{aligned} \mathrm{tr}_3(|\mathrm{W}\rangle\langle\mathrm{W}|) &= \begin{pmatrix} \frac{1}{3} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ &= \frac{1}{3}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \frac{2}{3}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ &= \frac{1}{3}|00\rangle\langle 00| + \frac{2}{3}\left(\frac{|01\rangle + |10\rangle}{\sqrt{2}}\right)\left(\frac{\langle 01| + \langle 10|}{\sqrt{2}}\right) \end{aligned} \tag{11.73}$$

The indicated factorization of the reduced density matrix can be interpreted as including a component in $\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$, which is one of the Bell states.

11.3.6 Maximally Entangled Mixed States

How does the concept of a maximally entangled state generalize to the case of mixed states? In what sense can a mixed state be said to be maximally entangled?

A superficially reasonable definition of a maximally entangled mixed state is a state that, for a given level of mixedness, attains the highest possible value for entanglement. Unfortunately, it turns out that such a definition is problematic without further qualification. This is because, by the above definition, the identity of the mixed states that are deemed maximally entangled will *change* depending on the measures one chooses with which to quantify the degree of mixedness and quantify the degree of entanglement in the state! This problem appears to be fundamental and unavoidable [525]. Nevertheless, once one pins down the measures for mixedness and entanglement, certain mixed states do pop out as special. These are "frontier" states in a scatter plots of where mixed states lie in an entanglement-mixedness plane.

Given the practical utility of maximally entangled pure states in ideal (i.e., noise-free) quantum information processing, it is possible these maximally entangled mixed states would find similar application in more noisy quantum information processing, as they possess the maximum amount of entanglement possible for a given degree of mixedness.

Bill Munro and collaborators have identified a class of mixed states that deserve to be called maximally entangled as they lie on the frontier in the tangle (entanglement measure) versus linear-entropy (mixedness measure) plane. The structure of the density matrices corresponding to these maximally entangled mixed states is:

$$\rho_{\text{max-ent-mixed}} = \begin{array}{ll} \begin{pmatrix} \frac{1}{3} & 0 & 0 & \frac{r}{2} \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{r}{2} & 0 & 0 & \frac{1}{3} \end{pmatrix} & 0 \leq r \leq \frac{2}{3} \\ \begin{pmatrix} \frac{r}{2} & 0 & 0 & \frac{r}{2} \\ 0 & 1-r & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{r}{2} & 0 & 0 & \frac{r}{2} \end{pmatrix} & \frac{2}{3} \leq r \leq 1 \end{array} \tag{11.74}$$

For a given value of the linear entropy (mixedness) these density matrices give the highest value of concurrence (entanglement). As tangle and entropy of formation are also both monotonic functions of concurrence, such density matrices also saturate the maximum possible degree of entanglement by these measures too.

11.3.7 The Schmidt Decomposition of a Pure Entangled State

Although we cannot write an entangled state of two quantum systems as the direct product of a state for each system we can, however, write it as a *sum* of such states. That is, if A is d_A-dimensional Hilbert space, and if B is d_B-dimensional Hilbert space, then for any entangled pure state $|\psi_{AB}\rangle$ in the Hilbert space of dimension $d_A \times d_B$, we can always find amplitudes such that:

$$|\psi_{AB}\rangle = \sum_{j=0}^{d_A-1} \sum_{k=0}^{d_B-1} a_{jk} |j_A\rangle \otimes |k_B\rangle \tag{11.75}$$

where $|j_A\rangle$ is a complete eigenbasis (i.e., set of orthonormal eigenvectors) for space A and $|k_B\rangle$ is a complete eigenbasis for space B. Notice, in particular, that to describe the state $|\psi_{AB}\rangle$ it is necessary to use a *double sum* over indices j and k.

The Schmidt decomposition, by contrast, allows us to re-express $|\psi_{AB}\rangle$ as a sum over a *single* index. And, moreover, the number of terms in this single sum is the lesser of d_A and d_B. This is rather counter-intuitive to most people when they first see this. Nevertheless, it is formally correct and allows you to interpret a given state in an interesting new way.

So how does the Schmidt decomposition work? Well it is actually rather simple. Everything hinges on using the *singular value decomposition* of a matrix built from the amplitudes that appear in the double sum description of $|\psi_{AB}\rangle$.

Schmidt Decomposition of a Pure State Given a (generally) entangled pure state of a composite quantum system, $|\psi_{AB}\rangle$, which can be thought of as composed of an n_A-qubit sub-system and an n_B-qubit sub-system, we can compute the Schmidt decomposition of $|\psi_{AB}\rangle$ as follows:

1. Sub-system A is in a $d_A = 2^{n_A}$ dimensional Hilbert space. Likewise, sub-system B is in a $d_B = 2^{n_B}$ dimensional Hilbert space. Let an eigenbasis for A be $\{|j_A\rangle\}_{j=0}^{d_A-1}$ and let an eigenbasis for B be $\{|k_B\rangle\}_{k=0}^{d_B-1}$.
2. Given the decomposition of $|\psi_{AB}\rangle$ in terms of the eigenbases $\{|j_A\rangle\}_{j=0}^{d_A-1}$ and $\{|k_B\rangle\}_{k=0}^{d_B-1}$ as:

$$|\psi_{AB}\rangle = \sum_{j=0}^{d_A-1} \sum_{k=0}^{d_B-1} a_{jk} |j_A\rangle \otimes |k_B\rangle \tag{11.76}$$

   Re-group the amplitudes a_{jk} into a $d_A \times d_B$ dimensional array, and call this $\{a_{jk}\}$. That is, take the linear sequence of amplitudes a_{jk} and make a matrix by starting a new row after every d_B amplitudes.
3. Now compute the singular value decomposition (SVD) of the matrix $\{a_{jk}\}$ you just obtained. Specifically, we can write:

$$\text{SVD}(\{a_{jk}\}) = U \cdot \Sigma \cdot V = \{u_{ji}\} \cdot \{\sigma_{ii}\} \cdot \{v_{ik}\} \tag{11.77}$$

   where $U = \{u_{ji}\}$ is a $d_A \times \min(d_A, d_B)$ dimensional unitary matrix, $V = \{v_{ik}\}$ is a $\min(d_A, d_B) \times d_B$ dimensional unitary matrix, and $\Sigma = \{\sigma_{ii}\}$ is a $\min(d_A, d_B) \times \min(d_A, d_B)$ diagonal matrix whose elements are the *singular values* of the matrix $\{a_{jk}\}$.
4. Now create new eigenbases as follows:

$$\{|i_A\rangle\}_{i=0}^{\min(d_A-1, d_B-1)} := \sum_{j=0}^{d_A-1} U_{j+1,i+1} |j_A\rangle \tag{11.78}$$

   and

$$\{|i_B\rangle\}_{i=0}^{\min(d_A-1, d_B-1)} := \sum_{k=0}^{d_B-1} V_{i+1,k+1} |k_B\rangle \tag{11.79}$$

5. Pick out the subset of the singular values:

$$\{\lambda_i\}_{i=0}^{\min(d_A-1, d_B-1)} := \{\sigma_{ii}\}_{i=0}^{\min(d_A-1, d_B-1)} \tag{11.80}$$

6. Then the (generally entangled) pure state $|\psi_{AB}\rangle$ that is describable as the double sum in (11.75) is equally well describable as the *single* sum:

$$|\psi_{AB}\rangle = \sum_{i=0}^{\min(d_A-1, d_B-1)} \lambda_i |i_A\rangle |i_B\rangle \tag{11.81}$$

   which is the Schmidt decomposition of $|\psi_{AB}\rangle$.

11.3.7.1 Example: Schmidt Decomposition

We illustrate the procedure for constructing the Schmidt decomposition using a simple 3-qubit pure state $|\psi_{AB}\rangle$. Here, we assume that A is a $d_A = 2$ dimensional sub-space and B is as $d_B = 4$ dimensional sub-space. To begin, we start with the state $|\psi_{AB}\rangle$ which we have defined to be:

$$\begin{aligned}|\psi_{AB}\rangle = {} & (-0.1661 - 0.17i)|0_A\rangle|00_B\rangle - (0.2982 + 0.0497i)|0_A\rangle|01_B\rangle \\ & + (0.3471 + 0.3943i)|0_A\rangle|10_B\rangle - (0.2667 + 0.432i)|0_A\rangle|11_B\rangle \\ & - (0.0293 + 0.2317i)|1_A\rangle|00_B\rangle + (0.1217 + 0.2168i)|1_A\rangle|01_B\rangle \\ & + (0.2162 - 0.1238i)|1_A\rangle|10_B\rangle - (0.183 + 0.3263i)|1_A\rangle|11_B\rangle \end{aligned} \tag{11.82}$$

Here we see the eigenbasis for A is $\{|j_A\rangle\} \equiv \{|0_A\rangle, |1_A\rangle\}$ and that of B is $\{|k_B\rangle\} \equiv \{|00_B\rangle, |01_B\rangle, |10_B\rangle, |11_B\rangle\}$. Next we re-group the sequence of amplitudes appearing in (11.82) to form a new matrix $\{a_{jk}\}$. As $d_B = 4$ we start a new row of this matrix after every 4 (i.e., d_B) elements. This gives us the matrix:

$$\{a_{jk}\} = \begin{pmatrix} -0.1661 - 0.17i & -0.2982 - 0.0497i & 0.3471 + 0.3943i & -0.2667 - 0.432i \\ -0.0293 - 0.2317i & 0.1217 + 0.2168i & 0.2162 - 0.1238i & -0.183 - 0.3263i \end{pmatrix} \tag{11.83}$$

Next we compute the singular value decomposition SVD($\{a_{jk}\}$) to give:

$$\text{SVD}(\{a_{jk}\}) = U \cdot \Sigma \cdot V \tag{11.84}$$

where

$$\begin{aligned} U &= \begin{pmatrix} 0.8876 & -0.4606 \\ 0.3806 - 0.2594i & 0.7334 - 0.4999i \end{pmatrix} \\ \Sigma &= \begin{pmatrix} 0.9031 & 0 \\ 0 & 0.4295 \end{pmatrix} \\ V &= \begin{pmatrix} -0.109 - 0.2732i & -0.3041 + 0.0775i & 0.4678 + 0.3975i & -0.2455 - 0.6147i \\ 0.3978 - 0.2475i & 0.2754 + 0.5651i & 0.1409 - 0.3826i & 0.3533 - 0.3069i \end{pmatrix} \end{aligned} \tag{11.85}$$

From the SVD we then construct the new bases, $\{|i_A\rangle\}$ and $\{|i_B\rangle\}$ (we use an overbar symbol to distinguish these bases from the earlier ones):

$$|\bar{0}_A\rangle := 0.8876|0_A\rangle + (0.3806 - 0.2594i)|1_A\rangle \tag{11.86}$$

$$|\bar{1}_A\rangle := -0.4606|0_A\rangle + (0.7334 - 0.4999i)|1_A\rangle \tag{11.87}$$

Likewise,

$$\begin{aligned}|\bar{0}_B\rangle := {} & (-0.109 - 0.2732i)|00_B\rangle - (0.3041 - 0.0775i)|01_B\rangle \\ & + (0.4678 + 0.3975i)|10_B\rangle - (0.2455 + 0.6147i)|11_B\rangle \end{aligned} \tag{11.88}$$

$$\begin{aligned}|\bar{1}_B\rangle := {} & (0.3978 - 0.2475i)|00_B\rangle + (0.2754 + 0.5651i)|01_B\rangle\rangle \\ & + (0.1409 - 0.3826i)|10_B + (0.3533 - 0.3069i)|11_B\rangle \end{aligned} \tag{11.89}$$

where here “$|\bar{0}_B\rangle$” and “$|\bar{0}_B\rangle$” represent 2-qubit states. Notice that we only need $\min(d_A, d_B)$ eigenvectors for each basis even though the dimensions of sub-space A is (in this example) less than that of sub-space B.

Finally, we pick out the Schmidt coefficients from the singular values to give:

$$\lambda_0 = 0.9031 \tag{11.90}$$

$$\lambda_1 = 0.4295 \tag{11.91}$$

Hence our Schmidt decomposition is predicted to be:

$$|\psi_{AB}\rangle = \lambda_0|\bar{0}_A\rangle|\bar{0}_B\rangle + \lambda_1|\bar{1}_A\rangle|\bar{1}_B\rangle \tag{11.92}$$

11.3.8 Entanglement Distillation

In most applications of quantum communications and distributed quantum computing it is necessary to establish noise-free maximally entangled pairs of particles, such as pure Bell states, between the ends of a quantum communications channel. Invariably, when one sends quantum particles down real channels those particles will be affected by noise sources in the channel. Thus, what starts off as pure maximally entangled states will not end up as pure maximally entangled states by the time they reach the ends of the channel. This usually causes a failure of the protocol for which the sharing of maximal entanglement was necessary.

The solution is to perform “entanglement distillation” whereby a few maximally entangled bi-partite pure states are obtained from a larger number of non-maximally entangled bi-partite states. Convention has it that if the original states are pure, the process is called “entanglement concentration”, whereas if they are mixed the process is called “entanglement purification”. Either way the principle is the same—one sacrifices some of the non-maximally entangled states in order to distill out a smaller number of maximally entangled ones. There has now been a great deal of research invested in entanglement distillation reflecting its importance as a quantum information processing primitive.

11.3.8.1 Distilling Entanglement from Pure States: Entanglement Concentration

In entanglement concentration we distill out several maximally entangled bi-partite pure states (e.g., states of the form $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle))$ from a larger number of non-maximally entangled bi-partite pure states (e.g. states of the form $\alpha|00\rangle + \beta|11\rangle$ with $|\alpha| \neq |\beta|$). Thus, entanglement concentration can also be thought of as a kind of error correction wherein several “buggy” Bell states are distilled into fewer perfect Bell states.

Entanglement concentration was first proposed by Charles Bennett, Herbert Bernstein, Sandu Popescu, and Benjamin Schumacher [51], but their scheme was later improved upon by Phillip Kaye and Michele Mosca [268] and it is the latter version we describe here.

Suppose Alice and Bob share an entangled pair of qubits in the state:

$$|\Psi\rangle = a_0|00\rangle + a_1|01\rangle + a_2|10\rangle + a_3|11\rangle \tag{11.93}$$

where $\sum_{i=0}^{3}|a_i|^2 = 1$. Using the Schmidt decomposition of Sect. 11.3.7, such a state can always be re-expressed in the form

$$|\Psi\rangle = \alpha|\psi_0\rangle|\phi_0\rangle + \beta|\psi_1\rangle|\phi_1\rangle \tag{11.94}$$

for positive reals α and β, a $\{|\psi_0\rangle, |\psi_1\rangle\}$-basis for Alice and a $\{|\phi_0\rangle, |\phi_1\rangle\}$-basis for Bob. Alice and Bob could simply agree by convention to call their bases $\{|\psi_0\rangle, |\psi_1\rangle\} \equiv \{\overset{A}{|0\rangle}, \overset{A}{|1\rangle}\}$ and $\{|\phi_0\rangle, |\phi_1\rangle\} \equiv \{\overset{B}{|0\rangle}, \overset{B}{|1\rangle}\}$. Thus, whatever the entangled state Alice and Bob share, we can think of it as a "buggy" Bell state:

$$|\Psi\rangle = \alpha\overset{A}{|0\rangle}\overset{B}{|0\rangle} + \beta\overset{A}{|1\rangle}\overset{B}{|1\rangle} \tag{11.95}$$

where the over set letters indicate whether we are talking about Alice's qubit or Bob's. If $|\alpha| = |\beta|$ we would be dealing with a maximally entangled state. But in general this is not the case. Yet it is the maximally entangled states we need routinely in quantum information protocols. So the question arises how do we distill out a few maximally entangled Bell states from a greater number of non-maximally entangled ones?

Let us imagine we begin with n of these buggy Bell pairs. Thus, our starting state can be written as $|\Psi\rangle = \big(\alpha\overset{A}{|0\rangle}\overset{B}{|0\rangle} + \beta\overset{A}{|1\rangle}\overset{B}{|1\rangle}\big)^{\otimes n}$. Expanding the definition gives us a state in which Alice's and Bob's qubits are scrambled together. For example, if $n = 2$, $|\Psi\rangle$ is equal to:

$$\begin{aligned} |\Psi\rangle &= \big(\alpha\overset{A}{|0\rangle}\overset{B}{|0\rangle} + \beta\overset{A}{|1\rangle}\overset{B}{|1\rangle}\big)^{\otimes 2} \\ &= \alpha^2\overset{AB}{|00\rangle}\overset{AB}{|00\rangle} + \alpha\beta\overset{AB}{|00\rangle}\overset{AB}{|11\rangle} + \alpha\beta\overset{AB}{|11\rangle}\overset{AB}{|00\rangle} + \beta^2\overset{AB}{|11\rangle}\overset{AB}{|11\rangle} \end{aligned} \tag{11.96}$$

However, it is apparent that a simple bit-permutation applied to the qubits will allow us to group Alice and Bob's qubits separately. After this permutation of qubits we can see $|\Psi\rangle$ is equivalent to:

$$|\Psi\rangle = \alpha^2\overset{AA}{|00\rangle}\overset{BB}{|00\rangle} + \alpha\beta\overset{AA}{|01\rangle}\overset{BB}{|01\rangle} + \alpha\beta\overset{AA}{|10\rangle}\overset{BB}{|10\rangle} + \beta^2\overset{AA}{|11\rangle}\overset{BB}{|11\rangle} \tag{11.97}$$

Generalizing, the state we obtain with n non-maximally entangled states is:

$$|\Psi\rangle = \sum_{j=0}^{n}\alpha^{n-j}\beta^{j}\left(\sum_{\text{HammingWeight}(\mathbf{x})=j}|\overset{AA\cdots A}{\mathbf{x}}\rangle|\overset{BB\cdots B}{\mathbf{x}}\rangle\right) \tag{11.98}$$

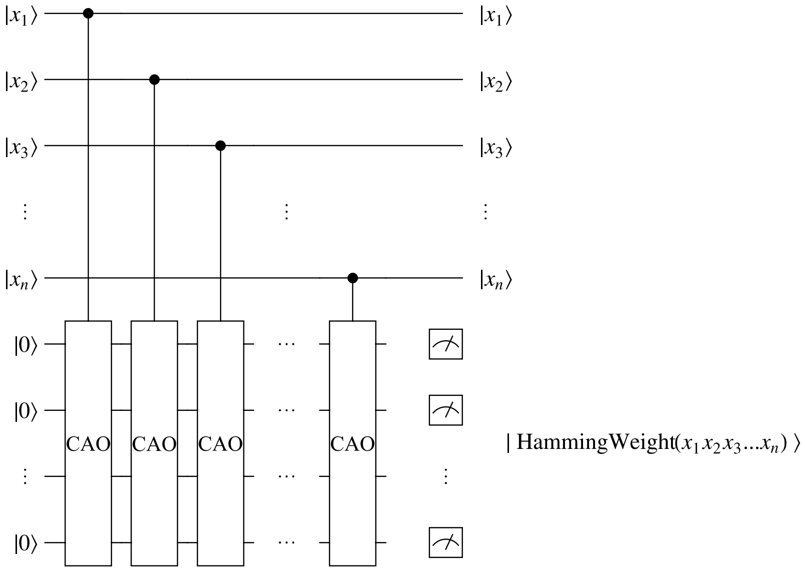

Fig. 11.4 Quantum circuit for measuring the Hamming weight of string of qubits. The quantum state whose Hamming weight we want is in the upper register, and a set of n ancillae qubits is in the lower register. Each qubit that is set to $|1\rangle$ in the upper register adds 1 to the Hamming weight via a controlled-add-one gate. However, each qubit set to $|0\rangle$ adds nothing to the Hamming weight. By initializing the ancillae qubits to $|00\ldots0\rangle$ we accumulate the Hamming weight in the lower register. If the upper register is a superposition state that has eigenstates of different Hamming weights (as we intend it to be) the act of measuring the Hamming weight projects the state of the upper register into a superposition of only those eigenstates consistent with the observed Hamming weight

Now suppose Alice measures the Hamming weight of $|\Psi\rangle$, i.e., she determines how many of her qubits are in state $|1\rangle$. By the structure of the state $|\Psi\rangle$ Bob would be guaranteed to obtain the same result if he were to measure the Hamming weight of his qubits. To measure the Hamming weight, Alice can use a quantum circuit like that shown in Fig. 11.4. This circuit consists of two registers: an upper n-qubit register holding the superposition $|\Psi\rangle$ and a lower n-qubit register holding n ancillae prepared initially in state $|00\ldots0\rangle$. These registers are connected via a cascade of controlled-add-one gates. If the i-th qubit in the upper register is set to $|1\rangle$ it adds 1 to the Hamming weight and 0 otherwise. Via the linearity of quantum mechanics, the circuit produces a superposition of Hamming weights in the lower register. When the Hamming weight in the lower register is measured the upper register is projected into a state whose component eigenstates then have the Hamming weight that was measured in the lower register. Thus, if Alice and Bob each measure the Hamming weight to be $|j\rangle$ they will project the upper register into a state of the form:

$$\frac{1}{\sqrt{\binom{n}{j}}} \sum_{\text{HammingWeight}(\mathbf{x})=j} \overset{AA\cdots A}{|\mathbf{x}\rangle} \overset{BB\cdots B}{|\mathbf{x}\rangle} \tag{11.99}$$

which is a superposition of $\binom{n}{j}$ bit strings.

Next define a function that maps each of these bits strings (arranged in lexicographic order) into a unique integer from 0 to $\binom{n}{j} - 1$. Specifically, we have:

$$\begin{aligned} f(00\ldots 00\underbrace{11\ldots 1}_{j}) &\rightarrow (0)_{10} = 00\ldots 0 \\ f(00\ldots 010\underbrace{1\ldots 1}_{j-1}) &\rightarrow (1)_{10} = 00\ldots 1 \\ &\vdots \\ f(\underbrace{11\ldots 1}_{j}00\ldots 0) &\rightarrow \left(\binom{n}{j} - 1\right)_{10} \end{aligned} \tag{11.100}$$

For example, if there are $n = 6$ qubits with Hamming weight 4, the mapping f would be:

$$\begin{aligned} f(0,0,1,1,1,1) &\rightarrow f(15) \rightarrow 0 \\ f(0,1,0,1,1,1) &\rightarrow f(23) \rightarrow 1 \\ f(0,1,1,0,1,1) &\rightarrow f(27) \rightarrow 2 \\ f(0,1,1,1,0,1) &\rightarrow f(29) \rightarrow 3 \\ f(0,1,1,1,1,0) &\rightarrow f(30) \rightarrow 4 \\ f(1,0,0,1,1,1) &\rightarrow f(39) \rightarrow 5 \\ f(1,0,1,0,1,1) &\rightarrow f(43) \rightarrow 6 \\ f(1,0,1,1,0,1) &\rightarrow f(45) \rightarrow 7 \\ f(1,0,1,1,1,0) &\rightarrow f(46) \rightarrow 8 \\ f(1,1,0,0,1,1) &\rightarrow f(51) \rightarrow 9 \\ f(1,1,0,1,0,1) &\rightarrow f(53) \rightarrow 10 \\ f(1,1,0,1,1,0) &\rightarrow f(54) \rightarrow 11 \\ f(1,1,1,0,0,1) &\rightarrow f(57) \rightarrow 12 \\ f(1,1,1,0,1,0) &\rightarrow f(58) \rightarrow 13 \\ f(1,1,1,1,0,0) &\rightarrow f(60) \rightarrow 14 \end{aligned} \tag{11.101}$$

We further extend f in any way that keeps it reversible and hence implementable as a permutation matrix, i.e. we extend the definition of f so that it maps each of the other bit strings (which do not have Hamming weight j) to unique indices too.

If we define $r = \left\lceil \log_2 \binom{n}{j} \right\rceil$ we can write:

$$\begin{aligned}
&\frac{1}{\sqrt{\binom{n}{j}}} \sum_{\text{HammingWeight}(x)=j} \overset{AA\cdots A}{|\mathbf{x}\rangle}\overset{BB\cdots B}{|\mathbf{x}\rangle} \\
&\xrightarrow{f} \frac{1}{\sqrt{\binom{n}{j}}} \sum_{\text{HammingWeight}(x)=j} |\overset{AA\cdots A}{f(\mathbf{x})}\rangle|\overset{BB\cdots B}{f(\mathbf{x})}\rangle \\
&= \frac{1}{\sqrt{\binom{n}{j}}} \sum_{y=0}^{\binom{n}{j}-1} |\underbrace{\overset{AA\cdots A}{\mathbf{0}}}_{n-r}\rangle|\underbrace{\overset{AA\cdots A}{\mathbf{y}}}_{r}\rangle|\overset{BB\cdots B}{\mathbf{0}}\rangle|\overset{BB\cdots B}{\mathbf{y}}\rangle
\end{aligned} \tag{11.102}$$

If $\binom{n}{j} = 2^r$, ignoring the first $n-r$ qubits in each register then gives:

$$\frac{1}{\sqrt{2^r}} \sum_{y=0}^{2^r-1} |\overset{AA\cdots A}{\mathbf{y}}\rangle|\overset{BB\cdots B}{\mathbf{y}}\rangle \tag{11.103}$$

which as before can, under a permutation of the qubits, be recognized as r pristine Bell state pairs, and the entanglement in $|\Psi\rangle$ has been concentrated.

Of course, in general $\binom{n}{j} \neq 2^r$. In this case, one can still distill out some perfect Bell state pairs, but their number is not certain a priori. See [268] for details.

Thus, to sum up, the entanglement concentration procedure can be described as follows:

Entanglement Concentration

1. Alice and Bob start off with n copies of a non-maximally entangled state $|\Psi\rangle = (\alpha|00\rangle + \beta|11\rangle)^{\otimes n}$ with $\alpha \neq \beta$, and they each hold one member of each non-maximally entangled pair.
2. Next Alice and Bob perform a qubit-permutation sufficient to group all Alice's qubits together and all Bob's qubits together, creating a state of the form

$$|\Psi\rangle = \sum_{j=0}^{n} \alpha^{n-j}\beta^{j} \left(\sum_{\text{HammingWeight}(\mathbf{x})=j} |\overset{AA\,\cdots A}{\mathbf{x}}\rangle|\overset{BB\,\cdots B}{\mathbf{x}}\rangle \right) \tag{11.104}$$

3. Alice and Bob each measure the Hamming weight of their set of particles, i.e., they each determine how many of their qubits are in state $|1\rangle$. Given the structure of the state, their results will always agree. If they each determine the Hamming weight is $|j\rangle$, this measurement operation has the effect of projecting Alice and

Bob's joint state into the form

$$\frac{1}{\sqrt{\binom{n}{j}}}\sum_{\text{HammingWeight}(\mathbf{x})=j} \overset{AA\cdots A}{|\mathbf{x}\rangle}\overset{BB\cdots B}{|\mathbf{x}\rangle} \tag{11.105}$$

where the labels "A" and "B" specify whether the qubits are in Alice's possession or Bob's possession.

4. Alice and Bob each apply the transformation f to the state obtained in the last step. In the simplest case when $\binom{n}{j}=2^r$ ignoring the first $n-r$ qubits gives the state $\frac{1}{\sqrt{2^r}}\sum_{y=0}^{2^r-1}|\overset{AA\cdots A}{\mathbf{y}}\rangle|\overset{BB\cdots B}{\mathbf{y}}\rangle$.
5. By inverting the qubit permutation performed at step (2) above, this state becomes that of r perfect Bell pairs.
6. The procedure can extended to deal with the case $\binom{n}{j}\neq 2^r$, and a quantum circuit can be defined which allows the number of perfect Bell pairs distilled out to be measured (see [268]). The expected number of pairs obtainable when $\binom{n}{j}\neq 2^r$ is at least $\sum_{j=0}^{n}|\alpha^2|^{n-j}(1-|\alpha^2|)^j\binom{n}{j}\left(\left\lfloor\log_2\binom{n}{j}\right\rfloor-1\right)$.

Entanglement concentration is of practical importance in many quantum communications protocols as well as in distributed quantum computing (see Sect. 15.2). It can be extended to the case of distilling bi-partite maximally entangled pure states from non-maximally entangled mixed states, and is then known as *entanglement purification* [53, 54, 139]. This is distinct from the concept of the purification of a mixed state discussed in Sect. 11.2.4 whereby a mixed state, ρ_B, is re-cast as the partial trace of a pure state, $|\psi_{AB}\rangle$, in a higher dimensional Hilbert space, i.e., state purification finds the $|\psi_{AB}\rangle$ such that $\rho_B=\text{tr}_A(|\psi_{AB}\rangle\langle\psi_{AB}|)$. By contrast, in entanglement purification we distill out a set of maximally entangled bi-partite pure states from a larger number of non-maximally entangled bi-partite mixed states.

11.3.9 Entanglement Swapping

Thus far, the schemes we have looked at for creating entanglement have all worked by causing pairs of initially unentangled qubits to interact *directly* and then separating them spatially. However, it is also possible to entangle two particles that have *never* interacted directly. The trick is to start with two maximally entangled pairs of particles, and to arrange for one member of each pair to be measured in a Bell basis using a device known as a "Bell State Analyzer" (BSA). This sounds fancy, and experimentally it is challenging to build one, but theoretically speaking it requires nothing more than the Bell state synthesizer circuit run in reverse followed by single qubit measurements in the computational basis. The net effect is that we can swap

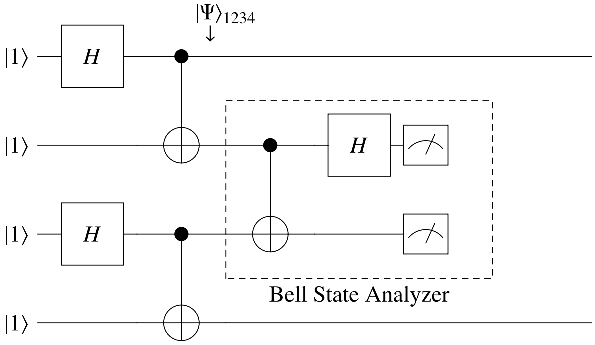

Fig. 11.5 Entanglement swapping provides a means to entangle two parties that have never interacted with one another directly. Alice and Bob each prepare a maximally entangled pair of particles. They each retain one of these particles and pass the other to Cerys. Cerys performs a complete Bell state analysis on the two particles she received, which results in classifying them as being in one of the four Bell states $|\beta_{00}\rangle$, $|\beta_{01}\rangle$, $|\beta_{10}\rangle$, $|\beta_{11}\rangle$. Thereafter, the two particles that remain in Alice and Bob's possession will be maximally entangled in some Bell state the identity of which depends on result of the Bell state analysis. Entanglement swapping is a key ingredient of quantum repeaters, distributed quantum computing, and heralded entangled photon sources

initial entanglement between particles 1 and 2 and initial entanglement between particles 3 and 4 into entanglement between particles 1 and 4, even though particles 1 and 4 never interacted directly. The procedure that does this is therefore called *entanglement swapping* and was originally conceived of by Marek Zukowski, Anton Zeilinger, Michael Horne, and Artur Ekert in 1993 [565]. The scheme is illustrated in Fig. 11.5.

Entanglement swapping works as follows: Alice and Bob each prepare matching maximally entangled pairs of particles. For example, they may each prepare their own Bell singlet pair $|\beta_{11}\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$. Such states can be prepared by feeding a $|11\rangle$ state into a Bell state synthesizer circuit, which consists of a single Walsh-Hadamard gate followed by a CNOT gate. Let us say that Alice starts off in possession of qubits 1 and 2, and Bob starts off in possession of qubits 3 and 4. Using these particle labels as subscripts to avoid ambiguity, the input to the entanglement swapping circuit is therefore the state $|11\rangle_{12} \otimes |11\rangle_{34} = |1111\rangle_{1234}$. Upon applying the double Bell state synthesizer circuits as show in Fig. 11.5, the following transformation occurs:

$$\begin{aligned}|\Psi\rangle_{1234} &= (\text{CNOT} \otimes \text{CNOT}) \cdot (H \otimes \mathbb{1} \otimes H \otimes \mathbb{1})|11\rangle_{12}|11\rangle_{34} \\ &= \left(\frac{1}{\sqrt{2}}(|01\rangle_{12} - |10\rangle_{12})\right) \otimes \left(\frac{1}{\sqrt{2}}(|01\rangle_{34} - |10\rangle_{34})\right) \qquad (11.106)\end{aligned}$$

However, $|\psi\rangle_{1234}$ can also be re-expressed in the Bell basis by imagining the qubits to be permuted as follows. Swap qubits 2 and 4 (to take the qubit ordering 1234 into 1432) and then swap qubits 3 and 2 (to take the qubit ordering 1432 into 1423). An operator sufficient to perform qubit permutation is $(\mathbb{1}_4 \otimes \text{SWAP}) \cdot (\mathbb{1}_2 \otimes \text{SWAP} \otimes$

$\mathbb{1}_2) \cdot (\mathbb{1}_4 \otimes \text{SWAP}) \cdot (\mathbb{1}_2 \otimes \text{SWAP} \otimes \mathbb{1}_2)$. In the "1423" basis, we can write $|\psi\rangle_{1234}$ as the equivalent $|\psi\rangle_{1423}$ where:

$$|\psi_{1423}\rangle = \frac{1}{2}(-|\beta_{00}\rangle_{14}|\beta_{00}\rangle_{23} + |\beta_{01}\rangle_{14}|\beta_{01}\rangle_{23} + |\beta_{10}\rangle_{14}|\beta_{10}\rangle_{23} - |\beta_{11}\rangle_{14}|\beta_{11}\rangle_{23}) \tag{11.107}$$

Hence, in this Bell basis representation, we can see immediately that if we perform a complete Bell state analysis of qubits 2 and 3 (i.e., if we figure out which Bell state they are in), then qubits 1 and 4 will then be projected into the identical Bell state (up to an overall phase factor), even though qubits 1 and 4 had, at no time, interacted directly.

To perform a complete Bell-basis measurement we need only *invert* the operation that synthesizes the Bell states starting from the computational basis states and then measure the result in the computational basis. In terms of a quantum circuit such an inversion is achieved by reversing the order of the gates and using the inverse (or, since they are unitary, the conjugate transpose) of each gate operator. Therefore, as the Bell state synthesizer is the operator, CNOT $\cdot$ $(H \otimes \mathbb{1})$, the complete Bell state analyzer is the operator $(H \otimes \mathbb{1})^\dagger \cdot \text{CNOT}^\dagger = (H \otimes \mathbb{1}) \cdot \text{CNOT}$ (as shown in the dashed box in Fig. 11.5). So defined, the Bell state analyzer accepts a Bell state and returns $|00\rangle$, $|01\rangle$, $|10\rangle$, or $|11\rangle$, according to whether the input Bell state was $|\beta_{00}\rangle$, $|\beta_{01}\rangle$, $|\beta_{10}\rangle$, or $|\beta_{11}\rangle$. A *complete* Bell state analyzer has been demonstrated experimentally by Yoon-Ho Kim, Sergei Kulik, and Yanhua Shih in 2001 [278].

Entanglement swapping is a very useful trick in quantum information science. It is a crucial building block in quantum repeaters (used to extend the range of quantum key distribution in optical fibers) [103, 160, 295, 428], in distributed quantum computing [552], and as a means to have a heralded source of entangled photon pairs [565].

11.3.10 Entanglement in "Warm" Bulk Matter

One of the most exciting developments in our understanding of entanglement in recent years has come from the realization that entanglement can persist in macroscopic amounts of matter at room temperature. This came as a complete surprise. Just a few years ago creating and sustaining entangled states of even a handful of quantum particles required exquisitely delicate experiments, and ideal laboratory conditions. Indeed, great suspicion fell on anyone suggesting that entanglement might play a role in the brain and biological structures mainly on the grounds that they were too warm and noisy to sustain such effects. However, old-school thinking about entanglement should no longer be taken as conclusive.

We now know that entanglement can be found in macroscopic systems [19], even relatively warm ones [513], and in fact plays an *essential* role in determining how such matter behaves, such as the anomalously low magnetic susceptibility of certain magnetic systems [85]. This is quite extraordinary. Such developments are very exciting because they could mark the beginning in an entirely new direction for materials science and solid state physics. Who knows what miracle materials await discovery if entanglement can persist and play a role in shaping their properties at temperatures well above absolute zero.

Similarly, other studies have provided evidence for the existence of quantum effects in certain biological structures. For example, quantum transport is believed to occur in the Fenna-Matthews-Olson (FMO) light harvesting complex of purple bacteria [172, 306]. At low temperatures the excitons created after photon absorption are found to propagate through the FMO complex *coherently*, and in fact, their transport is *enhanced* by the presence of a small amount of noise, perhaps allowing the phenomenon to persist up to biologically relevant temperatures. Likewise, it has been hypothesized that magnetoreception in birds works by interconverting singlet/triplet excited states of the cryptochrome protein [256]. And a recent model of olfaction replaces the standard shape-based theory with the notion that phonon-assisted tunneling is used to sense the vibrational spectra of odorant molecules [500]. All these results are intriguing and may point to more sophisticated ways of harnessing quantum effects in structures that are at relatively high temperatures.

11.4 Compressing Quantum Information

In classical information theory we describe messages as sequences of symbols drawn from some finite alphabet, such that each symbol may appear with a different probability. The obvious quantum analog of this is to treat a source of quantum information as a device that generates a sequence of quantum states, each potentially with some different probability. Thus, quantum states that are known only as some statistical mixture of pure states arise naturally when we extend information theory into the quantum realm.

Whereas Shannon information theory regards a classical source as a device that generates a sequence of classical symbols (i.e., distinguishable states) picked from a finite alphabet according to different probabilities, quantum information theory regards a quantum source as a device that generates a sequence of quantum symbols (i.e., not necessarily distinguishable states) picked from a finite alphabet according to different probabilities. Thus, we find ourselves having to model quantum states that are only specified in exactly the statistical sense mentioned above. Hence, the introduction of density operators is absolutely necessary.

If the quantum states used for the alphabet of symbols are all orthogonal to one another then, in principle, we can measure them without disturbing them, and hence to all intents and purposes they are essentially just classical symbols in disguise. Hence, we would could characterize such a source in terms of its Shannon entropy. In particular, if a source produces a stream of orthogonal (i.e., unambiguously distinguishable) states in which the i-th state occurs with probability p_i the source is characterized by its Shannon entropy $H(\{p_i\}) = -\sum_i p_i \log_2 p_i$.

The situation becomes more interesting if we assume that the quantum states encoding the symbols are not necessarily all orthogonal to one another.

11.4.1 Quantum Sources: The von Neumann Entropy

Let us imagine we have a device for outputting one of d not necessarily orthogonal quantum states at random. In particular, let the device output state $|\psi_i\rangle$ with corresponding probability p_i. The density operator for such a source would be:

$$\rho = \sum_{i=1}^{d} p_i |\psi_i\rangle\langle\psi_i| \tag{11.108}$$

and we can characterize its entropy using the techniques introduced in Sect. 11.2.5.2. There we saw that the entropy of a quantum source can be related to the Shannon entropy of a corresponding fictitious classical source. Specifically, the von Neumann entropy of a quantum source having density operator ρ is defined via its representation in a diagonal basis as:

$$S_V(\rho) = -\sum_i \lambda_i \log_2 \lambda_i = -\text{tr}(\rho \log_2 \rho) \tag{11.109}$$

As expected, the von Neumann entropy so-defined then equals to the Shannon entropy when the quantum states emitted by the source are unambiguously distinguishable.

The von Neumann entropy has many interesting uses and properties. For example, if ρ is a pure state, $S_V(\rho) = 0$. Hence, the von Neumann entropy can be used to decide whether or not a given density operator corresponds to that of a pure state. In addition, the von Neumann entropy of a state does not change under unitary evolution, i.e., $S_V(\rho) = S_V(U\rho U^\dagger)$, because the von Neumann entropy only depends upon the eigenvalues and these are not changed under unitary evolution. These and other important properties of the von Neumann entropy are summarized in Table 11.3

In analogy to the Shannon noiseless coding theorem wherein the n bit classical messages from a classical source with Shannon entropy $H(\{p_i\})$ can be compressed into at most $nH(\{p_i\})$ classical bits, n qubit quantum messages from a quantum source with von Neumann entropy $S_V(\rho)$ can be compressed into at most $nS_V(\rho)$ qubits. However, this tells us nothing about how to accomplish the compression. That is the domain of quantum data compression.

11.4.2 Schumacher-Jozsa Quantum Data Compression

Suppose Alice chooses real numbers α and β such that $\alpha^2 + \beta^2 = 1$, and creates a quantum message consisting of sequences of the states $|\psi_+\rangle$ and $|\psi_-\rangle$ defined as:

$$|\psi_+\rangle = \alpha|0\rangle + \beta|1\rangle \tag{11.110}$$

$$|\psi_-\rangle = \alpha|0\rangle - \beta|1\rangle \tag{11.111}$$

Table 11.3 Properties of the von Neumann entropy

Property	Formula	Condition
Purity	$S_V(\rho) = 0$	If ρ is a pure state, i.e., $\rho = \lvert\psi\rangle\langle\psi\rvert$
Invariance	$S_V(\rho) = S_V(U\rho U^\dagger)$	If U is a unitary transformation
Maximum	$S_V(\rho) \leq \log_2 k$	If ρ has k non-zero eigenvalues. Equality holds when all the non-zero eigenvalues are equal
Concavity	$S_V(\sum_i p_i\rho_i) \geq \sum_i p_i S_V(\rho_i)$	Provided $p_i \geq 0$ and $\sum_i p_i = 1$. This result shows that the less we know about how a state is prepared the greater its von Neumann entropy
Boundedness	$S_V(\rho) \leq H(\{p_i\})$	For an ensemble of quantum states $\lvert\psi_i\rangle$ occurring with probabilities p_i, and having density operator $\rho = \sum_i p_i\lvert\psi_i\rangle$ its von Neumann entropy is never greater than the Shannon entropy of the corresponding classical ensemble. Equality holds when all the quantum states are orthogonal and hence unambiguously distinguishable
Subadditivity	$S_V(\rho_{AB}) \leq S_V(\rho_A) + S_V(\rho_B)$	Equality holds when $\rho_{AB} = \rho_A \otimes \rho_B$. That is, the von Neumann entropies of independent systems add, but will be lowered if they are correlated
Strong subadditivity	$S_V(\rho_{ABC}) + S_V(\rho_B)$ $\leq S_V(\rho_{AB}) + S_V(\rho_{BC})$	For two systems AB and BC having a common subsystem B the sum of the von Neumann entropies of their union and intersection is less than the sum of their von Neumann entropies
Araki-Lieb inequality	$S_V(\rho_{AB}) \geq \lvert S_V(\rho_A) - S_V(\rho_B)\rvert$	A bipartite state ρ_{AB} can be completely known (zero entropy) even though its parts are not, such as when $S_V(\rho_A) = S_V(\rho_B) \neq 0$

The overlap $\langle\psi_+|\psi_-\rangle = 2\alpha^2 - 1$ is non-zero, and hence $|\psi_+\rangle$ and $|\psi_-\rangle$ are non-orthogonal, for most values of α. This means that the quantum "symbols" in Alice's message are not entirely distinguishable for most choices of α. There is the potential, therefore, for some added redundancy in messages encoded using (non-orthogonal) quantum symbols that is not present in messages encoded using (orthogonal) classical symbols. Ultimately, this is what allows quantum messages to be compressed beyond the Shannon bound.

If the two states, $|\psi_+\rangle$ and $|\psi_-\rangle$, appear with equal probability, the von Neumann entropy of Alice's source is:

$$S(\rho) = -\alpha^2 \log_2 \alpha^2 - \beta^2 \log_2 \beta^2 \tag{11.112}$$

Thus, if the states are orthogonal (which occurs when $\alpha^2 = \beta^2 = \frac{1}{2}$) the von Neumann entropy of the source reduces to exactly the Shannon entropy.

As shown by Mitsumori et al. [356] we can compress our quantum message in blocks of three qubits at a time.

$$|B_{\mathbf{L}}\rangle = |\psi_{L_1}\rangle \otimes |\psi_{L_2}\rangle \otimes |\psi_{L_3}\rangle \tag{11.113}$$

$\mathbf{L} = (L_1, L_2, L_3)$ and $L_i \in \{+, -\}$.

Index L corresponds to one of eight possible configurations for the 3-qubit block, namely $|\psi_+\rangle|\psi_+\rangle|\psi_+\rangle$, $|\psi_+\rangle|\psi_+\rangle|\psi_-\rangle$, $|\psi_+\rangle|\psi_-\rangle|\psi_+\rangle, \ldots, |\psi_-\rangle|\psi_-\rangle|\psi_-\rangle$.

Alice applies the "compressor" operation, U, which is defined via its action on computational basis states as follows:

$$U := \begin{array}{l} |000\rangle \to |000\rangle \\ |001\rangle \to |001\rangle \\ |010\rangle \to |010\rangle \\ |011\rangle \to |100\rangle \\ |100\rangle \to |011\rangle \\ |101\rangle \to |101\rangle \\ |110\rangle \to |110\rangle \\ |111\rangle \to |111\rangle \end{array}$$

The state of a block of three qubits after this compressor has been applied is as follows:

$$U|B_{\mathbf{L}}\rangle = \alpha^2\sqrt{1+2\beta^2}|0\rangle|\mu_{\mathbf{L}}\rangle + \beta^2\sqrt{1+2\alpha^2}|\nu_{\mathbf{L}}\rangle \tag{11.115}$$

where

$$|\mu_{\mathbf{L}}\rangle = \frac{1}{1+2\beta^2}(\alpha|00\rangle + \beta_1|11\rangle + \beta_2|10\rangle + \beta_3|01\rangle) \tag{11.116}$$

$$|\nu_{\mathbf{L}}\rangle = \frac{1}{\beta^2\sqrt{1+2\alpha^2}}[\alpha(\beta_1\beta_2|10\rangle + \beta_1\beta_3|01\rangle + \beta_2\beta_3|00\rangle) + \beta_1\beta_2\beta_3|11\rangle] \tag{11.117}$$

where $\beta_i = L_i\beta$ which will either be $+\beta$ or $-\beta$.

Next Alice measures the first qubit of the compressed state in the computational basis, to obtain the value $|0\rangle$ or $|1\rangle$ [356]. What happens next depends on whether Alice wants to pursue a "Discard-on-Fail" (see Fig. 11.6) or an "Augment-on-Fail" (see Fig. 11.7) quantum data compression protocol. Let us take a look at each of these protocols in turn.

11.4.3 "Discard-on-Fail" Quantum Data Compression Protocol

Discard-on-Fail Quantum Data Compression

1. Partition the data in blocks of three qubits at a time, apply the compressor, U, to each block, i.e., $|B_{\mathbf{L}}\rangle \to U|B_{\mathbf{L}}\rangle$.

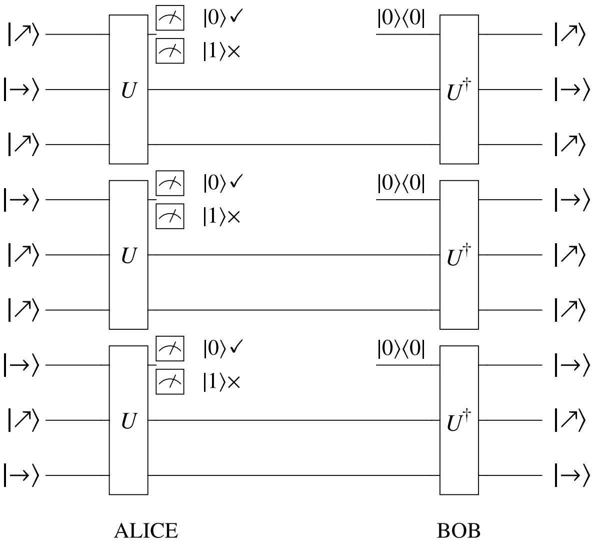

Fig. 11.6 First quantum data compression protocol. Alice encodes a sequence of non-orthogonal qubits in blocks of three qubits using the compressor U. She reads the first qubit obtaining the result $|0\rangle$ or $|1\rangle$. If Alice obtains $|0\rangle$ she will have prepared the second and third qubits in the same block in the state $|\mu_L\rangle$, and she sends these qubits to Bob. Upon receipt, Bob augments these qubits with a new first qubit prepared in state $|0\rangle$ and sends all three qubits through the decompressor $U^\dagger$. The output triplet of qubits is now restored close to their original values even though only two qubits (rather than three) passed through the channel between Alice and Bob. If, instead, when Alice had measured the first qubit Alice she had found it in state $|1\rangle$ she would have regarded this as a "failure" and would have sent nothing to Bob

2. Alice measures the first qubit in each block output from the compressor, and obtains $|0\rangle$ or $|1\rangle$.
3. If Alice obtains $|0\rangle$ she retains the measured qubit and passes the remaining two qubits, now in state $\rho_{\mathbf{L}}^{(1)} = |\mu_{\mathbf{L}}\rangle\langle\mu_{\mathbf{L}}|$, to Bob. This event occurs with probability $p = \alpha^4(1+2\beta^2)$. If Alice obtains $|1\rangle$ she regards this as a "failure" and sends nothing to Bob. This event, which Bob sees a data drop out in the stream from Alice, occurs with probability $1-p$.
4. If Bob does receive qubits from Alice, he prepares a new qubit in the state $|0\rangle\langle 0|$ to create the state $(|0\rangle\langle 0| \otimes \rho_{\mathbf{L}}^{(1)})$, and then feeds this expanded state into the 3-qubit decompressor, $U^\dagger$. This operation produces the state

$$\Phi_{\mathbf{L}}^{(1)} = U^\dagger(|0\rangle\langle 0| \otimes \rho_{\mathbf{L}}^{(1)})U \tag{11.118}$$

5. The result is that for each block, Bob either receives nothing from Alice or a pair of qubits which he can expand and decompress. Hence, the fidelity of the overall quantum data compression process is

$$F^{(1)} = \sum_{\mathbf{L}} \frac{1}{8}\langle B_{\mathbf{L}}|\Phi_{\mathbf{L}}^{(1)}|B_{\mathbf{L}}\rangle = \alpha^8(1+2\beta^2)^2 \tag{11.119}$$

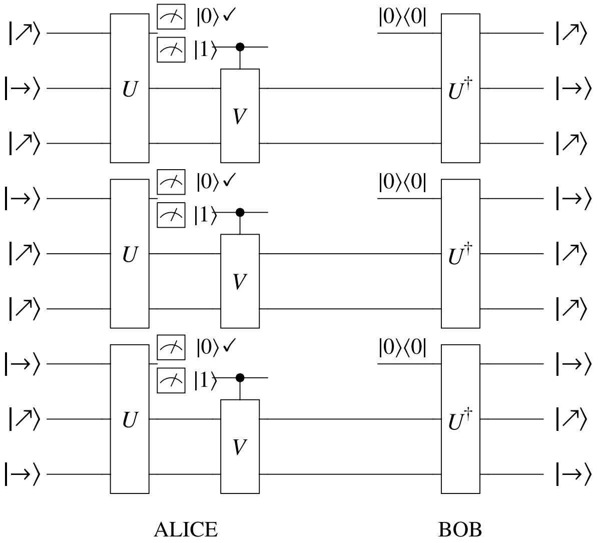

Fig. 11.7 Second quantum data compression protocol. Alice encodes a sequence of non-orthogonal qubits in blocks of three qubits using the compressor U. She reads the first qubit obtaining the result $|0\rangle$ or $|1\rangle$. If Alice obtains $|0\rangle$ she will have prepared the second and third qubits in the same block in the state $|\mu_L\rangle$, and she sends these qubits to Bob. If, however, Alice had obtained $|1\rangle$ when she had measured the first qubit, she would have modified the state of the second and third qubits before passing them to Bob. Upon receipt, Bob augments the qubits transmitted from Alice with a new first qubit prepared in state $|0\rangle$ and sends all three qubits through the decompressor $U^\dagger$. The output triplet of qubits is now restored close to their original values even though only two qubits (rather than three) passed through the channel between Alice and Bob

11.4.4 "Augment-on-Fail" Quantum Data Compression Protocol

Augment-on-Fail Quantum Data Compression

1. Partition the data in blocks of three qubits at a time, apply the compressor, U, to each block, i.e., $|B_{\mathrm{L}}\rangle \rightarrow U|B_{\mathrm{L}}\rangle$.
2. Alice measures the first qubit in each block output from the compressor, and obtains $0\rangle$ or $|1\rangle$.
3. If Alice obtains $|0\rangle$ she retains the measured qubit and passes the remaining two qubits to Bob. If Alice obtains $|1\rangle$ she applies a unitary operation V to the two unmeasured qubits and then sends them to Bob.
4. When Bob receives a pair of qubits from Alice, he prepares a new qubit in the state $|0\rangle$ to create the state $|0\rangle|\mu_L\rangle$, and then feeds this expanded state into the 3-qubit decompressor, $U^\dagger$.
5. The result is that for each block, Bob either receives nothing from Alice or a pair of qubits which he can expand and decompress.

The fidelity of the "augment-on-fail" quantum data compression protocol exceeds that of the "discard-on-fail" quantum data compression protocol. However, the "augment-on-fail" protocol is more challenging to implement in physical hardware due to the conditional correction that Alice must apply to the unmeasured qubits in each block prior to their transmission to Bob.

11.4.5 Quantum Circuit for Schumacher-Jozsa Compressor

The final step in understanding quantum data compression is to construct explicit quantum circuits for the compressor, U, and the decompressor, $U^\dagger$.

First, we can make our life easier by recognizing that once we know an efficient quantum circuit for U we know an efficient quantum circuit for $U^\dagger$ too. To see this, consider a unitary matrix, U, which can be factored in terms of a dot product of unitary matrices A and B i.e., $U = A \cdot B$. This implies that the unitary matrix $U^\dagger$ can be factored as $U^\dagger = U^{-1} = (A \cdot B)^{-1} = B^{-1} \cdot A^{-1} = B^\dagger \cdot A^\dagger$. Thus given a quantum circuit for the compressor, U, we can obtain a quantum circuit for the decompressor, $U^\dagger$, by inverting and reversing the gates in the quantum circuit for U. Hence, we need only find a quantum circuit for just the compressor U.

In order to realize the truth table (i.e., basis transformation) we want U to have, the matrix for U must take the form:

$$U := \begin{array}{c} \\ 000 \\ 001 \\ 010 \\ 011 \\ 100 \\ 101 \\ 110 \\ 111 \end{array} \!\!\! \begin{array}{c} \begin{array}{cccccccc} 000 & 001 & 010 & 011 & 100 & 101 & 110 & 111 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \tag{11.120}$$

which is a permutation matrix, similar to that of a TOFFOLI gate:

$$\text{TOFFOLI} := \begin{array}{c} \\ 000 \\ 001 \\ 010 \\ 011 \\ 100 \\ 101 \\ 110 \\ 111 \end{array} \!\!\! \begin{array}{c} \begin{array}{cccccccc} 000 & 001 & 010 & 011 & 100 & 101 & 110 & 111 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array} \tag{11.121}$$

except that the "NOT" part is shifted up the diagonal. This suggests that we can obtain U from TOFFOLI by shifting the "NOT" part using the permutation matrix Q:

$$Q := \begin{array}{c|cccccccc} & 000 & 001 & 010 & 011 & 100 & 101 & 110 & 111 \\ \hline 000 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 001 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 010 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 011 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 100 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 101 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 110 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 111 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \tag{11.122}$$

applied three times to TOFFOLI. Thus we obtain our first clue about how to construct U from the factorization:

$$U = Q \cdot Q \cdot Q \cdot \text{TOFFOLI} \cdot Q^{\dagger} \cdot Q^{\dagger} \cdot Q^{\dagger} \tag{11.123}$$

Next, we need to reduce TOFFOLI and Q to simpler forms. The TOFFOLI gate is well-studied and we already know of an efficient quantum circuit for implementing TOFFOLI using 1-qubit gates and CNOT gates (see Sect. 2.5.7. However, the Q gate is a new (and pretty useful) gate in the toolbox of the quantum circuit designer. So how do we factor Q into more familiar quantum gates?

The trick is to realize that for the general n-qubit case:

$$Q_{2^n} = \text{QFT}_{2^n}^{-1} \cdot T_{2^n} \cdot \text{QFT}_{2^n} \tag{11.124}$$

where T_{2^n} is defined by:

$$T_{2^n} = \bigotimes_{k=n-1}^{0} \begin{pmatrix} 1 & 0 \\ 0 & \exp(-\frac{2\pi i}{2^n} k) \end{pmatrix} \tag{11.125}$$

Consequently, in our 3-qubit example case, Q^3 reduces to:

$$Q^3 = \text{QFT}^{-1} \cdot T^3 \cdot \text{QFT} \tag{11.126}$$

Once, we recognize this basic structure further reductions become pretty easy to spot:

$$\begin{aligned} U &= Q^3 \cdot \text{TOFFOLI} \cdot {Q^{\dagger}}^3 \\ &= \text{QFT}^{-1} \cdot T^3 \cdot \text{QFT} \cdot \text{TOFFOLI} \cdot (\text{QFT}^{-1} \cdot T^3 \cdot \text{QFT})^{-1} \\ &= \text{QFT}^{-1} \cdot T^3 \cdot \text{QFT} \cdot \text{TOFFOLI} \cdot \text{QFT}^{-1} \cdot {T^{\dagger}}^3 \cdot \text{QFT} \end{aligned} \tag{11.127}$$

which can be further simplified by recognizing that $T^3 = Z \otimes R_z(\frac{\pi}{2}) \otimes R_z(-\frac{3\pi}{4})$.

Hence, we have succeeded in factorizing the compressor U in terms of TOFFOLI and Q gates, which in turn can both be reduced explicitly to 1-qubit and CNOT gates. Hence, our quantum circuit for the compressor, U, for the 3-qubit example block-size used, is shown in Fig. 11.8.

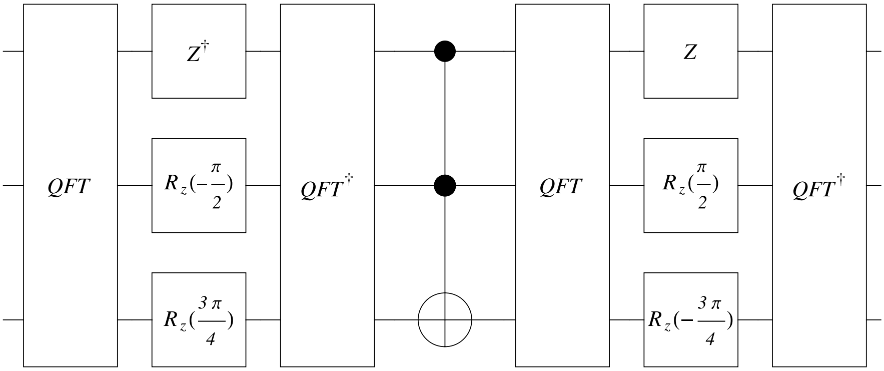

Fig. 11.8 Quantum circuit for the data compressor, U, used in the Schumacher-Jozsa quantum data compression protocols. The example given is appropriate for a block size of three qubits. The quantum circuit for the decompressor, $U^\dagger$, uses the inverse versions of the same gates applied in reverse order

11.4.6 Is Exponential Compression Possible?

A final thought on information compression in the quantum context is worthwhile. Let us compare the storage capacity of the Library of Congress to that a single qubit. Imagine, for example, that we translate each book in the Library of Congress into a bit string and concatenate them together. Then the entire Library, the entire repository of humankind's literary work product, is equivalent to some (very long) binary string. Let us call this string, s say. Ok... perhaps such an important bit string deserves a more grandiose letter. You've convinced me—let's call it Σ instead.

Now let's imagine affixing a period to the front of Σ to make $.\Sigma$. Having done so "$.\Sigma$" can be regarded as a binary *fraction* $0.j_1 j_2 \ldots j_n$. This represents a real number between 0 and 1 specifically $0 \leq \phi = j_1 2^{-1} + j_2 2^{-2} + \cdots + j_n 2^{-n} \leq 1$. Thus, in principle, we could imagine creating a single qubit state of the form

$$|\psi\rangle_\Sigma = |0\rangle + \exp(i\phi)|1\rangle \tag{11.128}$$

and so this single qubit state contains (in some sense) the entire body of human knowledge! So, on the face of it, it may appear possible to compress information into a single qubit by an exponential factor. Unfortunately, this is not possible. To encode all the bits needed to specify the complete contents of the Library of Congress would require a physically unrealistic precision in setting the angle ϕ. Moreover, any single attempt to perform a measurement on $|\psi\rangle_\Sigma$, or any transformed version thereof, will only reveal at most one bit of information. It is neither practically possible to cram the Library or Congress into a single qubit, nor to extract more than one bit of information from a single qubit state.

11.5 Superdense Coding

We know from Sect. 11.1.2 that if Alice wants to send Bob a *classical* message over a *classical* communications channel, the maximum extent to which she can compress her message is set by Shannon's Source Coding Theorem. This states that, if Alice wants negligible risk of information loss, a message comprising a string of n bits in which symbol 0 occurs with probability p_0, and symbol 1 occurs with probability $p_1 = 1 - p_0$ cannot be compressed into less than $nH(\{p_0, p_1\})$ bits, where $0 \leq H(\{p_0, p_1\}) \leq 1$ is the Shannon entropy of the source. The question of interest is whether Alice can compress her classical message beyond this Shannon bound if she is able to send it over a *quantum* communications channel?

At first sight it seems impossible for Alice to do any better that what is allowed by the Source Coding Theorem. Even if we allow Alice to use quantum states to encode her classical bits, the fact that we require those quantum states to be unambiguously distinguishable, consistent with our commonsense view of what it means to be a classical "symbol", forces Alice to have to use *orthogonal* quantum states to do the encoding. Thus, Alice could choose quantum state $|0\rangle$ to represent a classical bit 0, and quantum state $|1\rangle$ to represent classical bit 1. However, if Alice does this, the resulting von Neumann entropy, $S_V(\rho) = -\text{tr}(\rho \log_2 \rho)$, of her "quantum" source, described by density operator $\rho = p_0|0\rangle\langle 0| + p_1|1\rangle\langle 1|$, will be identical to the Shannon entropy of her equivalent classical source, having a probability distribution $\{p_0, p_1\} = \{p_0, 1 - p_0\}$ over the "symbols" 0 and 1. Hence, the maximum compression that is allowed quantum mechanically, i.e., $nS_V(\rho)$ qubits per n qubit message, will be identical to the maximum compression Alice can achieve classically, i.e., $nH(\{p_0, p_1\})$ bits per n-bit message. It would seem, therefore, that Alice can realize no benefit whatsoever from having access to a quantum channel over which to send her classical message.

It turns out, however, that there *is* a way of using a quantum communications channel to compress a stream of classical bits—*at communications time*—beyond that allowed by Shannon's Source Coding Theorem. The trick is to allow for the possibility of creating, distributing and storing certain entangled qubits (or "ebits" as they are called) over the quantum channel, prior to any "message" communications taking place. Then, when a classical message of n-bits needs to be communicated, it can be encoded in only $n/2$ qubits, sent over the quantum channel, and the measured jointly with some of the previously shared ebits already at the destination end of the channel in a such a ways as to re-constitute as the full classical message.

In fact, one could take a maximally compressed classical message, e.g., as given by a turbo code or low density parity check code, and then *further* compress this maximally compressed classical message into quantum message, at communications time, by an additional factor of two! As the result is, at communications time, a quantum message needing only half as many qubits as the (perhaps already maximally compressed) classical message, this trick is called "superdense coding" and is only possible using quantum information resources.

It is important to note that this scheme does not violate Shannon's Source Coding Theorem because it requires certain quantum states to be created, distributed, and stored across the quantum communications channel prior to any actual classical message being sent. When one takes account of the communication resources needed to distribute these shared prior states, and add it to the communications resources required to transmit the quantum-encoding of the classical message itself, the net efficiency is again identical to the Shannon bound. However, in many practical circumstances, it is possible to create, distribute, and store the ebits at leisure, so that an urgent classical message can be transmitted at double density at some critical communications time. That is the main benefit of superdense coding.

To understand how superdense coding works, we must first discuss the Bell states and how it is possible to interconvert between them by acting on only one end of a Bell state.

11.5.1 Bell States

The starting point for superdense coding is to begin with 2-qubit maximally entangled states such as the Bell states.

$$\begin{aligned}
|\beta_{00}\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \\
|\beta_{01}\rangle &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \\
|\beta_{10}\rangle &= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \\
|\beta_{11}\rangle &= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)
\end{aligned} \tag{11.129}$$

which can be summarized as:

$$|\beta_{xy}\rangle = \frac{1}{\sqrt{2}}(|0, y\rangle + (-1)^x|1, 1-y\rangle) \tag{11.130}$$

Each of these Bell states can be synthesized from a different starting computational basis state in a quantum circuit consisting of a single Walsh-Hadamard gate and a CNOT. Specifically, we have:

$$|\beta_{xy}\rangle = \text{CNOT} \cdot (H \otimes \mathbb{1})|xy\rangle \tag{11.131}$$

where x and y can each be 0 or 1.

For superdense coding Alice is going to create Bell states in this manner, store one member of each pair, and transmit the other member to Bob, which he also indexes and stores. Provided neither qubit is measured the entanglement between the qubits in each Bell state is preserved. This shared prior entanglement becomes the resource upon which we will draw to achieve superdense coding of a subsequent classical message.

11.5.2 Interconversion Between Bell States by Local Actions

The Bell states have the curious property that they can be converted into one another by performing single qubit operations on just one of the qubits in each Bell pair. Moreover, this capability persists even if the two qubits in a Bell state become separated spatially over an arbitrarily large distance, provided neither of them is measured during the separation process.

To see this, suppose Alice and Bob each hold one member of a Bell state. The single qubit operation Alice needs to perform on her qubit, in order to convert the joint state into some other Bell state are as follows:

$$\begin{aligned}
|\beta_{00}\rangle &\xrightarrow{\mathbb{1}\otimes\mathbb{1}} |\beta_{00}\rangle \\
|\beta_{00}\rangle &\xrightarrow{X\otimes\mathbb{1}} |\beta_{01}\rangle \\
|\beta_{00}\rangle &\xrightarrow{Z\otimes\mathbb{1}} |\beta_{10}\rangle \\
|\beta_{00}\rangle &\xrightarrow{Z\cdot X\otimes\mathbb{1}} |\beta_{11}\rangle
\end{aligned} \tag{11.132}$$

$$\begin{aligned}
|\beta_{01}\rangle &\xrightarrow{X\otimes\mathbb{1}} |\beta_{00}\rangle \\
|\beta_{01}\rangle &\xrightarrow{\mathbb{1}\otimes\mathbb{1}} |\beta_{01}\rangle \\
|\beta_{01}\rangle &\xrightarrow{Z\cdot X\otimes\mathbb{1}} |\beta_{10}\rangle \\
|\beta_{01}\rangle &\xrightarrow{Z\otimes\mathbb{1}} |\beta_{11}\rangle
\end{aligned} \tag{11.133}$$

$$\begin{aligned}
|\beta_{10}\rangle &\xrightarrow{Z\otimes\mathbb{1}} |\beta_{00}\rangle \\
|\beta_{10}\rangle &\xrightarrow{X\cdot Z\otimes\mathbb{1}} |\beta_{01}\rangle \\
|\beta_{10}\rangle &\xrightarrow{\mathbb{1}\otimes\mathbb{1}} |\beta_{10}\rangle \\
|\beta_{10}\rangle &\xrightarrow{Z\cdot X\cdot Z\otimes\mathbb{1}} |\beta_{11}\rangle
\end{aligned} \tag{11.134}$$

$$\begin{aligned}
|\beta_{11}\rangle &\xrightarrow{X\cdot Z\otimes\mathbb{1}} |\beta_{00}\rangle \\
|\beta_{11}\rangle &\xrightarrow{Z\otimes\mathbb{1}} |\beta_{01}\rangle \\
|\beta_{11}\rangle &\xrightarrow{Z\cdot X\cdot Z\otimes\mathbb{1}} |\beta_{10}\rangle \\
|\beta_{11}\rangle &\xrightarrow{\mathbb{1}\otimes\mathbb{1}} |\beta_{11}\rangle
\end{aligned} \tag{11.135}$$

11.5.3 Superdense Coding Protocol

We now have all the pieces needed to understand superdense coding. The protocol is surprisingly straight forward.

Superdense Coding Suppose Alice wishes to send Bob a classical message comprising a string of bits. If Alice and Bob have a quantum channel, and quantum memories available to them, they can halve the required number of communicative acts needed at the time the message is sent, but exploiting entanglement resources created, shared, and stored, at an earlier time. The superdense coding protocol works as follows:

1. Before any information-bearing message is communicated, Alice creates several pairs of entangled qubits (i.e., ebits), each in the state $|\beta_{00}\rangle$, indexes and stores one member of each pair and passes the other member of the same pair to Bob.
2. Upon receipt Bob indexes and stores each ebit he receives from Alice. The result is that Alice and Bob come to possess matching members of pairs of entangled qubits each in the state $|\beta_{00}\rangle$ stored at matching index locations in some quantum memory.
3. Subsequently, when Alice wants to send Bob a two bit classical message, presented as the quantum state $|x\rangle|y\rangle$, she performs one of four possible operations on the next indexed ebit in her possession. By acting on her end of an entangled pair of qubits, Alice is able to convert the joint state of the entangled pair into any of the four Bell states. In particular, if $|x\rangle|y\rangle = |0\rangle|0\rangle$ Alice applies $\mathbb{1}$ (the identity) to her ebit. If $|x\rangle|y\rangle = |0\rangle|1\rangle$ she applies X (the Pauli-X gate) to her ebit. If $|x\rangle|y\rangle = |1\rangle|0\rangle$ she applies Z (the Pauli-Z gate) to her ebit. Finally, if $|x\rangle|y\rangle = |1\rangle|1\rangle$ she applies $Z \cdot X$ to her ebit. These operations transform the entangled state (initially $|\beta_{00}\rangle$) shared between Alice and Bob as follows:

$$\begin{aligned}
|00\rangle|\beta_{00}\rangle &\overset{\mathbb{1}\otimes\mathbb{1}\otimes\mathbb{1}\otimes\mathbb{1}}{\longrightarrow} |00\rangle|\beta_{00}\rangle \\
|01\rangle|\beta_{00}\rangle &\overset{\mathbb{1}\otimes\mathbb{1}\otimes X\otimes\mathbb{1}}{\longrightarrow} |01\rangle|\beta_{01}\rangle \\
|10\rangle|\beta_{00}\rangle &\overset{\mathbb{1}\otimes\mathbb{1}\otimes Z\otimes\mathbb{1}}{\longrightarrow} |10\rangle|\beta_{10}\rangle \\
|11\rangle|\beta_{00}\rangle &\overset{\mathbb{1}\otimes\mathbb{1}\otimes Z\cdot X\otimes\mathbb{1}}{\longrightarrow} |11\rangle|\beta_{11}\rangle
\end{aligned} \tag{11.136}$$

4. Alice then sends her "treated" ebit to Bob over the quantum communications channel.
5. Upon receipt, Bob performs a joint Bell state analysis on the ebit he receives from Alice together with the correspondingly indexed ebit from his quantum memory.
6. The Bell state analysis allows Bob to determine unambiguously which Bell state he has ($|\beta_{00}\rangle$, $|\beta_{01}\rangle$, $|\beta_{10}\rangle$, or $|\beta_{11}\rangle$) and hence what bit value pair Alice intended to send. Thus, if Alice and Bob share prior entanglement, then to send a two-bit message subsequently, Alice need only send a single "treated" ebit to Bob.

It is important to appreciate that superdense coding does not result in a *net* reduction in the communications resources needed to transmit n classical bits. However, it does allow us to time-shift when channel capacity is available. In essence, superdense coding can use an under-utilized channel at one time to share and store successive members of EPR pairs so that, at a later time, a classical n bit message

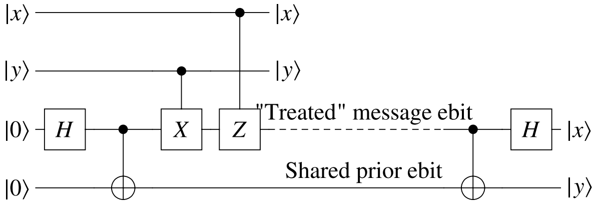

Fig. 11.9 Quantum circuit for superdense coding. Alice (using qubits 3 and 4 in the figure) prepares maximally entangled pairs of qubits (called “ebits”), keeps one member of each pair, and passes the other to Bob (qubit 4 in the figure). Subsequently, if Alice wants to send the bits xy encoded in the quantum state $|x\rangle|y\rangle$, she performs conditional operations on her retained ebit. This causes the entangled state shared between Alice and Bob to be set in the Bell state $|\beta_{xy}\rangle$. Next Alice transmits her “treated” ebit to Bob (qubit 3 in the figure). Upon receipt, Bob performs a complete Bell state analysis which allows him to determine which Bell state qubits 3 and 4 are in. This tells him what bit values Alive had intended to send. Thus, at communications time, Alice need only send one qubit to achieve the effect of sending two classical bits. Overall, superdense coding does not do any better than classical communications because of the communicative acts needed to establish the shared prior entanglement. Nevertheless, it does allow channel capacity available at one time to effectively be time-shifted to a later time

can be transmitted using the transmission of only $n/2$ qubits, consuming one EPR pair per qubit transmitted. Thus, the extra factor of two compression of the classical message can only be achieved for as long at the supply of EPR entangled particles lasts. However, this added factor of two additional compression is even possible if the classical message has already been maximally compressed (classically) using a turbo code or low density parity check code.

A quantum circuit for superdense coding is shown in Fig. 11.9

11.6 Cloning Quantum Information

One of the most useful aspects of classical information is our ability to copy, or “clone”, it reliably without any noticeable error. A photocopier, for example, can reproduce sheets of papers that are almost indistinguishable from the original. Digital computer files can be copied with even higher fidelity, in fact, perfectly. The ability to make perfect copies of classical data is also the curse of the entertainment and software industries because it also allows bootleggers to make illicit copies of digital music files, movies, and computer programs. As quantum information applications become more widespread it therefore behooves us to understand what can and cannot be done in terms of copying quantum information.

11.6.1 Historical Roots and Importance of Quantum Cloning

“*I was the referee who approved the publication of Nick Herbert’s FLASH paper, knowing perfectly well that it was wrong. I explain why my decision was the correct one*, [...]”

– Asher Peres [389]

The roots of quantum cloning can be traced back to a controversial paper written by Nick Herbert in 1981 describing an idea for a superluminal communicator based on the presumption that it is possible to make perfect copies (or clones) of an unknown quantum state. In 2002 Asher Peres revealed that he and Gian Carlo Ghirardi had been the "anonymous" reviewers of Herbert's FLASH paper and that Ghirardi had recommended its rejection on the grounds that the linear nature of quantum mechanics meant that the supposed copying process could not exist. Peres likewise realized the paper was flawed but nevertheless recommended its publication in the hopes of stimulating others to find the flaw and thereby draw more attention to the emerging field of quantum information theory.

It turned out Peres was correct. Soon after Herbert's paper was published William Wootters and Wojciech Zurek published a paper in Nature entitled "A Single Quantum Cannot be Cloned", which basically re-discovered Ghirardi's argument opposing Herbert's paper [547]. Around the same time Dennis Dieks published a paper arguing that the claims of superluminal communications in Herbert's paper were also flawed [142]. Thus the publication of the FLASH paper, and the reaction to it, went a long way towards stimulating more careful analyses of the properties of quantum information.

Recently, a more pragmatic motivation for studying quantum cloning has arisen from the need to understand how well an unscrupulous eavesdropper might be able to tap a quantum communications channel, whilst remaining undetected. If exact deterministic quantum cloning of unknown quantum states were possible (which luckily it isn't), then an eavesdropper would be able to tap a quantum channel, forward perfect copies of the qubits to the intended recipient, and examine the copies they made at leisure. Fortunately, as we will show below, such exact deterministic quantum copying is physically impossible. Nevertheless, the practical question is how well can an eavesdropper do? How much information from a quantum channel can they extract without their presence being detected? With what fidelity *can* they copy unknown quantum states? And if they cannot copy states deterministically, can they do so probabilistically? These questions and others demonstrate the practical need to understand what physics permits one to do in terms of cloning quantum information.

11.6.2 Impossibility of Exact Deterministic Quantum Cloning

As in the classical case, an ideal universal quantum copy machine, or ideal universal quantum "cloning" machine as it is sometimes called, would be able to make a *perfect* copy of *any* quantum state it was handed. In particular, the action of an ideal universal quantum cloning machine, U_{clone}, on an arbitrary pure state $|\psi\rangle$ would be described as:

$$|\psi\rangle_A|0\rangle_B \xrightarrow{U_{\text{clone}}} |\psi\rangle_A|\psi\rangle_B \tag{11.137}$$

which we read as "particle A starts off in state $|\psi\rangle$, and particle B starts off in state $|0\rangle$, and after cloning the state of particle A, i.e., $|\psi\rangle$, is replicated on particle B." This is more clearly seen to be a cloning procedure by suppressing the particle labels as in $|\psi\rangle|0\rangle \overset{U_{\text{clone}}}{\longrightarrow} |\psi\rangle|\psi\rangle$.

Likewise, the ideal behavior of a quantum cloner when handed an arbitrary mixed state, ρ, would be:

$$\rho_A \otimes |0\rangle_B\langle 0|_B \overset{U_{\text{clone}}}{\longrightarrow} \rho_A \otimes \rho_B \tag{11.138}$$

which we read as "particle A starts off in state ρ, and particle B starts off in state $|0\rangle\langle 0|$, and after cloning the state of particle A, i.e., ρ, is replicated on particle B." As above, this is more clearly seen to be a cloning procedure by suppressing the particle labels as in $\rho \otimes |0\rangle\langle 0| \overset{U_{\text{clone}}}{\longrightarrow} \rho \otimes \rho$.

The question is, does Nature permit such an ideal exact deterministic quantum cloning operation? To proceed, let us assume that U_{clone} *is* a perfect quantum cloning machine, i.e., a unitary operation such that whatever quantum state is given as input, two perfect copies of it are returned after U_{clone} has acted. In particular, U_{clone} will clone (say) the computational basis states perfectly. Thus, we would have:

$$\begin{aligned} |0\rangle|0\rangle &\overset{U_{\text{clone}}}{\longrightarrow} |0\rangle|0\rangle \\ |1\rangle|0\rangle &\overset{U_{\text{clone}}}{\longrightarrow} |1\rangle|1\rangle \end{aligned} \tag{11.139}$$

So far so good. But now let's assume the same machine was handed the states $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ instead, which are rotated with respect to the computational basis states. In this case, a *proper* quantum cloning machine is required to act as follows:

$$\begin{aligned} \tfrac{1}{\sqrt{2}}(|0\rangle + |1\rangle)|0\rangle &\overset{U_{\text{clone}}}{\longrightarrow} \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)|\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \\ \tfrac{1}{\sqrt{2}}(|0\rangle - |1\rangle)|0\rangle &\overset{U_{\text{clone}}}{\longrightarrow} \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \end{aligned} \tag{11.140}$$

But this not what our *supposed* quantum cloning machine U_{clone} does! If U_{clone} clones the computational basis states $(\{|0\rangle, |1\rangle\})$ correctly then, *by the linearity of quantum mechanics*, U_{clone} will transform the input states $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ as follows:

$$\begin{aligned} \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)|0\rangle &\overset{U_{\text{clone}}}{\longrightarrow} \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \\ \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)|0\rangle &\overset{U_{\text{clone}}}{\longrightarrow} \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \end{aligned} \tag{11.141}$$

In neither case is the output a product state of clones of the input state. Hence, if U_{clone} is a unitary procedure that clones computational basis states perfectly, then it is guaranteed to clone states non-orthogonal to these imperfectly, and vice versa. This echoes the argument Ghirardi and Wootters and Zurek found against the FLASH paper. Hence, U_{clone} cannot be an ideal universal quantum cloning machine as we had supposed, and in fact the foregoing argument proves that ideal universal quantum cloning is *physically impossible* using any unitary operation whatsoever! Thus we arrive at the so-called "no-cloning" theorem for quantum information.

No Cloning Theorem *There is no deterministic quantum procedure by which an unknown pure quantum state can be cloned exactly.*

11.6.3 Universal Approximate Quantum Cloning

Although the quantum no-cloning theorem proves that it is impossible to clone an unknown quantum state perfectly deterministically it leaves open the possibility of cloning an unknown quantum state *approximately* deterministically, or perfectly *non-deterministically*. We will consider approximate deterministic cloning first.

If we are able to make an approximate clone, our main concerns are going to be how good an approximation can we obtain; whether the quality of the approximation can be made independent of the state we are trying to clone; and whether the resulting approximate clones can be used freely in subsequent quantum computations as proxies for the state that was cloned. The latter concern arises because if the approximate clones are entangled, then is may not matter how good they are individually, because using one of them could mess up the other one. This last point is often neglected but is, in fact, crucial to the whom concept of the utility of the clones.

These concerns were well appreciated by Vladimir Bužek and Mark Hillery. In 1996 they devised the first quantum cloning machine that produced high quality clones, whose fidelities were input independent, and which were practical to use in lieu of the original state in subsequent quantum computations [93]. Their elegant scheme for cloning a single qubit can be described as follows.

Imagine a 3-qubit quantum memory register with the qubits labeled A, B, and C. Qubit A is to hold the qubit whose state we wish to clone, and the outputs of qubits B and C are to hold the approximate clones. The quantum cloning machine will be unitary operation, $\widetilde{U}_{\text{clone}}$, able to perform at least the following transformation on the computational basis states of qubit A, i.e., $|0\rangle_A$ and $|1\rangle_A$, augmented with a pair of ancillae prepared in the state $|00\rangle_{BC}$:

$$\begin{aligned}
|0\rangle_A|0\rangle_B|0\rangle_C &\overset{\widetilde{U}_{\text{clone}}}{\longrightarrow} \sqrt{\frac{2}{3}}|000\rangle_{ABC} + \frac{1}{\sqrt{3}}|1\rangle_A\left[\frac{1}{\sqrt{2}}(|01\rangle_{BC} + |10\rangle_{BC})\right] \\
|1\rangle_A|0\rangle_B|0\rangle_C &\overset{\widetilde{U}_{\text{clone}}}{\longrightarrow} \sqrt{\frac{2}{3}}|111\rangle_{ABC} + \frac{1}{\sqrt{3}}|0\rangle_A\left[\frac{1}{\sqrt{2}}(|01\rangle_{BC} + |10\rangle_{BC})\right]
\end{aligned} \tag{11.142}$$

Now imagine what the approximate quantum cloning transformation, $\widetilde{U}_{\text{clone}}$, does to an arbitrary superposition state on qubit A, i.e., $|\psi\rangle_A = \alpha|0\rangle_A + \beta|1\rangle_A$.

A little algebra shows that $\widetilde{U}_{\text{clone}}$ will transform a general superposition of qubit A as:

$$\begin{aligned}|\Psi_{ABC}\rangle &= \widetilde{U}_{\text{clone}}(\alpha|0\rangle_A + \beta|1\rangle_A)|0\rangle_B|0\rangle_C \\ &= \sqrt{\frac{2}{3}}\alpha|000\rangle + \frac{\beta}{\sqrt{6}}|001\rangle + \frac{\beta}{\sqrt{6}}|010\rangle + \frac{\alpha}{\sqrt{6}}|101\rangle \\ &\quad + \frac{\alpha}{\sqrt{6}}|110\rangle + \sqrt{\frac{2}{3}}\beta|111\rangle \end{aligned} \tag{11.143}$$

where we have dropped the qubit labels in the final output state. We can write $|\Psi_{ABC}\rangle$ equivalently as the density operator:

$$\begin{aligned}\rho_{ABC} &= |\Psi_{ABC}\rangle\langle\Psi_{ABC}| \\ &= \begin{pmatrix} \frac{2}{3}|\alpha|^2 & \frac{1}{3}\alpha\beta^* & \frac{1}{3}\alpha\beta^* & 0 & 0 & \frac{1}{3}|\alpha|^2 & \frac{1}{3}|\alpha|^2 & \frac{2}{3}\alpha\beta^* \\ \frac{1}{3}\beta\alpha^* & \frac{1}{6}|\beta|^2 & \frac{1}{6}|\beta|^2 & 0 & 0 & \frac{1}{6}\beta\alpha^* & \frac{1}{6}\beta\alpha^* & \frac{1}{3}|\beta|^2 \\ \frac{1}{3}\beta\alpha^* & \frac{1}{6}|\beta|^2 & \frac{1}{6}|\beta|^2 & 0 & 0 & \frac{1}{6}\beta\alpha^* & \frac{1}{6}\beta\alpha^* & \frac{1}{3}|\beta|^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3}|\alpha|^2 & \frac{1}{6}\alpha\beta^* & \frac{1}{6}\alpha\beta^* & 0 & 0 & \frac{1}{6}|\alpha|^2 & \frac{1}{6}|\alpha|^2 & \frac{1}{3}\alpha\beta^* \\ \frac{1}{3}|\alpha|^2 & \frac{1}{6}\alpha\beta^* & \frac{1}{6}\alpha\beta^* & 0 & 0 & \frac{1}{6}|\alpha|^2 & \frac{1}{6}|\alpha|^2 & \frac{1}{3}\alpha\beta^* \\ \frac{2}{3}\beta\alpha^* & \frac{1}{3}|\beta|^2 & \frac{1}{3}|\beta|^2 & 0 & 0 & \frac{1}{3}\beta\alpha^* & \frac{1}{3}\beta\alpha^* & \frac{2}{3}|\beta|^2 \end{pmatrix}\end{aligned} \tag{11.144}$$

This density operator ρ_{ABC} is therefore the output from our quantum cloning machine.

Next we determine the state of the clones individually by tracing out the unwanted qubits to obtain:

$$\rho_A = \text{tr}_{BC}(\rho_{ABC}) = \begin{pmatrix} \frac{2}{3}|\alpha|^2 + \frac{1}{3}|\beta|^2 & \frac{1}{3}\beta\alpha^* \\ \frac{1}{3}\alpha\beta^* & \frac{1}{3}|\alpha|^2 + \frac{2}{3}|\beta|^2 \end{pmatrix} \tag{11.145}$$

$$\begin{aligned}\rho_B = \text{tr}_{AC}(\rho_{ABC}) &= \begin{pmatrix} \frac{5}{6}|\alpha|^2 + \frac{1}{6}|\beta|^2 & \frac{2}{3}\alpha\beta^* \\ \frac{2}{3}\beta\alpha^* & \frac{1}{6}|\alpha|^2 + \frac{5}{6}|\beta|^2 \end{pmatrix} \\ &= \frac{5}{6}|\psi\rangle\langle\psi| + \frac{1}{6}|\psi^{\perp}\rangle\langle\psi^{\perp}| \end{aligned} \tag{11.146}$$

$$\begin{aligned}\rho_C = \text{tr}_{AB}(\rho_{ABC}) &= \begin{pmatrix} \frac{5}{6}|\alpha|^2 + \frac{1}{6}|\beta|^2 & \frac{2}{3}\alpha\beta^* \\ \frac{2}{3}\beta\alpha^* & \frac{1}{6}|\alpha|^2 + \frac{5}{6}|\beta|^2 \end{pmatrix} \\ &= \frac{5}{6}|\psi\rangle\langle\psi| + \frac{1}{6}|\psi^{\perp}\rangle\langle\psi^{\perp}| \end{aligned} \tag{11.147}$$

where $|\psi^\perp\rangle = \alpha^*|1\rangle - \beta^*|0\rangle$ is a state orthogonal to $|\psi\rangle$, which is the antipodal point to $|\psi\rangle$ on the Bloch sphere. Thus, we see that the reduced density operators for the clones contain the state being cloned plus some extra stuff we did not want.

To assess how close the clones are to the original state, we compute the fidelity of the clones, i.e., $\rho_B = \rho_C = \frac{5}{6}|\psi\rangle\langle\psi| + \frac{1}{6}|\psi^\perp\rangle\langle\psi^\perp|$, with respect to the original state, i.e., $\rho_{\text{ideal}} = |\psi\rangle\langle\psi|$. The formula for the fidelity with which one density operator, ρ, approximates another, σ, was given in Sect. 11.2.2.4 as:

$$\mathcal{F}(\rho,\sigma) = \left[\text{tr}\left(\sqrt{\sqrt{\rho}\,\sigma\sqrt{\rho}}\right)\right]^2 \tag{11.148}$$

Plugging the relevant density operators into this formula for fidelity we have:

$$\begin{aligned}\rho_{\text{ideal}} &= |\psi\rangle\langle\psi| = (\alpha|0\rangle + \beta|1\rangle)(\alpha^*\langle 0| + \beta^*\langle 1|)\\ &= \begin{pmatrix} |\alpha|^2 & \alpha\beta^* \\ \beta\alpha^* & |\beta|^2 \end{pmatrix}\end{aligned} \tag{11.149}$$

and

$$\rho_B = \rho_C = \begin{pmatrix} \frac{5}{6}|\alpha|^2 + \frac{1}{6}|\beta|^2 & \frac{2}{3}\alpha\beta^* \\ \frac{2}{3}\beta\alpha^* & \frac{1}{6}|\alpha|^2 + \frac{5}{6}|\beta|^2 \end{pmatrix} \tag{11.150}$$

which gives $\mathcal{F}(\rho_{\text{ideal}}, \rho_B) = \mathcal{F}(\rho_{\text{ideal}}, \rho_C)$ as:

$$\begin{aligned}\mathcal{F}(\rho_{\text{ideal}}, \rho_B) &= \left[\text{tr}(\sqrt{\rho_{\text{ideal}}}\cdot\rho_B\cdot\sqrt{\rho_{\text{ideal}}})\right]^2\\ &= \left[\text{tr}(\rho_{\text{ideal}}\cdot\rho_B\cdot\rho_{\text{ideal}})\right]^2\\ &= \left[\text{tr}\left(\begin{pmatrix} |\alpha|^2 & \alpha\beta^* \\ \beta\alpha^* & |\beta|^2 \end{pmatrix}\cdot\begin{pmatrix} \frac{5}{6}|\alpha|^2 + \frac{1}{6}|\beta|^2 & \frac{2}{3}\alpha\beta^* \\ \frac{2}{3}\beta\alpha^* & \frac{1}{6}|\alpha|^2 + \frac{5}{6}|\beta|^2 \end{pmatrix}\right.\right.\\ &\qquad \left.\left.\cdot\begin{pmatrix} |\alpha|^2 & \alpha\beta^* \\ \beta\alpha^* & |\beta|^2 \end{pmatrix}\right)\right]^2\\ &= \frac{5}{6}\end{aligned} \tag{11.151}$$

where we used the fact that, as ρ_{ideal} is pure, $\sqrt{\rho_{\text{ideal}}} = \rho_{\text{ideal}}$. The same result holds for the second clone ρ_C. In both cases the fidelity with which quantum cloning is achieved is $\frac{5}{6}$. Moreover, as the fidelity we obtain is a numerical constant and does involve α or β it must, therefore, be *independent* of the input state being cloned. So our cloning transform is a state independent cloner. However, whereas the original state was pure, the clones are mixed. We can understand how these two states are related in terms of the Bloch sphere/Bloch ball picture of a qubit (see Sect. 11.2.2.3). The pure state $|\psi\rangle$ is represented by a point on the surface of the Bloch sphere. If you imagine a vector drawn from center of the Bloch sphere to the point representing $|\psi\rangle$, then the clone is the mixed state obtained by shrinking the length of this vector radially without changing its direction. This may help you to visualize the physical meaning of an approximate quantum clone.

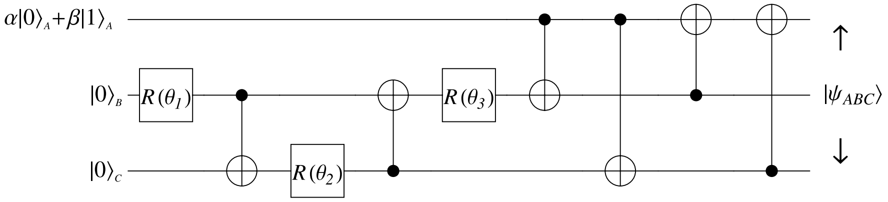

Fig. 11.10 Quantum circuit for cloning an unknown quantum state $|\psi\rangle_A = \alpha|0\rangle_A + \beta|1\rangle_A$. The clones appear in the output on qubits B and C. Their states are given by tracing out the other two qubits. That is, $\rho_B = \text{tr}_{AC}(|\psi_{ABC}\rangle\langle\psi_{ABC}|)$, and $\rho_C = \text{tr}_{AB}(|\psi_{ABC}\rangle\langle\psi_{ABC}|)$. Note that $\rho_B = \rho_C = \frac{5}{6}|\psi_A\rangle\langle\psi_A| + \frac{1}{6}|\psi_A^\perp\rangle\langle\psi_A^\perp|$, showing that the fidelity of the copies with respect to the original state is $\frac{5}{6}$

11.6.4 Circuit for Quantum Cloning

A quantum circuit that accomplishes our desired cloning transformation $|\psi\rangle_A|0\rangle_B|0\rangle_C \overset{\widetilde{U}_{\text{clone}}}{\longrightarrow} |\Psi_{ABC}\rangle$ to shown in Fig. 11.10. Here the 1-qubit gate $R(\theta)$ is defined to be:

$$R(\theta) := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{11.152}$$

and the particular angles used are set at:

$$\begin{aligned} \theta_1 &= \frac{\pi}{8} \\ \theta_2 &= -\arcsin\sqrt{\frac{1}{6}(3-2\sqrt{2})} \\ \theta_3 &= \frac{\pi}{8} \end{aligned} \tag{11.153}$$

With these angle values, the quantum cloning circuit induces a (fixed) unitary transformation described by the matrix:

$$\begin{pmatrix}
\sqrt{\frac{2}{3}} & 0 & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & 0 & \sqrt{\frac{2}{3}} \\
0 & 0 & 0 & 0 & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \sqrt{\frac{2}{3}} & 0 \\
0 & \sqrt{\frac{2}{3}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \sqrt{\frac{2}{3}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\
\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \sqrt{\frac{2}{3}} & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & 0 & \sqrt{\frac{2}{3}} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \sqrt{\frac{2}{3}} & 0 & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}}
\end{pmatrix} \tag{11.154}$$

As a check, it is easy to verify that this circuit transforms the basis states as follows:

$$\begin{aligned}\widetilde{U}_{\text{clone}}|000\rangle &= \sqrt{\frac{2}{3}}|000\rangle + \frac{1}{\sqrt{6}}|101\rangle + \frac{1}{\sqrt{6}}|110\rangle \\ \widetilde{U}_{\text{clone}}|100\rangle &= \frac{1}{\sqrt{6}}|001\rangle + \frac{1}{\sqrt{6}}|010\rangle + \sqrt{\frac{2}{3}}|111\rangle\end{aligned} \tag{11.155}$$

and hence transforms a superposition state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ as:

$$\begin{aligned}\widetilde{U}_{\text{clone}}|\psi\rangle|0\rangle|0\rangle = \sqrt{\frac{2}{3}}\alpha|000\rangle + \frac{\beta}{\sqrt{6}}|001\rangle + \frac{\beta}{\sqrt{6}}|010\rangle + \frac{\alpha}{\sqrt{6}}|101\rangle \\ + \frac{\alpha}{\sqrt{6}}|110\rangle + \sqrt{\frac{2}{3}}\beta|111\rangle\end{aligned} \tag{11.156}$$

which is exactly what is called for in (11.143).

11.6.5 Usability of the Quantum Clones

In an ideal universal cloning machine, the output clones would be *perfect* copies of the unknown state $|\psi\rangle$, and they would be unentangled from each other and the top qubit in the cloning circuit shown in Fig. 11.10. If these conditions hold, then the clones would clearly be useful as they could serve as perfect proxies for the state $|\psi\rangle$ in subsequent quantum computations. Unfortunately, the clones we obtain are neither perfect copies of the original state nor are they unentangled from each other and the top qubit of the cloning circuit. It is not immediately clear, therefore that cloning has achieved anything practically useful because, if the clones are entangled, operations performed on one of them might mess up the other. Furthermore, is a fidelity of $\frac{5}{6}$ really high enough to allow us to compute expectation values of observables that will be close enough to the true values to be useful? We will now address these issues by showing that although the clones are indeed entangled, they are nevertheless usable in subsequent quantum computations.

In an *ideal* cloning machine, an input state of the form $|\psi\rangle_A|0\rangle_B|0\rangle_C$ to be mapped into an output state of the form $|?\rangle_A|\psi\rangle_B|\psi\rangle_C$. Here perfect clones appear on qubits B and C, and they are unentangled from each other and from qubit A. Alas, we know from the no-cloning theorem, that quantum mechanics does not allow such perfection. Nevertheless, we can produce approximate clones on qubits B and C but these are no longer guaranteed to be unentangled form each or unentangled from qubit A. If they are entangled then, potentially, subsequent operations on one clone could perturb the other clone (not to mention the ancilla). So we need to understand whether or not the clones are entangled.

11.6.5.1 Are the Clones Entangled?

First let us determine whether or not the clones are entangled with each other. That is, we test whether the joint density operator of the clones, ρ_{BC}, is separable or inseparable. To test this, we can use the Peres-Horodecki criterion of Sect. 11.3.3. As you will recall this test is based on checking whether there is at least one negative eigenvalue in the partial transpose of the density operator whose entanglement status is sought—in our case ρ_{BC}.

Starting with $\rho_{ABC} = |\Psi_{ABC}\rangle\langle\Psi_{ABC}|$ we obtain ρ_{BC} by tracing over qubit A (the top qubit in the circuit shown in Fig. 11.10), to obtain:

$$\rho_{BC} = \mathrm{tr}_A(\rho_{ABC}) = \begin{pmatrix} \frac{2}{3}|\alpha|^2 & \frac{1}{3}\alpha\beta^* & \frac{1}{3}\alpha\beta^* & 0 \\ \frac{1}{3}\beta\alpha^* & \frac{1}{6} & \frac{1}{6} & \frac{1}{3}\alpha\beta^* \\ \frac{1}{3}\beta\alpha^* & \frac{1}{6} & \frac{1}{6} & \frac{1}{3}\alpha\beta^* \\ 0 & \frac{1}{3}\beta\alpha^* & \frac{1}{3}\beta\alpha^* & \frac{2}{3}|\beta|^2 \end{pmatrix} \tag{11.157}$$

Then, we compute the partial transpose of ρ_{BC} taken over the space "B" i.e., the space corresponding to the first of the two qubits in ρ_{BC}. This gives

$$\rho_{BC}^{T_B} = \begin{pmatrix} \frac{2}{3}|\alpha|^2 & \frac{1}{3}\alpha\beta^* & \frac{1}{3}\beta\alpha^* & \frac{1}{6} \\ \frac{1}{3}\beta\alpha^* & \frac{1}{6} & 0 & \frac{1}{3}\beta\alpha^* \\ \frac{1}{3}\alpha\beta^* & 0 & \frac{1}{6} & \frac{1}{3}\alpha\beta^* \\ \frac{1}{6} & \frac{1}{3}\alpha\beta^* & \frac{1}{3}\beta\alpha^* & \frac{2}{3}|\beta|^2 \end{pmatrix} \tag{11.158}$$

The eigenvalues of the partial transpose $\rho_{BC}^{T_B}$ can be obtained from the characteristic polynomial[3] of the partial transpose $\rho_{BC}^{T_B}$, i.e., as the roots of:

$$\det(\rho_{BC}^{T_B} - \lambda\mathbb{1}) = \frac{(6\lambda-1)^2(36\lambda^2-24\lambda-1)}{1296} = 0 \tag{11.159}$$

Amazingly, after simplifying $\det(\rho_{BC}^{T_B} - \lambda\mathbb{1})$ by using the fact that $|\alpha|^2 + |\beta|^2 = 1$ and $|\alpha| \leq 1$, the resulting characteristic polynomial does not contain any mention of α and β! This means that the eigenvalues of $\rho_{BC}^{T_B}$ are independent of the state being cloned, and are in fact equal to $\frac{1}{6}$, $\frac{1}{6}$, $\frac{1}{6}(2-\sqrt{5})$, and $\frac{1}{6}(2+\sqrt{5})$. As $\sqrt{5} > 2$, we see that the third eigenvalue is assuredly negative. Thus, by the Peres-Horodecki criterion ρ_{BC}, which is the joint state of the clones, must be an entangled.[4] Rats!

[3] The characteristic polynomial of a square matrix U is the left hand side of the equation $\det(U - \lambda\mathbb{1}) = 0$ where $\mathbb{1}$ is the identity matrix. The roots of the characteristic polynomial are the eigenvalues of the matrix U.

[4] N.B. If we had computed, instead, the partial transpose over the space "C" i.e., the space corresponding to the second of the two qubits in ρ_{BC}, we would have obtained a different matrix for the partial transpose, $\rho_{BC}^{T_C}$, but its eigenvalues would have been the same as those of $\rho_{BC}^{T_B}$, and therefore one would have still been negative.

11.6.5.2 How Entangled are the Clones?

Just how entangled are the clones? To quantify the degree to which the clones are entangled we can compute the tangle of ρ_{BC}. Tangle, as a measure of entanglement for pure states, was introduced in Sect. 2.8.1. However, it generalizes readily to the case of mixed states. Define $\widetilde{\rho}$ as the "spin-flipped" version of a density operator ρ:

$$\widetilde{\rho} = (Y \otimes Y) \cdot \rho \cdot (Y \otimes Y) \tag{11.160}$$

then the tangle of ρ, tangle(ρ), is related to the eigenvalues of the operator $\rho \cdot \widetilde{\rho}$. Specifically, if the four eigenvalues of $\rho \cdot \widetilde{\rho}$ are arranged in decreasing order so that $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$, then:

$$\text{tangle}(\rho) = \left[\max(\sqrt{\lambda_1} - \sqrt{\lambda_2} - \sqrt{\lambda_3} - \sqrt{\lambda_4}, 0)\right]^2 \tag{11.161}$$

For the density operator, ρ_{BC}, the spin-flipped version is:

$$\begin{aligned}\widetilde{\rho_{BC}} &= (Y \otimes Y) \cdot \rho_{BC} \cdot (Y \otimes Y) \\ &= \begin{pmatrix} \frac{2}{3}|\beta|^2 & -\frac{1}{3}\alpha\beta^* & -\frac{1}{3}\alpha\beta^* & 0 \\ -\frac{1}{3}\beta\alpha^* & \frac{1}{6} & \frac{1}{6} & -\frac{1}{3}\alpha\beta^* \\ -\frac{1}{3}\beta\alpha^* & \frac{1}{6} & \frac{1}{6} & -\frac{1}{3}\alpha\beta^* \\ 0 & -\frac{1}{3}\beta\alpha^* & -\frac{1}{3}\beta\alpha^* & \frac{2}{3}|\alpha|^2 \end{pmatrix}\end{aligned} \tag{11.162}$$

and so the eigenvalues of $\rho_{BC} \cdot \widetilde{\rho_{BC}}$ are the roots of the corresponding characteristic polynomial:

$$\det(\rho_{BC} \cdot \widetilde{\rho_{BC}} - \lambda \mathbb{1}) = \lambda^3 \left(\lambda - \frac{1}{9}\right) = 0 \tag{11.163}$$

Amazingly again, after simplifying $\det(\rho_{BC} \cdot \widetilde{\rho_{BC}} - \lambda \mathbb{1})$ by using the fact that $|\alpha|^2 + |\beta|^2 = 1$ and $|\alpha| \leq 1$, the resulting characteristic polynomial does not contain any mention of α and β. This means that the eigenvalues of $\rho_{BC} \cdot \widetilde{\rho_{BC}}$ are independent of the state being cloned, and are in fact equal to 0, 0, 0, and $\frac{1}{9}$. Thus, arranging the eigenvalues in decreasing order so that $\lambda_1 = \frac{1}{9}$, $\lambda_2 = \lambda_3 = \lambda_4 = 0$, and taking square roots, the tangle is then given by:

$$\begin{aligned}\text{tangle}(\rho_{BC}) &= \left[\max(\sqrt{\lambda_1} - \sqrt{\lambda_2} - \sqrt{\lambda_3} - \sqrt{\lambda_4}, 0)\right]^2 \\ &= \left[\max\left(\sqrt{\frac{1}{9}} - \sqrt{0} - \sqrt{0} - \sqrt{0}, 0\right)\right]^2 = \frac{1}{9}\end{aligned} \tag{11.164}$$

This is actually not that bad. A maximally entangled 2-qubit state has a tangle of 1, so tangle$(\rho_{BC}) = \frac{1}{9}$ is fairly small. So the clones are far from being maximally entangled. However, the fact that the clones are entangled at all could spell trouble because when one uses one of the clones, the operations performed on it, could change the other clone. Hence, we might wonder whether we can use the two clones freely in subsequent quantum computations. Furthermore, the fidelity of the clones, $\frac{5}{6}$ is noticeably less than 1. Is this good enough to learn anything trustworthy about ρ_{ideal} by subsequent observations on the clones ρ_B and ρ_C? This issues are resolved in the next two sections.

11.6.5.3 Expectation Value of an Observable Based on Ideal State

To assess how useful the clones really are, we need to examine how the expectation value of a general operator Ω, when the system is in a clone state ρ_B or ρ_C, differs from the expectation value of the same operator when the system is in the original state $\rho_{\text{ideal}} = |\psi\rangle\langle\psi|$.

The ideal state is just the original state $|\psi\rangle$ we are trying to clone. Thus we have:

$$\rho_{\text{ideal}} = |\psi\rangle\langle\psi| = \begin{pmatrix} |\alpha|^2 & \alpha\beta^* \\ \beta\alpha^* & |\beta|^2 \end{pmatrix} \tag{11.165}$$

Without loss of generality, the general form for an arbitrary 1-qubit observable operator, Ω, can be defined symbolically as:

$$\Omega = \begin{pmatrix} p & z \\ z^* & q \end{pmatrix} \tag{11.166}$$

where p and q are *real* numbers and z is (in general) a *complex* number. Any 1-qubit observable operator has to adhere to this form to be hermitian.

Now we can compute the expectation value of the observable Ω when the system is in state ρ_{ideal}. Using the formula given in Table 11.1 for computing the expectation value of an observable of a state defined by a density operator we have:

$$\langle\Omega\rangle = \text{tr}(\rho_{\text{ideal}}\Omega) = (p\alpha + z\beta)\alpha^* + (q\beta + \alpha z^*)\beta^* \tag{11.167}$$

This result is therefore our "gold standard" against which the quality of our clones can be judged.

11.6.5.4 Expectation Value of an Observable Based on a Clone

Now let us re-derive the expectation value $\langle\Omega\rangle$ this time using our clones. We want to tell two things. First, given that the clone is imperfect, what is the relationship between an operator expectation value for a clone state compared to that of the ideal state? Second, given that the clones are entangled, can we still use both clones in determining expectation values or does the use of one of them, render the other useless?

The state of each single clone is given by tracing over the other two qubits in the output state $|\Psi_{ABC}\rangle\langle\Psi_{ABC}|$. We calculated the reduce density matrices of the clones in (11.146) and (11.147). We found that:

$$\begin{aligned}\rho_B = \text{tr}_{AC}(\rho_{ABC}) &= \begin{pmatrix} \frac{5}{6}|\alpha|^2 + \frac{1}{6}|\beta|^2 & \frac{2}{3}\alpha\beta^* \\ \frac{2}{3}\beta\alpha^* & \frac{1}{6}|\alpha|^2 + \frac{5}{6}|\beta|^2 \end{pmatrix} \\ \rho_C = \text{tr}_{AB}(\rho_{ABC}) &= \begin{pmatrix} \frac{5}{6}|\alpha|^2 + \frac{1}{6}|\beta|^2 & \frac{2}{3}\alpha\beta^* \\ \frac{2}{3}\beta\alpha^* & \frac{1}{6}|\alpha|^2 + \frac{5}{6}|\beta|^2 \end{pmatrix}\end{aligned} \tag{11.168}$$

These reduced density matrices for the clones are telling. These are the states we will appear to have regardless of what happens to the other clone. So provided we can milk some useful information out of ρ_B and ρ_C we do not need to worry further about the fact that the clones are actually entangled. So can we extract useful information?

Well surprisingly, although the clones ρ_B and ρ_C are only *approximations* to the ideal state ρ_{ideal} we can, in principle, use them to obtain the *exact* expectation values for any operator, Ω! This is remarkable. The trick is to write "1" in the form "$|\alpha|^2 + |\beta|^2$" to see that the following identity holds:

$$\rho_B = \rho_C = \frac{2}{3}\rho_{\text{ideal}} + \left(\frac{|\alpha|^2}{6} + \frac{|\beta|^2}{6}\right)\mathbb{1} \tag{11.169}$$

where $\mathbb{1}$ is the identity matrix. It then follows that:

$$\langle\Omega\rangle = \text{tr}(\rho_{\text{ideal}} \cdot \Omega) = \frac{3}{2}\left(\text{tr}(\rho_B \cdot \Omega) - \frac{1}{6}\text{tr}(\Omega)\right) \tag{11.170}$$

Thus we can use the clones to obtain the exact value of any observable, even though they are only approximations to the ideal clone, and even though they are entangled. I find this really a most amazing result!

11.6.6 Universal Probabilistic Quantum Cloning

Recall that the no-cloning theorem proves the impossibility of cloning an unknown state exactly deterministically. Yet it does not preclude the possibility of cloning an unknown state *approximately* deterministically, or cloning one exactly *non-deterministically*. In the preceding sections we showed that approximate deterministic quantum cloning machines are feasible. These are quantum circuits that use only unitary quantum gates to produce approximate clones that are described by reduced density matrices corresponding to mixed states. So even if the input state is pure the approximate clone is mixed.

In this section we show that the alternative strategy of exact albeit non-deterministic cloning machines are also feasible. We call such devices "probabilistic cloning machines" because they might not produce clones every time they run but when they do the fidelity of those clones is higher than what can be achieved deterministically. The quantum circuits corresponding to probabilistic cloning machines use measurements, in addition to unitary gates, to achieve the desired state transformation. The success of the exact probabilistic cloning procedure is signalled by obtaining a specific outcome for these measurements.

The first design for a probabilistic cloning machine is due to Lu-Ming Duan and Guang-Can Guo [152]. They showed that if states are selected secretly from a set $\{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle\}$ they can be cloned exactly probabilistically if and only if the $\{|\psi_i\rangle\}$ are linearly independent. In other words, probabilistic cloning does not work for arbitrary states—they must be linearly independent—but their precise identity does not need to be known so long as the promise holds that they are linearly independent. If this condition holds, then Duan and Guo showed that there exists as unitary operation U and measurement M such that the following transformation is possible:

$$|\psi_i\rangle|\Sigma\rangle \xrightarrow{U\&M} |\psi_i\rangle|\psi_i\rangle \tag{11.171}$$

Here the measurement M means that the transformation is non-unitary overall, which is what allows it to appear to circumvent the no-cloning theorem.

To obtain such a transformation we need to design a unitary transformation and a measurement that does the trick. We begin by imaging there are three sub-spaces to our system A, B, and C. Sub-space A holds the state to be cloned. Sub-space B will hold the clone. And sub-space C will hold ancillae states that we intend to measure.

We can begin by defining an orthonormal set of $(n+1)$ states of a so-called measurement probe $\{|P_0\rangle, |P_1\rangle, \dots, |P_n\rangle\}$. These states can serve as an unambiguous measurement basis provided $\langle P_i|P_j\rangle = 0$ for $i \neq j$ and $\langle P_i|P_i\rangle = 1$. Given such basis states, and a state selected secretly from our linearly independent set $\{|\psi_i\rangle\}$, probabilistic cloning works by creating a unitary evolution of the form:

$$|\psi_i\rangle_A|\Sigma\rangle_B|P_0\rangle_C \xrightarrow{U} \sqrt{p_i}|\psi_i\rangle_A|\psi_i\rangle_B|P_0\rangle_C + \sum_{j=1}^{n} c_{ij}|\Phi_j\rangle_{AB}|P_j\rangle_C \tag{11.172}$$

followed by a measurement of sub-system C in the $\{|P_1\rangle, |P_2\rangle, \dots, |P_n\rangle\}$ basis. In this transformation $|\Phi_1\rangle_{AB}, |\Phi_2\rangle_{AB}, \dots, |\Phi_n\rangle_{AB}$ are n normalized states of subsystems A and B combined, but they are not necessarily orthogonal. Given the structure of the state produced under the action of U on an input $|\psi_i\rangle_A|\Sigma\rangle_B|P_0\rangle_C$ we can see immediately that exact cloning will be achieved whenever the measurement on sub-space C in the $\{|P_0\rangle, |P_1\rangle, \dots, |P_n\rangle\}$-basis yields the result $|P_0\rangle$. Moreover, this event will occur with probability p_i, which we can think of as the "cloning efficiency".

The simplest case is when we want to clone one of only two linearly independent states $\{|\psi_1\rangle, |\psi_2\rangle\}$. In this case Duan and Guo show that the cloning efficiencies p_1 and p_2 must satisfy the inequality:

$$\frac{1}{2}(p_1 + p_2) \leq \frac{1}{1 + \langle\psi_1|\psi_2\rangle} \tag{11.173}$$

This result can be generalized to bound all the probabilities $p_1, p_2, \ldots, p_n$ based on a certain matrix having to be positive semi-definite.

Optimal probabilistic cloning is closely related to the task of optimal unambiguous quantum state discrimination [109, 248, 254, 386].

11.6.7 Broadcasting Quantum Information

Extending the notion of quantum cloning to mixed states requires a little thought, because a complication arises that we do not have in the case of pure states. Given that we don't directly "see" the quantum state produced by cloning, but rather only experience it through the statistical properties it displays, we might wonder whether our goal is to clone a given mixed state literally, or merely produce clones that replicate the statistical properties of the given mixed state? This distinction can be best appreciated in terms of the two possible ways we could set up the notion of cloning for mixed states. These are usually distinguished by contrasting them as "cloning" versus "broadcasting".

$$\rho \otimes |\Phi\rangle\langle\Phi| \xrightarrow{\text{Cloner}} \rho \otimes \rho \tag{11.174}$$

$$\rho \otimes |\Phi\rangle\langle\Phi| \xrightarrow{\text{Broadcaster}} \rho_{AB} : \text{tr}_A(\rho_{AB}) = \rho \text{ and } \text{tr}_B(\rho_{AB}) = \rho \tag{11.175}$$

The problem is that there are many density operators that can mimic the statistical behavior of the true clones. Hence, merely obtaining output density operators that display the same statistical properties as the true clones, is not entirely enough to allow is to conclude we really have true clones.

11.7 Negating Quantum Information

> "*The process of optimal quantum cloning is closely connected to another impossible operation in quantum mechanics, the so-called universal NOT gate for qubits.*"
> – Nicolas Cerf [99]

An ideal classical NOT gate, NOT, is able to negate any bit it is handed even if the bit value is unknown. That is, if $b \in \{0, 1\}$, NOT$b = 1 - b = \neg b$ regardless of value of b.

Similarly, an *ideal* universal[5] *quantum* NOT gate (if it existed) would be able to negate any 1-qubit state it is handed. That is, for $|\psi\rangle = a|0\rangle + b|1\rangle$,

$$U_{\text{NOT}}^{\text{ideal}}|\psi\rangle = b^*|0\rangle - a^*|1\rangle \equiv |\psi^\perp\rangle \tag{11.176}$$

[5] Here "universal" means "input state independent."

In terms of the Bloch sphere, $|\psi^{\perp}\rangle$ is the antipodal point to $|\psi\rangle$ on the opposite side of the Bloch sphere along a straight line through its center. Hence $|\psi\rangle$ and $|\psi^{\perp}\rangle$ are orthogonal quantum states, i.e., $\langle\psi|\psi^{\perp}\rangle = 0$.

Unfortunately, such an *ideal* universal quantum NOT operation requires that $U_{\text{NOT}}^{(\text{ideal})}$ be described by an anti-unitary matrix, whereas deterministic quantum gates are always described by unitary matrices. Hence it is impossible to achieve $U_{\text{NOT}}^{(\text{ideal})}$ exactly deterministically as purely a rotation on the Bloch sphere. Nevertheless, as in the case of quantum cloning, we can define a universal quantum NOT as the best approximation to the ideal NOT operation on qubits.

11.7.1 Universal Quantum Negation Circuit

Surprisingly, as the alert reader will have noticed, the desired negated state $|\psi^{\perp}\rangle$ happens to be produced as an "unwanted" side effect of using a universal quantum cloning circuit! In (11.146) and (11.147) we see the negated state appears as the "distortion" that prevents the clone for being exact. Specifically, we have:

$$\begin{aligned}\rho_B &= \frac{5}{6}|\psi\rangle\langle\psi| + \frac{1}{6}|\psi^{\perp}\rangle\langle\psi^{\perp}| \\ \rho_C &= \frac{5}{6}|\psi\rangle\langle\psi| + \frac{1}{6}|\psi^{\perp}\rangle\langle\psi^{\perp}|\end{aligned} \tag{11.177}$$

However, although we did not show this earlier, the contribution of the negated state, $|\psi^{\perp}\rangle$, to the top qubit A, turns out to be even greater. We can see this by factoring the reduced density operator ρ_A in terms of $|\psi\rangle$ and $|\psi^{\perp}\rangle$ as follows:

$$\begin{aligned}\rho_A = \text{tr}_{BC}(\rho_{ABC}) &= \begin{pmatrix} \frac{2}{3}|\alpha|^2 + \frac{1}{3}|\beta|^2 & \frac{1}{3}\beta\alpha^* \\ \frac{1}{3}\alpha\beta^* & \frac{1}{3}|\alpha|^2 + \frac{2}{3}|\beta|^2 \end{pmatrix} \\ &= Y \cdot \left(\frac{1}{3}|\psi\rangle\langle\psi| + \frac{1}{3}|\psi^{\perp}\rangle\langle\psi^{\perp}|\right) \cdot Y\end{aligned} \tag{11.178}$$

where

$$|\psi\rangle\langle\psi| = \begin{pmatrix} |\alpha|^2 & \alpha\beta^* \\ \beta\alpha^* & |\beta|^2 \end{pmatrix} \tag{11.179}$$

$$|\psi^{\perp}\rangle\langle\psi^{\perp}| = \begin{pmatrix} |\beta|^2 & -\alpha\beta^* \\ -\beta\alpha^* & |\alpha|^2 \end{pmatrix} \tag{11.180}$$

In fact, it turns out that the *optimal* universal negating circuit is exactly the same as the optimal universal cloning circuit! The only difference, when we want to use the cloning circuit as a negating circuit, is that we pay attention to a different output qubit, namely the top qubit that contains ρ_A. Thus, a circuit for universal quantum negation is shown in Fig. 11.11.

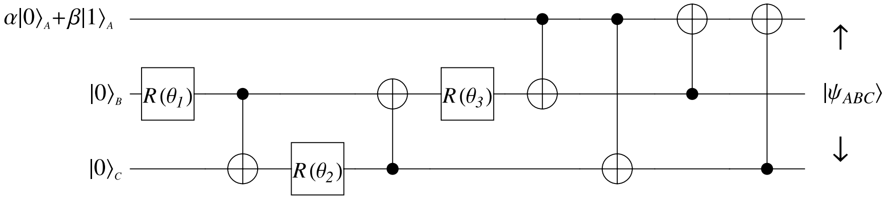

Fig. 11.11 Quantum circuit for universal quantum negation. In an ideal universal quantum negation circuit an unknown quantum state $|\psi\rangle_A = \alpha|0\rangle_A + \beta|1\rangle_A$ would be transformed into $\beta^*|0\rangle_A - \alpha^*|1\rangle_A$. This is not possible deterministically using unitary gates. Instead the best we can do is given by monitoring the output of the top qubit (A). The reduced density matrix of this qubit, $\rho_A = \text{tr}_{BC}(|\psi_{ABC}\rangle\langle\psi_{ABC}|)$ gives the best approximation to the negated state. This shows the fidelity of the negated state with respect to the ideal negated state is $\frac{1}{6}$

11.7.2 Expectation Value of an Observable Based on the Negated State

We can ask a similar question for universal negation that we asked for universal cloning: is the negated state, ρ_A, close enough to the ideal negated state, $|\psi^\perp\rangle\langle\psi^\perp|$, to be on use in subsequent quantum computations?

Using (11.178) we can express the ideal negated state, $\rho_{\text{ideal}}^{\text{UNOT}} = |\psi^\perp\rangle\langle\psi^\perp|$, in terms of the original state and the output on the top qubit of the quantum cloning (or equally, quantum "negating") circuit:

$$\rho_{\text{ideal}}^{\text{UNOT}} = \frac{3}{2}\left(Y \cdot \rho_A \cdot Y - \frac{1}{3}|\psi\rangle\langle\psi|\right) \tag{11.181}$$

where $|\psi\rangle\langle\psi| = \rho_{\text{ideal}}^{\text{CLONE}}$. So for any observable operator Ω we would have:

$$\langle\Omega\rangle = \text{tr}(\rho_{\text{ideal}}^{\text{UNOT}} \cdot \Omega) = \frac{3}{2}\left(\text{tr}(Y \cdot \rho_A \cdot Y \cdot \Omega) - \frac{1}{3}\text{tr}(|\psi\rangle\langle\psi| \cdot \Omega)\right) \tag{11.182}$$

So we can obtain the exact expectation value of an operator on the true negated state, by using the approximation to the negated state on qubit A in conjunction with $\rho_{\text{ideal}}^{\text{CLONE}}$.

11.8 Summary

In Shannon's view, information is equated to the representation of knowledge rather than the content of the knowledge per se. This view of "information" is alien to many people when they first encounter it. However, it turns out to be very useful in practice because it allows us to make concrete predictions on such matters as the degree to which an information bearing message can be compressed while ensuring the original message is recoverable, and the amount of redundancy to build into a communication to ensure it can be transmitted reliably through a noisy channel.

In the quantum context, the notion of information is extended in the obvious way by replacing classical streams of bits with quantum streams of qubits (possibly in non-orthogonal states). We found that the probability distribution by which we characterize a classical source is replaced by the density operator by which we characterize a corresponding quantum source. We introduced a new kind of entropy, the von Neumann entropy, which matches the Shannon entropy only when the quantum states are orthogonal and hence unambiguously distinguishable (like classical symbols). But when the quantum states are non-orthogonal, the von Neumann entropy exceeds the Shannon entropy. This allows certain operations on quantum information to exceed the bounds for corresponding operations on classical information. For example, we can compress quantum messages comprising non-orthogonal states over some probability distribution to a degree that is greater than that of classical messages over symbols that occur with the same probability distribution. We gave examples of two variants of such quantum compression protocols—discard-on-fail and augment-on-fail. More interestingly, we also found that we can use quantum information to compress a classical message by a factor of two beyond the Shannon bound at communication time provided we have already established and stored matching pairs of entangled qubits between the two ends of the communications channel. Thus, overall, Shannon's bound is not exceed. However, at communication time, we can temporarily appear to exceed the Shannon by a factor of two for as long as the supply of matching entangled pairs remain.

Some operations that we take for granted on classical information are not so easy with quantum information. For example, whereas we can copy classical information perfectly deterministically, we cannot do so for quantum information in an unknown quantum state. Similarly, whereas we can negate classical information perfectly deterministically, we cannot negate quantum information in an unknown state. In both cases, however, we can find approximate quantum protocols that do as well as Nature allows. Surprisingly, we can use the approximate clones and approximate negated states to obtain *exact* expectation values of observable operators based on them. So in this sense, they are almost as useful as having perfect clones and perfect negated states.

The main difference between quantum information and classical information is the ability of the former to use non-orthogonal states to represent symbols, and for those non-orthogonal states to be entangled. Neither of these options exists for classical information, and this difference is the root of the dissimilarities between quantum and classical information. We introduced the formalism of density operators to describe quantum sources. We showed how the partial trace was used to describe a part of a composite quantum system. We highlighted the difference between pure and mixed states and focussed on the maximally entangled variants of both kinds of quantum states. We introduced a measure of the degree of entanglement in a quantum state via the tangle, and showed that deciding whether or not a quantum was entangled could be answered using so-called entanglement witnesses or the Peres-Horodecki criterion.

11.9 Exercises

11.1 Calculate the density matrices for the following ensembles.

1. An ensemble of quantum states that are all $\frac{1}{2}|0\rangle - \frac{\sqrt{3}}{2}|1\rangle$.
2. An ensemble of quantum states that are all $\frac{1}{3\sqrt{3}}|01\rangle + \frac{1}{3}\sqrt{\frac{26}{3}}|10\rangle$.
3. An ensemble of quantum states that are $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ with probability 0.3, $\frac{1}{2}|0\rangle - \frac{\sqrt{3}}{2}|1\rangle$ with probability 0.4, and $|0\rangle$ with probability 0.3.

11.2 Compute the density operator for an ensemble that is 30% $|\psi_1\rangle = \frac{1}{\sqrt{3}}|0\rangle + \sqrt{\frac{2}{3}}|1\rangle$ and 70% $|\psi_2\rangle = \frac{2}{3}|0\rangle + \frac{\sqrt{5}}{3}|1\rangle$, and write its elements as decimal numbers. Now compute the density operators for the following ensembles:

1. An ensemble that is 50% $|\psi_1\rangle = 0.680082|0\rangle + 0.733136|1\rangle$ and 50% $|\psi_2\rangle = 0.599759|0\rangle + 0.800181|1\rangle$.
2. An ensemble that is 25% $|\psi_1\rangle = 0.568532|0\rangle + 0.822661|1\rangle$ and 75% $|\psi_2\rangle = 0.66363|0\rangle + 0.748061|1\rangle$.

What do you notice? Can you devise any experimental test to distinguish between these ensembles? Justify your answer.

11.3 What test on a density operator, ρ, tells you whether the state is pure or mixed? According to this test, does the density operator given by

$$\rho = \begin{pmatrix} \frac{1}{9} & 0 & -\frac{2}{9} & \frac{2}{9} \\ 0 & 0 & 0 & 0 \\ -\frac{2}{9} & 0 & \frac{4}{9} & -\frac{4}{9} \\ \frac{2}{9} & 0 & -\frac{4}{9} & \frac{4}{9} \end{pmatrix}$$

correspond to that of a pure state or a mixed state?

11.4 Under what conditions is a 2-qubit state said to be *separable*? Your definition should cover both pure states and mixed states.

11.5 Under what conditions is a 2-qubit density operator said to be that of a *pure* state?

11.6 Which of the following simultaneous conditions of a quantum state are possible? There may be more than one correct answer.

1. A state can be simultaneously pure and mixed.
2. A state can be simultaneously separable and entangled.
3. A state can be simultaneously entangled and mixed.
4. A state can simultaneously mixed and separable.

5. A state can be simultaneously entangled and pure.

11.7 What is the linear entropy of the density operator, ρ, defined by:

$$\rho = \begin{pmatrix} \frac{1}{8} & 0 & 0 & -\frac{\sqrt{3}}{8} \\ 0 & \frac{3}{8} & \frac{1}{8} & 0 \\ 0 & \frac{1}{8} & \frac{1}{8} & 0 \\ -\frac{\sqrt{3}}{8} & 0 & 0 & \frac{3}{8} \end{pmatrix} \tag{11.183}$$

Is linear entropy a good measure of the mixedness or the entanglement within a state? Explain your answer.

11.8 Exhibit a 2-qubit (i.e., 4 × 4) density operator having a linear entropy less than $\frac{8}{9}$ which is entangled.

11.9 Exhibit a 2-qubit (i.e., 4 × 4) density operator having a linear entropy less than $\frac{8}{9}$ which is separable.

11.10 What test based on the linear entropy of a density operator, ρ, tells you whether the state is entangled or separable? According to this test, does the density operator given by

$$\rho = \begin{pmatrix} 0.375003 & 0.0403853 & 0.0634155 & 0.00682943 \\ 0.0403853 & 0.126466 & 0.00682943 & 0.0213862 \\ 0.0634155 & 0.00682943 & 0.372806 & 0.0401487 \\ 0.00682943 & 0.0213862 & 0.0401487 & 0.125725 \end{pmatrix} \tag{11.184}$$

correspond to an entangled state or a separable state?

11.11 What is the von Neumann entropy of a mixed state described by a density operator ρ? Is the von Neumann entropy a good measure of the mixedness or entanglement within a state? Calculate the von Neumann entropies of the following density operators:

1. The maximally mixed state

$$\rho = \begin{pmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{pmatrix}$$

2. The typical mixed state

$$\rho = \begin{pmatrix} 0.314815 & -0.165635i & 0 & 0.166667 \\ 0.165635i & 0.372685 & -0.165359 & 0 \\ 0 & -0.165359 & 0.145833 & 0 \\ 0.166667 & 0 & 0 & 0.166667 \end{pmatrix}$$

3. The maximally entangled mixed state

$$\rho = \begin{pmatrix} \frac{1}{3} & 0 & 0 & \frac{1}{10} \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{10} & 0 & 0 & \frac{1}{3} \end{pmatrix}$$

11.12 Prove that the expectation value of an observable, $\mathcal{O}$, for a quantum system in state ρ, given by (11.8) can be re-expressed in trace form as given by (11.9). That is, prove $\langle \mathcal{O} \rangle = \sum_{i=1}^{N} p_i \langle \psi_i | \mathcal{O} | \psi_i \rangle = \text{tr}(\rho \cdot \mathcal{O})$ where $\mathcal{O}$ is an hermitian matrix, and ρ is a density operator.

11.13 It is possible to inter-convert between Bell states by applying single qubit operation to one member of a Bell state pair. What Bell state transformations do the following 1-qubit gates bring about?

$$|\beta_{00}\rangle \quad \xrightarrow{R_y(-\pi)\otimes \mathbb{1}} \quad ???$$

$$|\beta_{01}\rangle \quad \xrightarrow{Ph(\pi/2)\cdot R_y(\pi)\cdot R_z(\pi)\otimes \mathbb{1}} \quad ???$$

$$|\beta_{10}\rangle \quad \xrightarrow{Ph(\pi/2)\cdot R_y(-\pi)\cdot R_z(\pi)\otimes \mathbb{1}} \quad ???$$

$$|\beta_{11}\rangle \quad \xrightarrow{Ph(-\pi/2)\cdot R_z(-\pi)\otimes \mathbb{1}} \quad ???$$

11.14 Recall that the quantum No-Cloning theorem asserts that “An unknown quantum state cannot be cloned”. Thus, it is supposed to be impossible to find a unitary transformation that can accomplish the transformation $|\psi\rangle|0\rangle \longrightarrow |\psi\rangle|\psi\rangle$ for $|\psi\rangle$ unknown. However, you see an article that challenges the veracity of the No-Cloning theorem based on the following argument:

(a) A bit, by definition, can be only 0 or 1.
(b) If you are given a bit but not told its value, then the bit is, by definition, unknown to you. So let’s call the bit value b, but leave the value unspecified.
(c) Conceptually, you could use the bit value b to control the settings of a device such as a Pockels cell (see Chap. 13), that outputs a horizontally polarized photon if $b = 0$ and a vertically polarized photon if $b = 1$. Thus, without loss of generality, we can convert our unknown bit to an unknown quantum state, which we can represent as $|b\rangle$, without ever revealing the value of b.
(d) Now imagine augmenting the output from the Pockels cell, the *unknown* state $|b\rangle$, with another photon in a *known* state $|0\rangle$ (horizontally polarized photon) and push them through some optical apparatus that implements a CNOT gate, i.e., compute CNOT$|b\rangle|0\rangle$. Clearly, b has to be either 0 or 1 so the only two cases we need to consider are CNOT$|0\rangle|0\rangle = |0\rangle|0\rangle$ and CNOT$|1\rangle|0\rangle = |1\rangle|1\rangle$.
(e) Either way, the unknown quantum state $|b\rangle$ has been successfully cloned!
(f) Therefore, the No-Cloning theorem must be wrong, because here we have successfully cloned an unknown quantum state $|b\rangle$!

What is wrong with this argument? Why does it not disprove the No-Cloning theorem? Justify you answer by critiquing each step in the aforementioned argument.

11.15 Given the density matrix:

$$\rho = \begin{pmatrix} \frac{4}{49} & -\frac{6}{35} & \frac{4\sqrt{314}}{735} & -\frac{4i}{21} \\ -\frac{6}{35} & \frac{9}{25} & -\frac{2\sqrt{314}}{175} & \frac{2i}{5} \\ \frac{4\sqrt{314}}{735} & -\frac{2\sqrt{314}}{175} & \frac{1256}{11025} & -\frac{4i\sqrt{314}}{315} \\ \frac{4i}{21} & -\frac{2i}{5} & \frac{4i\sqrt{314}}{315} & \frac{4}{9} \end{pmatrix} \tag{11.185}$$

prove that its two partial transposes, ρ^{T_A} and ρ^{T_B}, have the same set of eigenvalues.

11.16 What are the density matrices corresponding to the four pure Bell states, $|\beta_{00}\rangle$, $|\beta_{01}\rangle$, $|\beta_{10}\rangle$, or $|\beta_{11}\rangle$ as defined in (11.69)? Are they the same or different? Now compute the reduced density matrices obtained by tracing over each of the qubits in each of these Bell states. Are these reduced density matrices the same or different? If your results are different, use them to find a single qubit observable, $\Omega = \begin{pmatrix} a & c \\ c^* & b \end{pmatrix}$, which is able to distinguish between the four Bell states. Alternatively, if your results are the same, use them to prove no such observable exists.

11.17 One way to measure the similarity between a pair of density matrices, σ and ρ, is via their fidelity:

$$\mathcal{F}(\sigma, \rho) = \left[\text{tr}(\sqrt{\sqrt{\sigma} \cdot \rho \cdot \sqrt{\sigma}}) \right]^2 \tag{11.186}$$

Show that if σ is the density matrix of an arbitrary single qubit pure state, i.e., if $\sigma = |\psi\rangle\langle\psi|$ where $\psi = a|0\rangle + \sqrt{1-|a|^2}|1\rangle$ (with $|a| \leq 1$ and $a \in \mathbb{C}$), and $\rho = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|$ (with $0 \leq p \leq 1$ and $p \in \mathbb{R}$) then the fidelity $\mathcal{F}(\sigma, \rho)$ can be written as:

$$\mathcal{F}(\sigma, \rho) = \langle\psi|\rho|\psi\rangle = 1 - p - (1-2p)|a|^2 \tag{11.187}$$

Notice that if $p = \frac{1}{2}$ the fidelity is then independent of a. What is so special about the state $\rho = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|$ when $p = \frac{1}{2}$? Why should the fidelity between ρ when $p = \frac{1}{2}$ and any pure state be independent of the form of that pure state?

11.18 Consider the pair of entangled states $|\psi_{\text{W}}\rangle$ and $|\psi_{\text{GHZ}}\rangle$ defined on the three qubits A, B, and C as follows:

$$\begin{aligned} |\psi_{\text{W}}\rangle &= \frac{1}{\sqrt{3}}(|001\rangle + |010\rangle + |100\rangle); \qquad \rho_{ABC}^{\text{W}} = |\psi_{\text{W}}\rangle\langle\psi_{\text{W}}| \\ |\psi_{\text{GHZ}}\rangle &= \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle); \qquad \rho_{ABC}^{\text{GHZ}} = |\psi_{\text{GHZ}}\rangle\langle\psi_{\text{GHZ}}| \end{aligned} \tag{11.188}$$

Prove the following:

(a) The states $|\psi_{\rm W}\rangle$ and $|\psi_{\rm GHZ}\rangle$ are orthogonal, i.e., $\langle\psi_{\rm W}|\psi_{\rm GHZ}\rangle = 0$. What does this tell you about the degree to which $|\psi_{\rm W}\rangle$ is similar to $|\psi_{\rm GHZ}\rangle$?

(b) $|\psi_{\rm W}\rangle$ and $|\psi_{\rm GHZ}\rangle$ are both entangled states.

(c) The 2-qubit sub-systems of $|\psi_{\rm W}\rangle$ are identical, i.e. ignoring indices, we obtain the same state whether we trace over the first, second, or third qubit:

$$\rho^{\rm W}_{BC} = {\rm tr}_A(|\psi_{\rm W}\rangle\langle\psi_{\rm W}|) = \rho^{\rm W}_{AC} = {\rm tr}_B(|\psi_{\rm W}\rangle\langle\psi_{\rm W}|) = \rho^{\rm W}_{AB} = {\rm tr}_C(|\psi_{\rm W}\rangle\langle\psi_{\rm W}|) \tag{11.189}$$

(d) The 2-qubit sub-systems of $|\psi_{\rm GHZ}\rangle$ are identical, i.e. ignoring indices, we obtain the same state whether we trace over the first, second, or third qubit:

$$\begin{aligned}\rho^{\rm GHZ}_{BC} &= {\rm tr}_A(|\psi_{\rm GHZ}\rangle\langle\psi_{\rm GHZ}|) = \rho^{\rm GHZ}_{AC} = {\rm tr}_B(|\psi_{\rm GHZ}\rangle\langle\psi_{\rm GHZ}|) \\ &= \rho^{\rm GHZ}_{AB} = {\rm tr}_C(|\psi_{\rm GHZ}\rangle\langle\psi_{\rm GHZ}|)\end{aligned} \tag{11.190}$$

(e) For any pair of indices $\{x, y\} \subset \{A, B, C\}$, the fidelity between the reduced density matrices $\rho^{\rm W}_{xy}$ and $\rho^{\rm GHZ}_{xy}$ is $\frac{1}{6}$. That is, prove $\mathcal{F}(\rho^{\rm W}_{xy}, \rho^{\rm GHZ}_{xy}) = \frac{1}{6}$.

(f) The 1-qubit sub-systems of $|\psi_{\rm W}\rangle$ are identical, i.e. ignoring indices, we obtain the same state whether we trace over the second and third, first and third, or first and second qubits:

$$\rho^{\rm W}_{A} = {\rm tr}_{BC}(|\psi_{\rm W}\rangle\langle\psi_{\rm W}|) = \rho^{\rm W}_{B} = {\rm tr}_{AC}(|\psi_{\rm W}\rangle\langle\psi_{\rm W}|) = \rho^{\rm W}_{C} = {\rm tr}_{AB}(|\psi_{\rm W}\rangle\langle\psi_{\rm W}|) \tag{11.191}$$

(g) The 1-qubit sub-systems of $|\psi_{\rm GHZ}\rangle$ are identical, i.e. ignoring indices, we obtain the same state whether we trace over the second and third, first and third, or first and second qubits:

$$\begin{aligned}\rho^{\rm GHZ}_{A} &= {\rm tr}_{BC}(|\psi_{\rm GHZ}\rangle\langle\psi_{\rm GHZ}|) = \rho^{\rm GHZ}_{B} = {\rm tr}_{AC}(|\psi_{\rm GHZ}\rangle\langle\psi_{\rm GHZ}|) = \rho^{\rm GHZ}_{C} \\ &= {\rm tr}_{AB}(|\psi_{\rm GHZ}\rangle\langle\psi_{\rm GHZ}|)\end{aligned} \tag{11.192}$$

(h) For any index $x \in \{A, B, C\}$, the fidelity between the reduced density matrices $\rho^{\rm W}_{x}$ and $\rho^{\rm GHZ}_{x}$ is $\frac{1}{6}(3 + 2\sqrt{2})$. That is prove, $\mathcal{F}(\rho^{\rm W}_{x}, \rho^{\rm GHZ}_{x}) = \frac{1}{6}(3 + 2\sqrt{2})$.

(i) What is the fidelity between the original pair of states $|\psi_{\rm W}\rangle$ and $|\psi_{\rm GHZ}\rangle$ in comparison to the fidelities of its 2-qubit and 1-qubit sub-systems?

11.19 Consider the state $|\psi_{\rm W}\rangle$ defined in (11.188). Use the Schmidt decomposition to "automatically" discover the (trivial) factorization of $|\psi_{\rm W}\rangle$ in the form:

$$|\psi_{\rm W}\rangle = \sqrt{\frac{2}{3}}|0\rangle \otimes \left(\frac{|01\rangle + |10\rangle}{\sqrt{2}}\right) + \frac{1}{\sqrt{3}}|1\rangle|00\rangle \tag{11.193}$$

11.20 Consider the 3-qubit state $|\psi_{ABC}\rangle = \frac{1}{\sqrt{2}}(|\psi_{\text{W}}\rangle + |\psi_{\text{GHZ}}\rangle)$ where $|\psi_{\text{W}}\rangle$ and $|\psi_{\text{GHZ}}\rangle$ are defined as in (11.188). Suppose you wish to write $|\psi_{ABC}\rangle$ in the form:

$$|\psi_{ABC}\rangle = \sum_{i=0}^{\min(d_A-1, d_{BC}-1)} \lambda_i |i_A\rangle |i_{BC}\rangle \tag{11.194}$$

(single index summation) where A is a 2-dimensional subspace, and BC is a 4-dimensional subspace. Demonstrate how to apply the Schmidt decomposition to find suitable values for the Schmidt coefficients (λ_i) and the eigenvectors ($\{|i_A\rangle\}$ and $\{|i_{BC}\rangle\}$). Verify that your solution yields a Schmidt decomposition for $|\psi_{ABC}\rangle$ of the form:

$$|\psi_{ABC}\rangle = \lambda_1 |1_A\rangle |1_{BC}\rangle + \lambda_2 |2_A\rangle |2_{BC}\rangle \tag{11.195}$$

where:

$$\begin{aligned}
\lambda_1 &= \sqrt{\frac{1}{2} + \frac{\sqrt{7}}{12}} \\
\lambda_2 &= \frac{1}{2}\sqrt{\frac{1}{3}\left(6 - \sqrt{7}\right)} \\
|1_A\rangle &= -\sqrt{\frac{1}{2} + \frac{1}{2\sqrt{7}}}\,|0\rangle - \sqrt{\frac{1}{14}\left(7 - \sqrt{7}\right)}\,|1\rangle \\
|2_A\rangle &= -\sqrt{\frac{1}{14}\left(7 - \sqrt{7}\right)}\,|0\rangle + \sqrt{\frac{1}{2} + \frac{1}{2\sqrt{7}}}\,|1\rangle \\
|1_{BC}\rangle &= -\sqrt{\frac{17}{58} + \frac{43}{58\sqrt{7}}}\,|00\rangle - \sqrt{\frac{1}{203}\left(35 - \sqrt{7}\right)}\,|01\rangle \\
&\quad - \sqrt{\frac{1}{203}\left(35 - \sqrt{7}\right)}\,|10\rangle - \sqrt{\frac{3}{406}\left(49 - 13\sqrt{7}\right)}\,|11\rangle \\
|2_{BC}\rangle &= \sqrt{\frac{1}{406}\left(119 - 43\sqrt{7}\right)}\,|00\rangle - \sqrt{\frac{5}{29} + \frac{1}{29\sqrt{7}}}\,|01\rangle \\
&\quad - \sqrt{\frac{5}{29} + \frac{1}{29\sqrt{7}}}\,|10\rangle + \sqrt{\frac{3}{406}\left(49 + 13\sqrt{7}\right)}\,|11\rangle
\end{aligned} \tag{11.196}$$

11.21 Alice and Bob wish to perform a quantum mechanical experiment over a distance of 400 km. The experiment requires that Alice and Bob have corresponding members of maximally entangled pairs of particles. However, if they transmit a particle over 100 km they can no longer guarantee its state is pristine. How, in principle, can Alice and Bob establish the required entangled pairs of particles over a distance of 400 km? Explain, by describing the sequence of state changes, how they could use this scheme to establish shared pairs of particles each in the state $\beta_{01} = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$.

11.22 Suppose Alice and Bob have access to ideal quantum memories, i.e., they are able to store quantum information without any loss of fidelity indefinitely. In addition, assume Alice and Bob are connected by a fiber optic communications network, which can support both quantum and classical communications but is shared with other users. This network is idle for approximately 20% of the time, underutilized for 70% of the time and at peak congestion for 10% of the time. Explain how Alice and Bob can exploit quantum information to boost their effective communications capacity at times of peak congestion. At such times, by what factor can they, in principle, increase their effective communications rate? Can this enhanced communications rate be maintained indefinitely? Explain your answer.

11.23 The states that have the maximal possible amount of entanglement for a given amount of mixedness (as measured by linear entropy) can be written in the form:

$$\rho = \begin{cases} \begin{pmatrix} \frac{1}{3} & 0 & 0 & \frac{r}{2} \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{r}{2} & 0 & 0 & \frac{1}{3} \end{pmatrix} & 0 \leq r \leq \frac{2}{3} \\ \begin{pmatrix} \frac{r}{2} & 0 & 0 & \frac{r}{2} \\ 0 & 1-r & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{r}{2} & 0 & 0 & \frac{r}{2} \end{pmatrix} & \frac{2}{3} < r \leq 1 \end{cases} \tag{11.197}$$

Show when $0 \leq r \leq \frac{2}{3}$ that ρ can be factored in the form:

$$\rho = p_1|\psi_1\rangle\langle\psi_1| + p_2|\psi_2\rangle\langle\psi_2| + (1 - (p_1 + p_2))|\psi_3\rangle\langle\psi_3| \tag{11.198}$$

where

$$\begin{aligned} p_1 &= \frac{1}{12}\left(4 - 9r^2\right) \\ p_2 &= \frac{1}{3} \\ |\psi_1\rangle &= |00\rangle \\ |\psi_2\rangle &= |01\rangle \\ |\psi_3\rangle &= \frac{3r}{\sqrt{4+9r^2}}|00\rangle + \frac{2}{\sqrt{4+9r^2}}|11\rangle \end{aligned} \tag{11.199}$$

Likewise, show when $\frac{2}{3} < r \leq 1$ that ρ can be factored in the form:

$$\rho = (1-r)|\psi_1\rangle\langle\psi_1| + r|\psi_2\rangle\langle\psi_2| \tag{11.200}$$

where

$$|\psi_1\rangle = |01\rangle$$
$$|\psi_2\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{11.201}$$

11.24 We can always regard a mixed state as the reduced density matrix of a larger pure state within some sub-system of interest. The procedure for finding such an encompassing pure state is called "purification of a mixed state", and was described in this Chapter. Review the purification procedure and apply it to show that the state

$$|\psi_{AB}\rangle = \left(\frac{1}{4}\sqrt{\frac{7}{6}} - \frac{1}{4}\sqrt{\frac{3}{2}}\right)|0000\rangle + \left(\frac{1}{4}\sqrt{\frac{7}{6}} + \frac{1}{4}\sqrt{\frac{3}{2}}\right)|0011\rangle + \frac{1}{\sqrt{3}}|0101\rangle$$
$$- \frac{1}{\sqrt{6}}|1100\rangle + \frac{1}{\sqrt{6}}|1111\rangle \tag{11.202}$$

is a purification of the mixed state

$$\rho_A = \begin{pmatrix} \frac{1}{3} & 0 & 0 & \frac{1}{4} \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & \frac{1}{3} \end{pmatrix} \tag{11.203}$$

Note that a state such as ρ_A has the maximum possible value of entanglement for the degree of mixedness (as measured by linear entropy) in ρ_A. Verify that $|\psi_{AB}\rangle$ is a purification of ρ_A by showing $\text{tr}_B(|\psi_{AB}\rangle\langle\psi_{AB}|) = \rho_A$, where sub-space A corresponds to the first and second qubits, and sub-space B corresponds to the third and fourth qubits.

11.25 Show how to construct the purification:

$$|\psi_{AB}\rangle = \frac{1}{2}\sqrt{\frac{3}{2}}|0001\rangle + \frac{1}{2\sqrt{2}}|0100\rangle + \frac{1}{2\sqrt{2}}|0111\rangle + \frac{1}{2}\sqrt{\frac{3}{2}}|1101\rangle \tag{11.204}$$

of the mixed state:

$$\rho_A = \begin{pmatrix} \frac{3}{8} & 0 & 0 & \frac{3}{8} \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{3}{8} & 0 & 0 & \frac{3}{8} \end{pmatrix} \tag{11.205}$$

Note that a state such as ρ_A has the maximum possible value of entanglement for the degree of mixedness (as measured by linear entropy) in ρ_A. Verify that $|\psi_{AB}\rangle$ is a purification of ρ_A by showing $\text{tr}_B(|\psi_{AB}\rangle\langle\psi_{AB}|) = \rho_A$.

11.26 Use the Peres-Horodecki criterion to decide whether each of the following states is or is not entangled:

1.

$$|\psi\rangle = \frac{1}{\sqrt{6}}\,|00\rangle + \frac{1}{\sqrt{3}}\,|01\rangle + \frac{1}{4}\,|10\rangle + \frac{1}{4}\sqrt{7}\,|11\rangle \tag{11.206}$$

2.

$$|\psi\rangle = \frac{1}{5}\sqrt{3}\,|00\rangle + \frac{1}{5}\sqrt{6}\,|01\rangle + \frac{4}{5\sqrt{3}}\,|10\rangle + \frac{4}{5}\sqrt{\frac{2}{3}}\,|11\rangle \tag{11.207}$$

3.

$$\rho = \begin{pmatrix} \frac{5}{14} & 0 & 0 & \frac{5}{14} \\ 0 & \frac{2}{7} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{5}{14} & 0 & 0 & \frac{5}{14} \end{pmatrix} \tag{11.208}$$

4.

$$\begin{pmatrix} \frac{1}{90}(15-4\sqrt{3}) & -\frac{i}{15\sqrt{3}} & -\frac{1}{90}i(-15+4\sqrt{3}) & \frac{1}{15\sqrt{3}} \\ \frac{i}{15\sqrt{3}} & \frac{2}{15\sqrt{3}} & -\frac{1}{15\sqrt{3}} & \frac{2i}{15\sqrt{3}} \\ \frac{1}{90}i(-15+4\sqrt{3}) & -\frac{1}{15\sqrt{3}} & \frac{5}{6}-\frac{2}{3\sqrt{3}} & -\frac{i}{3\sqrt{3}} \\ \frac{1}{15\sqrt{3}} & -\frac{2i}{15\sqrt{3}} & \frac{i}{3\sqrt{3}} & \frac{2}{3\sqrt{3}} \end{pmatrix} \tag{11.209}$$

5.

$$\rho = \begin{pmatrix} \frac{2}{3}(1-\frac{2}{\sqrt{5}}) & 0 & 0 & \frac{1}{15}\sqrt{2}(-5+2\sqrt{5}) \\ 0 & \frac{1}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & 0 \\ 0 & -\frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & 0 \\ \frac{1}{15}\sqrt{2}(-5+2\sqrt{5}) & 0 & 0 & \frac{1}{15}(5-2\sqrt{5}) \end{pmatrix} \tag{11.210}$$

11.27 There are many possible entanglement monotones that can be used to quantify the degree of entanglement within a quantum state. Two popular ones for 2-qubit states are "negativity" and "concurrence" (which in turn is just the square root of the tangle). Look up the definitions of negativity and concurrence (i.e., tangle) and then answer the following questions:

1. Compute the negativity and concurrence for each of the quantum states listed in Problem 11.26.
2. What do you notice about the values of negativity and concurrence when the states are pure?
3. What do you notice about the values of negativity and concurrence when the states are determined, e.g., by the Peres-Horodecki criterion, to be separable?