Short notes on Lecture
Strathclyde Summer School

Chapter 1

Classical information

1.1 The bit

Let me start my lecture on classical information trying to build an intuitive understanding of the concept of classical information. A more quantitative approach will be taken in section 1.2, but for the full blown mathematical apparatus I have to refer you to textbooks, e.g. [1].

Imagine that you are holding an object, be it an array of cards, geometric shapes or a complex molecule and we ask the following question: *what is the information content of this object?* To answer this question, we introduce another party, say a friend, who shares some background knowledge with us (e.g. the same language or other sets of prior agreements that make communication possible at all), but who does not know the state of the object. We define the *information content* of the object as the size of the set of instructions that our friend requires to be able to identify the object, or better the state of the object. For example, assume that the object is a spin-up particle and that we share with the friend the background knowledge that the spin is oriented either upwards or downwards along the z direction with equal probability (see fig. 1.1 for a slightly more involved example). In this case, the only instruction we need to transmit to another party to let him recreate the state is whether the state is spin-up $\uparrow$ or spin-down $\downarrow$. This example shows that in some cases the instruction transmitted to our friend is just a choice between two alternatives. More generally,

Figure 1.1: An example for a decision tree. Two binary choices have to be made to identify the shape (triangle or square) and the orientation (horizontal or rotated). In sending with equal probability one of the four objects, one therefore transmits 2 bits of information.

we can reduce a complicated set of instructions to n binary choices. If that is done we readily get a measure of the information content of the object by simply counting the number of binary choices. In classical information theory, a variable which can assume only the values 0 or 1 is called a *bit*. Instructions to make a binary choice can be given by transmitting 1 to suggest one of the alternative (say arrow up $\uparrow$) and 0 for the other (arrow down $\downarrow$).

To sum up, we say that n bits of information can be encoded in a system when instructions in the form of n binary choices need to be transmitted to identify or recreate the state of the system. In the following we will turn this idea into a more precise form.

1.2 Quantifying classical information

In 1948 Shannon developed a rigorous framework for the description of information and derived an expression for the information content of the message which depends on the probability of each letter occurring and results in the Shannon entropy. We will illustrate Shannon's reasoning in the context of the example above. Shannon invoked the law of large numbers and stated that, if the message is composed of N letters where N is very large, then the *typical* messages will be composed of Np_1 1's and Np_0 0's. For simplicity, we assume that N is 8 and that p_1 and p_0 are $\frac{1}{8}$ and $\frac{7}{8}$ respectively. In this case the typical messages are the 8 possible sequences composed of 8 binary digits of which only one is equal to 1 (see left side of figure 1.2). As the length of the message increases (i.e. N gets large) the probability of getting a message which is all 1's or any other message that differs significantly from a typical sequence is negligible so that we can safely ignore them. But how many distinct typical messages are there? In the previous example the answer was clear: just 8. In the general case one has to find in how many ways the Np_1 1's can be arranged in a sequence of N letters? Simple

Figure 1.2: The idea behind classical data compression. The most likely sequences are relabeled using fewer bits while rare sequences are discarded. The smaller number of bits still allows the reconstruction of the original sequences with very high probability.

combinatorics tells us that the number of distinct typical messages is

$$\left(\begin{array}{c} N \\ Np_1 \end{array} \right) = \frac{N!}{(Np_1)!(Np_0)!} \tag{1.1}$$

and they are all equally likely to occur. Therefore, we can label each of these possible messages by a binary number. If that is done, the number of binary digits I we need to label each typical message is equal to $log_2 \frac{N!}{Np_1!Np_0!}$. In the example above each of the 8 typical message can be labeled by a binary number composed by $I = log_2 8 = 3$ digits (see figure 1.2). It therefore makes sense that the number I is also the number of bits encoded in the message, because Alice can unambiguously identify the content of each typical message if Bob sends her the corresponding binary number, provided they share the background knowledge on the labeling of the typical messages. All other letters in the original message are really redundant and do not add any information! When the message is very long almost any message is a typical one. Therefore, Alice can reconstruct with arbitrary precision the original N bits message Bob wanted to send her just by receiving I bits. In the example above, Alice can compress an 8 bits message down to 3 bits. Though, the efficiency of this procedure is limited when the message is only 8 letters long, because the approximation of considering only typical sequences is not that good. We leave to the reader to show that the number of bits I contained in a large N-letter message can in general be written, after using Stirling's formula, as

$$I = -N(p_1 log_2 p_1 + p_0 log_2 p_0) \ . \tag{1.2}$$

If we plug the numbers $\frac{1}{8}$ and $\frac{7}{8}$ for p_0 and p_1 respectively in equation 1.2, we find that the information content per symbol $\frac{I}{N}$ when N is very large is approximately 0.5436 bits. On the other hand, when the binary letters 1 and 0 appear with equal probabilities, then compression is not possible, i.e. the message has no redundancy and each letter of the message contains one full bit of information per symbol. These results match nicely the intuitive arguments given above.

Equation 1.2 can easily be generalized to an alphabet of n letters ρ_i each occurring with probabilities p_i. In this case, the average information in bits transmitted per symbol in a message composed of a large number N of letters is given by the Shannon entropy:

$$\frac{I}{N} = H\{p_i\} = -\sum_{i=1}^{n} p_i log p_i \; . \tag{1.3}$$

We remark that the information content of a complicated classical system composed of a large number N of subsystems each of which can be in any of n states occurring with probabilities p_i is given by $N \times H\{p_i\}$.

1.2.1 Classical Correlations

Now that we have a good idea how to quantify classical information, we can proceed to understand and quantify what classical correlations are. Most of you will already have a good intuitive feeling about this because you will have learnt about correlated random variables in your university studies.

Let us begin by considering two strings of messages with a joint probability distribution $p(x, y)$ and marginal distributions $p(x) = \sum_y p(x, y)$ and $q(y) = \sum_x p(x, y)$.

String 1 : 00100010111000101110011000100010110001101000101000
String 2 : 01100111111001101110111101110110110111111011101000

When you look very carefully at the two strings, then you will see that a $'1'$ in string 1 always has a corresponding digit $'1'$ in sequence 2. Therefore clearly the two sequences are correlated: If we look at one sequence, we learn something about the other.

But how strongly are the two messages correlated? Well, a sensible measure seems the following. First let us consider message 2 alone. Its information content is given by $H(Y) = -q(0)logq(0) - q(1)logq(1)$. If we read message 1 then we already obtain information about message 2 and the remaining information content will be decreased. This decrease will depend on which letter we find in message 1 and we have

$$\begin{aligned} H(Y|X) &= \sum_x p(x) \sum_y -p(y|x) \log p(y|x) \\ &= -\sum_{x,y} p(x)p(y|x) \log \frac{p(x, y)}{p(x)} \end{aligned}$$

$$
\begin{aligned}
&= -\sum_{x,y} p(x)p(y|x)(\log p(x,y) - \log p(x)) \\
&= -\sum_{x,y} p(x,y)\log p(x,y) - \left[-\sum_{x} p(x)\log p(x)\right] \\
&= H(X,Y) - H(X)
\end{aligned}
$$

The amount of information that we obtain about message 2 on average is therefore the difference its information content before and after the measurement of sequence 1, ie

$$I = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y)$$

This appears like a well motivated correlation measure and is called the mutual information between the two messages. Its quite satisfying to see that the mutual information is symmetric with respect to interchange of X and Y.

Before I move on from here let me consider two examples, one for fully correlated sequences and one for totally uncorrelated sequences.

Example 1: Consider the sequences with a joint probability distribution $p(0,0) = 0.5 = p(1,1)$ and $p(0,1) = 0 = p(1,0)$, ie

String 1 : 001000101110001011100110001000101100011010001
String 2 : 001000101110001011100110001000101100011010001

In this case, reading message 1 already reveals message 2 completely and therefore it makes sense to say that the two messages are fully correlated. If we compute the mutual information, then we indeed find $I = 1$.

Consider the sequences with a joint probability distribution $p(0,0) = 1$ and $p(1,1) = p(0,1) = 0 = p(1,0)$, ie

String 1 : 000000000000000000000000000000000000000000000
String 2 : 000000000000000000000000000000000000000000000

This example is more subtle than it appears. Not so few people would say that these two sequences are fully correlated, as a '0' in

> message 1 always corresponds to a ’0’ in message 2. However, we have to be careful about this argument. We have defined correlations as the amount of information that we gain about sequence 2 if we measure sequence 1. Now, if we compute what the information content of sequence 2 is, then we find that it vanishes. Thats not surprising, after all, the message seems rather boring as all the digits are actually the same. After measuring sequence 1 the information content is still zero. Therefore the correlations between the two messages are zero and this is indeed the result that you obtain when you compute the mutual information between the two sequences.

Before I now move on to discuss quantum information, let me refine the concept of classical correlations a little bit. This will allows me to introduce the classical equivalent of teleportation and entanglement.

In fact, there are two forms of classical correlations. Firstly there are public correlations and there are secret correlations. What is the difference. Well, imagine parties Alice and Bob talk to each other publicly and the whenever Alice says $'0'('1')$ then Alice and Bob note down a $'0'('1')$ in their notebook. Of course everybody has heard their communication and they also know about their correlations - these correlations are public. On the other hand Alice and Bob may sit down together and secretly create a set of two codebooks in which they write perfectly correlated messages. These correlations are not accessible to anybody else and they are therefore secret. In fact, these correlations have a name which is one-time-pad.

Now assume that Alice and Bob are on different sides of the world. Now I want to show you that Alice and Bob can use these secret correlations to transmit secret messages between each other using only public communication.

The procedure works as follows (see the figure below). Alice takes a bit from here message and one bit from her codebook. She adds the two numbers modulo 2 and sends the result to Bob. He then takes the fully correlated bit from his codebook and adds it (modulo 2) to the digit that he has received from Alice. The result is the digit from Alices original message.

Message Alice	Codebook Alice	Transmit	Codebook Bob	Result Bob
0	0	0	0	0
1	1	0	1	1
1	0	1	0	1
⋮	⋮	⋮	⋮	⋮

This is really a secret information transmission, if the digits in Alice and Bobs codebook are completely random but prefectly correlated. Then the transmitted bit from Alice to Bob is completely random and not correlated to Alices message. Note that a full unit of secret correlations has to be consumed for each classical bit that is to be transmitted. In fact, this could have been used as an alternative way for defining the unit of correlations, namely one bit of correlations is that amount of correlations that allow the secret transmission of one classical bit of information. We will use this approach lateron to define a unit of entanglement.

Now let us see how we can move all these ideas from the classical world to the quantum world.

1.3 Quantum bits

Now I will explain how, in close analogy to the classical compression procedure, an arbitrary quantum state of a composite system comprised of n 2-level atoms, can be compressed and transmitted by sending a number $m < n$ of qubits. This procedure defines the use of the qubit as the unit of quantum information and by analogy with classical data compression partly justifies the name *qubit*. I proceed in close *mathematical* analogy to the classical case studied in section 1.2 and see how well one can compress quantum states, ie. how many qubits are needed to describe a quantum state. I first give a simple example, that illustrates the key ideas, and then reiterate these ideas in a slightly more general and formal way.

Quantum data compression: a simple example

Let me begin with the following very simple example, which appears essentially classical, but displays all the relevant ideas of the more general case. Consider a quantum source that emits two-level systems with probability $p_0 = 0.95$ in state $|0\rangle$ and with probability $p_1 = 1-p_0 = 0.05$ in the orthogonal state $|1\rangle$. Our knowledge of this preparation procedure for a single qubit is represented by the density operator $\hat{\rho}$ given by

$$\hat{\rho} = \frac{6}{7}|0\rangle\langle 0| + \frac{1}{7}|1\rangle\langle 1| \tag{1.4}$$

Note, that the two states generated by the oven have been chosen to be orthogonal for simplicity. We will consider the more general case later. For the time being, let us consider blocks of 7 qubits generated by the source described above. Clearly any sequence of qubits in states $|0\rangle$ and $|1\rangle$ is possible, but some are more likely than others. In fact, typically you will find either a sequence that contains only qubits in state $|0\rangle$ or sequences with a single qubit in state $|1\rangle$ and all others in state $|0\rangle$, or sequences with only two qubits in state $|1\rangle$ as shown below:

$$\begin{aligned}
|\psi_{00000}\rangle &= |0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle \\
|\psi_{00001}\rangle &= |0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|1\rangle \\
|\psi_{00010}\rangle &= |0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|1\rangle|0\rangle \\
|\psi_{00011}\rangle &= |0\rangle|0\rangle|0\rangle|0\rangle|1\rangle|0\rangle|0\rangle \\
|\psi_{00100}\rangle &= |0\rangle|0\rangle|0\rangle|1\rangle|0\rangle|0\rangle|0\rangle \qquad (1.5) \\
|\psi_{00101}\rangle &= |0\rangle|0\rangle|1\rangle|0\rangle|0\rangle|0\rangle|0\rangle \\
|\psi_{00110}\rangle &= |0\rangle|1\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle \\
|\psi_{00111}\rangle &= |1\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle \\
|\psi_{01000}\rangle &= |1\rangle|1\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle \\
&\vdots \qquad (1.6)
\end{aligned}$$

The probability that you will get one of the above sequences is $p_{likely} = (\frac{6}{7})^7 + 7(\frac{6}{7})^6(\frac{1}{7}) + 21(\frac{6}{7})^5(\frac{1}{7})^2 = 0.935$. Of course, these 'typical' states can be enumerated using just three binary digits, i.e. 5 binary digits are sufficient to enumerate 93.5% of all occurring sequences. This procedure is analogous to labeling the typical sequences of 0s and 1s shown in figure 1.2 except that we now 'enumerate' the typical sequences with 'quantum states'. Now, let us see how we can use this fact quantum mechanically. We define a unitary transformation that has the following effect:

$$\begin{array}{rcl}
U|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle & = & |0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle \\
U|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|1\rangle & = & |0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|1\rangle \\
U|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|1\rangle|0\rangle & = & |0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|1\rangle|0\rangle \\
U|0\rangle|0\rangle|0\rangle|0\rangle|1\rangle|0\rangle|0\rangle & = & |0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|1\rangle|1\rangle \\
U|0\rangle|0\rangle|0\rangle|1\rangle|0\rangle|0\rangle|0\rangle & = & |0\rangle|0\rangle|0\rangle|0\rangle|1\rangle|0\rangle|0\rangle \qquad (1.7) \\
U|0\rangle|0\rangle|1\rangle|0\rangle|0\rangle|0\rangle|0\rangle & = & |0\rangle|0\rangle|0\rangle|0\rangle|1\rangle|0\rangle|1\rangle \\
U|0\rangle|1\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle & = & |0\rangle|0\rangle|0\rangle|0\rangle|1\rangle|1\rangle|0\rangle \\
U|1\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle & = & |0\rangle|0\rangle|0\rangle|0\rangle|1\rangle|1\rangle|1\rangle \\
U|1\rangle|1\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle & = & |0\rangle|0\rangle|0\rangle|1\rangle|0\rangle|0\rangle|0\rangle \qquad (1.8) \\
 & \vdots & \qquad (1.9)
\end{array}$$

In this case the unitary transformation is a matrix that maps a set of 29 orthogonal column vectors on another set of 29 orthogonal vectors in a complex vector space of dimension 2^7. The effect of this unitary transformation is to compress the information about the typical sequences into the last three qubits, while the first four qubits are always in the same pure state $|0\rangle$ and therefore do not carry any information. However, when U acts on other, less likely, sequences it will generate states that have some of the first four qubits in state $|1\rangle$. Now comes the crucial step, we throw away the first two qubits and obtain a sequence of five qubits:

$$\begin{array}{rcl}
|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle & \rightarrow & |0\rangle|0\rangle|0\rangle|0\rangle|0\rangle \\
|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|1\rangle & \rightarrow & |0\rangle|0\rangle|0\rangle|0\rangle|1\rangle \\
|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|1\rangle|0\rangle & \rightarrow & |0\rangle|0\rangle|0\rangle|1\rangle|0\rangle \\
|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|1\rangle|1\rangle & \rightarrow & |0\rangle|0\rangle|0\rangle|1\rangle|1\rangle \\
|0\rangle|0\rangle|0\rangle|0\rangle|1\rangle|0\rangle|0\rangle & \rightarrow & |0\rangle|0\rangle|1\rangle|0\rangle|0\rangle \\
|0\rangle|0\rangle|0\rangle|0\rangle|1\rangle|0\rangle|1\rangle & \rightarrow & |0\rangle|0\rangle|1\rangle|0\rangle|1\rangle \\
|0\rangle|0\rangle|0\rangle|0\rangle|1\rangle|1\rangle|0\rangle & \rightarrow & |0\rangle|0\rangle|1\rangle|1\rangle|0\rangle
\end{array}$$

$$|0\rangle|0\rangle|0\rangle|0\rangle|1\rangle|1\rangle|1\rangle \rightarrow |0\rangle|0\rangle|1\rangle|1\rangle|1\rangle \tag{1.10}$$
$$|0\rangle|0\rangle|0\rangle|1\rangle|0\rangle|0\rangle|0\rangle \rightarrow |0\rangle|1\rangle|0\rangle|0\rangle|0\rangle \tag{1.11}$$

Therefore we have compressed the 7 qubits into 5 qubits. Of course we need to see whether this compression can be undone again. This is indeed the case, when these three qubits are passed on to some other person, this person then adds four qubits all in the state $|0\rangle$ and then applies the inverse unitary transformation U^{-1} and obtains the states in equation 1.7 back. This implies that this person will reconstruct the correct quantum state in at least 93.5% of the cases and he has achieved this sending only 3 qubits. As we showed in the classical case (see equation 1.3), in the limit of very long blocks composed of n qubits, our friend will be able to reconstruct almost all quantum states by sending only $nH(\frac{6}{7}) = 0.59n$ qubits. Note that this procedure also works when we have a superposition of states. For example, the state

$$|\psi\rangle = \alpha|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle + \beta|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle|1\rangle \tag{1.12}$$

can be reconstructed perfectly if we just send the state of three qubits given below:

$$|\psi\rangle = \alpha|0\rangle|0\rangle|0\rangle|0\rangle|0\rangle + \beta|0\rangle|0\rangle|0\rangle|0\rangle|1\rangle \tag{1.13}$$

Therefore not only the states in equation 1.7 are reconstructed perfectly, but also all superpositions of these states.

A very similar procedure would work also when we have a source that emits quantum states $|\psi_i\rangle$ with probabilities p_i, giving rise to an arbitrary density operator $\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|$. Unlike the example in equation 1.4, the states $|\psi_i\rangle$ can be $non-orthogonal$ states of a two level system so the resulting density matrix is *not* in diagonal form. In this slightly more complicated case, the first step consists in finding the eigenvectors and eigenvalues of ρ. As the eigenvectors to different eigenvalues are orthogonal, we are then in the situation of equation 1.4. We can immediately see that the number of qubits that need to be sent, to ensure that the probability with which we can reconstruct the quantum state correctly is arbitrarily close to unity, is given by n times the Shannon entropy of the eigenvalues of ρ which is in turn equal to the von Neumann entropy $S(\rho)$. Since we can reconstruct the quantum state $\rho^{\otimes n}$ of a system composed of n qubits by sending only $nS(\rho)$ qubits, we say that $nS(\rho)$ is the quantum information content of the composite system.

1.4 Quantum State Teleportation

The procedure we will analyse is called quantum teleportation and can be understood as follows. The naive idea of teleportation involves a protocol whereby an object positioned at a place A and time t first "dematerializes" and then reappears at a distant place B at some later time $t + T$. Quantum teleportation implies that we wish to apply this procedure to a quantum object. However, a genuine quantum teleportation differs from this idea, because we are not teleporting the whole object but rather its state from particle A to particle B. As quantum particles are indistinguishable anyway, this amounts to 'real' teleportation.

One way of performing teleportation (and certainly the way portrayed in various science fiction movies, e.g. The Fly) is first to learn all the properties of that object (thereby possibly destroying it). We then send this information as a classical string of data to B where another object with the same properties is re-created. One problem with this picture is that, if we have a single quantum system in an unknown state, we cannot determine its state completely because of the uncertainty principle. More precisely, we need an infinite ensemble of identically prepared quantum systems to be able completely to determine its quantum state. So it would seem that the laws of quantum mechanics prohibit teleportation of single quantum systems. However, the very feature of quantum mechanics that leads to the uncertainty principle (the superposition principle) also allows the existence of entangled states. These entangled states will provide a form of quantum channel to conduct a teleportation protocol. It will turn out that there is no need to learn the state of the system in order to teleport it. On the other hand, there is a need to send some classical information from A to B, but part of the information also travels down an entangled channel. This then provides a way of distinguishing quantum and classical correlations, which we said was at the heart of quantifying entanglement. After the teleportation is completed, the original state of the particle at A is destroyed (although the particle itself remains intact) and so is the entanglement in the quantum channel. These two features are direct consequences of fundamental laws in information processing. I cannot explain these here as I do not have enough time, but if you are interested you should have a look at the article M.B. Plenio and V. Vedral, Contemp. Physics **39**, 431 (1998) which has been written for final year students and first year PhD students.

1.4.1 A basic description of teleportation

Let us begin by describing quantum teleportation in the form originally proposed by Bennett, Brassard, Crepeau, Jozsa, Peres, and Wootters in 1993. Suppose that Alice and Bob, who are distant from each other, wish to implement a teleportation procedure. Initially they need to share a maximally entangled pair of quantum mechanical two level systems. Unlike the classical bit, a qubit can be in a superposition of its basis states, like $|\Psi\rangle = a|0\rangle + b|1\rangle$. This means that if Alice and Bob both have one qubit each then the joint state may for example be

$$|\Psi_{AB}\rangle = (|0_A\rangle|0_B\rangle + |1_A\rangle|1_B\rangle)/\sqrt{2}\ , \tag{1.14}$$

where the first ket (with subscript A) belongs to Alice and second (with subscript B) to Bob. This state is entangled meaning, that it cannot be written as a product of the individual states (like e.g. $|00\rangle$). Note that this state is different from a statistical mixture $(|00\rangle\langle 00| + |11\rangle\langle 11|)/2$ which is the most correlated state allowed by classical physics. Now suppose that Alice receives a qubit in a state which is unknown to her (let us label it $|\Phi\rangle = a|0\rangle + b|1\rangle$) and she has to teleport it to Bob. The state has to be unknown to her because otherwise she can just phone Bob up and tell him all the details of the state, and he can then recreate it on a particle that he possesses. If Alice does not know the state, then she cannot measure it to obtain all the necessary information to specify it. Therefore she has to resort to using the state $|\Psi_{AB}\rangle$ that she shares with Bob. To see what she has to do, we write out the total state of all three qubits

$$|\Phi_{AB}\rangle := |\Phi\rangle|\Psi_{AB}\rangle = (a|0\rangle + b|1\rangle)(|00\rangle + |11\rangle)/\sqrt{2}\ . \tag{1.15}$$

Figure 1.3: A schematic picture of quantum state teleportation. A qubit in an unknown quantum state is entered into a machine which consumes one unit of entanglement (ebit) and a local measurement whose four possible outcomes are transmitted to the receiver. As a result the original state of the qubit is destroyed at the senders location and appears at the receivers end. The mathematical details can be found in [2, 3]

However, the above state can be written in the following convenient way (here we are only rewriting the above expression in a different basis, and there is no physical process taking place in between)

$$\begin{aligned} |\Phi_{AB}\rangle &= (a|000\rangle + a|011\rangle + b|100\rangle + b|111\rangle)/\sqrt{2} \\ &= \frac{1}{2}\Big[|\Phi^+\rangle(a|0\rangle + b|1\rangle) + |\Phi^-\rangle(a|0\rangle - b|1\rangle) \\ &\quad +|\Psi^+\rangle(a|1\rangle + b|0\rangle) + |\Psi^-\rangle(a|1\rangle - b|0\rangle)\Big] \ , \qquad (1.16) \end{aligned}$$

where

$$\begin{aligned} |\Phi^+\rangle &= (|00\rangle + |11\rangle)/\sqrt{2} \qquad (1.17) \\ |\Phi^-\rangle &= (|00\rangle - |11\rangle)/\sqrt{2} \qquad (1.18) \\ |\Psi^+\rangle &= (|01\rangle + |10\rangle)/\sqrt{2} \qquad (1.19) \\ |\Psi^-\rangle &= (|01\rangle - |10\rangle)/\sqrt{2} \qquad (1.20) \end{aligned}$$

form an ortho-normal basis of Alice's two qubits (remember that the first two qubits belong to Alice and the last qubit belongs to Bob). The above basis is frequently called the Bell basis. This is a very useful way of writing the state of Alice's two qubits and Bob's single qubit because it displays a high degree of correlations between Alice's and Bob's parts: to every state of Alice's two qubits (i.e. $|\Phi^+\rangle, |\Phi^-\rangle, |\Psi^+\rangle, |\Psi^-\rangle$) corresponds a state of Bob's qubit. In addition the state of Bob's qubit in all four cases looks very much like the original qubit that Alice has to teleport to Bob. It is now straightforward to see how to proceed with the teleportation protocol:

1. Upon receiving the unknown qubit in state $|\Phi\rangle$ Alice performs projective measurements on her two qubits in the Bell basis. This means that she will obtain one of the four Bell states randomly, and with equal probability.

2. Suppose Alice obtains the state $|\Psi^+\rangle$. Then the state of all three qubits (Alice + Bob) collapses to the following state

$$|\Psi^+\rangle(a|1\rangle + b|0\rangle)\ . \tag{1.21}$$

   (the last qubit belongs to Bob as usual). Alice now has to communicate the result of her measurement to Bob (over the phone, for example). The point of this communication is to inform Bob how the state of his qubit now differs from the state of the qubit Alice was holding previously.

3. Now Bob knows exactly what to do in order to complete the teleportation. He has to apply a unitary transformation on his qubit which simulates a logical NOT operation: $|0\rangle \rightarrow |1\rangle$ and $|1\rangle \rightarrow |0\rangle$. He thereby transforms the state of his qubit into the state $a|0\rangle + b|1\rangle$, which is precisely the state that Alice had to teleport to him initially. This completes the protocol. It is easy to see that if Alice obtained some other Bell state then Bob would have to apply some other simple operation to complete teleportation. We leave it to the reader to work out the other two operations (note that if Alice obtained $|\Phi^+\rangle$ he would not have to do anything). If $|0\rangle$ and $|1\rangle$ are written in their vector form then the operations that Bob has to perform can be represented by the Pauli spin matrices, as depicted in Fig. 1.4.

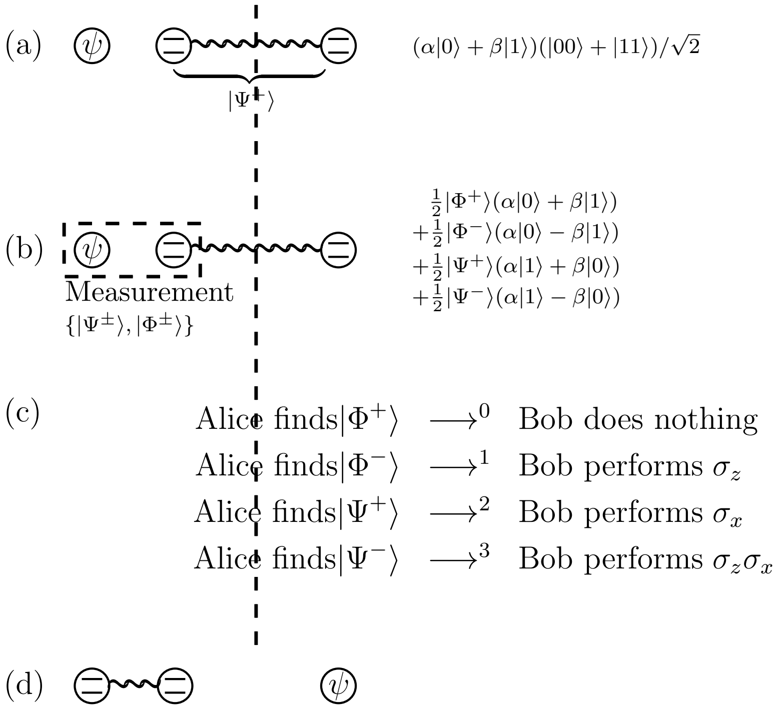
(a)
ψ
$(\alpha|0\rangle + \beta|1\rangle)(|00\rangle + |11\rangle)/\sqrt{2}$
$|\Psi^+\rangle$
(b)
ψ
$\frac{1}{2}|\Phi^+\rangle(\alpha|0\rangle + \beta|1\rangle)$
$+\frac{1}{2}|\Phi^-\rangle(\alpha|0\rangle - \beta|1\rangle)$
$+\frac{1}{2}|\Psi^+\rangle(\alpha|1\rangle + \beta|0\rangle)$
$+\frac{1}{2}|\Psi^-\rangle(\alpha|1\rangle - \beta|0\rangle)$
Measurement
$\{|\Psi^\pm\rangle, |\Phi^\pm\rangle\}$
(c)
Alice finds$|\Phi^+\rangle$ $\longrightarrow^0$ Bob does nothing
Alice finds$|\Phi^-\rangle$ $\longrightarrow^1$ Bob performs σ_z
Alice finds$|\Psi^+\rangle$ $\longrightarrow^2$ Bob performs σ_x
Alice finds$|\Psi^-\rangle$ $\longrightarrow^3$ Bob performs $\sigma_z\sigma_x$
(d)
ψ

Figure 1.4: The basic steps of quantum state teleportation. Alice and Bob are spatially separated, Alice on the left of the dashed line, Bob on the right. (a) Alice and Bob share a maximally entangled pair of particles in the state $(|00\rangle + |11\rangle)/\sqrt{2}$. Alice wants to teleport the unknown state $|\psi\rangle$ to Bob. (b) The total state of the three particles that Alice and Bob are holding is rewritten in the Bell basis Eqs. (1.17-1.20) for the two particles Alice is holding. Alice performs a measurement that projects the state of her two particles onto one of the four Bell states. (c) She transmits the result encoded in the numbers $0, 1, 2, 3$ to Bob, who performs a unitary transformation $\mathbf{1}, \sigma_z, \sigma_x, \sigma_z\sigma_x$ that depends only on the measurement result that Alice obtained but **not** on the state $|\psi\rangle$! (d) After Bob has applied the appropriate unitary operation on his particle he can be sure that he is now holding the state that Alice was holding in (a).

An important fact to observe in the above protocol is that all the operations (Alice's measurements and Bob's unitary transformations) are *local* in nature. This means that there is never any need to perform a (global) transformation or measurement on all three qubits simultaneously, which is what allows us to call the above protocol a genuine teleportation. It is also important that the operations that Bob performs are independent of the state that Alice tries to teleport to Bob. Note also that the classical communication from Alice to Bob in step 2 above is crucial because otherwise the protocol would be impossible to execute (there is a deeper reason for this: if we could perform teleportation without classical communication then Alice could send messages to Bob faster than the speed of light, remember that I explained this in a previous lecture.

Important to observe is also the fact that the initial state to be teleported is at the end destroyed, i.e it becomes maximally mixed, of the form $(|0\rangle\langle 0| + |1\rangle\langle 1|)/2$. This has to happen since otherwise we would end up with two qubits in the same state at the end of teleportation (one with Alice and the other one with Bob). So, effectively, we would clone an unknown quantum state, which is impossible by the laws of quantum mechanics (this is the no-cloning theorem of Wootters and Zurek).

1.5 Two conclusions

After you have learnt about teleportation we can now use the idea of teleportation to define the unit of entanglement. A unit of entanglement (ebit) is that amount of quantum correlations that allow us to teleport one qubit of quantum information perfectly. While this definition is perfectly acceptable there are other ways to define entanglement and those will be discussed by Ignacio Cirac in the following set of lectures.

The other idea that will be elaborated on in the lectures by Rolf Tarrach is that of quantum cryptography. If Alice and Bob share a maximally entangled state, then they can use this state for teleportation of an unknown quantum state. But they can also complete a simpler task, namely Alice can instead take a bit of a classical message and encode it in two orthogonal quantum states: the letter 0 is encoded in state $|0\rangle$ and the letter 1 is encoded in state $|1\rangle$. Of course she cannot simply transmit the state from to Bob as an eavesdropper could just measure the state and determine which one it is. However, Alice can use her ebit that she shares with Bob and teleport her state to Bob. Now an eavesdropper gets useless classical bits of information that do not tell him anything about the message that Alice wishes to transmit. Therefore Alice has sent a secret message to Bob.

This first set of lectures is based on a number of articles and lecture notes from which I have taken parts. These are my own lecture notes on Quantum Mechanics from which the teleportation section originates [4]. Furthermore I have used two introductory articles that I have co-written for Contemporary Physics and which aim at third year physics students. These are [3] and [5]. if you want to dive deeper into quantum information theory you may want to have a look at the excellent book of Nielsen and Chuang [6].

Chapter 2

Distinguishability and entanglement measures

2.1 Typical and atypical messages

In the first set of lectures I have introduced the key ideas on classical and quantum information and in particular I have demonstrated how to quantify classical information, classical correlations and quantum information. However, the goal of quantifying quantum correlations has not been dealt with thoroughly. Ignacio Cirac gave you some first insights into the problem and he defined some entanglement measures. He approached the problem via local operations and classical communication and the constraints arising from this. In these lectures I will approach the problem from a different viewpoint which may initially seems a bit surprising and unrelated. We will attempt to develop a notion of how different or how similar two quantum states are. Then the key idea will be that we would like to be able to say that a state is more entangled the further it is away from all unentangled states. This basic idea will then be turned into a quantitative form and I will discuss some properties of the new entanglement measure. In doing so I will introduce a quantity that has proven to be very important in quantum information theory, namely the relative entropy of entanglement.

As usual I will start my considerations with the classical case as this I technically easier. Then I will transfer these ideas to quantum mechanics and use them to quantify entanglement.

2.1.1 Typical and atypical sequences

In my first lectures you have learnt that messages can often be compressed and the reason for that was that usually messages have a very typical form. Remember what this meant. If you take a coin and it is not quite a fair coin, then say the probability that the coin shows up is $p_{up} = 0.7$ and that it shows down $p_{down} = 0.3$. Now let us throw this coin N times. Clearly you would expect that normally you will find roughly $0.7N$ times up and roughly $0.3N$ times down. In fact you would expect that the variation around the mean number is at most of the order of $\sqrt{N0.70.3}$ which you obtain from the binomial distribution. Therefore finding $0.7N \pm \sqrt{N0.70.3}$ times up would be the typical result. Of course as you know from life or the law of large numbers the typical result will become more and more likely the more often you throw the coin, ie deviations from this typical behaviour will become less and less frequent. Now you may ask quite naturally how infrequent do these deviations become? Is there a simple law governing this? These are questions I am going to answer in the following. I do not want to bother you with mathematical details which can for example be found in [1] but I will rather give a slightly handwaving derivation of the correct result.

2.1.2 Distinguishing coins and the relative entropy

Let us make some experiments. I will now give you a coin but I will not tell you what the probabilities for up and down are. You will have to estimate this by making experiments, ie throwing the coin. Let us say that you throw the coin N times. You will then find some number of up, say $q_{up}N$, and down, say $q_{down}N = (1-q_{up})N$. Now I ask you to make a guess as to which are actually the probabilities for up and down, ie from a limited set of data you have to conclude the true properties of the coin. Well, the best guess you can make is that indeed the coin has probabilities q_{up} for 'up' and q_{down} for 'down'. But this may not be the correct guess, simply because we have statistical fluctuations. I give you an extreme example:

Imagine the following situation. I have given you a fair coin, ie $p_{up} = 1/2 = p_{down}$. Now you throw it N times. Clearly, there is a small chance, namely 2^{-N}, that you will find in all your throws that the coin shows 'up'. Then you would guess that the coin is in fact totally unfair. But thats a mistake because I had passed you a fair coin.

Now in general we would like to know how likely it is that a coin which has probability q for showing up will, in N tries, actually produce pN up!? We find by Stirlings formula $(n! \cong n^n e^{-n})$

$$\begin{aligned}
\ln Q &= \ln\left(q^{pN}(1-q)^{(1-p)N}\begin{pmatrix} N \\ pN \end{pmatrix}\right) \\
&\cong pN\ln q + (1-p)N\ln(1-q) + N\ln N - N \\
&\quad -(Np\ln Np - Np) - ((1-p)N\ln N(1-p) - N(1-p)) \\
&\cong N(-p\ln p - (1-p)\ln(1-p) + p\ln q + (1-p)\ln(1-q))
\end{aligned}$$

or when written for logarithm to base 2

$$\begin{aligned}
\log_2 Q &= N(-p\log_2 p - (1-p)\log_2(1-p) + p\log_2 q + (1-p)\log_2(1-q)) \\
&\quad \text{or equivalently} \\
Q &= 2^{-NS(p||q)}
\end{aligned}$$

where we defined the relative entropy (also sometimes called the Kullback-Leibler distance)

$$S(p||q) = p\log_2 p + (1-p)\log_2(1-p) - p\log_2 q - (1-p)\log_2(1-q)$$

Now you can calculate for any situation the probability for making the wrong guess that the coin has probability q for up if in fact it has probability p. So the relative entropy is indeed some measure that tells you how likely it is that a certain coin, characterized by a probability distribution $\{p_i\}$ behaves like a coin with probability distribution $\{q_i\}$ and therefore we can use the relative entropy as a measure of separation between the two probability distributions, ie two probability distributions are more different if it is less likely that one will behave like the other in an experiment.

Now look at the relative entropy again! It is actually not symmetric with respect to interchange of p and q! Thats a bit weird as we would expect it to be as likely to falsely assume q if p is correct as the other way round. Thats wrong, but why? Lets look at our extreme example again but with interchanged p and q. Now I pass you a coin that has probability $q = 1$ to show 'up'. You throw it N times and indeed you find N times 'up' and your inference is that the coin has $q = 1$ and you are right - always. The reason for this behaviour is that different sequences have different probabilities to behave atypical, in fact the totally unfair coin never behaves atypical! This justifies the asymmetry of the relative entropy. In fact if for the two examples you plug the number into our little formula for Q, you get exactly the right answers.

2.1.3 The quantum version of all this

Now you will of course ask whether there is a quantum version of all this and indeed there is. However, now things are a lot more difficult to see partly because there are so many more kinds of measurements that you can do. Therefore I will not present you with any details of derivations, but I will just state the essential result again in terms of making guesses about quantum states.

The probability of falsely inferring that one holds N copies of a quantum state σ if one is in fact given N copies of a quantum state ρ is for large N given by

$$p(\rho \to \sigma) = 2^{-NS(\sigma||\rho)} \tag{2.1}$$

where the quantum relative entropy is defined as

$$S(\sigma||\rho) = tr\{\sigma \log \sigma - \sigma \log \rho\}$$

Clearly, when both σ and ρ commute, then this definition goes over to the classical relative entropy of entanglement.

So, also in quantum mechanics we can use the relative entropy as a measure of how different two density operators are. Before I will proceed to study entanglement measures again, I will give you a few more properties of the relative entropy.

Firstly the relative entropy is always non-negative, ie

$$\forall \sigma, \rho : S(\sigma||\rho) \geq 0$$

and

$$S(\sigma||\rho) = 0 \Leftrightarrow \sigma = \rho$$

The interpretation of this is quite clear. The first statement says that the probability for falsely assuming σ given ρ is always smaller or equal to 1 and decreases with increasing N, which implies that the relative entropy is non-negative. The second statement says that exactly when $\sigma = \rho$ then the probability to confuse them is 1.

Worryingly though the relative entropy does not have any other properties that you would normally expect from a mathematical distance measure. The relative entropy is not symmetric and it does not satisfy a triangular inequality.

The relative entropy is a jointly convex function which means that for all $p_i \geq 0$ and $\sum_i p_i = 1$ we have

$$\forall \rho_i, \sigma_i : S(\sum p_i \sigma_i || \sum p_i \rho_i) \leq \sum p_i S(\sigma_i | \rho_i) \tag{2.2}$$

What does this mean? If we have systems in different states and we mix them together, ie we forget which system is in which individual state, then we discard information. Discarding this information makes systems less distinguishable.

After these few properties, let us now move on to study the application of the relative entropy to entanglement measures.

2.1.4 Another entanglement measure

As you have learnt from Ignacio Cirac, there are various entanglement measures. Two basic ones are the entanglement of formation and the entanglement of distillation. These two measures are however not the only ones. But before I introduce the new measure let us first think briefly what sort of property an entanglement measure should possess. Clearly, if the state is not entangled then the measure should indicate this by taking the value zero. Furthermore, it cannot make a difference in which local basis one considers the state. As any local basis change is implemented by a local unitary transformation which is reversible, the amount of entanglement should remain unchanged under these operations. Finally, general local operations and classical communication with subselection can not increase entanglement and therefore any entanglement measure should not increase under these operations. Finally you have learnt from Ignacio Ciracs lectures that the unique measure of entanglement for pure states is given by the entropy of entanglement. Therefore we would hope that any measure of entanglement reduces to the entropy of entanglement for pure states. Let me state these constraints more formally.

E1. $E(\sigma) = 0$ iff σ is separable (or if it is ppt).

E2. Local unitary operations leave $E(\sigma)$ invariant, i.e. $E(\sigma) = E(U_A \otimes U_B \sigma U_A^\dagger \otimes U_B^\dagger)$.

E3. The expected entanglement cannot increase under LGM+CC+PS given by $\sum V_i^\dagger V_i = \mathbb{1}$, i.e.

$$\sum tr(\sigma_i)\ E(\sigma_i/tr(\sigma_i)) \le E(\sigma)\ \ , \tag{2.3}$$

where $\sigma_i = V_i \sigma V_i^\dagger$.

E4. For pure states the entanglement is given by the entropy of entanglement, ie the von Neumann entropy of the reduced density operator.

All these properties are satisfied by both the entanglement of formation and the entanglement of distillation. Now I would like to introduce another entanglement measure, the relative entropy of entanglement, which satisfies all these criteria.

I do not want this measure to drop from the sky. Therefore I will explain to you how it was actually found. As I have explained to you in the last lecture, a good measure for the correlations in classically correlated random variables is given by the mutual information

$$I = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y)$$

Now I would like to show you something remarkable, namely that this expression can be rewritten in terms of the classical relative entropy.

Given a joint probability distribution $p(x,y)$ and its marginal distributions $p(x)$ and $p(y)$ then we have

$$\begin{aligned} S(p(x,y)|p(x)\cdot p(y)) &= \sum_{x,y}[p(x,y)\log_2 p(x,y) - p(x,y)\log_2(p(x)p(y))] \\ &= -H(X,Y) - \sum_{x,y} p(x,y)(\log_2 p(x) + \log_2 p(y)) \\ &= -H(X,Y) - \sum_x p(x)\log_2 p(x) - \sum_y p(y)\log_2 p(y) \\ &= H(X) + H(Y) - H(X,Y) \end{aligned}$$

Now let us carry this over to quantum mechanics. One could expect that the quantum version of the mutual information is a measure of all the correlations in a state σ and it would be given by

$$I_{QM} := S(\sigma_A) + S(\sigma_B) - S(\sigma)$$

where σ_A is the reduced density operator of system A and correspondingly σ_B is the reduced density operator of system B. But again this can be expressed in terms of the relative entropy and we find

$$I_{QM} := S(\sigma|\sigma_A \otimes \sigma_B)$$

Now this quantity clearly does not satisfy all the criteria that I have demanded from an entanglement measure, so that we still have to amend it a bit. What is going wrong is that with local operations and classical communication we can create classically correlated states, but we are comparing our state σ with a completely uncorrelated product state. In addition we are not really interested in all the correlations, we really would like to know amount of quantum correlations. So maybe it would make more sense to compare our state σ to a classically correlated state and see how different they are. But which classically correlated state? Well, we do not know, so the best we can do is to compare to all the classically correlated states and then find the closest one (see figure). The idea behind this is to take account of all the classical correlations in a state so that the remaining correlations should somehow be of quantum nature. Therefore we would define our entanglement measure as

$$E_R(\sigma) = \min_{\rho\in\mathcal{D}} S(\sigma||\rho)$$

Well, this is the idea, which you can also see in the figure. But now, having seen the idea one really has to sit down and prove that the so defined measure actually satisfies all the requirements that I have written down. Well, this is in fact quite tricky, especially when one does it the first time round. Therefore I am not going to demonstrate this. For those of you who want to see the proofs, have a look some of the original papers [7].

Theorem $E(x_1\sigma_1+x_2\sigma_2) \leq x_1E(\sigma_1)+x_2E(\sigma_2)$, where $x_1+x_2=1$. **Proof**. This property follows from the convexity of the quantum relative entropy in both arguments

$$S(x_1\sigma_1 + x_2\sigma_2||x_1\rho_1 + x_2\rho_2) \leq x_1S(\sigma_1||\rho_1) + x_2S(\sigma_2||\rho_2) \quad . \tag{2.4}$$

Now,

$$\begin{aligned} E(x_1\sigma_1 + x_2\sigma_2) &\leq S(x_1\sigma_1 + x_2\sigma_2||x_1\rho_1^* + x_2\rho_2^*) \\ &\leq x_1S(\sigma_1||\rho_1^*) + x_2S(\sigma_2||\rho_2^*) \\ &= x_1E(\sigma_1) + x_2E(\sigma_2) \enspace , \end{aligned} \tag{2.5}$$

which completes our proof of convexity □. This is physically a very satisfying property of an entanglement measure. It says that when we mix two states having a certain amount of entanglement we cannot get a more entangled state, i.e. succinctly stated “mixing does not increase entanglement”. This is what is indeed expected from a measure of entanglement to predict.

As a last property we state that the entanglement of creation E_c is never smaller than the Relative Entropy of Entanglement E. We will show later that this property has the important implication that the amount of entanglement that we have to invest to create a given quantum state is usually larger than the entanglement that you can recover using quantum state distillation methods.

Figure 2.1: The set of all density matrices, $\mathcal{T}$ is represented by the outer circle. Its subset, a set of disentangled states $\mathcal{D}$ is represented by the inner circle. A state σ belongs to the entangled states, and ρ^* is the disentangled state that minimizes the distance $D(\sigma||\rho)$, thus representing the amount of quantum correlations in σ. State $\rho^*_A \otimes \rho^*_B$ is obtained by tracing ρ^* over A and B. $D(\rho^*||\rho^*_A \otimes \rho^*_B)$ represent the classical part of the correlations in the state σ.

Theorem $E(\sigma) \leq E_c(\sigma) = \min_{\rho\in\mathcal{D}} S(\sigma||\rho)$.
Proof. Given a state σ then by definition of the entanglement of creation there is a convex decomposition $\sigma = \sum p_i\sigma_i$ with pure states σ_i such that

$$E_c(\sigma) = \sum p_i E_c(\sigma_i) \ . \tag{2.6}$$

As the entanglement of creation coincides with our entanglement for pure states and as our entanglement is convex it follows that

$$E_c(\sigma) = \sum p_i E_c(\sigma_i) = \sum p_i E(\sigma_i) \geq E(\sum p_i\sigma_i) = E(\sigma) \tag{2.7}$$

and the proof is completed $\square$.

The physical explanation of the above result lies in the fact that a certain amount of additional knowledge is involved in the entanglement of formation which gives it a higher value to the Relative Entropy of Entanglement. Let me add that the relative entropy of entanglement $E(\sigma)$ can be calculated easily for Bell diagonal states [7]. Comparing the result to those for the entanglement of formation [8] one finds that, in fact, strict inequality holds.

2.2 Thermodynamics of Entanglement: Purification Procedures

There are two ways to produce an upper bound to the efficiency of any purification procedure. Using condition E3 and the fact that the Relative Entropy of Entanglement is additive, we can immediately derive this bound. However, this bound can be derived in an entirely different way. In this section we now abandon conditions E1-E3 and use only methods of the previous section to put an upper bound to the efficiency of purification procedures. In particular, we show that the entanglement of formation is in general larger than the entanglement of distillation. This is in contrast with the situation for pure states where both quantities coincide. The Quantum Relative Entropy is seen to play a distinctive role here, and is singled out as a 'good' generator of a measure of entanglement from among other suggested candidates.

Figure 2.2: Comparison of the entanglement of creation and the Relative Entropy of Entanglement for the Werner states (these are are Bell diagonal states of the form $W = \text{diag}(F, (1-F)/3, (1-F)/3, (1-F)/3.)$ One clearly sees that the entanglement of creation is strictly larger than the Relative Entropy of Entanglement for $0 < F < 1$

2.2.1 Distinguishability and Purification Procedures

In the previous section we presented a statistical basis to the Relative Entropy of Entanglement by considering distinguishability of two (or more) quantum states encapsulated in the form of the Quantum Sanov Theorem. We now use this Quantum Sanov Theorem to put an upper bound on the amount of entanglement that can be distilled using any purification procedure. This line of reasoning follows from the fact that any purification scheme can be viewed as a measurement to distinguish entangled and disentangled quantum states. Suppose that there exist a purification procedure with the following property

- Initially there are n copies of the state σ. If σ is entangled, then the end product is $0 < m \leq n$ singlets and $n - m$ states in $\rho \in \mathcal{D}$. Otherwise, the final state does not contain any entanglement, i.e. $m = 0$ (in fact, there is nothing special about singlets: the final state can be any other known, maximally entangled state because these can be converted into singlets by applying local unitary operations).

Note that we can allow the complete knowledge of the state σ. We also allow that purification procedures differ for different states σ. Perhaps there is a "universal" purification procedure independent of the initial state. However, in reality, this property is hard to fulfill. At present the best that can be done is to purify a certain class of entangled states. The above is therefore an idealization that might never be achieved. Now, by calculating the upper bound on the efficiency of a procedure described above we present an absolute bound for any particular procedure. We ask: "What is the largest number of singlets that can be produced (distilled) from n pairs in state σ"? Suppose that we produce m pairs. This has to be compared to a disentangled state. Let us now project them *non–locally* onto the singlet state. The procedure will yield positive outcomes (1) with certainty so long as the state we measure indeed is a singlet. The best that you can obtain from a disentangled state is a positive result with probability of $\frac{1}{2}$. Suppose that after performing singlet projections onto all m particles we get a string of m 1's. From this we conclude that the final state is a singlet (and therefore the initial state σ was entangled). However, we could have made a mistake. But with what probability? The answer is as follows: the largest probability of making a wrong inference is 2^{-m} (if the state that we were measuring had an overlap with a singlet state of 1/2). On the other hand, if we were measuring σ from the very beginning (without performing the purification first), then the probability (i.e. the lower bound) of the wrong inference would be $2^{-nE(\sigma)}$. But, purification procedure might waste some information (i.e. it is just a particular way of distinguishing entangled from disentangled states, not necessarily the best one), so that the following has to hold

$$2^{-nE(\sigma)} \leq 2^{-m} \; , \tag{2.8}$$

which implies that

$$nE(\sigma) \geq m \ , \tag{2.9}$$

i.e. we cannot obtain more entanglement than is originally present. This, of course, is also directly guaranteed by our condition E3. The above, however, was a deliberate exercise in deriving the same result from a different perspective, abandoning conditions E1–E3. Therefore the relative entropy of entanglement can be used to provide an upper bound on the efficiency of any purification procedure.

Actually, in the above considerations we implicitly assumed that the entanglement of n pairs, equivalently prepared in the state σ, is the same as $n \times E(\sigma)$. We already indicated that this is a conjecture with a strongly supported basis in the case of the Quantum Relative Entropy.

Now let us make up a formal analogy between thermodynamical engines and purification procedures. This analogy has the advantage that it becomes clear why there is a difference between the entanglement of formation and the entanglement of distillation.

Let us first consider entanglement purification. In this process we take pairs of particles which have a low degree of entanglement. Then we apply local measurements and other quantum mechanical operations to convert the pairs into some pairs with a high degree of entanglement and some pairs with no entanglement. But this is not the only thing the machine produces. It also generates a measurement record (all the outcomes of the measurement). These outcomes will have to be erased and therefore heat will be created (This is Landauers principle which is in fact equivalent to the second law of thermodynamics, for more info look at [5]). This introduces a degree of irreversibility into the game, as the generation of heat is an irreversible thermodynamic process. From Ignacio Ciracs lectures you also learnt about the reverse process, namely that of creating an entangled state by local operations from a resources of pure singlet states. Again, if we wish to create a mixed entangled state from this resource we will have to erase some information and this again leads to a generation of heat and thereby to some irreversibility. Therefore you can see that both processes, distillation and formation, are potentially irreversible as they may create heat. This is the physical origin of the difference between entanglement of formation and distillation.

Here I have to stop, but of course there would be a lot more to tell. If you have become interested, then I recommend to you to read some of the references in the back of these notes. Maybe the best recommendation would be to start with the two introductory articles [3, 5] and then move to more advanced articles some of which will be cited in [5].

Bibliography

[1] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 1991.

[2] C.H. Bennett, G. Brassard, C. Crepeau, R. Jozsa, A. Peres, W.K. Wootters, Physical Review Letters **70**, 1895 (1993).

[3] M.B. Plenio and V. Vedral, *Teleportation, Entanglement and Thermodynamics in the Quantum World*, Contemporary Physics **39**, 431 (1998).

[4] http://www.lsr.ph.ic.ac.uk/ plenio/lecture.html

[5] M.B. Plenio and V. Vitelli, *The Physics of Forgetting: Landauer's erasure principle and information theory*, Contemporary Physics **42**, 25 (2001)

[6] M.A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information Theory*, Cambridge University Press, Cambridge 2000.

[7] V. Vedral and M.B. Plenio, *Entanglement Measures and Purification Procedures*, Physical Review A **57**, 1619 (1998); M.B. Plenio, S. Virmani and P. Papadopoulos, *Operator monotones, the reduction criterion and the relative entropy*, J. Phys. A **33**, 193 (2000)

[8] C.H. Bennett, D.P. DiVincenzo, J.A. Smolin, and W.K. Wootters, Phys. Rev. A **54**, 3824 (1996).