UNIVERSITY OF LONDON
IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

Department of Physics

Optics Section

QUANTUM INFORMATION THEORY OF ENTANGLEMENT

by

Vlatko Vedral

A Thesis submitted for the degree of Doctor of Philosophy of the University of London and for the Diploma of Membership of the Imperial College

March 1998

To the Memory of my Grandfather ...

Every man gets a narrower and narrower field of knowledge in which he must be an expert in order to compete with other people. The specialist knows more and more about less and less and finally knows everything about nothing.

Konrad Lorenz

Abstract

Classical correlations are described consistently within classical information theory. This thesis presents a consistent quantum information theory of purely quantum correlations, i.e. entanglement. The main problem arises when we consider mixed states, for which it is difficult to separate quantum from purely classical correlations. This problem is the main subject of the thesis and is undertaken from two different perspectives. The first approach follows Shannon's own approach, where we define a number of intuitively clear and physically sound conditions that a "good" measure of entanglement has to satisfy, and then search for measures satisfying these conditions. Our second approach is to extend the classical idea of distinguishing two probability distributions to quantum physics. The amount of entanglement will then determine the experimental ability to distinguish a given entangled state from a classical, disentangled state. We show that these two approaches have a number of features in common, leading to the same measures of entanglement.

Classical information can be spoilt due to interactions with the environment. Classical information theory has a branch dealing with methods for protecting information called classical error correction. Quantum information is even more fragile and here we develop the quantum analogue of error correction. We develop a code that protects quantum states in the presence of spontaneous emission. We then show how to protect entanglement using this method.

We also present a cavity QED implementation of various schemes aiming at increasing and protecting entanglement between two cavities using the standard Jaynes-Cummings interaction model between an atom and a cavity.

Acknowledgements

I would like to thank Prof. Peter Knight for his supervision, patience and encouragement during the entire length of my PhD research, for reading my thesis drafts and for giving me many opportunities. His help in finding financial support enabled me to undertake and complete my PhD.

I would also like to thank Dr. Artur Ekert for discussing many interesting ideas with me and helping me in my research in more than one way. He introduced me to all the most exciting aspects of what he calls "quantum schmantum" business and a number of my successful ideas originated in our conversations. I can say without exaggeration that his role has been that of my "second" supervisor.

My research has benefitted extremely from conversations and collaborations with Martin Plenio. I thank him for patiently listening to all my succesful and unsuccesful ideas. Collaborations and discussions with him, and above all his friendship and encouragement, have made my PhD a very exciting and enjoyable adventure.

Very special thanks goes to my best man Marcos Protopapas. I will never forget the coffee breaks on level 5, Huxley building, where we have frequently discussed so many ideas related to physics, life and universe in general. His friendship and help will always be of greatest value to me. His crazy ideas have kept me sane during some of the darkest moments of my life.

I also thank the following people on various collaborations resulting in a number of publications: Sougato Bose, Vladimir Bužek, Mark Hillery, Kurt Jacobs, Mio Murao, Sandu Popescu and Mike Rippin.

I would like to thank my home team on many enjoyable discussions: Konrad Banaszek, Emma Coghill, Barry Garraway, Thornton Greenland, Philip Gou, Florence Gauthey, Daniel Jonathan, Christoph Keitel, Manos Paspalakis, Akshay Patel, Richard Marshall, Aliki Ritsataki, Jörg Steinbach, Jason Twamley and David Zdravković. I have learnt so much from so many people throughout the last three

years. Here I mention only some of those who had considerable influence upon me: Adriano Barenco, Mathew Donald, Chris Fuchs, Peter Landsberg, Chiara Macchiavello, Masonao Ozawa and Andy Steane.

I am deeply indebted to my parents for their patience and support. They are responsible for my great love for science and for my belief that scientific reasoning can and should be applied to everyday life. They taught me always to look forward and without them I would never have made it.

My wife Ivona has been with me through thick and thin. She showed me that hounesty is the greatest virtue of all and love the deepest form of human understanding. Relationship with her is without any doubt the most meaningful and precious thing I have. Her belief in me, her patience and encouragement, and countless hours spent listening to my ideas and stories from my work have been invaluable to me. This thesis belongs to her as much as it does to me.

I would also like to thank the Knight trust for supporting the last year and a half of my studies, the Science and Engineering Research Council (EPSRC) for its financial support in the first year, and the Abdus Salam Trust for additional financial help during my first year.

Contents

1 Overview **2**

2 Theoretical Background **8**

2.1 Classical Information Theory 8

2.1.1 Measures of Uncertainty and Correlations 9

2.1.2 Classical Communication Theory 16

2.1.3 Classical Error Correction 18

2.2 Comparing Thermodynamical and Shannon's Entropy 23

2.3 Information Theory and Statistics 27

2.3.1 The Theory of Types 27

2.3.2 Data Compression and Sanov's theorem 29

3 Quantum Information Theory **32**

3.1 Quantum Correlations 32

3.1.1 Entanglement and Schmidt Decomposition 33

3.1.2 Quantum Measures of Uncertainty and Correlations 36

3.1.3 Bell's Inequalities 44

3.2 Quantum Communication Theory 49

3.2.1 Disentangled Channels are Additive 52

3.2.2 Entangled Channels are Superadditive 53

3.3 Quantum Computation 55

3.3.1 Quantum Gates 55

3.3.2 Shor's Algorithm 58
3.3.3 Practical Realisations of Quantum Computers 61

4 Entanglement Measures and Purification Procedures 71
4.1 Introduction 71
4.2 Mathematical Prelude 75
4.2.1 Purification Procedures 75
4.2.2 Purification of Pure States 78
4.2.3 Quantification of Entanglement 80
4.2.4 Proofs 83
4.2.5 Two Realisations of $D(\sigma, \rho)$ 85
4.3 Numerics for Two Spin 1/2 Particles 100
4.4 Statistical Basis of Entanglement Measure 108
4.5 Thermodynamics of Entanglement 112
4.6 More Than two Subsystems 117
4.7 Conclusions 118

5 Quantum Error Correction 120
5.1 Introduction 120
5.2 General Conditions 122
5.3 Error Correction in the Presence of Spontaneous Emission 126
5.3.1 Spontaneous Emission Dynamics 127
5.3.2 Single error correcting codes 129
5.3.3 Correcting spontaneous emission 132
5.4 Reliable Quantum Computation from Unreliable Components . . . 138
5.5 Conclusions 141

6 Cavity QED Implementations of Purification Procedures 142
6.1 Introduction 142
6.2 Jaynes-Cummings Model 143

6.2.1 Quantization of EM Field 143
6.2.2 Spin–Boson Interaction Dynamics 145
6.3 Atom–Cavity Models 146
6.3.1 Cavity Models With Local Feedback 148
6.3.2 Increasing Entanglement Non-Locally 155
6.4 Local Error Correction Preserves Correlations 157
6.4.1 Theoretical Considerations 157
6.4.2 Example With Cavities 158
6.5 Conclusion 161

7 Conclusions **163**
7.1 Summary of the Thesis 163
7.2 Further Work 166

Bibliography **169**

Publications **181**

List of Figures

3.1 Truth tables and graphical representations of the elementary quantum gates used for the construction of more complicated quantum networks. The control qubits are graphically represented by a dot, the target qubits by a cross. i) NOT operation. ii) Control–NOT. This gate can be seen as a *bitwise* "copy operation" in the sense that a target qubit (b) initially in the state 0 will be after the action of the gate in the same state as the control qubit. iii) TOFFOLI gate. This gate can also be seen as a Control–control–NOT: the target bit (c) undergoes a NOT operation only when the two controls (a and b) are in state 1. 56

3.2 Schematic picture of a linear ion trap computer. Electrodes generate a time dependent electric field which generates an effective potential such that a string of ions (the blue dots in the middle of the trap) is trapped. The motion of the ions, and in particular the centre of mass mode, has to be cooled to its ground state. The centre of mass mode then acts as a bus that allows us to generate interactions between any two ions. 63

3.3 In part a) the centre-of-mass mode is illustrated. All the ions oscillate with the same phase. In part b) a mode of higher frequency is given. Here the ions have different phases and their relative distances change. 64

3.4 The vertical axis gives the energy while the horizontal axis gives the degree of excitation of the centre-of-mass mode. In $|xy\rangle$ the first number x denotes the internal degree of freedom of the ion, while the second number y denotes the degree of excitation of the centre of mass mode. 67

4.1 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. 74

4.2 Local general measurements (LGM) by Alice and Bob and classical communication (CC) between Alice and Bob which correlates local general measurements. 75

4.3 Subselection (SS) according to measurement results: the original ensemble of entangled pairs E is split under the action of LGM+CC into a number of subensembles which contain pairs with different degrees of entanglement. 77

4.4 Comparison of the entanglement of formation and the Relative Entropy of Entanglement for the Werner states (these are are Bell diagonal states of the form $W = \mathrm{diag}(F, (1-F)/3, (1-F)/3, (1-F)/3.)$ One clearly sees that the entanglement of formation is strictly larger than the relative entropy of entanglement for $0 < F < 1$. 116

5.1 The encoding network: R describes a one bit rotation which takes $|0\rangle \rightarrow (|0\rangle + |1\rangle)/\sqrt{2}$ and $|1\rangle \rightarrow (|0\rangle - |1\rangle)/\sqrt{2}$. An encircled cross denotes a NOT operation while a dot denotes a control bit. For a filled circle the operation is carried out if the control bit is 1; for an empty circle the operation is carried out if the control bit is 0. A circle with a π represents multiplication with phase $\exp(i\pi)$. Qubit 3 is in the state $|\psi\rangle$ that we wish to encode, while all other qubits are initially in their ground state $|0\rangle$. 134

6.1 The experimental setup for local interactions: two cavities are initially entangled in a state of the form $|\Psi\rangle_{AB} = \alpha|n\rangle_A|m\rangle_B + \beta|n'\rangle_A|m'\rangle_B$, and atoms are sent through cavity A only. 150

6.2 The encoding network for protecting against amplitude errors is shown in the upper diagram: the encircled cross denotes a NOT operation while a dot denotes a control bit, together making a CNOT operation. The atoms are initially in their ground states, and the order in which the gates are executed is irrelevant. The lower diagram gives a truth table for the CNOT operation; here, 'C' and 'T' represent *control* and *target* bits respectively. 159

List of Tables

5.1 We obtain an error syndrome, i.e. the state of all qubits except qubit 3, depending on the error that occurred and the place in which it occurred. P_i indicates a sign change of the upper level of qubit i, A_i an amplitude error which is given by the transformation $|0\rangle \leftrightarrow |1\rangle$. The product of both applied to the same qubit gives the third kind of error. Note that the error syndrome is not able to distinguish between P_i and P_{9-i} which leads to global phases in some of the corrected states. This table does not take into account that before and after the error a conditional time evolution takes place. 136

Chapter 1

Overview

Information theory initially developed as a branch of communication theory dealing with the fundamental questions of information manipulation. In a nutshell, it provides answers to the two fundamental questions concerning the ultimate limit of data compression (given by the entropy, S) and the ultimate limit of communication (given by the maximum amount of mutual information, I, over all possible inputs to a communication channel) [1]. Today, information theory is an extremely successful and widely used theoretical tool, having important implication in thermodynamics and statistical physics [2], probability theory, statistics [3], economics, mathematics, and computer science [4]. When we say information theory, we generally mean *classical* information theory, which implies that the information that we manipulate is written into binary digits–bits, for short. A bit represents two different logical states, conventionally labelled as 0 and 1. When information is processed in any way, we have to use physical systems in order to represent bits. Take as an example a modern computer: the electrical circuits that make up the building blocks of modern computers, electronic chips, consist of wires conducting electrical current. These wires have two basic states of existence: either there is a current flowing through them, which is taken as representing a 1, or there is no current, meaning that we have a 0. Therefore, in reality 0 and 1 are not only two different logical states, but must also be two different (i.e. *distinguishable*) physical

states. Any information processing is therefore performed using physical systems, and consequently, is directly dependent on the laws of physics that these systems obey. Electronic wires in the above example obey laws of classical physics (Newtonian mechanics, Maxwell's equations, statistical mechanics and thermodynamics and Einstein's general relativity). This is why the information laws arising from using classical systems to encode information lead to *classical* information theory. Suppose, however, that the information is instead written into an atom: choose, for instance, two electronic energy levels of this atom and label them as 0 and 1. Then, since atoms obey the laws of quantum mechanics, the nature of information processing, its efficiency and the ultimate bounds will be different to those predicted by classical information theory. Studies of quantum information processing, its limits and efficiency are a part of quantum information theory.

The basic difference between the two lies in the superposition principle. Quantum systems, unlike their classical counterparts, can be in a coherent superposition of their basic states. This gives rise to the notion of a quantum bit, or (*qubit*) for short [5], which is in general in the state

$$\alpha|0\rangle + \sqrt{1-|\alpha|^2}|1\rangle \,, \tag{1.1}$$

having more possibilities than a classical bit. When we consider two qubits, the coherent superposition property of quantum systems leads to the notion of entanglement, i.e. the fact that two qubits can be correlated to an extent not accessible to two bits. The term *entanglement* was coined by Schrödinger [6] who used it to exemplify the strange nature of quantum mechanics (he actually used a German word *Verschränkung* which roughly translated means entanglement). An example is the often quoted Einstein, Podolsky, Rosen (EPR) state [7, 8]

$$\frac{|0,1\rangle - |1,0\rangle}{\sqrt{2}} \,, \tag{1.2}$$

which has absolutely no classical analogue. We can readily see that the above state displays a high degree of correlations: if we measure the first system, and learn its

state, we then immediately know the state of the second system. However, this can also be true for classical systems. Take as an example a person who wears either two blue or two red socks. Then if we observe the colour of one of the socks, we know that the colour of the other sock must be the same. Nevertheless, it is known that the EPR pair is more correlated than the socks described above, or, indeed, any other two classical systems can ever be, since it violates the so called Bell inequalities [8, 9] which are always satisfied by classical systems. Bell's inequalities will be introduced in the following chapter and serve to put an upper bound on the amount of correlations that can be possessed by two correlated classical systems. In general, two correlated quantum systems will have both quantum and classical correlations. The separation of these two contributions is a central problem solved in this thesis. Classical information theory has developed a method for quantifying the amount of classical correlations. The core of this thesis presents the development of its quantum counter-part, i.e. a quantum information theory of purely quantum correlation (or entanglement; we will use both expressions, but they will always have the same meaning throughout this thesis).

As far as quantum information theory is concerned there is already a developed quantum theory of data compression [5], and there exist preliminary results providing bounds to certain quantum communication protocols [10, 11, 12, 13]. Also quantum computation has been developed [14, 15] and shown to be more efficient in principle than its classical counter-part [16, 17]. In addition quantum communication is in principle more efficient [18] and also more secure [19] than classical communication. All these features and advantages of quantum information processing use the superposition principle and are a direct consequence of the existence of entangled states mentioned above. However, as soon as the entanglement is destroyed, which happens when the system storing the information interacts with the environment, the quantum computer inevitably reduces to the classical one. Therefore in addition to creating a consistent theory of quantum correlations we will show in this thesis how to protect them from negative and unwanted influences from the

"outside". This will be the analogue of classical information protection and will therefore be called quantum error correction. Quantum error correction is already a well established field [20, 21, 22, 23, 24], but our exposition will be different to the above and hence an original contribution to the field.

In this thesis we will be describing quantities that measure uncertainties in classical and quantum information theory which will hence assume different forms. The general convention used throughout is that Shannon's name will be always associated with the classical quantities, whereas von Neumann's name will always be written together with their quantum analogues (e.g. the Shannon entropy *vs* the von Neumann entropy). Note that this does not imply that classical information theoretic quantities cannot be used in quantum mechanics. For example, the Shannon entropy can be used to quantify the uncertainty in the spectrum of an observable pertaining to a certain quantum system. In contrast, the von Neumann entropy will be basis independent, and will refer to the state of that system as a whole.

My contribution to the field of quantum information included in this thesis has been to generalize the entanglement measures to mixed quantum systems containing two and more subsystems. I have realized that the main principle leading to entanglement quantification is that "the amount of entanglement cannot be increased by local operations" (I have linked this to the classical analogue stating that the relative entropy does not increase under stochastic evolution, as proven in Chapter 2). In the spirit of Shannon's formulation of entropy, I have postulated two additional, physically intuitive conditions that any measure of entanglement has to satisfy (Chapter 4). These three conditions alone do not lead to a unique measure, but imply a whole class of different measures which I presented in Chapter 4. I have provided statistical interpretations for two of the measures, i.e. the relative entropy of entanglement and the Bures metric of entanglement. This statistical way of interpreting entanglement, which is also my original contribution to the quantum information field, can naturally lead to an upper bound to the amount of entangle-

ment that can be locally distilled from a given ensemble of quantum states. I have then provided an alternative derivation to Knill's and Laflamme's, and Bennett's, DiVincenzo's, Smolin's and Wootters', of the conditions that quantum error correcting codes have to satisfy in order to be successful (Chapter 5). In addition I have worked on constructing the first quantum code to correct for spontaneous emission. I have then presented a practical cavity QED implementation of purification protocols that can be directly translated into ion-trap settings. Within this scheme I have illustrated the fact that the non-local quantum properties (entanglement) can be preserved by local error correcting methods.

The remainder of the thesis is organized as follows:

Chapter 2. This presents the mathematical background for the thesis. The first part concerns the basics of classical information theory putting emphasis on the formalism describing classical correlations. We show how this is reflected in classical communication theory and expose the relationship between information theory and thermodynamics, and statistics via the theory of types. This leads to the idea of distinguishing between different classical probability distributions.

Chapter 3. We then present some basic results of quantum information theory. We show how a procedure called Schmidt decomposition leads to an easy understanding of quantum correlations of a system in a joint *pure* state of two entangled quantum subsystems. Bell's inequalities are then derived and shown not to be entirely adequate for quantifying quantum correlations in general. We briefly review quantum communication theory and quantum computation emphasising the central role of entanglement.

Chapter 4. We generalize the classical theory of correlations to quantum mechanics. We offer two different ways of understanding quantum correlations. One is through purification procedures which aim at increasing, or strictly speaking, compressing quantum correlations by the means of local measurement and including the possibility of classical communication. The other one is by creating a quantum theory of types, and looking at the distinguishability of quantum states (which are

the equivalent of classical probability distributions). These two ways are then compared and lead to the same idea of how to quantify entanglement. We show how this way of measuring entanglement can naturally be generalised to more than two quantum subsystems.

Chapter 5. Basic conditions for quantum error correction are introduced. We then show how to preserve information written in a single atom which is spontaneously radiating into vacuum.

Chapter 6. We propose cavity QED implementations of the purification procedures. It is then shown how to preserve entanglement between two modes in two different cavities by using quantum error correction methods introduced in Chapter 5 and involving atoms.

Chapter 7. Here we summarize the thesis and briefly review the topics which can be illuminated by the results from this thesis. We also indicate various possibilities for future investigations.

Chapter 2

Theoretical Background

The enormous usefulness of mathematics in natural sciences is something bordering on the mysterious, and there is no rational explanation for it. It is not at all natural that 'laws of nature' exist, much less that man is able to discover them. The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve.

Eugene P. Wigner

2.1 Classical Information Theory

Classical information theory is a very wide subject encompassing three basic mathematical disciplines: the theory of communication, theory of computation and theory of error correction. In this chapter we present basic results from classical information theory with emphasis on the correlations between two random variables. The notion of correlations forms the basis of classical theory of communications and its quantum generalization will be the main subject of this thesis.

2.1.1 Measures of Uncertainty and Correlations

In this subsection we introduce various classical information measures [4]. Quantum analogues are then defined in the following chapter. Fundamental to our understanding of correlations is the measure of *uncertainty* in a given probability distribution. This uncertainty can be quantified by introducing the idea of "*surprise*" (first realized by the Roman Petronius Arbitrer *c* 60 AD [25]). Suppose that a certain event happens with a probability p. Then we would like to quantify how surprised we are when that event does happen. The first guess would be $1/p$: the smaller the probability of an event, the more surprised we are when the event happens. However, an event might be composed of two independent events which happen with probabilities p_1 and p_2 respectively, so that $p = p_1 \times p_2$. Now, we would intuitively expect that the surprise of p is the same as the surprise of p_1 plus the surprise of p_2. So, $1/p$ is not really a satisfactory definition from this perspective. However, if we define surprise as $\ln(1/p)$ then the above property called *additivity* is satisfied since $-\ln p_1 p_2 = -\ln p_1 - \ln p_2$. Now if we have a probability distribution $\sum_n p_n = 1$, then the total uncertainty is just the average of all the surprises, which brings us to our first definition.

Definition. The uncertainty in a collection of possible states a_i with corresponding probability distribution $p(a_i)$ is given by an *entropy*

$$S(p) := -\sum_i p(a_i) \ln p(a_i) \tag{2.1}$$

called the *Shannon entropy*. We note that there is no Boltzmann constant term in this expression, as there is for the physical entropy, since k_B is by convention set to unity.

We frequently require a means of comparing two different probability distributions, and for this reason we introduce the notion of *relative entropy* (first introduced by Kullback and Leibler in [26]).

Definition. Suppose that we have two sets of discrete events a_i and b_j with the corresponding probability distributions, $p(a_i)$ and $p(b_j)$. *The Shannon relative entropy*

between these two distributions is defined as

$$S(p(a)\,||\,p(b)) := \sum_i p(a_i) \ln \frac{p(a_i)}{p(b_i)} \,. \tag{2.2}$$

This function is a good measure of the 'distance' between $p(a_i)$ and $p(b_j)$, even though, strictly speaking, it is not a mathematical distance since $S(p(a)\,||\,p(b)) \neq S(p(b)\,||\,p(a))$. Its information–theoretic significance becomes apparent through the notion of mutual information.

Definition. *The Shannon mutual information* between two random variables A and B, having a joint probability distribution $p(a_i, b_j)$, and marginal probability distributions $p(a_i)$ and $p(b_j)$ is defined as

$$I_S(A : B) := S(p(a)) + S(p(b)) - S(p(a, b)) \,. \tag{2.3}$$

We now present two very instructive ways of looking at this quantity, which will form a basis for the work in this thesis. Mathematically, I_S can be written in terms of the Shannon relative entropy. In this sense it would represent a distance between the distribution $p(a, b)$ and the product of the marginals $p(a) \times p(b)$. As such, it is intuitively clear that this is a good measure of correlations, since it shows how far a joint distribution is from the product one in which all the correlations have been destroyed. So, we have that

$$I_S(A : B) = S(p(a, b)\,||\,p(a) \times p(b)) \,. \tag{2.4}$$

Let us now view this from another angle. Suppose that we wish to know the probability of observing b_j *if* a_i *has been observed.* This is called a conditional probability and is given by:

$$p_{a_i}(b_j) := \frac{p(a_i, b_j)}{p(a_i)} \,. \tag{2.5}$$

This motivates us to introduce a conditional entropy, $S_A(B)$, as:

$$\begin{aligned} S_A(B) &= -\sum_i p(a_i) \sum_j p_{a_i}(b_j) \ln p_{a_i}(b_j) & (2.6) \\ &= -\sum_{ij} p(a_i, b_j) \ln p_{a_i}(b_j) \,. & (2.7) \end{aligned}$$

This quantity tells us how uncertain we are about the value of B once we have learned about the value of A. Now the Shannon mutual information can be rewritten as

$$I_S(A:B) = S(B) - S_A(B) = S(A) - S_B(A)\,. \tag{2.8}$$

So, the Shannon mutual information, as its name indicates, measures the quantity of information conveyed about the random variable A (B) through measurements of the random variable B (A). This quantity, being positive, tells us that the initial uncertainty in $B(A)$ can in no way be increased by making observations on $A(B)$. Note also that, unlike the Shannon relative entropy, the Shannon mutual information is symmetric. Let us briefly go back to our original idea of a surprise to interpret the Shannon mutual information as a measure of correlations. Suppose that one of our friends likes to wear socks of two colours only: red and blue. In addition we know that her socks are always the same colour and that when she gets up in the morning, she randomly chooses the colour, but we know that she prefers blue to red with the ratio 3 : 1. So, when we meet our friend, before we have looked at the colour of her socks, we know that she wears blue socks with the probability $p(b) = 0.75$ and red socks with the probability $p(r) = 0.25$. However, when we look at one sock and observe, say, blue colour, we immediately know that the other colour must be blue, too. This means that the colours of her two socks are correlated. So, before we look at one of the socks, we are uncertain about the colour of the other sock by an amount of $-0.75\ln 0.75 - 0.25\ln 0.25$. But then, when we look at one of them the uncertainty immediately disappears. So, we expect that the information we gain about one sock by looking at the colour of the other is given by $-0.75\ln 0.75 - 0.25\ln 0.25$. The Shannon mutual information predicts exactly the same thing. We see that the largest correlations would be if $p = q = 0.5$ and would be $\ln 2$. This, of course, agrees with our intuitive notion of surprise, since then, before looking at her one sock, we would be completely uncertain about the colour of the other sock. Therefore by observing its colour we obtain the largest possible amount of information (i.e. remove the largest possible uncertainty in this

case).

Although it will be seen that the Shannon mutual information is a good measure of correlations between two random variables, its natural generalization to three and more random variables fails. It is easy to see that for three random variables the Shannon mutual information should be of the following form:

$$\begin{aligned} I_S(A:B:C) &= S(A,B,C) - S(A,B) - S(A,C) - S(B,C) \\ &+ S(A) + S(B) + S(C)\,. \end{aligned} \tag{2.9}$$

However, there exist A, B, C such that $I_S(A:B:C) < 0$ [27], and since we regard the amount of correlation as being strictly positive, this is automatically ruled out as a good measure of correlation. Curiously, the measures of entanglement proposed in this thesis in Chapter 4, and which are quantum generalization of the above classical ideas, will not suffer from this problem, and can thus naturally be defined for any number of correlated subsystems.

We mention that there are many other measures of correlations, but that the above are the most suitable for the purpose of this thesis. It is their generalization to the quantum case that will represent the solution to the problem of quantification of the amount of quantum correlations in a given quantum state. Among other, more significant measures of classical correlations, we have

- *Coefficient of correlations.* Suppose that we have two random variables $\mathbf{x}$ and $\mathbf{y}$. Let $\langle\ \rangle$ denote the expectation value, then the coefficient of correlation is defined as:

$$r := \frac{\langle(\mathbf{x} - \langle\mathbf{x}\rangle)(\mathbf{y} - \langle\mathbf{y}\rangle)\rangle}{\langle\mathbf{x} - \langle\mathbf{x}\rangle\rangle\langle\mathbf{y} - \langle\mathbf{y}\rangle\rangle}\,. \tag{2.10}$$

  It can be seen that $|r| \leq 1$ being 0 when $\mathbf{x}$ and $\mathbf{y}$ are independent. This quantity is unfortunately only appropriate for measuring linear correlations. Namely, if $\mathbf{x}$ and $\mathbf{y}$ are independent then $r = 0$; however, the converse is not true [27], since the two variables can be non-linearly dependent and still have $r = 0$.

In general, as soon as we have a function that measures some kind of distance between two distributions, we can immediately define correlations to be that distance between a joint state and the product of the marginals. Here we present two examples:

- *The Rényi relative entropy* is a generalization of the Shannon relative entropy (reducing to it when $\alpha \to 1$) given by

$$S_\alpha(p(a)||p(b)) := \frac{1}{\alpha - 1} \ln \left\{ \sum_i p(a_i)^\alpha p(b_i)^{1-\alpha} \right\} . \tag{2.11}$$

  Therefore a quantity that would measure correlations would be the Rényi mutual information

$$I_\alpha(A:B) \;=\; S_\alpha(p(a,b) \,||\, p(a) \times p(b)) \, . \tag{2.12}$$

  (Note that the relative entropy formulation of mutual information is indispensable here. A formula of the type $S_\alpha(p(a)) + S_\alpha(p(b)) - S_\alpha(p(a,b))$ cannot be derived in a consistent manner.)

- *The Rényi overlaps* are defined as

$$R_\alpha(p(a)||p(b)) := \sum_i p(a_i)^\alpha p(b_i)^{1-\alpha} \tag{2.13}$$

  The measure of correlation is then defined in a completely analogous fashion to the Rényi relative entropy.

Quantum generalizations of these two measures (for a detailed account see [28]) will be useful to us in Chapter 4. In the remainder of this section we confine ourselves only to Shannon's measures of entropy and correlations.

One very important property of any measure that aims at quantifying the amount of correlations between two random variables is the following: if either or both of the variables undergo a *local* stochastic evolution, then the amount of correlations cannot increase (in fact, it usually decreases). We now prove this in

the case of the Shannon mutual information, following an approach similar to that given by Everett in [29].

First, we establish without proof some inequalities following from the convex properties of the logarithmic functions.

Lemma1. $\sum_i P_i x_i \ln \sum_i P_i x_i \leq \sum_i P_i x_i \ln x_i$, where $x_i \geq 0$, $P_i \geq 0$ and $\sum_i P_i = 1$.

Lemma 2. $\sum_i x_i \ln \frac{\sum_i x_i}{\sum_i a_i} \leq \sum_i x_i \ln \frac{x_i}{a_i}$, where $x_i \geq 0$ and $a_i \geq 0$ for all i. (Proof follows from Lemma 1.)

We first show that the Shannon relative entropy between two probability distributions decreases when the same two undergo a stochastic process. This is a very satisfying property from a physical point of view, where two probability distributions undergoing stochastic changes can represent two evolving physical systems. It says that two probability distributions are in some sense closer to each other (i.e. "harder to distinguish") after a stochastic process, or analogously, that two physical systems become more alike.

So, we consider a sequence of transition-probability matrices T_{ij}^n, where $\sum_j T_{ij}^n = 1$ for all n, i, and $0 \leq T_{ij}^n \leq 1$, and a sequence of positive measures a_i^n having the property that

$$a_j^{n+1} = \sum_i a_i^n T_{ij}^n \,. \tag{2.14}$$

We further suppose that we have a sequence of probability distributions P_i^n generated by the action of the above stochastic process, such that

$$P_j^{n+1} = \sum_i P_i^n T_{ij}^n \,. \tag{2.15}$$

For each of these probability distributions the relative information I^n is defined as

$$S^n(P||a) := S(P^n||a^n) = \sum_i P_i^n \ln \frac{P_i^n}{a_i^n} \,. \tag{2.16}$$

We prove the following theorem:

Theorem. $S^{n+1}(P||a) \leq S^n(P||a)$.

Proof. Expanding $S^{n+1}(P||a)$ we obtain:

$$S^{n+1}(P||a) = \sum_j P_j^{n+1} \ln \frac{P_j^{n+1}}{a_j^{n+1}} = \sum_j \{\sum_i P_i^n T_{ij}^n\} \ln \frac{\sum_i P_i^n T_{ij}^n}{\sum_i a_i^n T_{ij}^n} \,. \tag{2.17}$$

However, using the concave property of the logarithmic function we have the following inequality

$$\sum_i P_i^n T_{ij}^n \ln \frac{\sum_i P_i^n T_{ij}^n}{\sum_i a_i^n T_{ij}^n} \leq \sum_i P_i^n T_{ij}^n \ln \frac{P_i^n T_{ij}^n}{a_i^n T_{ij}^n} \quad . \tag{2.18}$$

From the above two it follows that

$$S^{n+1}(P||a) \leq \sum_j \left\{ \sum_i P_i^n T_{ij}^n \ln \frac{P_i^n}{a_i^n} \right\} = \sum_i P_i^n T_{ij}^n \ln \frac{P_i^n}{a_i^n} \tag{2.19}$$

$$= \sum_i P_i^n \ln \frac{P_i^n}{a_i^n} = S^n(P||a) \tag{2.20}$$

and the proof is completed □.

This proof can be immediately specialized to the cases when T is stationary, i.e. T is independent of n, and when T is *doubly stochastic*, i.e. $\sum_i T_{ij} = 1$ for all j. A corollary to this important lemma is the following

Corollary. If we take $P = p(a, b)$, and $a = p(a)p(b)$, and suppose that the stochastic process acting separately on A and B are uncorrelated, we see that the Shannon mutual information does not increase under these local stochastic processes (by local we mean that they act separately on A and B).

Proof. Obvious.

This is a very important, and physically intuitive, property of any measure of correlations; its quantum analogue will be of central importance for quantifying quantum correlations between entangled subsystems which is the main subject of this thesis. This corollary, in fact, can be taken as a guidance for a "good" measure of correlations. We can state that any measure of correlations has to be non-increasing under local stochastic processes. The nature of quantum local stochastic processes will form the physical basis for our argument in the next chapter. A condition similar to property above, but employing quantum stochastic processes will be a key element in our search for measures of entanglement. When we go to quantum mechanics, the notion of a probability distribution will be replaced by a quantum state (i.e. density matrix), and a stochastic process will become a measurement process in quantum theory. A formulation of the probability theory which

is then most naturally generalized to quantum states is provided by Kolmogorov [30], and the quantum generalization expressing similarities with von Neumann's Hilbert Space formulation [31] can be found in [32] (c.f. [33]). However, knowledge of this approach will not be necessary for the rest of the thesis. At the end, it is important to stress that if the local stochastic processes are correlated, the correlations between the systems can increase as well as decrease. For more results on the behaviour of the Shannon relative entropy under stochastic processes it is instructive to read the work by Cohen et al in [34]; however this goes far beyond the subject of this thesis.

Correlations between two subsystems are sometimes very important to maintain. In classical communication theory, we wish the input and the output of a communication channel to be maximally correlated, since this implies that the receiver has obtained maximum information about what has been sent. The channel, on the other hand, introduces some noise through stochastic processes and disrupts communications. In order to maintain a low level of errors we have to use methods of error correction. We first describe the mathematical basis of classical communication and error correction theory. The second aim of this thesis is to generalize these methods to the quantum case in order to protect entanglement, on which both quantum communication and computation very much depend.

2.1.2 Classical Communication Theory

Let us suppose that the sender (source) of information, usually called Alice, encodes words into strings of the type $a_1 a_2 \dots a_N$ where each symbol a_j appears with the probability p_j. The Shannon entropy of the source is $S(A)$. Now the message goes through a channel which introduces various errors. If the output is the string $b_1 b_2 \dots b_N$, then the channel is completely specified by giving the probabilities of the type: the probability of receiving b_j if a_i was sent for all i and j. The receiver, usually called Bob, now tries to decode the original information. Bob's aim is to

obtain as much information as possible about A from the measurements conducted on B; this is, as we have seen, described by the Shannon mutual information, $I(A:B)$. The channel capacity is now defined as

$$\mathbf{C} = \max_{A} I(A:B)\,. \tag{2.21}$$

Since B is related to A via the above described channel's transition matrix, the channel capacity depends only on the channel's characteristics and is independent of the input and output of the channel. This quantity is of a fundamental importance in classical communication theory due to the following result proved by Shannon [1].

Theorem (Shannon [1]). If R is the rate of information production, then providing that $R < \mathbf{C}$ the information can be transmitted with an arbitrary reliability.

Here we only present an intuitive reasoning to justify the above form of the capacity (for a mathematically rigorous theory of communication see for example [35]). We stress that this proof is valid only for ergodic, stationary sources for which most sequences of n bits of a source with an entropy S have a probability of about e^{-nS}. Loosely speaking a source is stationary if the probabilities of emitting states do not change over time; it is ergodic if each subsequence of states appears in longer sequences with a frequency equal to its probability (the physical meaning of this statement will be analysed in section 2.3). This statement is then an information theoretic analogue of the Law of Large Numbers in probability theory. The source with entropy S will generate about $e^{TS(A)}$ sequences in a time interval T (this result also follows from the Law of Large Numbers and 'about' indicates that this is only true asymptotically–this will be explained in a greater detail in the section 2.3 Information Theory and Statistics). Now, each of these will be measured at the output and each output could then be produced by about $e^{TS_B(A)}$ inputs, since $S_B(A)$ represents the entropy of A once B has been measured. Therefore the total number of useful (*non-redundant*) messages (as they are called in communications)

is

$$N = e^{T(S(A)-S_B(A))} \tag{2.22}$$

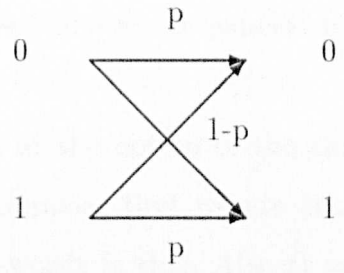
and therefore for the capacity we choose a source with the entropy that maximizes $S(A)-S_B(A)$, as stated before. If we instead chose a source whose entropy produces a larger quantity than the channel capacity, then that particular channel will not be able to handle the input and inevitably errors will result. The mutual information between the input and the output of a communication channel is thus a very important quantity since it determines the maximum rate at which the information can be transmitted without errors occurring.

2.1.3 Classical Error Correction

In the above we saw that, at least in principle, it is possible to communicate at a rate arbitrarily close to the channel's capacity with an arbitrarily small error. The above argument is informal and does not show us the exact way how this can be achieved. In practice we would complete a number of transmissions and encode our information into a few which are "separated enough" so that the error in the channel does not confuse them at the output. This method is known under the name of classical error correction [36]. The word *classical* is important to emphasize because the quantum analogue, although based on the same idea, will be different due to the basic differences between quantum and classical physics. Now we review the basic principles of classical error correction, which will help us to preserve the information in an error-inducing environment. First we focus on single errors and then generalize to an arbitrary number of errors. The quantum analogue of this procedure will be presented in Chapter 5.

A string of bits 0010... is sent through a classical channel which then introduces errors. For simplicity, we suppose that the errors affect the bits independently. In this case we have the following possibilities (we take a simple case where the probabilities for 0 and 1 are the same, so that the name of the channel is the

binary symmetric channel)



Now suppose that we wish to protect a bit '0' from an error. The simplest way is to use the so called repetition code: we encode a 0 into three 0's, i.e. $0 \to 000$. Now, if we allow only a single error to happen we can use a "majority vote" to restore the original. By the same token we encode 1 as 111. So in this way we have eliminated all the first order (single) errors. The total probability of error is now easily calculated to be

$$P_{\text{error}} = 3p^2 - 2p^3 \tag{2.23}$$

of the second order. The strings 000 and 111 will be referred to as code-words. Let us now generalize this simple idea. Suppose that we have e errors. Let d be the Hamming distance between two code-words: this is defined as the number of places (bits) where the two codes differ, so that, for example, $d(101, 001) = 2$ and $d(000, 111) = 3$. For error correction to work we need that at least $d = 2e + 1$. The reason for this is that if each code-word suffers e errors they are still different at the end (by at least one bit). Then if n denotes the length (number of bits) in each code-word, we have the following upper bound on the number of code–words $A(n, d)$

$$A(n, d) \sum_{k=0}^{e} \binom{n}{k} \leq 2^n \tag{2.24}$$

called the sphere-packing, or Hamming bound. If we wish to send two letters 0 and 1 ($A = 2$) protected against a single error ($e = 1$) then the above says that:

$$\sum_{k=0}^{1} \binom{n}{k} \leq 2^{n-1} \,. \tag{2.25}$$

The smallest n satisfying this bound by saturating the equality is $n = 3$ which is the repetition code described above. In general a code satisfying the equality is called the *perfect* code.

Let us now briefly look at the notion of the channel's capacity from the error correction point of view. Suppose that we are encoding a string of k bits into n bits. The number of code-words is then $A(n, d) = 2^k$. Substituting this into eq. (2.24) and taking the limit for large n, k, e we obtain

$$\frac{k}{n} \leq 1 - H\left(\frac{e}{n}\right), \tag{2.26}$$

where $H(x) = -x \log x - (1 - x) \log(1 - x)$. Now we can define the capacity of the channel to be the number of useful bits (k) to the total number of bits needed for error correction (n), and take the limit $n \to \infty$. Therefore the quantity on the right hand side of eq. (2.26) is the channel capacity of the binary symmetric channel. Note that $e/n = p$ where p is the error probability. A simple calculation shows that the same result is obtained from the original definition of capacity in eq. (2.21).

We now generalize our discussion by introducing some basic facts about *linear* codes through a simple example. These will be used later when constructing quantum error correcting codes. A binary code is linear if and only if the bitwise sum (modulo 2) of any two code–words is yet another code–word from that code. Linear codes are convenient because they have a number of simple properties [37]. First of all, there is a *generator matrix* whose rows form a basis of a linear code. Suppose that we wish to encode $\mathbf{u} = \{00, 01, 10, 11\}$, i.e. two bits of information, to protect against a single error. In order to do this we need to add another two bits to obtain the corresponding codewords $\mathcal{C} = \{0000, 1011, 0101, 1110\}$. The generator matrix for this code is given by

$$G = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

In fact, once we have the generator we can easily obtain the code by bitwise addition of its codewords (including the addition of a codeword to itself). The encoding

process is now given by a multiplication of $\mathbf{u}$ with G. We now need to understand how to correct an error and then the resulting code-word can be decoded to produce the original message. For this we introduce the notion of a *coset* of $\mathcal{C}$. If $\mathbf{a}$ is a four bit string, then the coset of $\mathcal{C}$ is $\mathbf{a} \oplus \mathcal{C}$. In our example $\mathbf{a}$ consecutively takes the value of four different strings $\mathbf{a}_1 = 0000, \mathbf{a}_2 = 1000, \mathbf{a}_3 = 0100, \mathbf{a}_4 = 0010$ for $\mathbf{a}$ and this will generate the set of all the possible four bit strings $\sum_i \mathbf{a}_i \oplus \mathcal{C}$. For example, $1000 \oplus C = \{1000, 0011, 1101, 0110\}$, and altogether we have

$$\begin{array}{cccc} 0000 & 1011 & 0101 & 1110 \\ 1000 & 0011 & 1101 & 0110 \\ 0100 & 1111 & 0001 & 1010 \\ 0010 & 1001 & 0111 & 1100 \end{array} \tag{2.27}$$

Note that the coset leaders are in the first column and that rows contain the cosets generated by the corresponding coset leader. In order to error correct we construct a *parity check matrix*, H, which is a generator matrix for the code $C^{\perp} = \{1101, 0111, 1010, 0000\}$ so that we have the following

$$H = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}.$$

Suppose now that the initial state to encode is 10. The corresponding code-word is $\mathbf{u} = 1011$. Suppose that the error has occurred on the second bit and that the resulting state is $\mathbf{v} = 1111$. Then if we look at the table the corrected state is the one at the top of the column where 1111 is found. This state is indeed $\mathbf{u} = 1011$, and the error correction has been successful. However this search for the position of the word in the table is inefficient (time consuming), and this is where the parity check matrix offers help. The parity check matrix is used to generate *error syndrome*. This is done by multiplying the final state (after the error) by the transpose of the

parity check matrix $\mathbf{v}H^T$. So in our example if $\mathbf{v} = 1111$ then

$$(1111)\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = (01)\ .$$

The crucial observation now is that all the words in the same row have the same error syndrome, so that, for example

$$(0100)\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = (01)\ .$$

Therefore the error syndrome helps allocate the position of the code–word after undergoing an error, and thus speeds–up the error correction process. A simple way to understand this is to realize that the product of the generator and the parity check $G \cdot H^T = 0$. The state after the error can be represented as $\mathbf{v} = \mathbf{u} \oplus \mathbf{e}$ where $\mathbf{e}$ is the error vector. Now we have,

$$H(\mathbf{v}) = H(\mathbf{u} \oplus \mathbf{e}) = H(\mathbf{u}) \oplus H(\mathbf{e}) = 0 \oplus H(\mathbf{e}) = H(\mathbf{e}) \tag{2.28}$$

so that the value of $H(\mathbf{v})$ does not depend on $\mathbf{u}$ but only on the error $\mathbf{e}$. If $H(\mathbf{v})$ is different for all possible errors, we will be able to determine precisely what error occurred and will be able to correct it.

In Chapter 5 we present the basic rules of quantum error correction, providing a quantum analogue of the single error correcting perfect code, and designing a code to cope with the atomic spontaneous emission of radiation. We will see that the above concepts, such as linear codes and error syndromes, retain their basic meaning in quantum error correction and prove to be very useful.

2.2 Comparing Thermodynamical and Shannon's Entropy

The particular form of a measure of uncertainty depends on the physical conditions. What will briefly be discussed in this section is that the Shannon entropy is the *thermodynamically* appropriate measure of uncertainty only if the system under consideration is ergodic. If, however, this is not the case, we are forced to use a more general quantity which corresponds to the so called Lévy statistics. Before we quantify this, let us formally state the conditions which single out the Shannon entropy as a good measure of uncertainty.

We have seen that there is a number of different measures of uncertainty. Each one of these posesses different properties distinguishing them from one another. Let us consider the following four conditions [27]:

1. Continuity: $S(p_1, p_2, \ldots, p_n)$ continuous in p_k for all k.

2. Symmetry: $S(p_1, p_2, \ldots, p_n) = S(p_2, p_1, \ldots, p_n)$.

3. Extremal property: Maximum of S is

$$\max_{\{p\}} S(p_1, p_2, \ldots, p_n) = S\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) . \tag{2.29}$$

4. Additivity: Let $p_n = \sum_{k=1}^{m} q_k$. Define,

$$S_1(p_1, p_2, \ldots, p_n) \tag{2.30}$$

$$S_2(p_1, p_2, \ldots, p_{n-1}, q_1, q_2, \ldots, q_m) \tag{2.31}$$

$$S_3\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}, \ldots, \frac{q_m}{p_n}\right) \tag{2.32}$$

   Additivity requires that $S_2 = S_1 + p_n S_3$.

The physical basis of the first three conditions is immediately clear: firstly, if we change the probabilities slightly we do not want the uncertainty to be very much

different; secondly, all the probabilities are equally important, and thirdly, the uncertainty is maximum when we have absolutely *no* a priori knowledge to believe one outcome to be more likely than any other. The last property is the one mentioned before, at the beginning of this chapter. Loosely stated, if we combine two independent events their uncertainties must add up, a property sometimes referred to as *additivity*. Now the crucial result is that if we decide to use the natural logarithm and to set the overall multiplication constant to be equal to 1 (instead of Boltzmann's constant), then the only function satisfying the above four conditions is $S = -\sum_i p_i \ln p_i$ [27] (this was originally proven by Shannon).

Suppose that we wish to extremize S with the constraint $\sum_{i=1}^{W} p_i = 1$. Using Lagrangian multipliers we immediately see that S is extremized in the case of having the *equiprobable* distribution, i.e. when $p_i = 1/W$ (corresponding to a microcanonical ensemble treatment). In this case

$$S = \ln W \tag{2.33}$$

which, when multiplied by Boltzmann's constant, produces the celebrated Boltzmann expression.

Suppose now that we want to extremize S with the following two constraints (corresponding to a canonical ensemble treatment):

$$\sum_{i=1}^{W} p_i = 1 , \tag{2.34}$$

$$\sum_{i=1}^{W} p_i E_i = U . \tag{2.35}$$

The extremized S is given by

$$p_i = \frac{e^{\beta E_i}}{Z} \quad \text{where} \tag{2.36}$$

$$Z = \sum_{i=1}^{W} e^{\beta E_i} \tag{2.37}$$

and $\beta = 1/k_B T$, T being the temperature. So, the Shannon and Boltzmann entropy are very closely related. Now, the key result in classical thermodynamics is the

Second Law which states that the entropy of a closed (i.e. noninteracting) system never decreases. It is important to point out that, in the case of the Shannon entropy, this is not always true [38, 39]. What is true is that the Shannon relative entropy never increases, as shown in subsection 2.1. However, the Shannon relative entropy decrease does not always imply an entropy increase. A counter-example is provided by a Markov chain with a *non-uniform, stationary distribution* [4]. A uniform distribution is the one for which all the probabilities are equal and the stationary distribution is the one independent of n. If we start this Markov chain from a uniform distribution (maximal entropy), the distribution will tend to the stationary one (which has lower entropy) (a good discussion of this can be found in the first chapter of [40]).

If, however, the stationary distribution is uniform (as is required for the Boltzmann formula) the Shannon relative entropy is given by:

$$S(P^n||a) = \ln W - S(P^n) \, . \tag{2.38}$$

In this case the monotone decrease in the Shannon relative entropy implies a monotone increase in $S(P^n)$, as in the Second Law of Thermodynamics. Thus the crucial observation to remember is that, in information theory, the central result states that the Shannon relative entropy decreases under stochastic evolution [34] (rather than the erroneous law stating that the Shannon entropy increases under a stochastic evolution).

The second important observation we wish to make is that the physical circumstances might be such that the above four assumption categorising the measure of uncertainty (and hence correlations) are no longer appropriate. Many physical systems do not behave according to the Maxwell-Boltzmann exponential law described above. An extraordinary example is found in medicine: Peng and colleagues from Harvard medical school, found that the erratic patterns observed in the heartbeats of healthy subjects do not follow a Gaussian (Maxwell-Boltzmann) distribution, but the more general, so called *Lévy* statistics (curiously the heartbeats of unhealthy

subject more closely resemble a Gaussian) [41]. The Lévy distribution is given by [42]

$$p_i = \frac{(1-\beta(q-1)E_i)^{1/(q-1)}}{Z_q} , \tag{2.39}$$

where

$$Z_q = \sum_{i=1}^{W}(1-\beta(q-1)E_i)^{1/(q-1)} . \tag{2.40}$$

The parameter q describes how much a given distribution deviates from the Maxwell-Boltzmann law. When $q \to 1$ we recover the Maxwell-Boltzmann exponential distribution. The Shannon entropy is no longer appropriate to maximize under these circumstances, and a new, generalized entropy is required:

$$S_q(p) = \frac{1-\sum p_i^q}{q-1} \tag{2.41}$$

which is the so called Tsallis entropy [43] (although originally introduced by Daróczy in [44]). For $q \to 1$ this reduces to the Shannon entropy. Systems satisfying these statistics do not satisfy the law of additivity discussed before. Namely if we have two independent systems A and B then

$$S_q(AB) = S_q(A) + S_q(B) - (q-1)S_q(A)S_q(B) . \tag{2.42}$$

This law should now replace the condition 4 of additivity in order to derive the generalized entropy S_q as a unique measure of uncertainty. The physical basis of the difference between the systems obeying Maxwell-Boltzmann statistics and the systems obeying Lévy statistics lies in the *ergodic* property we mentioned in relation to the channel capacity. Namely, the former systems are ergodic, meaning that the long time average of physical quantities is equal to the ensemble average; the latter are, on the other hand, non-ergodic, where the long time average is not equal to the ensemble average (this is directly analogous to ergodic and non-ergodic information sources analysed in the previous subsection). Therefore, *measures of uncertainty and hence correlations, are by no means unique, and different physical circumstances give rise to different measures.* This will also be seen in the case of quantum measures of uncertainty and correlations (entanglement). In Chapter 4

we shall present a whole class of "good" measures of entanglement each one suitable to a different physical background. Now we turn to the idea which provides the physical, or strictly speaking *statistical*, interpretation behind the Shannon entropy and relative entropy.

2.3 Information Theory and Statistics

Here we present a fruitful connection between information theory and statistics. This will provide us with another interpretation of the Shannon entropy and the Shannon relative entropy, but this time from the statistical point of view. The generalization of this formalism to the quantum domain will be presented in the next section and we will offer an operational interpretation of the measures of quantum correlations to be introduced therein. We follow the approaches of Cover and Thomas in [4], and Csiszár and Körner in [45].

2.3.1 The Theory of Types

Let $X_1, X_2, ...X_n$ be a sequence of n symbols from an alphabet $A = \{a_1, a_2, ..., a_{|A|}\}$. We denote a sequence $x_1, x_2, ..., x_n$ by x^n or, equivalently, by $\mathbf{x}$. The type $P_{\mathbf{x}}$ of a sequence $x_1, x_2, ..., x_n$ will be called the relative proportion of occurances of each symbol of A, i.e. $P_{\mathbf{x}}(a) = N(a|\mathbf{x})/n$ for all $a \in A$, where $N(a|\mathbf{x})$ is the number of times the symbol a occurs in the sequence $\mathbf{x} \in A^n$. $\mathcal{P}_n$ will denote the set of types with denominator n. If $P \in \mathcal{P}_n$, then the set of sequences of length n and type P is called the *type class* of P, denoted by $T(P)$, i.e. mathematically

$$T(P) = \{\mathbf{x} \in A^n : P_{\mathbf{x}} = P\} . \tag{2.43}$$

We now approach the first theorem about types which is at the heart of success of this theory and states that the number of types increases only polynomially with n.

Theorem.

$$|\mathcal{P}_n| \leq (n+1)^{|A|} \tag{2.44}$$

Proof. Obvious □.

The most important point is that the number of sequences is exponential in n, so that at least one type has exponentially many sequences in its type class. Actually, the largest type class has essentially the same number of elements as the entire set of sequences (they become equivalent as $n \to \infty$). We now arrive at the most important theorem for us, which, in fact, present the basis of the statistical interpretation of the Shannon entropy and relative entropy.

Theorem. If $X_1, X_2, ...X_n$ are drawn according to $Q(x)$, then the probability of $\mathbf{x}$ depends only on its type and is given by

$$Q^n(\mathbf{x}) = e^{-n(S(P_\mathbf{x})+S(P_\mathbf{x}||Q))} \tag{2.45}$$

Proof.

$$\begin{aligned}
Q^n(\mathbf{x}) &= \prod_{i=1}^{n} Q(x_i) && (2.46)\\
&= \prod_{a\in A} Q(a)^{N(a|\mathbf{x})} && (2.47)\\
&= \prod_{a\in A} Q(a)^{nP_\mathbf{x}(a)} && (2.48)\\
&= \prod_{a\in A} e^{nP_\mathbf{x}(a)\ln Q(a)} && (2.49)\\
&= \exp\left\{n\sum_{a\in A} -P_\mathbf{x}(a)\ln\frac{P_\mathbf{x}(a)}{Q(a)} + P_\mathbf{x}(a)\ln P_\mathbf{x}(a)\right\} && (2.50)\\
&= e^{-n(S(P_\mathbf{x})+S(P_\mathbf{x}||Q))}\ \square && (2.51)
\end{aligned}$$

Corollary. If $\mathbf{x}$ is the type class of Q, then

$$Q^n(\mathbf{x}) = e^{-nS(Q)} \tag{2.52}$$

Proof. Obvious □.

The above theorem has very important implications in the theory of statistical inference and distinguishability of probability distributions. To see how this comes

about we state without proof two theorems that give bounds to the size of a type class and also bounds on the probability of a particular type class. The proofs follow directly from the above two theorems and the corollary [4, 45].

Theorem. For any type $P \in \mathcal{P}_n$,

$$\frac{1}{(n+1)^{|A|}} e^{nS(P)} \leq |T(P)| \leq e^{nS(P)} \tag{2.53}$$

Theorem. For any type $P \in \mathcal{P}_n$, and any distribution Q, the probability of the type class $T(P)$ under Q^n is $e^{-nS(P||Q)}$ to first order in the exponent. More precisely,

$$\frac{1}{(n+1)^{|A|}} e^{-nS(P||Q)} \leq Q^n(T(P)) \leq e^{-nS(P||Q)} \tag{2.54}$$

The above two results can be succinctly written in an exponential fashion that will be useful to us as

$$|T(P)| \rightarrow e^{-nS(P)} \tag{2.55}$$

$$Q^n(T(P)) \rightarrow e^{-nS(P||Q)} . \tag{2.56}$$

We have already made use of these statements when dealing with the channel capacities and proving the Shannon theorem, and now they are set on a firm theoretical basis. The first statement also leads to the idea of *data compression*, where a string of length n generated by a source with entropy S can be encoded into a string of length nS. The second statement says that if we are performing n experiments according to distribution Q, the probability that we will get something that looks as if it was generated by distribution P decreases exponentially with n depending on the relative entropy between P and Q. This idea immediately leads to Sanov's theorem, whose quantum analogue will provide a statistical interpretation of the measure of entanglement presented in the next chapter. Now we present examples of data compression and introduce Sanov's theorem.

2.3.2 Data Compression and Sanov's theorem

Suppose that we have a binary source generating 0's with twice as big a probability as that of 1's, so that the Shannon entropy is $S = \ln 3 - 2/3 \ln 2 = 0.64$. Imagine

that we have a string of 15 digits coming out of this source. Then, according to the above considerations (eq. (2.56)) , the most likely type will be the one with ten 0's and five 1's. But the size of this class is only $0.64 \times 15 \approx 10$. So we can use only 10 digits to encode all the above sequences of 15 numbers just by assigning the following conventional mapping: the first sequence of 15 numbers is to be encoded in 0000000000, the second sequence is to be encoded in 0000000001, ... , the e^{10}th sequence is to be encoded in 1111111111. This encoding is for obvious reasons called data compression. This, in fact, offers a statistical reason for employing the Shannon entropy as a measure of uncertainty.

Now we look at the distinguishability of two probability distributions. Suppose we would like to check if a given coin is "fair", i.e. if it generates a "head–tail" distribution of $f = (1/2, 1/2)$. When the coin is biased then it will produce some other distribution, say $uf = (1/3, 2/3)$. So, our question of the coin fairness boils down to how well we can differentiate between two given probability distributions given a finite, n, number of experiments to perform on one of the two distributions. In the case of a coin we would toss it n times and record the number of 0's and 1's. From simple statistics we know that if the coin is fair than the number of 0's, $N(0)$, will be roughly $n/2 - \sqrt{n} \leq N(0) \leq n/2 + \sqrt{n}$, for large n and the same for the number of 1's. So if our experimentally determined values do not fall within the above limits the coin is not fair. We can look at this from another point of view which is in the spirit of the method of types; namely, what is the probability that a fair coin will be mistaken for an unfair one with the distribution of $(1/3, 2/3)$ given n trials on the fair coin? For large n the answer is given in the previous subsection

$$p(\text{fair} \to \text{unfair}) = e^{-nS\ (uf||f)} \ , \tag{2.57}$$

where $S\ (uf||f) = 1/3 \ln 1/3 + 2/3 \ln 2/3 - 1/3 \ln 1/2 - 2/3 \ln 1/2$ is the Shannon relative entropy for the two distributions. So,

$$p(\text{fair} \to \text{unfair}) = 3^n 2^{-\frac{5}{3}n} \ , \tag{2.58}$$

which tends exponentially to zero with $n \to \infty$. In fact we see that already after

~ 20 trials the probability of mistaking the two distributions is vanishingly small, $< 10^{-10}$. Sanov's theorem [46] now states that if we have a probability distribution Q and a set of distributions $E \subset \mathcal{P}$ then

$$Q^n(E) \to e^{-nS(P^*||Q)} \tag{2.59}$$

where

$$P* = \min_{E \subset \mathcal{P}} S(P||Q) \tag{2.60}$$

is the distribution in E that is closest to Q in the Shannon relative entropy. This can also be rephrased in the language of distinguishability: when we are distinguishing a given distribution from a set of distributions, then what matters is how well we can distinguish that distribution from the closest one in the set. When we turn to the quantum case in the next chapter, the probability distributions will become quantum densities representing states of a quantum system, and the question will be how well we can distinguish between these states.

Chapter 3

Quantum Information Theory

3.1 Quantum Correlations

The main difference between quantum and classical physics is seen in the superposition principle which, when two or more systems are involved, leads to the phenomenon of entanglement. Quantum systems, unlike their classical counter-parts, can be in states involving superpositions of their basic states. This alone is responsible for the fact that information theory based on quantum mechanics is radically different from the classical information theory described previously. This basic difference is manifested in the fact that the amount of correlations in two entangled quantum subsystems can exceed the amount of "allowed" classical correlations. This excess of correlations enables quantum communications to be in a certain sense more efficient than classical communications and also quantum computation to be "faster" than its classical equivalent. Therefore quantum correlations have a central role in quantum information theory. This section introduces the basic ingredients of quantum information theory that will enable us to obtain a class of measures of quantum correlations and set the basis of quantum error correction.

3.1.1 Entanglement and Schmidt Decomposition

A composite quantum system is one that consists of a number of quantum subsystems. When those subsystems are entangled it is impossible to ascribe a definite state vector to any one of them. The most often quoted entangled system is a pair of two photons, being in the "EPR" state [7, 8]. The composite system is then mathematically described by

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle|\downarrow\rangle + |\downarrow\rangle|\uparrow\rangle) \tag{3.1}$$

where the first ket in either product belongs to one photon and the second to the other. The property that is described is the direction of spin or polarization along the z-axis, which can either be "up" ($|\uparrow\rangle$) or "down" ($|\downarrow\rangle$). A two level system of this type is a quantum analogue of a bit, which we shall henceforth call a *qubit*. We can immediately see that neither of the photons possesses a definite state vector. The best that one can say is that if a measurement is made on one photon, and it is found to be in the state "up" for example, then the other photon is certain to be in the state "down". This idea cannot be applied to a general composite system, unless the former is written in a special form. This motivates us to introduce the so called Schmidt decomposition [47], which not only is mathematically convenient, but also gives a deeper insight into correlations between the two subsystems.

According to the rules of quantum mechanics, the state vector of a composite system, consisting of subsystems U and V, is represented by a vector belonging to the tensor product of the two Hilbert Spaces $\mathcal{H}_U \otimes \mathcal{H}_V$. The general state of this system can be written as a linear superposition of products of individual states:

$$|\Psi\rangle = \sum_n \sum_m c_{nm}|u_n\rangle|v_m\rangle \tag{3.2}$$

where $\{|u_n\rangle\}_{n=1}^N$ and $\{|v_m\rangle\}_{m=1}^N$ are the orthonormal basis of the subsystems U and V respectively, whose dimensions are dim $U = N$ and dim $V = M$. We will now describe the procedure of Schmidt decomposition whereby the above state $|\Psi\rangle$ is re-expressed in terms of the so called Schmidt basis.

To that end, let us assume that $M > N$, which in no way affects our line of argument since the procedure is symmetric with respect to the subsystems. Then we have the following five steps:

1. First we construct a density matrix describing $|\Psi\rangle$. Once the density matrix is known all the properties of the system can be deduced from it. Moreover, ensembles which are prepared differently, but have the same density matrix are statistically indistinguishable and therefore equivalent (see [48] on how to construct all different ensembles given a density matrix). Generally, if we have a mixed state involving vectors $|\Psi_1\rangle, |\Psi_2\rangle, \ldots |\Psi_D\rangle$ with corresponding classical probabilities $w_1, w_2, \ldots, w_3$, then the density matrix is defined to be:

$$\rho = \sum_{d=1}^{D} w_d |\Psi_d\rangle\langle\Psi_d| \,. \tag{3.3}$$

   Since in our case $|\Psi\rangle$ is a pure state, the density matrix is a projection operator on to $|\Psi\rangle$, i.e.

$$\rho = |\Psi\rangle\langle\Psi| = \sum_{nm}\sum_{pq} \rho_{nmpq} |u_n\rangle\langle u_p| \otimes |v_m\rangle\langle v_q| \tag{3.4}$$

   where $\rho_{nmpq} = c_{nm} c^*_{pq}$. If we, however, wish to deal with one of the subsystems only, then we employ the concept of the reduced density matrix.

2. We find the reduced density matrix of the subsystem U, obtained by tracing ρ over all states of the subsystem V, so that

$$\rho_U = \sum_q \langle v_q|\rho|v_q\rangle = \sum_{nm}\sum_p \rho_{nmpm} |u_n\rangle\langle u_p| \,. \tag{3.5}$$

   The crucial step in the Schmidt decomposition is diagonalizing the above. We shall call the eigenvalues of ρ_U $|g_1|^2, |g_2|^2, \ldots, |g_N|^2$, and the corresponding eigenvectors $|u'_1\rangle, |u'_2\rangle, \ldots, |u'_N\rangle$.

3. Then we re-express the above in terms of $|u'\rangle$'s, i.e

$$|\Psi\rangle = \sum_n \sum_m c'_{nm} |u'_n\rangle |v_m\rangle \,. \tag{3.6}$$

4. Now, we construct a new orthonormal basis of the subsystem V such that each new vector is a "clever" linear superposition of the old ones, so that

$$|v'_i\rangle = \sum_m \frac{c'_{im}}{g_i}|v_m\rangle \ . \tag{3.7}$$

5. The Schmidt decomposition of $|\Psi\rangle$ is now given by

$$|\Psi\rangle = \sum_n g_n |u'_n\rangle|v'_n\rangle \ . \tag{3.8}$$

There are two important observations to be made, which are absolutely fundamental to understanding correlations between the two subsystems in a joint pure state:

- The reduced density matrices of both subsystems, written in the Schmidt basis, are diagonal and have the same positive spectrum. in particular, the overall density matrix is given by

$$\rho = \sum_{nm} g_n g_m^* |u'_n\rangle\langle u'_m| \otimes |v'_n\rangle\langle v'_m| \tag{3.9}$$

whereas the reduced ones are

$$\rho_U = \sum_m \langle v'_m|\rho|v'_m\rangle = \sum_n |g_n|^2 |u'_n\rangle\langle u'_n| \tag{3.10}$$

$$\rho_V = \sum_n \langle u'_n|\rho|u'_n\rangle = \sum_m |g_m|^2 |v'_m\rangle\langle v'_m| \ . \tag{3.11}$$

- If a subsystem is N dimensional it then can be entangled with no more than N orthogonal states of another one.

At the end we would like to point out that the Schmidt decomposition is, in general, impossible for more than two entangled subsystems. Mathematical details of this fact are exposed in [49]. To clarify it, however, we consider three entangled subsystem as an example. Here, our intention would be to write a general state such that by observing the state of the one of the subsystems we instantaneously and with certainty know the state of the other two. But, this is impossible in general, for the presence of the third system makes the prediction uncertain. Loosely speaking,

while we know the state of one of the subsystems, the other two might still be entangled and cannot have definite vectors associated with them (an exception to this general rule is, for example, a state of the Greenberger–Horne–Zeilinger (GHZ) type $(1/\sqrt{2})(|\uparrow\rangle|\uparrow\rangle|\uparrow\rangle + |\downarrow\rangle|\downarrow\rangle|\downarrow\rangle)$). Clearly, involvement of even more subsystems complicates this analysis even further and produces, so to speak, an even greater mixture and uncertainty. The same reasoning applies to mixed states of two or more subsystems (i.e. states whose density operator is not idempotent $\rho^2 \neq \rho$), for which we cannot have the Schmidt decomposition in general. This reason alone is responsible for the fact that the entanglement of two subsystems in a pure state is simple to understand and quantify, while for mixed states, or states consisting of more than two subsystems, the question is much more involved. The solution to the problem of understanding entanglement and quantifying its amount in a given, general, quantum state consisting of an arbitrary number of systems is the central theme of this thesis.

Next we turn to some natural quantum generalizations of the classical measures of uncertainty and correlations given in the previous section of this chapter. We talk about two subsystems mainly, and any generalizations will always be emphasised in particular.

3.1.2 Quantum Measures of Uncertainty and Correlations

When two subsystems become entangled, we saw that the composite state can be expressed as a superposition of the product of the corresponding Schmidt basis vectors. From eq. (3.8) it follows that the i-th vector of either subsystem has a probability of $|g_i|^2$ associated with it. We are, therefore, uncertain about the state of each subsystem, the uncertainty being larger if the probabilities are evenly distributed. Since the uncertainty in the probability distribution is naturally described by the Shannon entropy, this classical measure can also be applied in quantum theory. In an entangled system this entropy is related to a single observable. The

general state of a quantum system, as we have already remarked, is described by its density matrix ρ. If A is an operator pertaining to the system described by ρ, then by the spectral decomposition theorem $A = \sum_i a_i P_i$, where P_i is the projection onto the state with the eigenvalue a_i. The probability of obtaining the eigenvalue a_j is given by $p_j = \mathrm{Tr}(\rho P_j) = \mathrm{Tr}(P_j \rho)$. The uncertainty in a given observable can now be expressed through the Shannon entropy. Let the observables A and B, pertaining to the subsystems U and V respectively, have a discrete, non-degenerate spectrum, with corresponding probabilities $p(a_i)$ and $p(b_j)$ of observables A being a_i and B being b_j. Let also the joint probability be $p(a_i, b_j)$. Then,

$$S(A) = -\sum_i p(a_i) \ln p(a_i) = -\sum_{ij} p(a_i, b_j) \ln \sum_j p(a_i, b_j) \tag{3.12}$$

$$S(B) = -\sum_j p(b_j) \ln p(b_j) = -\sum_{ij} p(a_i, b_j) \ln \sum_i p(a_i, b_j) \tag{3.13}$$

$$S(A, B) = -\sum_{ij} p(a_i, b_j) \ln p(a_i, b_j) \tag{3.14}$$

where we have used the fact that $\sum_j p(a_i, b_j) = p(a_i)$ and $\sum_i p(a_i, b_j) = p(b_j)$. We have seen that a signature of correlations is that the sum of the uncertainties in the individual subsystems is greater than the uncertainty in the total state. So, the Shannon mutual information is a good indicator of how much the two given observables are correlated. However, this quantity describes the correlations between single observables only. The quantity that is related to the correlations in the overall state as a whole is the von Neumann mutual information. Since it is assigned to the state as a whole, it is of little surprise that it involves in its expression the density matrix. First, however, we define the von Neumann entropy [31], which can be considered as the proper quantum analogue of the Shannon entropy [50, 51, 52].

Definition. *The von Neumann entropy* of a quantum system described by a density matrix ρ is defined as

$$S_N(\rho) := -\mathrm{Tr}(\rho \ln \rho) \, . \tag{3.15}$$

(We will drop the subscript N whenever there is no possibility of confusion). The

Shannon entropy is equal to the von Neumann entropy only when it describes the uncertainties in the values of the Schmidt observables. Otherwise,

$$S(A) \geq S_N(\rho) \tag{3.16}$$

where A is any observable of a system described by ρ. This means that there is more uncertainty in a single observable than in the whole of the state, the fact which entirely contradicts our expectations.

We will now state without proof, a relation concerning the entropies of two subsystems. One part of it is somewhat analogous to its classical counterpart, but instead to referring to observables is related to the two states. This inequality is called the Araki-Lieb inequality [53] and is one of the most important results in the quantum theory of correlations. Let ρ_A and ρ_B be the reduced density matrices of subsystems A and B respectively, and ρ be the matrix of a composite system, then:

$$S_N(\rho_A) + S_N(\rho_B) \geq S_N(\rho) \geq |S_N(\rho_A) - S_N(\rho_B)| \,. \tag{3.17}$$

Physically, the left hand side implies that we have more information (less uncertainty) in an entangled state than if the two states are treated separately. This arises naturally, since by treating the subsystems separately we have neglected the correlations (entanglement). We note that if the composite system is in a pure state, then $S(\rho) = 0$, and from the right hand side it follows that $S(\rho_A) = S(\rho_B)$. To appreciate the extent to which this is a counter-intuitive result we consider the following example. Suppose a two level atom is interacting with a single mode of an EM field as in the Jaynes-Cummings model which will be described in detail in Chapter 6. If the overall state is initially pure, and the whole system is isolated then the entropies of the atom and the field are equally uncertain at all the times. But this is not expected since the atom has only two degrees of freedom and the field infinitely many [54]! This, however, is possible, as, by the second observation, the atom, as a two dimensional subsystem, is only entangled with two dimensions of the field.

We now present without proofs several properties of entropy which will be used in the later sections [52]. These are:

1. *additivity:* $S_N(\rho_A \otimes \rho_B) = S_N(\rho_A) + S_N(\rho_B)$; (3.18)
2. *concavity:* $S_N\left(\sum_i \lambda_i \rho_i\right) \geq \sum_i \lambda_i S_N(\rho_i)$; (3.19)
3. *strong subadditivity:* $S_N(\rho_{ABC}) + S_N(\rho_B) \leq S_N(\rho_{AB}) + S_N(\rho_{BC})$ (3.20)

(where $\rho_B = \mathrm{Tr}_{AC}\rho_{ABC}$ and similarly for the others). The first property is the same as in classical information theory, namely the entropies of independent systems add up. The concavity simply reflects the fact that "mixing increases uncertainty". It is also worth mentioning that the consequence of the strong subadditivity is the so called weak subadditivity described by the Araki–Lieb inequality introduced before.

Following the definition of the Shannon mutual information we introduce the von Neumann mutual information, which refers to the correlation between the whole subsystems rather than relating two observables only.

Definition. *The von Neumann mutual information* between the two subsystems ρ_U and ρ_V of the joint state ρ_{UV} is defined as

$$I_N(\rho_U : \rho_V; \rho_{UV}) = S_N(\rho_U) + S_N(\rho_V) - S_N(\rho_{UV}) \ . \tag{3.21}$$

As in the case of the Shannon mutual information this quantity can be interpreted as a distance between two quantum states. For this we first need to define the von Neumann relative entropy, in a direct analogy with the Shannon relative entropy (in fact, this quantity was first considered by Umegaki in [55], but for consistency reasons we name it after von Neumann).

Definition. *The von Neumann relative entropy* between the two systems σ and ρ is defined as

$$S_N(\sigma||\rho) = \mathrm{Tr}\sigma(\ln\sigma - \ln\rho) \ . \tag{3.22}$$

Now, the von Neumann mutual information can be understood as a distance of the state ρ_{UV} to the uncorrelated state $\rho_U \otimes \rho_V$,

$$I_N(\rho_U : \rho_V; \rho_{UV}) = S_N(\rho_{UV}||\rho_U \otimes \rho_V) \ . \tag{3.23}$$

The von Neumann relative entropy will be the most important quantity in classifying and quantifying quantum correlations in Chapter 3. It will be seen that this quantity does not increase under local general measurements, which are quantum analogues of the stochastic processes considered in subsection 2.1. Therefore, a natural concept to consider now is that of the general measurement in quantum mechanics.

3.1.2.1 Complete Measurement

In this subsection we present two different ways of describing the dynamical evolution of a quantum system. First we can look at the joint unitary evolution of the system, S, and its environment, E. The environment can be a similar quantum system to the one we observe, or much larger: we leave this choice completely open in order to be as general as possible. Let the joint 'S+E' state initially be disentangled, $|\psi\rangle_S|\psi\rangle_E$, after which we apply a unitary evolution U_{SE} on 'S+E' resulting in the state

$$U_{SE}|\psi\rangle_S|\psi\rangle_E \,. \tag{3.24}$$

Since we are interested in the system's evolution only, to obtain its final state, ρ_S, we have to trace over the environment, i.e.

$$\rho_S = \mathrm{Tr}_E(U_{SE}\,|\psi\rangle_S\langle\psi|_S \otimes |\psi\rangle_E\langle\psi|_E\, U_{SE}^\dagger) \,. \tag{3.25}$$

Another way to obtain the same result is to exclude the environment from the picture completely by defining operators of the 'complete measurement' [56, 57, 58] (sometimes also referred to as the Positive Operator Valued Measure, POVM)

$$\sum_i A^{i\dagger}A^i = \mathbb{1} \tag{3.26}$$

which act on the system alone, and therefore to be equivalent to the above system's evolution they must satisfy

$$\sum_i A^i|\psi\rangle_S\langle\psi|_S A^{i\dagger} = \rho_S \,. \tag{3.27}$$

Let us now derive the necessary form of A's using eq. (3.24). Let an orthonormal basis of E be $\{|\phi\rangle_E^i\}$. Then,

$$A^i = {}_E\langle\phi|_E^i U_{SE}|\psi\rangle_E \,. \tag{3.28}$$

It can easily be checked that the above $\{A^j\}$'s satisfy the completeness relations in eq. (3.26). Since the choice of basis for E is not unique, then neither is the choice of complete measurement operators. In fact, there is an infinite number of possibilities for the operators $\{A^j\}$. Note that the dimension of the complete measurement, A, is in general different to the dimension of the observed system, and is in fact equal to the dimension of E. Although this infinity at first sight appears to be creating problems for concrete calculations, it is, in fact, unnecessary. This is a consequence of the fact that a set of complete measurements is a convex set. Namely, if Φ_i are complete measurements, and $0 \leq \lambda_i \leq 1$ are such that $\sum_i \lambda_i = 1$, then $\sum_i \lambda_i \Phi_i$ is also a complete measurement. This implies that we need no more than d^2 terms in the sum, where $d = \dim \mathcal{H}_S$. This is a general result for convex sets, and is known under the name of Caratheodory's theorem [59]. We will prove this theorem in the following chapter, when this result will be of great importance in quantifying quantum correlations. Now, having developed a formalism for describing quantum stochastic evolution through measurements, we consider the problem of local increase of correlations in quantum mechanics.

3.1.2.2 Local Interactions Cannot Increase Correlations

The central problem addressed now, and described in the classical case in the previous section, is summarised in the following theorem:

Theorem. Correlations, as measured by the von Neumann mutual information, do not increase during local complete measurements carried on two entangled quantum systems.

We present here two quite separate, but mathematically rigorous proofs of this

theorem, the first using the notion of entropy, the second using the ideas of complete measurements and conditional entropy as a measure of relative information.

Proof 1. This proof is due to Partovi [60], who proved it as a general result, rather than applying it to increasing correlations by local operations. We decide to drop the subscript N for the von Neumann entropy since there is no possibility of confusion with the Shannon entropy. Consider three quantum systems A, B, C, initially in the state described by a density matrix of the form: $\rho_{ABC}(0) = \rho_{AB}(0)\rho_C(0)$, i.e. A and B are initially correlated and both are completely independent of C. We are now going to let B and C interact and evolve unitarily for time t, resulting in the state $\rho_{ABC}(t)$. The partial trace is defined in the usual fashion, e.g. $\rho_{AB}(t) = \mathrm{Tr}_C\rho_{ABC}(t)$, and similarly for all the other subsystems. Now we use the strong subadditivity [60] applied to $A + B + C$ at time t to obtain

$$S_{ABC}(t) + S_B(t) \leq S_{AB}(t) + S_{BC}(t) \ . \tag{3.29}$$

But $S_{ABC}(t) = S_{ABC}(0)$, as the whole system evolves unitarily. Also, $S_{ABC}(0) = S_{AB}(0) + S_C(0)$, since at the beginning C is independent of A, B. A is only a spectator in the evolution of B and C, so that, as shown above, $S_A(t) = S_A(0)$, $S_{BC}(t) = S_{BC}(0)$. Finally, there are no correlations between B and C at the beginning, implying: $S_{BC}(0) = S_B(0) + S_C(0)$. Invoking the definition in eq. (3.21) for the amount of correlations, and using the above properties and strong subadditivity in eq. (3.29), we arrive at the following

$$I(\rho_A : \rho_B; \rho_{AB})(t) \leq I(\rho_A : \rho_B; \rho_{AB})(0) \ . \tag{3.30}$$

Adding another system D to interact with A locally would lead to the same conclusion, hence completing the proof □.

Proof 2. This proof is a quantum analogue of the well known classical result that can loosely be stated as 'Stochastic processes cannot increase correlations' and which was presented in the previous chapter. We will now describe the interactions of $A+B$ with C and D in terms of complete measurements performed on $A+B$. Let

the state of $A+B$ be initially described by the density operator ρ, whose diagonal elements, ρ_{ii}, give the probabilities of being in various states, depending on the basis of the density matrix. Let this state undergo a complete measurement, described by operators A^j, such that

$$\sum A^{\dagger j} A^j = \mathbb{1} \,. \tag{3.31}$$

The new diagonal elements are then:

$$\rho'_{ii} = \sum_{nlm} A^n_{il} \rho_{lm} A^{\dagger n}_{\ mi} \,. \tag{3.32}$$

Let us introduce a relative information measure to ρ_{ii}: to each value of ρ_{ii} we assign a nonnegative number a_{ii}. We now wish to compare the distance [29] between ρ and a before and after (ρ' and a') the complete measurement, A (this will in fact be done using the Shannon relative entropy given in Definition 2.). We note that this measure of correlations is more appropriate in the classical case than in the quantum case, since for the latter it is not invariant under local unitary transformations and can in addition be infinite. The distance after the measurement is:

$$\begin{aligned}
\sum_i \rho'_{ii} \log \frac{\rho'_{ii}}{a'_{ii}} &= \sum_i \left(\sum_{nlm} A^n_{il}\, \rho_{lm}\, A^{\dagger n}_{\ mi} \right) \log \frac{\sum_{nlm} A^n_{il}\, \rho_{lm}\, A^{\dagger n}_{\ mi}}{\sum_{nlm} A^n_{il}\, a_{lm}\, A^{\dagger n}_{\ mi}} \\
&\leq \sum_i \left(\sum_{nlm} A^n_{il}\, \rho_{lm}\, A^{\dagger n}_{\ mi} \right) \log \frac{A^n_{il}\, \rho_{lm}\, A^{\dagger n}_{\ mi}}{A^n_{il}\, a_{lm}\, A^{\dagger n}_{\ mi}} \\
&= \sum_i \left(\sum_{nlm} A^n_{il}\, \rho_{lm}\, A^{\dagger n}_{\ mi} \right) \log \frac{\rho_{lm}}{a_{lm}} \\
&= \sum_{lm} \rho_{lm} \left(\sum_{in} A^n_{il}\, A^{\dagger n}_{\ mi} \right) \log \frac{\rho_{lm}}{a_{lm}} \\
&= \sum_{lm} \rho_{lm} \delta_{lm} \log \frac{\rho_{lm}}{a_{lm}} \\
&= \sum_l \rho_{ll} \log \frac{\rho_{ll}}{a_{ll}} \,,
\end{aligned} \tag{3.33}$$

where for the inequality in the second line we have used one of the consequences of the concave property of the logarithmic function [29, 52], and in the fifth line we used the completeness relation given in eq. (3.31). The locality of the complete measurement A is used in the fact that a' is disentangled if a is disentangled,

which is a necessary requirement for the above measure to be meaningful. This implies that the distance between the density matrix distribution and the relative information measure decreases by making a complete measurement. If we now consider the particular case where a is taken to be a distribution generated by the direct product of the reduced density matrices (i.e. if we assume no correlations), then the result above implies that the full density matrix becomes 'more like' the uncorrelated density matrix. From this, the theorem immediately follows □.

The Shannon mutual information, although having the above desired property, does not distinguish between the quantum and classical correlations. In order to do this we will have to introduce the possibility of classical communication between A and B. This will allow classical correlations to increase while leaving quantum correlations intact, as will be seen in the following chapter. We have emphasised a number of times that the quantum correlations can be higher that their classical counter-part, and a manifestation of this fact is described next through the statement of Bell's inequalities.

3.1.3 Bell's Inequalities

Bell's inequalities concern correlations between observables pertaining to two entangled quantum subsystems. We first derive the Clauser-Horne-Shimony-Holt (CHSH) form of Bell's inequality. Our derivation is based on the *Locality Principle* which we formulate following Redhead [61]:

Locality Principle. *A sharp value for an observable cannot be changed into another sharp value by altering the setting of a remote piece of apparatus.*

Consider an EPR pair of spin $\frac{1}{2}$ particles distributed between two observers: Alice and Bob. Let Alice perform a measurement on her particle of the value of spin in two different directions specified by vectors $\mathbf{a}$ and $\mathbf{a}'$. Let Bob perform the same kind of measurement on his particle, in directions given by $\mathbf{b}$ and $\mathbf{b}'$. The Locality Principle will reflect itself in our calculations as the fact that the measurements

of Bob will in no way affect the measurements of Alice, and vice versa. When expressed in units of $\hbar/2$ the result of any of these measurements belongs to the set $\{1, -1\}$. Consider now an observable constructed in the following way:

$$\gamma_n := a_n b_n + a_n b'_n + a'_n b_n - a'_n b'_n \tag{3.34}$$

where subscript n refers to result of the nth measurement of the corresponding observable. By writing γ in a different way

$$\gamma_n = a_n(b_n + b'_n) + a'_n(b_n - b'_n) \tag{3.35}$$

we observe that $|\gamma_n| \leq 2$. We now average γ over N measurements, and define correlation coefficients as:

$$c(a, b) := \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} a_n b_n \tag{3.36}$$

and similarly for other three expressions. Given this we can now state Bell's inequality as:

$$\left| \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \gamma_n \right| = |c(a, b) + c(a, b') + c(a', b) + c(a', b')| \leq 2 \,. \tag{3.37}$$

We emphasise that the *only* assumption that went into the above derivation was the Locality Principle. There exists, however, a class of states of EPR pairs that violate Bell's inequality. This happens because quantum mechanically the value of a_n, for instance, is generally not defined before the measurement and depends on whether Bob measures b or b', *i.e.* it is somehow influenced by *altering the setting of a remote piece of apparatus.* Note that this, nevertheless, does not allow for superluminal communications between Alice and Bob. This is related to the fact that whatever Alice or Bob do locally, *providing they do not communicate via a classical channel*, will not affect the other party's reduced density matrix. To prove this let us perform a complete measurement on A, defined by $\sum_i A^{i\dagger} A^i \otimes \mathbb{1}_B = \mathbb{1}$ where the identity in the direct product signifies that the other subsystem does not undergo any interaction. Let the overall state of 'A+B' be described by ρ. Then

after A has undergone a complete measurement, B's reduced density matrix is given by:

$$\begin{aligned} \rho'_B &= \mathrm{Tr}_A \left\{ \sum_i A^i \otimes \mathbb{1}_B \; \rho \; A^{i\dagger} \otimes \mathbb{1}_B \right\} \\ &= \sum_i \mathrm{Tr}_A \{ \rho A^{i\dagger} A^i \otimes \mathbb{1}_B \} = \mathrm{Tr}_A \left\{ \rho \sum_i A^{i\dagger} A^i \otimes \mathbb{1}_B \right\} \\ &= \mathrm{Tr}_A \{\rho\} = \rho_B \; . \end{aligned} \tag{3.38}$$

The impossibility of superluminal communication puts severe constraints on possible non-linear modifications of Schrödinger's equation to accommodate wave-function collapse in general (see e.g. [62]).

Following the Horodecki family [63] let us now analyse the situation from the quantum mechanical point of view. The quantum analogue of γ is

$$B = \hat{\mathbf{a}}\sigma \otimes (\hat{\mathbf{b}} + \hat{\mathbf{b}}')\sigma + \hat{\mathbf{a}}'\sigma \otimes (\hat{\mathbf{b}} - \hat{\mathbf{b}}')\sigma \tag{3.39}$$

where hats signify unit vectors and σ's represent Pauli matrices. If $\hat{\mathbf{a}} = (a_1, a_2, a_3)$ then $\hat{\mathbf{a}}\sigma = a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3$ is the operator representing the spin observable in the direction of $\mathbf{a}$. We can immediately write Bell's inequality as:

$$|\langle B \rangle_\rho| \leq 2 \tag{3.40}$$

where ρ is the density matrix describing the state of the EPR pair. The average of operator B is given by the well known formula

$$\langle B \rangle_\rho = \mathrm{Tr}(\rho B) \; . \tag{3.41}$$

We now introduce a necessary and sufficient condition for violating the inequality in eq. (3.40). Any two spin $\frac{1}{2}$ particle state can be described by a density matrix of the form

$$\rho = \frac{1}{4}(\mathbb{1} \otimes \mathbb{1} + \mathbf{r}\sigma \otimes \mathbb{1} + \mathbb{1} \otimes \mathbf{s}\sigma + \sum_{n,m=1}^{3} t_{nm} \sigma_n \otimes \sigma_m) \tag{3.42}$$

where I stands for the identity operator, σ's are Pauli matrices and $\mathbf{r}$ and $\mathbf{s}$ are vectors in R^3. Note that the above form already contains the normalisation condition

$tr(\rho) = 1$. Additional constraints on r_i, s_i and t_{nm} imposed by the non-negativity condition and by the fact that $\mathrm{Tr}(\rho^2) \leq 1$ have not been included in the above as they are not important for our present analysis. Let T_ρ be the real matrix whose elements are $t_{nm} = \mathrm{Tr}(\rho\sigma_n \otimes \sigma_m)$. Define $U_\rho := T_\rho^T T_\rho$, which, being a symmetric matrix, can always be diagonalized. Denoting by u and $\tilde{u}$ the two largest, by definition positive, eigenvalues of U_ρ, we define

$$M(\rho) := u + \tilde{u} \,. \tag{3.43}$$

Horodecki's theorem then reads:

Theorem. *The necessary and sufficient condition for ρ to violate Bell's inequality is that $M(\rho) > 1$.*

The knowledge of the proof of this theorem is not necessary for further discussion and we refer the interested reader to the Horodecki's original paper [63] for a detailed proof. It can now be seen that the product states $\rho_A \otimes \rho_B$ obey Bell's inequalities, and non-product pure states violate them [64]. In fact, any mixture of product states $\sum_i p_i \rho_A^i \otimes \rho_B^i$ likewise obeys Bell's inequalities [65]. A question is whether all other states can be considered to be entangled? The answer is positive for two qubits and a qubit and a three level system, since from any state not of this (separable) form we can "distill" a subensemble of maximally entangled states (e.g. singlets) by using only local operations and classical communication [66](these procedures are referred to as *purification procedures* and will be described in greater detail below and are mathematically formalized in Chapter 4). However, some of these *inseparable* states satisfy Bell's inequalities. This leads us to conclude that violation of Bell's inequalities is not a sufficient condition for having an entangled state in general (i.e. for mixed states). We now present a simple example.

The idea to increase nonlocality by local interactions originates from Gisin's analysis of Polarisation Dependent Losses (PDL) in optical fibres [67]. As its name indicates this effect reflects itself in the observation that the fibre absorption is dependent on the polarisation of the propagated light. Fibres of this kind have

already been manufactured, and the PDL observed. Gisin proposes the following experiment. Suppose that Alice and Bob each have an optical fibre with PDL. Let Alice's fibre attenuate only spin "up" photons ($|1\rangle$) and Bob's fibre only spin "down" photons ($|0\rangle$). Let the initial state of the photon pair, before propagating through the fibres, be:

$$\rho_i(\lambda,\alpha) = \lambda P_{\psi_{\alpha,\beta}} + \frac{1-\lambda}{2}(P_{\psi_{1,1}} + P_{\psi_{0,0}}) \tag{3.44}$$

where P's denote projections onto onto the following states: $\psi_{\alpha,\beta} = (\alpha|0\rangle_A|1\rangle_B + \beta|1\rangle_A|0\rangle_B)$, $\psi_{1,1} = |1\rangle_A|1\rangle_B$ and $\psi_{0,0} = |0\rangle_A|0\rangle_B$, and $\beta = \sqrt{1-\alpha^2}$. If we assume that $\lambda > \frac{1}{2(1-\alpha\beta)}$ the condition that $M(\rho) \leq 1$ becomes

$$\lambda < \frac{1}{1+\alpha^2\beta^2} \tag{3.45}$$

and no violation of Bell's inequality occurs. Let now the pair be distributed to Alice and Bob and be propagated through the fibres. According to Gisin the action of the fibres with PDL are described by two diagonal matrices:

$$T_A = \begin{pmatrix} \sqrt{\frac{\beta}{\alpha}} & 0 \\ 0 & 1 \end{pmatrix} \qquad T_B = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{\frac{\beta}{\alpha}} \end{pmatrix}$$

where the fibres are manufactured in such a way that the loss is equal to $\sqrt{\frac{\beta}{\alpha}}$ in both fibres. When the photons have propagated through, the final state is given by:

$$\rho_f(\lambda,\alpha) = \frac{1}{N}\left(2\alpha\beta\lambda P_{\psi_{\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}}} + \frac{1-\lambda}{2}(P_{\psi_{1,1}} + P_{\psi_{0,0}})\right) \tag{3.46}$$

where $\psi_{\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}} = \frac{1}{\sqrt{2}}(|0\rangle_A|1\rangle_B + |1\rangle_A|0\rangle_B)$, and $N = \frac{1}{2\alpha\beta\lambda+(1-\lambda)}$ is the normalization constant. Note that at this point Alice and Bob will have to communicate to each other about the presence or absence of their photons. They then subselect the ensemble in which both the photons are present. This is a crucial step, since without the communication and subselection the entire ensemble after the action of PDL fibres could not be more entangled. This (subselected) state now violates Bell's inequality iff

$$\lambda > \frac{1}{1+2\alpha\beta(\sqrt{2}-1)}\,. \tag{3.47}$$

Consider now the following values for our parameters: $\lambda = 0.9$, $\alpha\beta = 0.2$. Then $M(\rho_i) = 0.77 < 1$ and $M(\rho_f) = 1.22 > 1$. Thus the interaction of photons with their local environments apparently creates a nonlocal state out of a local one. This is, of course, incorrect and we have to conclude that the initial state was already entangled even though it satisfies Bell's inequalities. Thus Bell's inequalities are not entirely satisfactory for distinguishing between entangled and disentangled states. This idea of locally concentrating entanglement will be the basis of our quantification of entanglement in Chapter 4. Before considering this central problem, we review how entanglement affects quantum communications and how it makes quantum computation exponentially more efficient than classical computation. This will be a good illustration of the extent to which quantum correlations are fundamental in communications and computation and hence provide the main motivation behind this thesis.

3.2 Quantum Communication Theory

We do not intend an in-depth and rigorous presentation of this subject here, but will only focus on the role of entanglement in communications. The question of quantum channel capacities is an area not entirely understood at present, and to which this thesis can provide a valuable contribution. Open questions for further research will be presented in Chapters 4 and 7.

A quantum communication channel (QCC) consists of a number, N, of quantum systems prepared in states $\rho_1, \rho_2 \ldots \rho_N$. These states are then sent by Alice one after another with certain a priori probabilities, $p_1, p_2, \ldots p_N$, to the receiver, Bob, who then performs measurement to "decipher" the correct sequence of states comprising a message. If the states suffer no error on the way to the receiver, then the channel is called noiseless; otherwise it is noisy. First we consider the notion of capacity of a noiseless QCC, since the generalization to a noisy channel is then straightforward.

Let $S(\rho) = -\mathrm{Tr}\rho \ln \rho$ be the standard von Neumann entropy of a density matrix

ρ. Then, we define the capacity of a QCC as

$$\mathbf{C} := \max_{\{p\}} C(\{p\}, \rho) \tag{3.48}$$

where

$$C(\{p\}, \rho) = S(\sum_i p_i \rho_i) - \sum_i p_i S(\rho_i) \ , \tag{3.49}$$

is the so called Holevo bound. Note that the above can be expressed succinctly as

$$C(\{p\}, \rho) = \sum_i p_i S(\rho_i || \rho) \ , \tag{3.50}$$

where $S(\ ||\ )$ is the von Neumann relative entropy and $\rho = \sum_i p_i \rho_i$. When there is no possibility of confusion we write $C(\{p\}, \rho) \equiv C(\{p\})$. To see the physical motivation behind this consider only two states ρ_1 and ρ_2 sent by Alice to Bob according to probabilities p_1 and $p_2 = 1 - p_1$ respectively. Bob now performs a set of POVMs $\sum_i E_i = I$ in order to determine which state was sent to him. The *accessible information* to Bob is given by the mutual information between ρ_1 and ρ_2 [11, 57], defined as

$$\begin{aligned} I(\rho_1 : \rho_2) &= \max_E \Big\{ \sum_i -\mathrm{Tr}(\rho E_i) \ln(\mathrm{Tr}(\rho E_i)) + p_1 \mathrm{Tr}(\rho_1 E_i) \ln(\mathrm{Tr}(\rho_1 E_i)) \\ &+ p_2 \mathrm{Tr}(\rho_2 E_i) \ln(\mathrm{Tr}(\rho_2 E_i)) \Big\} \qquad (3.51) \\ &= \max E_i \sum_i \Big\{ p_1 tr(\rho_1 E_i) \ln \frac{(\mathrm{Tr}(\rho_1 E_i))}{(\mathrm{Tr}(\rho_0 E_i))} \\ &+ p_2 \mathrm{Tr}(\rho_2 E_i) \ln \frac{(\mathrm{Tr}(\rho_2 E_i))}{(\mathrm{Tr}(\rho_0 E_i))} \Big\} \ . \qquad (3.52) \end{aligned}$$

The Holevo bound is an upper bound of the above accessible information

$$S(\sum_i p_i \rho_i) - \sum_i p_i S(\rho_i) \geq I(\rho_1 : \rho_2) \ , \tag{3.53}$$

where the equality is saturated if and only if $[\rho_i, \rho_j] = 0$ for all i and j [28]. Therefore since the Holevo bound is an upper bound to information that Bob can gain about Alice's signal we identify its maximum over all possible initial probabilities with the classical capacity of a quantum channel. Note that one of the most profound implications of the Holevo bound is that a quantum bit cannot store more

information than a classical bit. In spite of this limitation, quantum computing is more efficient than its classical analogue due to different natrure of information, which is reflected in the existence of entanglement between qubits.

We would like to emphasise the similarity of this expression to the capacity of a classical communication channel [10]. This exists because the Holevo bound is defined in direct analogy with the classical notion of mutual information, whose maximization leads to the notion of classical channel capacity as described before. This, as we might expect, will happen when all ρ_is are diagonal in the same basis, i.e. they commute. Let us call this representation the B representation, with orthonormal eigenvectors $|b\rangle$. Then the probability that the measurement of the symbol represented by ρ_i will yield the value b is just $\langle b|\rho_i|b\rangle$. This we call the conditional probability $p_i(b)$, that if ρ_i was sent the result b was obtained. The conditional entropy is now given by

$$S_B(\rho_i) = \sum_i p_i \sum_b \langle b|\rho_i|b\rangle \ln\langle b|\rho_i|b\rangle = \sum_i p_i \rho_i \ . \tag{3.54}$$

Now the classical capacity would be

$$\mathbf{C} = S(\rho) - S_B(\rho_i) = S(\rho) - \sum_i p_i \rho_i \ , \tag{3.55}$$

which is identical to the Holevo bound. In general, the usual rule of thumb for obtaining quantum information theoretic quantities from their classical counterparts is by the convention

$$\sum \ \longrightarrow \ \text{Trace} \tag{3.56}$$

$$\sum p(a) \ln p(a) \ \longrightarrow \ \text{Tr}\rho_A \ln \rho_A \tag{3.57}$$

$$\sum p(a) \ln \frac{p(a)}{q(b)} \ \longrightarrow \ \text{Tr}\{\rho_A \ln \rho_A - \rho_A \ln \rho_B\} \ , \tag{3.58}$$

so that, for example, the Shannon mutual information $I(A : B) = S(p(a)) + S(p(b)) - S(p(a,b))$ now becomes the von Neumann mutual information $I(\rho_A, \rho_B; \rho_{A,B}) = S(\rho_A) + S(\rho_B) - S(\rho_{A,B})$. Next we prove a result which shows that without entanglement we cannot increase the classical capacity of a quantum channel. The

author is not aware of any similar proof of this type, although the general result is well known in the literature [18].

3.2.1 Disentangled Channels are Additive

First we show that if we have several quantum channels in parallel, through which we send a disentangled input state, the capacity will be equal to the sum of the capacities of the individual channels.

Proposition. Classical capacities of QCC with disentangled inputs are additive.

Proof. We show this in the case of two QCC; the general result follows immediately. Let the input to the two QCC be in the most general disentangled state

$$\rho^{1+2} = \sum_{ij} p_{ij} \rho_i^1 \otimes \rho_j^2 \tag{3.59}$$

where $\sum_i p_{ij} = p_j$ and $\sum_j p_{ij} = p_i$. From the Araki-Lieb inequality it follows that,

$$S(\sum_{ij} p_{ij} \rho_i^1 \otimes \rho_j^2) \leq S(\sum_i \rho_i^1) + S(\sum_j \rho_j^2) \; . \tag{3.60}$$

Thus,

$$\begin{aligned} C(\{p_{ij}\}) &:= S(\sum_{ij} p_{ij} \rho_i^1 \otimes \rho_i^2) - \sum_{ij} p_{ij} S(\rho_i^1 \otimes \rho_j^2) && (3.61) \\ &\leq S(\sum_i \rho_i^1) + S(\sum_j \rho_j^2) - \sum_i p_i S(\rho_i^1) - \sum_j p_j S(\rho_j^1) && (3.62) \\ &:= C_1(\{p_i\}) + C_2(\{p_j\}) \; . && (3.63) \end{aligned}$$

Taking the maximum of both of the sides of the above inequality we obtain

$$\mathbf{C_{1+2} \leq C_1 + C_2} \; . \tag{3.64}$$

However we know that the equality can be reached for $p_{ij} = p_i p_j$ which proves the above proposition □. This result is related to the result of classical information theory where two classical channels have capacity always less than the sum of individual capacity if there inputs are correlated. If we wish to achieve the maximum capcity of the two channels their inputs should be uncorrelated, a physically intuitive property [68].

3.2.2 Entangled Channels are Superadditive

In contrast with the above result is the fact that if the inputs to parallel channels are entangled, then the total capacity can be greater than the sum of the individual capacities – a property known as the superadditivity of classical capacities of QCC. There is no rigorous general proof of this fact, but there are a number of particular examples corroborating this result [69, 70]. We will explain the role of entanglement through the following discussion involving two channels. It is well known that the von Neumann relative entropy is superadditive (as opposed to the von Neumann entropy which is subadditive). Namely, we have that

$$S(\omega^{1+2}||\phi^1 \otimes \phi^2) \geq S(\omega^1||\phi^1) + S(\omega^2||\phi^2) \ , \tag{3.65}$$

where $\omega^1 = \mathrm{Tr}_2\omega^{1+2}$ and $\omega^2 = \mathrm{Tr}_1\omega^{1+2}$. So, imagine that states $\{\omega_i^1\}_{i=1}^N$ and $\{\omega_j^2\}_{j=1}^N$ are used as the respective inputs to the two QCC's such that

$$\sum p_i\omega_i^1 = \phi^1 \tag{3.66}$$

$$\sum p_j\omega_j^2 = \phi^1 \ , \tag{3.67}$$

and in addition

$$\sum_i p_i\omega_{12}^i = \phi^1 \otimes \phi^2 \ . \tag{3.68}$$

These three conditions will be needed so that the corresponding von Neumann relative entropies do represent channel capacities. Then it follows from eq. (3.65) that

$$\sum p_i S(\omega_i^{1+2}||\phi_i^1 \otimes \phi_i^2) \geq \sum p_i S(\omega_i^1||\phi_i^1) + \sum p_j S(\omega_j^2||\phi_j^2) \ . \tag{3.69}$$

Now using the von Neumann relative entropy representation of the channel's capacity in eq. (3.50) and taking the maximum of both sides in the above, we derive

$$\mathbf{C_{1+2}} \geq \mathbf{C_1} + \mathbf{C_2} \ , \tag{3.70}$$

which would imply that entanglement can increase classical capacity of a QCC. The problem with the last step is that it is not clear that, in the process of maximizing,

all three conditions in eq. (3.67) and eq. (3.68) can be maintained. Thus, the above line of thought, although not being a general proof, does offer a heuristic argument supporting the claim that entanglement can increase classical capacity of a QCC. An open question is whether we can use this difference $\mathbf{C_{1+2}} - (\mathbf{C_1} + \mathbf{C_2})$ to quantify the purely quantum capacity of a QCC. To that end let us explain what the quantum capacity of a QCC would be. This second, and different scenario of communication, is if Alice wishes to transmit a given (to her *unknown*) state of a quantum system to Bob as accurately as possible through a noisy QCC [71]. This general problem is closely related to the notion of entanglement and can be understood as follows. Suppose that Alice prepares an entangled state $\alpha|00\rangle + \beta|11\rangle$. She now sends one of the particle to Bob through a noisy channel described by a decoherent (POVM) measurement, Φ, such that

$$\alpha|00\rangle + \beta|11\rangle \longrightarrow \alpha|0\rangle\Phi(|0\rangle) + \beta|1\rangle\Phi(|1\rangle) \ , \tag{3.71}$$

where the positive map Φ describes the action of the QCC. Once they have established some entanglement between them Alice can use the standard teleportation protocol [72] to send her state to Bob. Teleportation is a protocol between Alice and Bob who share an entangled pair. Then Alice receives qubit in an *unknown* state and performs a measurement on her two qubits, the outcome of which she then communicates to Bob. Bob then performs an appropriate measurement on his qubit which, as a result, is in the state of Alice's original unknown qubit. The entangled pair they originally shared is destroyed at the end of the protocol. Since the efficiency of teleportation is directly dependent on the amount of entanglement shared between Alice and Bob, we first need to understand how to quantify entanglement in order to study this case. We turn to this problem in Chapter 4, but before that we review some basic concepts of quantum computation, where entanglement also plays a central role.

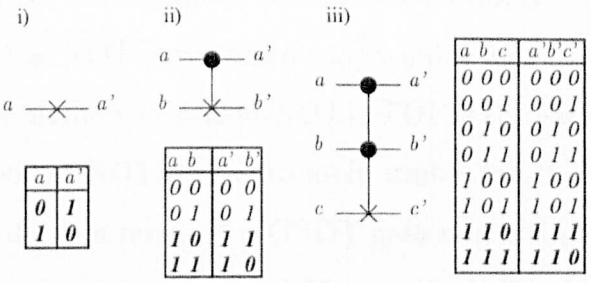
3.3 Quantum Computation

A quantum computer is a physical system that can accept input states which represent a coherent superposition, i.e. an entangled state, of many different possible basis states and subsequently evolve them into a corresponding superposition of outputs. Computation, *i.e.* a sequence of unitary transformations, affects simultaneously each element of the superposition, generating a massive parallel data processing albeit within one piece of quantum hardware [14]. This way quantum computers can efficiently solve some problems which are believed to be intractable on any classical computer [16, 17]. Therefore the advantage of a quantum computer lies in the exploitation of the phenomenon of entanglement. The great importance of the quantum theory of computation is in the fact that it reveals the fundamental connections between the laws of physics and the nature of computation and mathematics [73]. However, practical realization of a quantum computer is limited by the problem of decoherence, i.e. by the fact that the computer interacts with its environment and hence undergoes errors. Here, methods of quantum error correction are also useful and this area has been extensively studied recently. We will develop this subject in Chapter 5 and then apply it to studying entanglement in Chapter 6.

3.3.1 Quantum Gates

For the purpose of this thesis a quantum computer will be viewed as a quantum network (or a family of quantum networks) composed of quantum logic gates; each gate performing an elementary unitary operation on one, two or more two–state quantum systems called qubits [15]. Each qubit represents an elementary unit of information; it has a chosen "computational" basis $\{|0\rangle, |1\rangle\}$ corresponding to the classical bit values 0 and 1. Boolean operations which map sequences of 0's and 1's into another sequences of 0's and 1's are defined with respect to this computational basis.

Any unitary operation is reversible and that is why quantum networks effecting



a	a'
0	1
1	0

a b	a' b'
0 0	0 0
0 1	0 1
1 0	1 1
1 1	1 0

a b c	a'b'c'
0 0 0	0 0 0
0 0 1	0 0 1
0 1 0	0 1 0
0 1 1	0 1 1
1 0 0	1 0 0
1 0 1	1 0 1
1 1 0	1 1 1
1 1 1	1 1 0

Figure 3.1: Truth tables and graphical representations of the elementary quantum gates used for the construction of more complicated quantum networks. The control qubits are graphically represented by a dot, the target qubits by a cross. i) NOT operation. ii) Control–NOT. This gate can be seen as a *bitwise* "copy operation" in the sense that a target qubit (b) initially in the state 0 will be after the action of the gate in the same state as the control qubit. iii) TOFFOLI gate. This gate can also be seen as a Control–control–NOT: the target bit (c) undergoes a NOT operation only when the two controls (a and b) are in state 1.

elementary arithmetic operations such as addition, multiplication and exponentiation cannot be directly derived from their classical Boolean counterparts (classical logic gates such as AND or OR are clearly irreversible: reading 1 at the output of the OR gate does not provide enough information to determine the input which could be either $(0,1)$ or $(1,0)$ or $(1,1)$). Quantum arithmetic must be built from *reversible logical components*. It has been shown that reversible networks (a prerequisite for quantum computation) require some additional memory for storing intermediate results [74, 75]. Hence the art of building quantum networks is often reduced to minimising this auxiliary memory or to optimising the trade–off between the auxiliary memory and a number of computational steps required to complete a given operation in a reversible way (see [76] for some optimization techniques). Figure 3.1 presents three basic reversible gates used in quantum computing: NOT gate, Controlled NOT gate and TOFFOLI gate. Controlled NOT gate (CNOT, for short) is a two qubit gate, where the value of the first qubit (called control) determines

what will happen to the second qubit (called target) qubit. Namely if the control qubit is 1, we apply the NOT gate to the target qubit and otherwise nothing happens to it (hence the name Controlled NOT). TOFFOLI gate can be understood as Controlled–Controlled NOT. An extremely useful result is that any quantum computation can be done in terms of a CNOT gate and a single qubit gate [77, 78] (which varies), although, of course, it might sometimes be more convenient to use other gates as well [79]. These basic gates will be directly used in Chapter 5 and Chapter 6, where we will also present their cavity QED implementation. Another important one qubit gate is the so called Hadamard transformation whose action is the following (the normalization is omitted)

$$|0\rangle \longrightarrow |0\rangle + |1\rangle \tag{3.72}$$

$$|1\rangle \longrightarrow |0\rangle - |1\rangle \, . \tag{3.73}$$

This transformation will also be used frequently in the later chapters.

An interesting and important question is how to create an entangled EPR state starting form just a disentangled, say, $|01\rangle$ state. The required quantum computation is very simple: first we apply a Hadamard transformation to the first qubit, and then a Controlled Not between the first qubit and the second qubit, where the second qubit is the target. These two steps can be written as

$$|01\rangle \longrightarrow (|0\rangle + |1\rangle)|1\rangle \longrightarrow |01\rangle + |10\rangle \, . \tag{3.74}$$

We see that after the action of the Hadamard transformation the qubits are still disentangled. This is because this transformation acts on only one of the qubits, i.e. is applied *locally* and not *globally*, and therefore cannot create global features such as entanglement. This principle that local operations cannot create non-local features is the central theme of this thesis.

This computational way of representing unitary evolution of quantum systems will be very helpful in describing quantum error correction and its use in preserving entanglement in Chapters 5 and 6. Next we briefly review the biggest breakthrough

in quantum computing, Shor's quantum algorithm for factorization of natural numbers.

3.3.2 Shor's Algorithm

The algorithm for factorization dates back to the Ancient Greeks (the book by Knuth in [80] is a bible for algorithms, containing a number of important classical computational problems). It was probably known to Euclid, and it can be described simply as follows. We wish to find the prime factors of N. This amounts to finding the smallest r such that $a^r \equiv 1(\mathrm{mod}N)$, where a is chosen to be coprime to N, i.e. so that a and N have no common divisors apart from 1. In other words, we want to determine the period of the function $a^r(\mathrm{mod}N)$. Let us see how this works for, say, $N = 15$.

- We choose a=2. Then obviously $\gcd(2, 15) = 1$.
- Next we compute $2^0, 2^1, \ldots 2^i$ modulo 15, and this gives $1, 2, 4, 8, 1, 2, 4, 8, \ldots$.
- This sequence is periodic with the period $r = 4$, which also satisfies $2^4 \equiv 1(\mathrm{mod}15)$.
- Once r is obtained we find the factors of N by computing $\gcd(a^{r/2} \pm 1, 15)$, which in our case is $\gcd(4 \pm 1, 15) = 3, 5$.

Hence we have factorised 15 into 3×5. Now this algorithm (or some of its close variants) can be implemented on a classical or on a quantum computer. To be able to compare their efficiency we need to know that there are two basic classes of problems:

1. *easy problems*: the time of computation T is a polynomial function of the size of the input l, i.e. $T = c_n l^n + ... + c_1 l + c_0$, where the coefficients c are determined by the problem.

2. *hard problems*: the time of computation is an exponential function of the size of the input (e.g. $T = 2^{cl}$, where c is problem dependent).

The size of the input is always measured in bits (qubits). For example, if we are to factorize 15 then we need 4 bits to store this number. In general, to store a number N we need about $l = \log N$, where the base of the logarithm is 2. (this is easy to see: just ask yourself how many different numbers can be written with l bits). The easy problems are considered as computationally efficient, or tractable, whereas the hard problems are computationally inefficient, or intractable. Now the upshot of this discussion is that, for a given N, there is no *known* efficient classical algorithm to factorise it. Let us illustrate how the simplest factorization algorithm performs: suppose that we want to determine the factors of N by dividing it by 2, then 3 then 4 and so on up to $\sqrt{N}$. So the time of computation (which is in fact the number of elementary steps) is proportional to the number of divisions we have to perform, and this is $\sqrt{N} = 2^{l/2}$, i.e. it is exponential. However, using a quantum computer and the above–described Euclid's approach to factorization, we can factor any N efficiently in polynomial time involving a linear number of qubits. This is essence of Shor's algorithm.

There are two distinct stages in this algorithm [17] (for an extensive review of this algorithm see [81]). Initially, we have two registers (plus several other registers containing garbage, but these are irrelevant for explaining the basic principle of quantum factorization) at the input to the quantum computer. First, we prepare the first register in a superposition of consecutive natural numbers, while leaving the second register in 0 state to obtain (as usual we omit the normalization)

$$|\Psi\rangle = \sum_{n=0}^{M-1} |n\rangle|0\rangle \tag{3.75}$$

where $M = 2^m$ is some sufficiently large number. Now in the second register we compute the function $a^i \mathrm{mod} N$. This can be achieved unitarily and the result is

$$|\Psi_1\rangle = \sum_{n=1}^{M-1} |n\rangle|a^n \mathrm{mod} N\rangle \; . \tag{3.76}$$

Here again we see the famous quantum parallelism in action. This completes the first stage of the algorithm and the trick now is to extract the period r from the first register. To help us visualise this let us think of our previous example when $N = 15$ and $a = 2$. Then we would have

$$\begin{aligned} |\Psi_1\rangle &= |0\rangle|2^0 \text{mod} 15\rangle + |1\rangle|2^1 \text{mod} 15\rangle + |2\rangle|2^2 \text{mod} 15\rangle + |3\rangle|2^3 \text{mod} 15\rangle + \\ &+ |4\rangle|2^4 \text{mod} 15\rangle + |5\rangle|2^5 \text{mod} 15\rangle + \ldots + |2^{M-1} \text{mod} 15\rangle \qquad (3.77) \\ &= |0\rangle|1\rangle + |1\rangle|2\rangle + |2\rangle|4\rangle + |3\rangle|8\rangle + |4\rangle|1\rangle + \ldots + |2^{M-1} \text{mod} 15\rangle \; . \qquad (3.78) \end{aligned}$$

Let us recall that we do not need to extract all the values of $2^i \text{mod} 15$, but just the period of this function. This now sounds very much like Deutsch's problem, where only the knowledge of a property of f was important and not both its values. The solution is likewise similar, but is however much more computationally involved. Suppose that we now perform a measurement on the second register to determine its state. Suppose further that we obtain 4 as the result. The remaining state will be

$$|\Psi_2\rangle = (|2\rangle + |6\rangle + |10\rangle + \ldots)|4\rangle \qquad (3.79)$$

so that the first register contains numbers repeating periodically with the period 4. This is now what we have to extract by manipulating the first register. To see how this works suppose for simplicity that r divides M exactly. For general a and N this state is

$$|\Psi_2\rangle = \sum_{j=0}^{A} |jr + l\rangle|l\rangle \qquad (3.80)$$

where $A = M/r - 1$ and the second register is obviously irrelevant. Extracting r involves performing a Fast Fourier Transform on the first register, so that the final state becomes

$$|\Psi_3\rangle = \sum_{j=0}^{r-1} \exp(2\pi i l j/r)|jM/r\rangle \qquad (3.81)$$

We can now perform a measurement in the $y = jM/r$ basis where j is an integer. Therefore, once we obtain a particular y we have to solve the following equation $y/M = j/r$ where y and M are known. Assuming that j and r have no common

factors (apart from 1) we can determine r by cancelling y/M down to an irreducible fraction. Once r is obtained we can easily infer the factors of N.

In general, of course, Shor's algorithm is probabilistic. This means that r, and hence the factors of N that we obtain by running the above quantum computation, might sometimes not be the right answer. However, whether the answer is right or wrong can be easily checked by multiplying the factors to get N. Since multiplication is an easy computation this can be performed efficiently on a classical computer. If the result is not N, we then repeat the whole Shor's algorithm all over again, and we keep doing this until we get the right answer. Shor showed that even with this random element his algorithm is still efficient. In fact, the most time consuming part is the first one, where we have to obtain the state in eq. (3.76). Modular exponentiation takes of the order of $(\log N)^3$ elementary gates and this dominates the whole algorithm [76]. We should say that the memory space, i.e. the number of qubits needed for the entire computation, is of the order of $\log N$. For completeness we state that all the above networks for addition, multiplication and exponentiation can be improved using standard computational techniques (see e.g. [82]), however, this improvement is not substantial and does not change the fundamental conclusion about the efficiency of quantum factorization.

3.3.3 Practical Realisations of Quantum Computers

In the previous subsections we have seen that quantum computation is a fundamentally new concept that promises the solution of problems which are intractable on classical computers. We will now address the question of how to implement such a quantum computer in practice. An important question is whether a quantum computer requires fundamentally new experimental techniques for its realization or whether already known techniques would be sufficient. In fact, some of the early proposals had the disadvantage of using somewhat "futuristic" experimental techniques. Then, however, a very beautiful model for an ion-trap quantum computer

was proposed by Cirac and Zoller [83] which employed only experimental methods which were already realized or which were expected to be realizable in the near future. Subsequently, other realistic suggestions, such as quantum computation based on nuclear magnetic resonance methods, have been proposed [84, 85, 86, 87]. Although these new proposals are very interesting, we confine ourselves here to the description of the linear ion trap implementation of Cirac and Zoller which exhibits all the important features of any quantum computer. The experimental situation is given in Fig. 3.2. The basic element is a linear ion trap: electrodes generate a time dependent electric quadrupole field which generates an effective (time independent) potential which has a minimum along the axis of the trap where the ions are trapped. The equilibrium between the trapping force of the effective potential and the mutual electrostatic repulsion of the ions is achieved when the ions form a string where adjacent ions are separated by a few wavelengths of light. Therefore, they can be addressed individually by a laser. The idea of a linear ion trap is basically equivalent to that of a Paul trap [88] which is being used to trap single ions for very long times. Linear ion traps are already working and it is possible to trap strings of 30 ions or more in them [89].

The practical problem with this proposal is the mechanical degree of freedom of the ions. Although the ions are trapped along the axis of the linear ion trap they are not at rest, but oscillate around their equilibrium position. After having trapped the ions, the next step is then to cool them using methods of laser cooling [90, 91]. While it is fairly standard today to cool ions to temperatures of the order of milli Kelvin, it is very difficult to cool them to the necessary ground state of motion, i.e. to a state in which only the unavoidable motion due to the quantum mechanical uncertainty principle is present. While a single ion has been cooled to its ground state of motion [92], no laboratory in the world has yet cooled a string of only two ions to the ground state of motion.

To see why we need to cool the ions, remember that we want to implement quantum gates between different qubits. In the ion trap, these qubits are localized

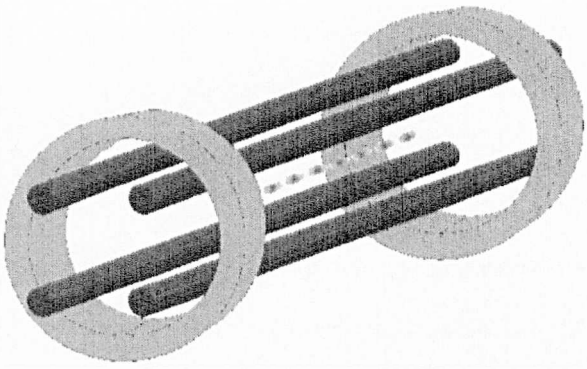


Figure 3.2: Schematic picture of a linear ion trap computer. Electrodes generate a time dependent electric field which generates an effective potential such that a string of ions (the blue dots in the middle of the trap) is trapped. The motion of the ions, and in particular the centre of mass mode, has to be cooled to its ground state. The centre of mass mode then acts as a bus that allows us to generate interactions between any two ions.

and we cannot really move them from one place to another. If we want to implement a quantum gate between two ions that are separated, e.g. one at the beginning of the string and one at the end, then we need some 'medium' that can be used for communication between these ions. Note that this communication is not classical but has to be quantum mechanical in nature as we want to establish quantum mechanical coherence between different ions. This communication is achieved by using the centre-of-mass mode of the ions. If we excite the centre-of-mass mode of the ions then all of the ions will oscillate and therefore all of them will feel this oscillation in the same way. Therefore even distant ions will be connected. This behaviour is illustrated in Fig. 3.3.

This idea of using the centre-of-mass mode as a 'bus' is the key ingredient in the ion trap quantum computer. It allows the implementation of two-bit gates such as a CNOT gate, for example. In the following we will briefly explain how one can implement a CNOT gate in a linear ion trap computer. More complicated gates can be constructed, but a CNOT gate together with one bit rotations is sufficient

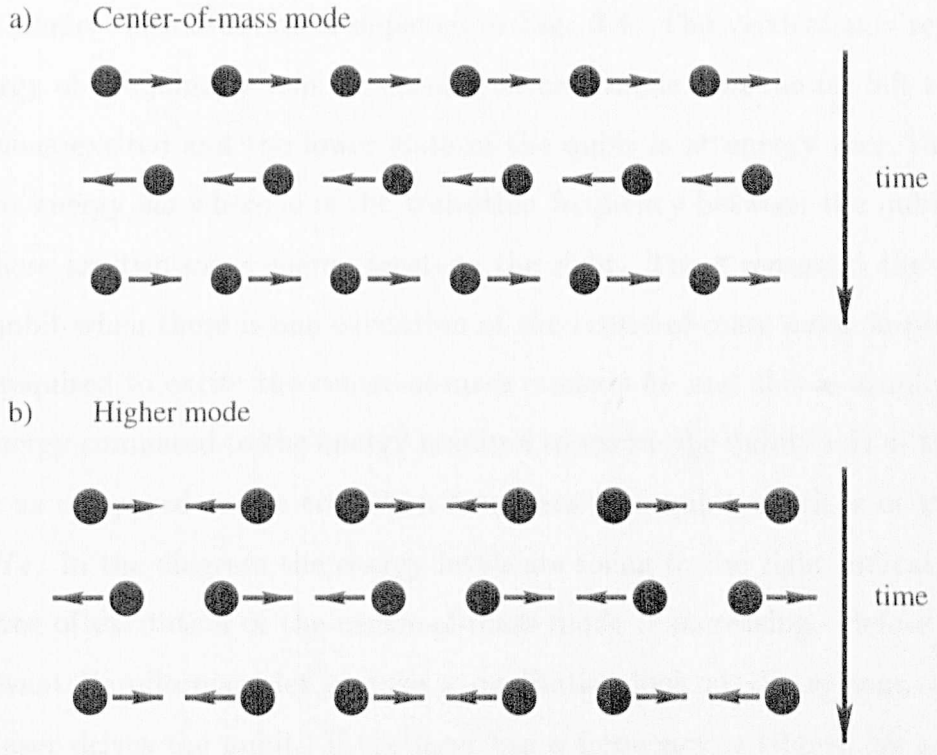


Figure 3.3: In part a) the centre-of-mass mode is illustrated. All the ions oscillate with the same phase. In part b) a mode of higher frequency is given. Here the ions have different phases and their relative distances change.

to implement any unitary operation between any number of quantum bits [77, 78]. Obviously single ion gates are simple as they are implemented by manipulating a single ion with a suitably made laser pulse. In a CNOT gate it is essential that the two qubits interact and this is achieved by exciting the centre of mass mode of all the ions in the trap. Therefore, before we describe how to implement a full quantum gate we first explain how we can excite the centre of mass mode of the ions. Let us first have a look at the energy levels of a single qubit and the centre of mass mode. The situation is depicted in Fig. 3.4. The vertical axis represents the energy of the joint system ion+centre-of-mass mode. On the far left there are no phonons excited and the lower state of the qubit is at energy zero, the upper state has energy $\hbar\omega$ where ω is the transition frequency between the qubit levels. Then there are two more energy levels to the right. These represent the energies of the qubit when there is one excitation of the centre-of-mass mode around. The energy required to excite the centre-of-mass mode is $\hbar\nu$ and this is usually a very small energy compared to the energy required to excite the qubit. ν is of the order of MHz as compared to the transition frequency in a qubit which is of the order of $10^{15} Hz$. In the diagram the energy levels are rising to the right indicating that the degree of excitation of the centre-of-mass mode is increasing. Before we give the relevant Hamiltonian, let us take a qualitative look at the system. Imagine that a laser drives the qubit. If the laser has a frequency ω (shown by a vertical arrow in Fig. 3.4), then the laser will be more likely to induce transitions between the lower and upper state of the qubit without affecting the centre of mass mode. This is simply because all other transitions are out of resonance. If, however, the frequency of the laser is $\omega - \nu$ (shown by the other arrow in Fig 3.4), then it generates transitions between the upper state of the ion and the vibrational state with n excitations in the centre-of-mass mode and the ground state of the ion and $n + 1$ excitations in the centre-of-mass mode. If the ion is in the ground state and the centre-of-mass mode is not excited then nothing at all happens. Therefore we can see two things. Firstly, a red detuned laser can change simultaneously the

electronic state of the qubit and the state of the centre of mass mode. Secondly the dynamics is conditional on the state of the qubit. If the ion is in the ground state then there is no dynamics, while there are Rabi oscillations if the system is excited. One can easily see that this would not be possible if the ions were not cooled to the ground state of the motion. If with high probability there is at least one phonon in the centre-of-mass mode then the red detuned laser would always affect ions that are in the ground state. This qualitative discussion neglects the importance of the position of the ion in the standing laser field. This is a very important factor as it can be shown that an ion localized at the anti–node of the standing wave will, in leading order, interact with the laser without changing the excitation of the centre-of-mass mode. If, on the other hand, the ion is localized at the node of the standing wave then in leading order both the internal degrees of freedom of the ion as well as the excitation of the centre-of-mass mode are changed. Qualitatively this can be understood in the following way. If the ion is at the anti–node of the field then it does not see any photons. Therefore in order to interact with the field it has to change position and therefore it either has to absorb or emit a photon. Hence a change in its internal degree of freedom always requires a change in the motional degree of freedom of the ion. If the ion is sitting at the node of the field then it is not necessary for it to move in order to see photons. This qualitative reasoning can be corroborated by a precise derivation of the interaction Hamiltonian between ion, laser and centre-of-mass mode. However, we refer the reader to the literature [93].

If the ion is localized at the node of the standing wave light field and if we use a a red detuned laser then in leading order the Hamiltonian is given by

$$H = \frac{\eta}{\sqrt{N}} \frac{\Omega}{2} \left[|1\rangle\langle 0| a e^{i\phi} + |0\rangle\langle 1| a^\dagger e^{-i\phi} \right] , \tag{3.82}$$

where N is the number of ions that are in the linear ion trap, Ω is the Rabi frequency of the laser, $\eta = (2\pi/\lambda)\sqrt{\hbar/2M\nu}$ is the Lamb-Dicke parameter which describes how well the ions are localized and $a^\dagger, a$ are creation and annihilation operator for excitations in the centre-of-mass mode. The phase of the lasers detuned to the

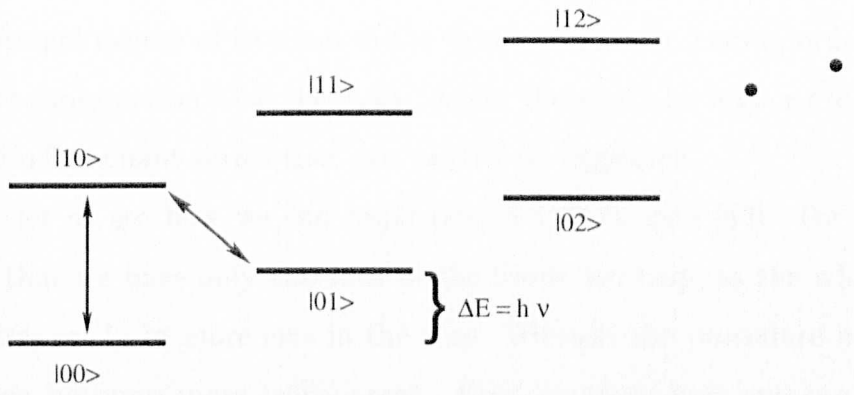


Figure 3.4: The vertical axis gives the energy while the horizontal axis gives the degree of excitation of the centre-of-mass mode. In $|xy\rangle$ the first number x denotes the internal degree of freedom of the ion, while the second number y denotes the degree of excitation of the centre of mass mode.

red side of electronic transition is ϕ. This Hamiltonian is an approximation and represents only the first term in an expansion of the true Hamiltonian in terms of η [94]. In addition, it does not describe the interaction with other modes than the centre-of-mass mode and off-resonant terms which can usually be neglected [95]. These are good approximations as η is much smaller than unity and because other modes of oscillation have different resonance frequencies and are therefore out of resonance with the laser. This interaction is known as the Jaynes–Cummings model [96] and will be described in a greater detail in Chapter 6 (the first derivation of the Jaynes–Cummings interaction in a quantized trap is in [97]). An advantage is that the same Hamiltonian is used in cavity QED to describe the interaction of cavity field with atoms. This means that when we implement a certain quantum computation using cavities and atoms to encode information, then exactly the same computation can be performed in a linear ion-trap.

If the ion is localized at the anti node of the standing wave of frequency ω then the Hamiltonian is given by

$$H = \frac{\Omega}{2}\left[|1\rangle\langle 0|e^{i\phi} + |0\rangle\langle 1|e^{-i\phi}\right] . \tag{3.83}$$

The motional degree of freedom of the ions remains (in leading order) unchanged during the interaction with the laser. Again there will be higher order corrections in η and off-resonant terms that can usually be neglected.

Now let us see how we can implement a CNOT gate [83]. For simplicity we assume that we have only two ions in the linear ion trap, as the whole procedure generalizes easily to more ions in the trap. We split the procedure into two parts, as it then becomes more transparent. First we show how one can implement a controlled phase gate, i.e. a gate that changes the phase of the upper state of the target bit only if the control bit is in the upper state.

Initially the centre-of-mass mode is in the ground state. Now we apply a laser pulse on the control bit which is described by the Hamiltonian eq. (3.82). The duration of the laser pulse is $t = \pi/(\Omega\eta/\sqrt{2})$ so that it produces a π-pulse. Note that the action of the laser depends on the state of the control ion. If the control bit is in the ground state then nothing happens, but if it is in the upper state then it goes to the ground state and simultaneously the centre-of-mass mode is excited. The effect is

$$\begin{aligned} |00\rangle|0\rangle &\rightarrow |00\rangle|0\rangle \\ |01\rangle|0\rangle &\rightarrow |01\rangle|0\rangle \\ |10\rangle|0\rangle &\rightarrow -i|00\rangle|1\rangle \\ |11\rangle|0\rangle &\rightarrow -i|01\rangle|1\rangle \ . \end{aligned} \tag{3.84}$$

The first number describes the state of the control qubit, the second the target qubit and the third one describes the state of the centre-of-mass mode which was initially in the ground state. Now we manipulate the target qubit. In this step we make use of the fact that we have more atomic levels available. We now couple the lower level of the qubit to a third level again using a Hamiltonian of the form eq. (3.82). We apply the laser for a time $t = 2\pi/(\Omega\eta/\sqrt{2})$ so that we effectively perform a full 2π rotation and the system ends up in the same state as it started, with one exception: The ground state of the target bit flips its phase if the centre-of-mass

mode is excited. Therefore we obtain

$$\begin{aligned} |00\rangle|0\rangle &\rightarrow |00\rangle|0\rangle \\ |01\rangle|0\rangle &\rightarrow |01\rangle|0\rangle \\ -i|00\rangle|1\rangle &\rightarrow i|00\rangle|1\rangle \\ -i|01\rangle|1\rangle &\rightarrow -i|01\rangle|1\rangle \ . \end{aligned} \tag{3.85}$$

In the next step we apply again a pulse of duration $t = \pi/(\Omega\eta/\sqrt{2})$ to the control bit using Hamiltonian eq. (3.82). This results in the total transformation

$$\begin{aligned} |00\rangle|0\rangle &\rightarrow |00\rangle|0\rangle \\ |01\rangle|0\rangle &\rightarrow |01\rangle|0\rangle \\ |10\rangle|0\rangle &\rightarrow |10\rangle|0\rangle \\ |11\rangle|0\rangle &\rightarrow -|11\rangle|0\rangle \ . \end{aligned} \tag{3.86}$$

Therefore we have implemented a controlled phase gate. Now let us see why this transformation is really a basic building block for a CNOT gate. Consider the different set of basis states

$$|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2} \ , \tag{3.87}$$

and let us look at the controlled phase transformation again, but now in the new basis. We find

$$\begin{aligned} |0+\rangle|0\rangle &\rightarrow |0+\rangle|0\rangle \\ |0-\rangle|0\rangle &\rightarrow |0-\rangle|0\rangle \\ |1+\rangle|0\rangle &\rightarrow |1-\rangle|0\rangle \\ |1-\rangle|0\rangle &\rightarrow |1+\rangle|0\rangle \ , \end{aligned} \tag{3.88}$$

and this is a version of a CNOT gate. All we need to do is to rotate the $\{|0\rangle, |1\rangle\}$ basis of the target bit into the $\{|+\rangle, |-\rangle\}$ basis, then perform the controlled phase gate, and finally we rotate the basis back. This then gives us a CNOT gate in

the $\{|0\rangle, |1\rangle\}$ basis. The rotation between the basis sets can be achieved using the Hamiltonian eq. (3.83), i.e. with a standing wave that has the target ion at its anti-node. We chose the phase of the laser to be $\phi = -\pi/2$ and we perform a $\pi/2$ pulse, i.e. a pulse with the length $t = \pi/(2\Omega)$ for the transformation from the $|0/1\rangle$ basis to the $|\pm\rangle$ basis. Going back from the $|\pm\rangle$ basis to the $|0/1\rangle$ basis is then done by a $-\pi/2 = 3\pi/2$ pulse. Therefore we are now able to generate a CNOT gate as well as single qubit gates and this is all we need to implement any unitary transformation between qubits. It should be said that a single CNOT gate has been demonstrated using a single ion in a trap by Monroe et al [98], albeit by a somewhat different scheme to the one we described above.

However, we have made a number of simplifying assumptions, some of which we have already mentioned. The Hamilton operators Eqs. (3.82,3.83) are only the leading order terms in an expansion with respect to the Lamb-Dicke parameter. In addition we have only taken into account the centre-of-mass mode although there are many other modes that might also get excited. This is a reasonable approximation because these modes have different frequencies and are therefore detuned from the laser so that their coupling is weak. We also neglected any spontaneous emission from the ions and losses of excitations of the centre-of mass mode. In addition to these more fundamental problems, we also have to face technical sources of errors which are, at the moment, the dominant source of error in experiments. We will see how to combat general errors in Chapter 5, and also how to protect against spontaneous emission of ions in an ion trap computer. Before that we turn to the main part of this thesis, analysis of the measures of quantum entanglement. The next chapter can be understood without any knowledge of quantum computation.

Chapter 4

Entanglement Measures and Purification Procedures

4.1 Introduction

In this chapter the problem of entanglement quantification is analysed. In the previous chapter we have seen that there is a number of good measures of entanglement for two entangled subsystems in a joint pure state (see also [99] for an extensive presentation). This is a consequence of the Schmidt decomposition procedure introduced earlier. However, for the mixed states of two subsystems, or for more than two subsystems this procedure does not exist in general. Therefore it is not immediately clear how to understand or quantify correlations for these states. Initially, it was thought that Bell's inequalities would provide a good criterion for separating quantum correlations (entanglement) from classical correlations in a given quantum state. However, we saw that, while it is true that a violation of Bell's inequalities is a signature of quantum correlations (non–locality), not all entangled states violate Bell's inequalities although some of them can be purified to singlets by a procedure due to Gisin [67]. So, it is clear that in order to completely separate quantum from classical correlations we need a new criterion. This is closely related to the question

of the amount of entanglement contained in a given quantum state. In an important paper Bennett et al [100] have recently proposed three measures of entanglement (we will discuss these in more detail later in this chapter). Their measures are based on concrete physical ideas and are intuitively easy to understand. They investigated many properties of these measures and calculated one of them, the entanglement of formation, for a number of states. More recently, Hill and Wootters [101] have conjectured a closed form for the entanglement of formation for two spin 1/2 particles and Wootters has recently proven this to be the correct form [102].

We adopt an entirely different approach to Bennett et al and show how to construct a whole class of measures of entanglement [103, 104], and in addition impose conditions that any candidate for such a measure has to satisfy [103, 104]. In short, we consider the disentangled states which form a convex subset of the set of all quantum states. Entanglement is then defined as a distance (not necessarily in the mathematical sense) from a given state to this subset of disentangled states (see Fig. 4.1). An attractive feature of our measure is that it is independent of the number of systems and their dimensionality, and is therefore completely general, [103, 104, 105]. We present here two candidates for measuring distances on our set of states and prove that they satisfy generalized conditions for a measure of entanglement.

It should be noted that, in much the same way, we can calculate the amount of classical correlations in a state. One would then define another subset, namely that of all product states which do not contain any classical correlations. Given a disentangled state one would then look for the closest uncorrelated state. The distance could be interpreted as a measure of classical correlations. The physical consequences of this definition of classical correlations is still under investigation. In addition to many analytical results we also explain how to calculate efficiently using numerical methods our measure of entanglement of two spin 1/2 particles. We present a number of examples and prove several important properties of our measure which have important physical consequences. To illuminate the physical

meaning behind the above ideas we present a statistical view of our entanglement measure in the case of the von Neumann relative entropy [105]. We then relate our measure to a purification procedure and use it to define a reversible purification. This reversible purification is then linked to the notion of entanglement through the idea of distinguishing two classes of quantum states. We also argue that the measures of entanglement that we propose give an upper bound for the number of singlet states that can be distilled from a given state. We find that in general the distillable entanglement is smaller than the entanglement of formation. This result was independently proven for Bell diagonal states using completely different methods [106].

The rest of the chapter is organized as follows. Section 4.2 introduces the basis of purification procedures, conditions for a measure of entanglement and our suggestion for a measure of entanglement. We also prove that the von Neumann relative entropy and the *modified* Bures metric (to be defined later) satisfy the imposed conditions and can therefore be used as generators of measures of entanglement. We compute our measure explicitly for some simple examples. In section 4.3 we introduce a simple numerical method to compute our measure of entanglement numerically and we apply it to the case of two spin 1/2 systems. We present a number of examples of entanglement computations using the von Neumann relative entropy. In section 4.4 we present a statistical basis for the von Neumann relative entropy as a measure of distinguishability between quantum states and hence of amount of entanglement. Based on this, in section 4.5 we derive an upper bound to the efficiency (number of maximally entangled pairs distilled) of any purification procedure. We also show how to extend our measure to more than two subsystems.

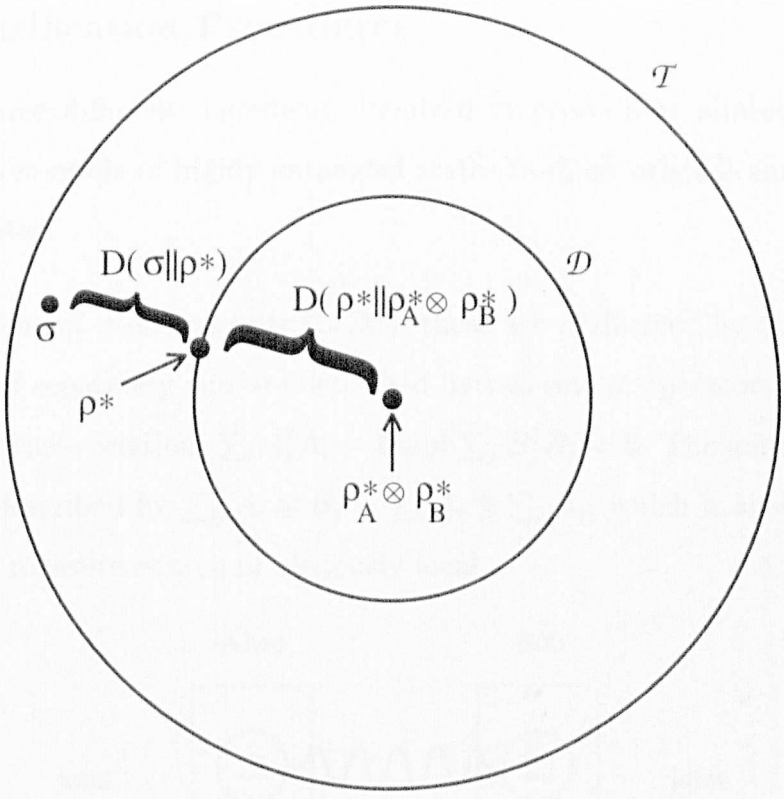


Figure 4.1: The set of all density matrices, $\mathcal{T}$ is represented by the outer circle. Its subset, a set of disentangled states $\mathcal{D}$ is represented by the inner circle. A state σ belongs to the entangled states, and ρ^* is the disentangled state that minimizes the distance $D(\sigma||\rho)$, 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.

4.2 Mathematical Prelude

4.2.1 Purification Procedures

There are three different ingredients involved in procedures aiming at distilling locally a subensemble of highly entangled states from an original ensemble of less entangled states.

1. *Local general measurements* (LGM): these are performed by the two parties A and B separately and are described by two sets of operators satisfying the completeness relations $\sum_i A_i^\dagger A_i = \mathbb{1}$ and $\sum_j B_j^\dagger B_j = \mathbb{1}$. The joint action of the two is described by $\sum_{ij} A_i \otimes B_j = \sum_i A_i \otimes \sum_j B_j$, which is again a complete general measurement, and obviously local.

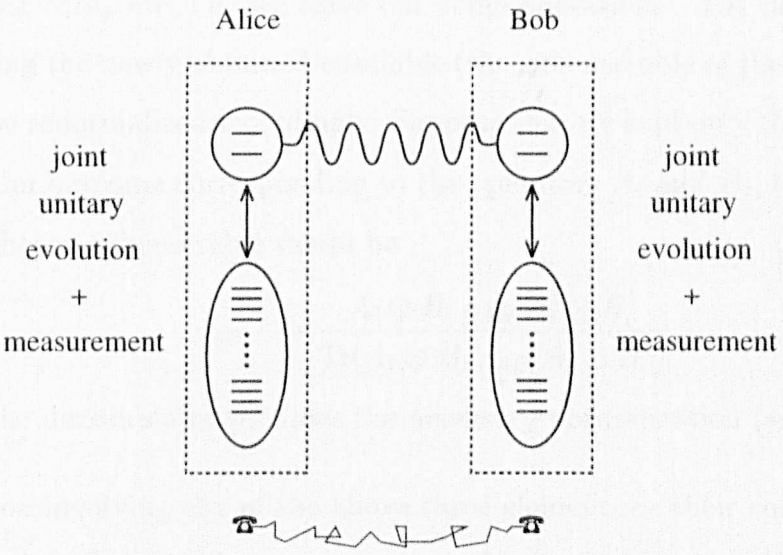


Figure 4.2: Local general measurements (LGM) by Alice and Bob and classical communication (CC) between Alice and Bob which correlates local general measurements.

2. *Classical communication* (CC): this means that the actions of A and B can be correlated. This can be described by a *complete measurement* on the whole space $A+B$ and is not necessarily decomposable into a sum of direct products of individual operators (as in LGM). If ρ_{AB} describes the initial state shared between A and B then the transformation involving 'LGM+CC' would look like

$$\Phi(\rho_{AB}) = \sum_i A_i \otimes B_i \, \rho_{AB} \, A_i^\dagger \otimes B_i^\dagger \ , \tag{4.1}$$

where $\sum_i A_i^\dagger A_i B_i^\dagger B_i = \mathbb{1}$, i.e. the actions of A and B are 'correlated' (see Fig. 4.2).

3. *Post-selection* (PS) is performed on the *final ensemble* according to the above two procedures. Mathematically this amounts to the general measurement not being complete, i.e. we leave out some operations. The density matrix describing the newly obtained ensemble (the subensemble of the original one) has to be renormalized accordingly. Suppose that we kept only the pairs where we had an outcome corresponding to the operators A_i and B_j, then the state of the chosen subensemble would be

$$\rho_{AB} \longrightarrow \frac{A_i \otimes B_i \, \rho_{AB} \, A_i^\dagger \otimes B_i^\dagger}{\mathrm{Tr}(A_i \otimes B_i \, \rho_{AB} \, A_i^\dagger \otimes B_i^\dagger)} \tag{4.2}$$

where the denominator provides the necessary normalization (see Fig. 4.3).

A manipulation involving any of the above three elements or their combination we shall henceforth call a *purification procedure*. It should be noted that the three operations described above are local, and that, strictly speaking, we can derive the whole formalism without assuming eq. (4.1). This implies that the entanglement of the total ensemble cannot increase under these operations. However, classical correlations between the two subsystems *can* be increased, even for the whole ensemble, if we allow classical communication. A simple example easily confirms this. Suppose that the initial ensemble contains states $|0_A\rangle \otimes (|0_B\rangle + |1_B\rangle)/\sqrt{2}$. The correlations (measured by e.g. the von Neumann mutual information) between A and

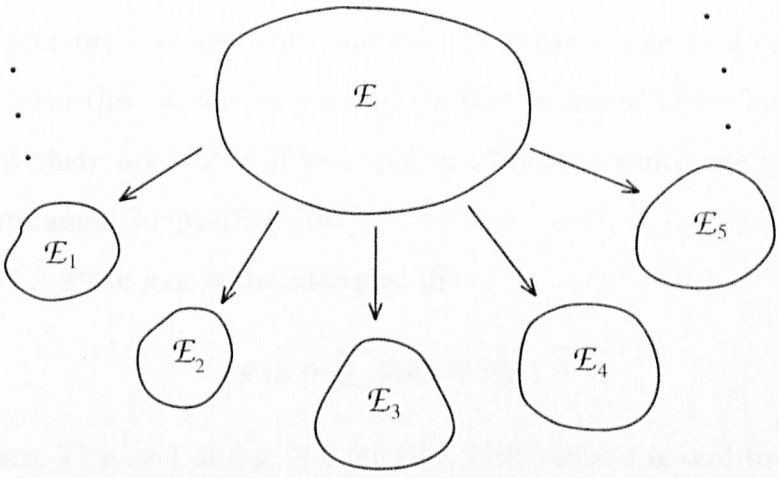


Figure 4.3: Subselection (SS) according to measurement results: the original ensemble of entangled pairs E is split under the action of LGM+CC into a number of subensembles which contain pairs with different degrees of entanglement.

B are zero. Suppose that B performs measurement of his particles in the standard 0, 1 basis. If 1 is obtained, B communicates this to A who then "rotates" his qubit to the state $|1_A\rangle$. Otherwise they do nothing. The final state will therefore be

$$\rho = \frac{1}{2}(|0_A\rangle\langle 0_A| \otimes |0_B\rangle\langle 0_B| + |1_A\rangle\langle 1_A| \otimes |1_B\rangle\langle 1_B|) \ , \tag{4.3}$$

where the correlations are now $\ln 2$ (i.e. nonzero). So, the classical content of correlations can be increased by performing local general measurements and classically communicating.

An important result was proved for pairs of spin–1/2 systems in [66]: All states that are not of the form $\rho_{AB} = \sum_i p_i \rho_A^i \otimes \rho_B^i$, where $\sum_i p_i = 1$ and $p_i \geq 0$ for all i, can be distilled to a subensemble of maximally entangled states using only operations 1, 2 and 3. (The states of the above form obviously remain of the same form under any purification procedure). The local nature of the above three operations implies that we define a disentangled state of two quantum systems A and B as a state from which, by means of local operations, no subensemble of entangled states can

be distilled. It should be noted that these states are sometimes called separable in the existing literature. We also note that it is not proven in general that if the state is not of this form then it can be purified (in fact, a recent letter by Horodecki et al. shows that there are states of two spin one systems which are inseparable, or entangled, but cannot be purified [107]).

Definition 1. A state ρ_{AB} is disentangled iff

$$\rho_{AB} = \sum_i p_i \rho_A^i \otimes \rho_B^i \ , \tag{4.4}$$

where, as before, $\sum_i p_i = 1$ and $p_i \geq 0$ for all i. Otherwise it is said to be entangled. Note that all the states in the above expansion can be taken to be pure. This is because each ρ^i can be expanded in terms of its eigenvectors. So, in the above sum we can in addition require that ${\rho_A^i}^2 = \rho_A^i$ and ${\rho_B^i}^2 = \rho_B^i$ for all i. This fact will be used later in this section and will be formalized further in section 4.3. For two entangled qubits there is a criterion [108, 109] determining whether a given density operator can be written in the above separable form in eq. (4.4): it says that iff either partial transposition of ρ_{AB} over the first or over the second qubit results in a negative operator then the state ρ_{AB} is entnagled.

4.2.2 Purification of Pure States

In Chapter 3 we provided an argument for using the von Neumann entropy as a measure of entanglement for pure bipartite systems. Now we present an additional argument to strengthen this conclusion, but from the point of view of pure states involving two entangled qubits. We consider the following problem first analysed by Bennett and coworkers in [110]: Alice and Bob share n entangled qubit pairs, where each pair is prepared in the state

$$|\Psi_{AB}\rangle = a|00\rangle + b|11\rangle \tag{4.5}$$

where we take $a, b \in \mathbf{R}$, and $a^2 + b^2 = 1$. How many maximally entangled states can they purify? It turns out that the answer is governed by the von Neumann

reduced entropy and is asymptotically given by $N \times S_N(\rho_A)$. To see how this is so, consider the state of n pairs given by

$$|\Psi_{AB}^{\otimes n}\rangle = (a|00\rangle + b|11\rangle) \otimes (a|00\rangle + b|11\rangle) \ldots (a|00\rangle + b|11\rangle) \tag{4.6}$$

$$= a^n|0000\ldots00\rangle + a^{(n-1)}b^2(|0000\ldots11\rangle + \ldots + |1100\ldots00\rangle)$$

$$+ \ldots b^n|1111\ldots11\rangle \,. \tag{4.7}$$

(The convention in the second and the third line is that the odd states in the large joint ket states belong to Alice and the even states belong to Bob). Alice can now perform projections (locally, of course) onto the subspaces which have no states $|1\rangle$, 2 states $|1\rangle$, 4 states $|1\rangle$, and so on, and communicates her results to Bob. The probability of having a successful projection onto a particular subspace with $2k$ states $|1\rangle$ can easily be seen for the above equation to be

$$p_{2k} = a^{2(n-k)}b^{2k}\binom{n}{k} \,. \tag{4.8}$$

On the other hand, the entanglement of a state in this subspace is equal to

$$E_{2k} = \ln\binom{n}{k} \,. \tag{4.9}$$

This can be seen by treating Alice's n particles as a single 2^n level system and Bob's particles likewise, and then applying the Schmidt decomposition to this state. In addition it can be shown that this state can be converted into singlets preserving the amount of entanglement [110]. Therefore, the total expected entanglement is seen to be

$$E = \sum_{k=0}^{n} a^{2(n-k)}b^{2k}\binom{n}{k}\ln\binom{n}{k} \,. \tag{4.10}$$

We wish to see how this sum behaves asymptotically as $n \to \infty$. From the theory of types presented in the previous chapter we know that the dominant term, i.e. the most likely outcome sequence will be such that

$$E \sim (a^2)^{na^2}(b^2)^{nb^2}\binom{n}{b^2n}\ln\binom{n}{b^2n} \,, \tag{4.11}$$

which can, in turn, be simplified using Stirling's approximation to obtain

$$\begin{aligned} E &\sim e^{-nS_N(\rho_A)} e^{n\ln n - a^2 n \ln a^2 n - b^2 n \ln b^2 n} (n\ln n - a^2 n \ln a^2 n - b^2 n \ln b^2 n) \\ &= e^{-nS_N(\rho_A)} e^{nS_N(\rho_A)} \times nS_N(\rho_A) = nS_N(\rho_A)\,. \end{aligned} \tag{4.12}$$

This now shows that for pure states the singlet yield of a purification procedure is determined by the von Neumann reduced entropy. It is also important to stress that the above procedure is *reversible*, i.e. starting from m singlets Alice and Bob can locally produce a given state $a|0,0\rangle + b|1,1\rangle$ with an asymptotic efficiency of $m\ln 2 = nS_N(\rho_A)$. This will be the basis of one of the measures of entanglement introduced by Bennett et al. [100]. Of course, Alice and Bob cannot do better than this limit, since both of them see the initial string of qubits as a classical 0, 1 string with the corresponding probabilities a^2 and b^2. This, we have seen, cannot be compressed to more than its Shannon entropy (which in this case coincides with the von Neumann entropy). In fact, it is worth mentioning that the von Neumann entropy describes how much a string of qubits can be compressed in the same way that the Shannon entropy shows by how much a string of bits can be compressed [5, 111]. Now we consider the quantification of entanglement in general, including the mixed states.

4.2.3 Quantification of Entanglement

In the previous section we have indicated that out of certain states it is possible to distill by means of LGM+CC+PS a subensemble of maximally entangled states (we call these states entangled). The question remains open about how much entanglement a certain state contains. Of course, this question is not entirely well defined unless we state what physical circumstances characterize the amount of entanglement. This suggests that there is no unique measure of entanglement. Before we define three different measures of entanglement we state three conditions that every measure of entanglement has to satisfy. The third condition represents a generalization of the corresponding one I presented with my co-workers in [103].

E1. $E(\sigma) = 0$ iff σ is separable.

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.

$$\sum \mathrm{Tr}(\sigma_i)\ E(\sigma_i/\mathrm{Tr}(\sigma_i)) \leq E(\sigma)\ , \tag{4.13}$$

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

Condition E1 ensures that disentangled and only disentangled states have a zero value of entanglement. Condition E2 ensures that a local change of basis has no effect on the amount of entanglement. Condition E3 is intended to remove the possibility of increasing entanglement by performing local measurements aided by classical communication. It is an improvement over the condition 3 in [103] which required that $E(\sum_i V_i \sigma V_i^\dagger) \leq E(\sigma)$. The condition E3 is physically more appropriate as it takes into account the fact that we have some knowledge of the final state. Namely, when we start with n states σ we know exactly which $m_i = n \times \mathrm{Tr}(\sigma_i)$ pairs will end up in the state σ_i after performing a purification procedure. Therefore we can separately access the entanglement in each of the possible subensembles described by σ_i. Clearly the total expected entanglement at the end should not exceed the original entanglement, which is stated in E3. This, of course, does not exclude the possibility that we can select a subensemble whose entanglement per pair is higher than the original entanglement per pair. We now introduce three different measures of entanglement which obey E1–E3.

First we discuss the entanglement of formation [100]. Bennett et al [100] define the entanglement of formation of a state ρ by

$$E_c(\rho) := \min \sum_i p_i S(\rho_A^i) \tag{4.14}$$

where $S(\rho_A) = -\mathrm{Tr}\rho_A \ln \rho_A$ is the von Neumann entropy and the minimum is taken over all the possible realisations of the state, $\rho_{AB} = \sum_j p_j |\psi_j\rangle\langle\psi_j|$ with $\rho_A^i =$

$\mathrm{Tr}_B(|\psi_i\rangle\langle\psi_i|)$. The entanglement of formation satisfies all the three conditions E1–E3 [100]. The physical basis of this measure presents the *minimum* number of singlets needed to be shared by Alice and Bob in order to create a given entangled state by local operations. The result in eq. (4.14) follows immediately from the entropic efficiency of reversible purification of pure state shown in the previous subsection in eq. (4.12). We will analyse the relationship between the entanglement of formation and measures proposed here in section 4.5.

Related to this measure is the entanglement of distillation [100]. It defines the amount of entanglement of a state σ as the proportion of singlets that can be distilled using a purification procedure. As such, it is dependent on the efficiency of a particular purification procedure. It can be made more general only by introducing some sort of universal purification procedure or asking for the best state dependent purification procedure. We investigate this in subection 4.5.

We now introduce our suggestion for a measure of an amount of entanglement. It is seen in subsection 3.5 that this measure is intimately related to the entanglement of distillation by providing an upper bound for it. If $\mathcal{D}$ is the set of all disentangled states, the measure of entanglement for a state σ is then defined as

$$E(\sigma) := \min_{\rho\in\mathcal{D}} \; D(\sigma||\rho) \tag{4.15}$$

where D is any measure of *distance* (not necessarily a metric) between the two density matrices ρ and σ such that $E(\sigma)$ satisfies the above three conditions E1–E3. Note that this, in fact, generates a whole class of measures depending on the form of $D(\sigma||\rho)$.

Now the central question is what condition a candidate for $D(\sigma||\rho)$ has to satisfy in order for E1–E3 to hold for the entanglement measure? We present here a set of sufficient conditions published in [104].

F1. $D(\sigma||\rho) \geq 0$ with the equality saturated iff $\sigma = \rho$.

F2. Unitary operations leave $D(\sigma||\rho)$ invariant, i.e. $D(\sigma||\rho) = D(U\sigma U^\dagger||U\rho U^\dagger)$.

F3. $D(\mathrm{Tr}_p\sigma||\mathrm{Tr}_p\rho) \leq D(\sigma||\rho)$, where Tr_p is a partial trace.

F4. $\sum p_i\, D(\sigma_i/p_i||\rho_i/q_i) \leq \sum D(\sigma_i||\rho_i)$, where $p_i = \mathrm{Tr}(\sigma_i)$, $q_i = \mathrm{Tr}(\rho_i)$ and $\sigma_i = V_i\sigma V_i^\dagger$ and $\rho_i = V_i\rho V_i^\dagger$ (note that V_i's are not necessarily local).

F5a. $D(\sum_i P_i\sigma P_i||\sum_i P_i\rho P_i) = \sum_i D(P_i\sigma P_i||P_i\rho P_i)$, where P_i is any set of orthogonal projectors such that $P_iP_j = \delta_{ij}P_i$.

F5b. $D(\sigma \otimes P_\alpha||\rho \otimes P_\alpha) = D(\sigma||\rho)$ where P_α is any projector.

Conditions F1 and F2 ensure that E1 and E2 hold; F2, F3, F4 and F5 ensure that E3 is satisfied. The argument for the former is trivial, while for the latter it is more lengthy and will be presented in the remainder of this section.

4.2.4 Proofs

We claim that F2, F3, F4 and F5 are sufficient for E3 to be satisfied and hence need to prove that $F2 - F5 \Rightarrow E3$. If F2, F3 and F5b hold, then we can prove the following statement,

Theorem 1. For any completely positive, trace preserving map Φ, given by $\Phi\sigma = \sum V_i\sigma V_i^\dagger$ and $\sum V_i^\dagger V_i = \mathbb{1}$, we have that $D(\Phi\sigma||\Phi\rho) \leq D(\sigma||\rho)$. [1]

Proof. It is well known that a complete measurement can *always* be represented as a unitary operation+partial tracing on an extended Hilbert Space $\mathcal{H} \otimes \mathcal{H}_n$, where $\dim \mathcal{H}_n = n$ [112, 113]. Let $\{|i\rangle\}$ be an orthonormal basis in $\mathcal{H}_n$ and $|\alpha\rangle$ be a unit vector. So we define,

$$W = \sum_i V_i \otimes |i\rangle\langle\alpha| \ . \tag{4.16}$$

Then, $W^\dagger W = \mathbb{1} \otimes P_\alpha$ where $P_\alpha = |\alpha\rangle\langle\alpha|$, and there is a unitary operator U in $\mathcal{H} \otimes \mathcal{H}_n$ such that $W = U(\mathbb{1} \otimes P_\alpha)$ [114]. Consequently,

$$U(A \otimes P_\alpha)U^\dagger = \sum_{ij} V_iAV_j^\dagger \otimes |i\rangle\langle j| \ , \tag{4.17}$$

[1] We frequently interchange the Φ and $\sum V^\dagger V$ notations for one another throughout this section.

so that,

$$\mathrm{Tr}_2\{U(A\otimes P_\alpha)U^\dagger\} = \sum_i V_i A V_i^\dagger \quad . \tag{4.18}$$

Now using F3, then F2, and finally F5b we find the following

$$\begin{aligned} D(\mathrm{Tr}_2\{U(\sigma\otimes P_\alpha)U^\dagger\} \;||\; & \mathrm{Tr}_2\{U(\rho\otimes P_\alpha)U^\dagger\}) \\ &\leq D(U(\sigma\otimes P_\alpha)U^\dagger||U(\rho\otimes P_\alpha)U^\dagger) \\ &= D(\sigma\otimes P_\alpha||\rho\otimes P_\alpha) \\ &= D(\sigma||\rho) \quad . \end{aligned} \tag{4.19}$$

This proves Theorem 1 □.

Corollary. Since for a complete set of orthonormal projectors P, $\sum_i P_i\sigma P_i$ is a complete positive trace preserving map, then

$$\sum_i D(P_i\sigma P_i||P_i\rho P_i) \leq D(\sigma||\rho) \quad . \tag{4.20}$$

(The sum can be taken outside as F5a requires that $D(\sum_i P_i\sigma P_i||\sum_i P_i\rho P_i) = \sum_i D(P_i\sigma P_i||P_i\rho P_i)$). Now from F2, F3, F5b and eq. (4.20) we have the following

Theorem 2. If $\sigma_i = V_i\sigma V_i^\dagger$ then $\sum D(\sigma_i||\rho_i) \leq D(\sigma||\rho)$.

Proof. Equations (4.16) and (4.17) are introduced as in the previous proof. From eq. (4.17) we have that

$$\mathrm{Tr}_2\{\mathbb{1}\otimes P_i U(A\otimes P_\alpha)U^\dagger \mathbb{1}\otimes P_i\} = V_i A V_i^\dagger \quad . \tag{4.21}$$

where $P_i = |i\rangle\langle i|$. Now, from F3, the Corollary and F5b it follows that

$$\begin{aligned} \sum_i \; D \; & (\mathrm{Tr}_2\{\mathbb{1}\otimes P_i U(\sigma\otimes P_\alpha)U^\dagger\mathbb{1}\otimes P_i\}||\mathrm{Tr}_2\{\mathbb{1}\otimes P_i U(\rho\otimes P_\alpha)U^\dagger\mathbb{1}\otimes P_i\}) \\ &\leq \sum_i D(\mathbb{1}\otimes P_i U(\sigma\otimes P_\alpha)U^\dagger\mathbb{1}\otimes P_i||\mathbb{1}\otimes P_i U(\rho\otimes P_\alpha)U^\dagger\mathbb{1}\otimes P_i) \\ &\leq D(U(\sigma\otimes P_\alpha)U^\dagger||U(\rho\otimes P_\alpha)U^\dagger) \\ &= D(\sigma\otimes P_\alpha||\rho\otimes P_\alpha) \\ &= D(\sigma||\rho) \quad . \end{aligned} \tag{4.22}$$

This proves Theorem 2 □.

From Theorem 2 and F4 we have,

$$\sum p_i \, D\Big(\frac{\sigma_i}{p_i}\Big\|\frac{\rho_i}{q_i}\Big) \leq D(\sigma||\rho) \quad . \tag{4.23}$$

Now let $E(\sigma) = D(\sigma||\rho^*)$, *i.e.* let the minimum of $D(\sigma||\rho)$ over all $\rho \in \mathcal{D}$ be attained at ρ^*. Then from eq. (4.23)

$$E(\sigma) := D(\sigma||\rho^*) \geq \sum p_i \, D\Big(\frac{\sigma_i}{p_i}\Big\|\frac{V_i^\dagger \rho^* V_i}{q_i}\Big) \geq \sum p_i \; E(\sigma_i/p_i) \tag{4.24}$$

and E3 is satisfied. Note that in all the proofs for $D(\sigma||\rho)$ we never use the fact that the completely positive, trace preserving map Φ is local. This is only used in the last inequality of eq. (4.24) where LGM (+CC+PS) maps disentangled states onto disentangled states. This ensures that ρ_i^* is disentangled and therefore $D(\sigma_i/p_i||\rho_i^*/q_i) \geq E(\sigma_i/p_i)$. So, the need for local Φ arises only in eq. (4.24); otherwise all the other proofs hold for a general Φ. Note also that we can derive, by the same methods, a slightly more general condition

*E3**. The expected entanglement of the initial state $\sigma^n = \sigma_1 \otimes \ldots \otimes \sigma_n$ cannot increase under LGM+CC+PS given by $\sum V_i^\dagger V_i = \mathbb{1}$, i.e.

$$E(\sigma^n) \equiv E(\sigma_1 \otimes \ldots \otimes \sigma_n) \geq \sum \mathrm{Tr}(V_i \sigma^n V_i^\dagger) \; E(V_i \sigma^n V_i^\dagger / \mathrm{Tr}(V_i \sigma^n V_i^\dagger)) \quad . \tag{4.25}$$

where V_i can be of the form $V_i^1 \otimes \ldots \otimes V_i^n$, but also can be of any other completely general form.

However, in the following we will not make use of this generalization until section 3.5 when we will use it to estimate the efficiency of purification procedures.

4.2.5 Two Realisations of $D(\sigma, \rho)$

In this section we show that F1–F5 hold for the von Neumann relative entropy and for the modified Bures metric, which as we have seen immediately renders them generators of a good measure of entanglement.

4.2.5.1 Von Neumann relative entropy

We first prove F1–F5 for the von Neumann relative entropy, i.e. when $D(\sigma||\rho) = S(\sigma||\rho) := \mathrm{Tr}\,\{\sigma(\ln\sigma - \ln\rho)\}$ (Note that the von Neumann relative entropy is not a true metric, as it is not symmetric and does not satisfy the triangle inequality. Such quantities are usually called *pseudo* metrics, and in the next section the reasons for this asymmetry will become clear. For further properties of the von Neumann relative entropy see [13, 115, 116].) Properties F1 and F2 are satisfied [52]. F3 follows from the strong subadditivity property of the von Neumann entropy [52, 112] (originally proven by Lieb and Ruskai [117]). Since $\sum S(\sigma_i||\rho_i) = \sum p_i\, S(\sigma_i/p_i||\rho_i/q_i) + \sum p_i \ln p_i/q_i$ and $\sum p_i \ln \frac{p_i}{q_i} \geq 0$ (see [4] for proof) F4 is also satisfied. Property F5 can be proved to hold by inspection [112]. Now, a question arises as to why the entanglement is not defined as $E(\sigma) = \min_{\rho\in\mathcal{D}} S(\rho||\sigma)$. Since the von Neumann relative entropy is asymmetric this gives a different result to the original definition. However, the major problem with this convention is that for all pure states this measure is infinite. Although this does have a sound statistical interpretation (see the next section) it is hard to relate it to any physically reasonable scheme (e.g. a purification procedure) and, in addition, it fails to distinguish between different entangled pure states. This is the prime reason for excluding this convention from any further considerations. The measure of entanglement generated by the von Neumann relative entropy will hereafter be referred to as the relative entropy of entanglement.

Properties of the relative entropy of entanglement

For pure, maximally entangled states we showed that the relative entropy of entanglement reduces to the von Neumann reduced entropy [103]. We also conjectured [103] that for general pure state this would be true. Now we present a proof of this conjecture. In short, our proof goes as follows: we already have a guess as to what the minimum for a pure state σ should be: say, it is a disentangled state ρ^*. Then

we show that the gradient $\frac{d}{dx}S(\sigma||(1-x)\rho^* + x\rho)$ for any $\rho \in \mathcal{D}$ is nonnegative. However, if ρ^* was not a minimum the above gradient would be strictly negative which is a contradiction. In addition, for a convex function on a convex set every local minimum is also global (to be proven in the following section), and we arrive at the result that ρ^* is indeed a true minimum. Now we present a more formal proof [118] that applies to arbitrary dimensions of the two subsystems. An alternative proof that also applies to arbitrary dimensions will be given in section 3.

Theorem 3. For pure states $\sigma = \sum_{n_1 n_2} \sqrt{p_{n_1} p_{n_2}} |\phi_{n_1}\psi_{n_1}\rangle\langle\phi_{n_2}\psi_{n_2}|$ the relative entropy of entanglement is equal to the Von Neumann reduced entropy, i.e. $E(\sigma) = -\sum_n p_n \ln p_n$.

Proof. For $a > 0$, $\log a = \int_0^\infty \frac{at-1}{a+t}\frac{dt}{1+t^2}$, and thus, for any positive operator A, $\log A = \int_0^\infty \frac{At-1}{A+t}\frac{dt}{1+t^2}$. Let $f(x,\rho) = S(\sigma||(1-x)\rho^* + x\rho)$. Then

$$\begin{aligned}
\frac{\partial f}{\partial x}(0,\rho) &= -\lim_{x\to 0} \mathrm{Tr}\left\{\frac{\sigma(\log((1-x)\rho^* + x\rho) - \log\rho^*)}{x}\right\} \\
&= \mathrm{Tr}(\sigma \int_0^\infty (\rho^* + t)^{-1}(\rho^* - \rho)(\rho^* + t)^{-1} dt) \\
&= 1 - \int_0^\infty \mathrm{Tr}(\sigma(\rho^* + t)^{-1}\rho(\rho^* + t)^{-1}) dt \\
&= 1 - \int_0^\infty \mathrm{Tr}((\rho^* + t)^{-1}\sigma(\rho^* + t)^{-1}\rho) dt \qquad (4.26)
\end{aligned}$$

Take $\rho^* = \sum_n p_n |\phi_n\psi_n\rangle\langle\phi_n\psi_n|$ (this is our guess for the minimum). Then

$$\begin{aligned}
(\rho^* + t)^{-1}\sigma(\rho^* + t)^{-1} &= \sum_{n_1,n_2,n_3,n_4} (p_{n_1} + t)^{-1} |\phi_{n_1}\psi_{n_1}\rangle\langle\phi_{n_1}\psi_{n_1}| \\
&\quad \sqrt{p_{n_2} p_{n_3}} |\phi_{n_2}\psi_{n_2}\rangle\langle\phi_{n_3}\psi_{n_3}| (p_{n_4} + t)^{-1} |\phi_{n_4}\psi_{n_4}\rangle\langle\phi_{n_4}\psi_{n_4}| \\
&= \sum_{n,n'} (p_n + t)^{-1} \sqrt{p_n p_{n'}} (p_{n'} + t)^{-1} |\phi_n\psi_n\rangle\langle\phi_{n'}\psi_{n'}|. \qquad (4.27)
\end{aligned}$$

Set $g(p,q) = \int_0^\infty (p+t)^{-1}\sqrt{pq}(q+t)^{-1}dt$. Then it follows that $g(p,p) = 1$ and, for $p < q$,

$$\begin{aligned}
g(p,q) &= \sqrt{pq} \int_0^\infty \left(\frac{1}{p+t} - \frac{1}{q+t}\right)\frac{1}{q-p} dt \\
&= \frac{\sqrt{pq}}{q-p} \log\frac{q}{p} \ . \qquad (4.28)
\end{aligned}$$

lemma $0 \leq g(p,q) \leq 1$ for all $p, q \in [0,1]$.

proof. We know that $g(p,q) = \sqrt{pq}\int_0^\infty (p+t)^{-1}(q+t)^{-1}dt$. But,

$$(p+t)(q+t) = pq + t(p+q) + t^2 \geq pq + 2t\sqrt{pq} + t^2 = (\sqrt{pq}+t)^2 \ , \qquad (4.29)$$

and so

$$g(p,q) \leq \sqrt{pq}\int_0^\infty (\sqrt{pq}+t)^{-2}dt = 1 \ . \qquad (4.30)$$

Let $\rho = |\alpha\rangle\langle\alpha| \otimes |\beta\rangle\langle\beta|$ where $|\alpha\rangle = \sum_n a_n|\phi_n\rangle$ and $\beta = \sum_n b_n\psi_n$ are normalized vectors. Then

$$\begin{aligned}
\frac{\partial f}{\partial x}(0,\rho) - 1 &= -\mathrm{Tr}(\int_0^\infty (\rho^* + t)^{-1}\sigma(\rho^* + t)^{-1}dt\rho) \\
&= -\mathrm{Tr}(\sum_{n_1,n_2,n_3,n_4,n_5,n_6} g(p_{n_1}, p_{n_2})|\phi_{n_1}\psi_{n_1}\rangle\langle\phi_{n_2}\psi_{n_2}| \\
&\quad a_{n_3}b_{n_4}\bar{a}_{n_5}\bar{b}_{n_6}|\phi_{n_3}\psi_{n_4}\rangle\langle\phi_{n_5}\psi_{n_6}|) \\
&= -\sum_{n_1,n_2} g(p_{n_1}, p_{n_2})a_{n_2}b_{n_2}\bar{a}_{n_1}\bar{b}_{n_1} \qquad (4.31)
\end{aligned}$$

and (employing standard well known inequalities [119])

$$\begin{aligned}
|\frac{\partial f}{\partial x}(0,\rho) - 1| &\leq \sum_{n_1,n_2} |a_{n_1}||b_{n_1}||a_{n_2}||b_{n_2}| \\
&= (\sum_n |a_n||b_n|)^2 \leq \sum_n |a_n|^2 \sum_n |b_n|^2 = 1 \ . \qquad (4.32)
\end{aligned}$$

Thus it follows that $\frac{\partial f}{\partial x}(0, |\alpha\beta\rangle\langle\alpha\beta|) \geq 0$.

But any $\rho \in \mathcal{D}$ can be written in the form $\rho = \sum_i r_i|\alpha^i\beta^i\rangle\langle\alpha^i\beta^i|$ and so

$$\frac{\partial f}{\partial x}(0,\rho) = \sum_i r_i \frac{\partial f}{\partial x}(0, |\alpha^i\beta^i\rangle\langle\alpha^i\beta^i|) \geq 0 \ . \qquad (4.33)$$

Proposition Let $\Phi \in H$ have the following Schmidt decomposition [47]

$$|\Phi\rangle = \sum_n \sqrt{p_n}|\varphi_n\psi_n\rangle \qquad (4.34)$$

and set $\sigma = |\Phi\rangle\langle\Phi|$. Then $E(\sigma) = -\sum_n p_n \log p_n$.

Proof. $S(\sigma||\rho^*) = -\sum_n p_n \log p_n$ so it is sufficient to prove that $S(\sigma||\rho) \geq S(\sigma||\rho^*)$ for all $\rho \in D$. Suppose that $S(\sigma||\rho) < S(\sigma||\rho^*)$ for some $\rho \in \mathcal{D}$. Then, for $0 < x \leq 1$,

$$f(x,\rho) = S(\sigma||(1-x)\rho^* + x\rho) \quad \leq \quad (1-x)S(\sigma||\rho^*) + xS(\sigma||\rho)$$

$$
\begin{aligned}
&= (1-x)f(0,\rho) + xf(1,\rho) \\
\Rightarrow \quad & \frac{f(x,\rho) - f(0,\rho)}{x} \leq \\
& f(1,\rho) - f(0,\rho) < 0 \, . \qquad (4.35)
\end{aligned}
$$

This is impossible since $\frac{\partial f}{\partial x}(0,\rho) = \lim_{x\to 0} \frac{f(x,\rho)-f(0,\rho)}{x} \geq 0$. This therefore proves the above proposition □.

Therefore we have shown that for arbitrary dimensions of the subsystems the entropy of entanglement reduces to the entropy of entanglement for pure states. This is, in fact, a very desirable property, as the entropy of entanglement is known to be a good measure of entanglement for pure states. In fact one might want to elevate Theorem 3 to a condition for any good measure of entanglement, i.e.

E4: For pure states the measure of entanglement reduces to the entropy of entanglement, i.e.

$$E(\sigma) = -\mathrm{Tr}\left\{\sigma_A \ln \sigma_A\right\} \quad , \qquad (4.36)$$

with $\sigma_A = \mathrm{Tr}_B\{\sigma\}$ being the reduced density operator of one subsystem of the entangled pair.

However, in subsection 2 we will see that measures which do not satisfy E4 can nevertheless contain useful information. We will discuss this point later in this chapter.

We would like to point out another property of the relative entropy of entanglement that helps us find the amount of entanglement. It gives us a method to construct from a density operator σ with known entanglement a new density operator σ' with known entanglement.

Theorem 4. If ρ^* minimizes $S(\sigma||\rho^*)$ over $\rho \in \mathcal{D}$ then ρ^* is also a minimum for any state of the form $\sigma_x = (1-x)\sigma + x\rho^*$.

Proof. Consider,

$$
\begin{aligned}
S(\sigma_x||\rho) - S(\sigma_x||\rho^*) &= \mathrm{Tr}\{\sigma_x \ln \rho^* - \sigma_x \ln \rho\} \\
&= -x\mathrm{Tr}(\sigma \ln \rho) - (1-x)\mathrm{Tr}(\rho^* \ln \rho)
\end{aligned}
$$

$$
\begin{aligned}
&+ \quad x\mathrm{Tr}(\sigma \ln \rho^*) + (1-x)\mathrm{Tr}(\rho^* \ln \rho^*) \\
&= \quad x\{S(\sigma||\rho) - S(\sigma||\rho^*)\} + (1-x)S(\rho^*||\rho) \geq 0. \quad (4.37)
\end{aligned}
$$

This is true for any ρ. Thus ρ^* is indeed a minimum of σ_x □. For completeness we now prove here that $E(\sigma)$ is convex. Namely,

Theorem 5. $E(x_1\sigma_1 + x_2\sigma_2) \leq x_1E(\sigma_1) + x_2E(\sigma_2)$, where $x_1 + x_2 = 1$.

Proof. This property follows from the convexity of the von Neumann relative entropy in both arguments [116]

$$
S(x_1\sigma_1 + x_2\sigma_2||x_1\rho_1 + x_2\rho_2) \leq x_1S(\sigma_1||\rho_1) + x_2S(\sigma_2||\rho_2) \ . \quad (4.38)
$$

Now,

$$
\begin{aligned}
E(x_1\sigma_1 + x_2\sigma_2) &\leq S(x_1\sigma_1 + x_2\sigma_2||x_1\rho_1^* + x_2\rho_2^*) \\
&\leq x_1S(\sigma_1||\rho_1^*) + x_2S(\sigma_2||\rho_2^*) \\
&= x_1E(\sigma_1) + x_2E(\sigma_2) \ , \quad (4.39)
\end{aligned}
$$

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 formation 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.

Theorem 6. $E_c(\sigma) \geq E(\sigma) = \min_{\rho\in\mathcal{D}} S(\sigma||\rho)$.

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

$$
E_c(\sigma) = \sum p_iE_c(\sigma_i) \ . \quad (4.40)
$$

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

$$E_c(\sigma) = \sum p_i E_c(\sigma_i) = \sum p_i E(\sigma_i) \geq E(\sum p_i \sigma_i) = E(\sigma) \ , \tag{4.41}$$

and the proof is completed □.

We add that the relative entropy of entanglement $E(\sigma)$ can be calculated easily for Bell diagonal states [103]. We define the density operators $\sigma_{1/2} = |e_{1/2}\rangle\langle e_{1/2}| = |\Psi^{\pm}\rangle\langle\Psi^{\pm}|$ and $\sigma_{3/4} = |e_{3/4}\rangle\langle e_{3/4}| = |\Phi^{\pm}\rangle\langle\Phi^{\pm}|$ where $|\Psi^{\pm}\rangle, |\Phi^{\pm}\rangle$ is the usual Bell basis. Then a Bell–diagonal state has the for $W = \sum_i \lambda_i \sigma_i$ We now prove the following

Theorem 7. For a Bell diagonal state $\sigma = \sum_i \lambda_i \sigma_i$ where all $\lambda_i \in [0, \frac{1}{2}]$ we find

$$E(\sigma) = 0 \tag{4.42}$$

while for $\lambda_1 \geq \frac{1}{2}$ we obtain

$$E(\sigma) = \lambda_1 \ln \lambda_1 + (1 - \lambda_1) \ln(1 - \lambda_1) + \ln 2 \tag{4.43}$$

and analogously for $\lambda_i \geq \frac{1}{2}$.

Proof: The first case is simple once we remember that a Bell diagonal state ρ is separable, i.e. $\rho \in \mathcal{D}$, iff its spectrum lies in $[0, \frac{1}{2}]$ [109]. Therefore $E(\sigma) = 0$.

To prove the theorem for $\lambda_1 \geq \frac{1}{2}$ we again utilize the fact that $f(x) = -\ln x$ is convex. We obtain

$$\begin{aligned} E(\sigma) &= \sum_i \lambda_i \ln \lambda_i + \min_{\rho \in \mathcal{D}} -\mathrm{Tr}\{\sigma \ln \rho\} \\ &\geq \sum_i \lambda_i \ln \lambda_i + \min_{\rho \in \mathcal{D}} -\sum_i \lambda_i \ln \langle e_i|\rho|e_i\rangle \ . \end{aligned} \tag{4.44}$$

We know that $\rho \in \mathcal{D}$ implies that all $\rho_{ii} \leq \frac{1}{2}$ (or otherwise the state can be purified [19, 66]). Therefore we can determine the minimum, not over the states from $\mathcal{D}$, but over the space $\mathcal{B}$ of all Bell diagonal states with spectrum in $[0, \frac{1}{2}]$. This gives a lower bound to eq. (4.44) because

$$\min_{\rho \in \mathcal{D}} -\sum_i \lambda_i \ln \langle e_i|\rho|e_i\rangle = \min_{\rho \in \mathcal{B}} -\sum_i \lambda_i \ln \langle e_i|\rho|e_i\rangle \ .$$

Defining $p_i = \langle e_i|\rho|e_i\rangle$ we have to minimize the function $f(p_1, p_2, p_3, p_4) = -\sum_i \lambda_i \ln p_i$, under the constraints $\sum_{i=1}^4 p_i = 1$ and $p_i \in [0, \frac{1}{2}]$. This minimization yields

$$p_1 = 1/2 \qquad p_i = \lambda_i/2(1-\lambda_1) \ . \tag{4.45}$$

The state $\rho = \sum_i p_i \sigma_i$ with the values from eq. (4.45) lies in $\mathcal{D}$ [66] and therefore the lower limit can be reached which proves eq. (4.43). □

Comparing the result to those for the entanglement of formation [100] one finds that, in fact, strict inequality holds. In general, we have unfortunately found no "closed form" for the relative entropy of entanglement and a computer search is necessary to find the minimum ρ^*, for each given σ. However, we can find numerically the amount of entanglement for two spin 1/2 subsystems very efficiently using general methods independent of the dimensionality and the number of subsystems involved which are described in the next section.

For completeness we would also like to state another useful result that might be used in estimating the value of the relative entropy of entanglement in a state σ.

Theorem 8: Let $\sigma = \sum_j s_j \sigma^j$. Then

$$E(\sigma) \geq \sum_j s_j \Big\{ E(\sigma^j) - S(\sigma^j||\sigma) \Big\} \ .$$

Proof. Recall that for any state ρ [116],

$$S(\sigma||\rho) + \sum_j s_j S(\sigma^j||\sigma) = \sum_j s_j S(\sigma^j||\rho).$$

Thus

$$E(\sigma) := S(\sigma||\hat{\rho}(\sigma)) = \sum_j s_j \Big\{ S(\sigma^j||\hat{\rho}^*) - S(\sigma^j||\sigma) \Big\} \geq \sum_j s_j \Big\{ E(\sigma^j) - S(\sigma^j||\sigma) \Big\} \ .$$

□.

We now turn to discussing other measures of entanglement generated by different measures of distance.

4.2.5.2 Bures Metric

Another distance measure that leads to a measure of entanglement that satisfies the conditions E1-E3 is generated by the modified Bures metric (the Bures metric was introduced in [120]). However, it will turn out that it does not satisfy the condition E4, i.e. it does not reduce to the von Neumann entropy for pure states. Based on this it could be said that it is not a measure of entanglement at all; however, I believe that this is not the case and that it all very much depends on the physical basis under considerations as will be clarified later. In this respect we have a situation similar to the one existing for classical uncertainty measures, where the Shannon entropy is by no means unique, but can be made so be asking for the particular properties to be satisfied.

We now prove F1-F5 for the modified Bures metric, i.e. when $D(\sigma||\rho) = D_B(\sigma||\rho) := 2 - 2F(\sigma,\rho)$, where $F(\sigma,\rho) := \left[\mathrm{Tr}\{\sqrt{\rho}\sigma\sqrt{\rho}\}^{1/2}\right]^2$ is the so called fidelity (or Uhlmann's transition probability). The true Bures metric is given by $2 - 2\sqrt{F(\sigma,\rho)})$. Property F1 follows from the fact that the Bures metric is a true metric and F2 is obvious. F3 is a consequence of the fact that D_B does not increase under a complete positive trace-preserving map [121]. We can also easily check that $p_i q_i F(\sigma_i/p_i, \rho_i/q_i) = F(\sigma_i, \rho_i)$, from where F4 immediately follows as $q_i \in [0,1]$; F5 also follows by inspection. As conditions F1-F5 are satisfied it immediately follows that conditions E1-E3 are satisfied too.

In the following present some properties of the Bures measure of entanglement $E_B(\sigma)$. First we show that for pure states we do not recover the entropy of entanglement.

Theorem 9: For a pure state $|\psi\rangle = \alpha|00\rangle + \beta|11\rangle$ one has

$$E_B(|\psi\rangle\langle\psi|) = 4\alpha^2(1-\alpha^2) \ . \tag{4.46}$$

Proof. To prove Theorem 7 we have to show that the closest disentangled state to $\sigma = |\psi\rangle\langle\psi|$ under the Bures metric is given by $\rho^* = \alpha^2|00\rangle\langle 00| + \beta^2|11\rangle\langle 11|$ To this end we consider a slight variation around ρ^* of the form $\rho_\lambda = (1-\lambda)\rho^* + \lambda\rho$ where

$\rho \in \mathcal{D}$. Now we need to show that

$$\frac{d}{d\lambda} D_B(\sigma||\rho_\lambda)|_{\lambda=0} = 2F_\lambda \frac{d}{d\lambda} \mathrm{Tr}\left\{\sqrt{\sqrt{\sigma}\rho_\lambda\sqrt{\sigma}}\right\} \leq 0 \,, \tag{4.47}$$

where F_λ is the fidelity between σ and ρ_λ and is always positive. Invoking the fact that $\sqrt{\sigma} = \sigma$ as σ is pure we obtain

$$\frac{d}{d\lambda} D_B(\sigma||\rho_\lambda)|_{\lambda=0} = 2F_\lambda \frac{d}{d\lambda} \sqrt{\alpha^4 + \beta^4 + \lambda(\langle\psi|\rho|\psi\rangle - 1)}|_{\lambda=0} \leq 0 \,. \tag{4.48}$$

Using now the closest state ρ^* one then obtains eq. (4.46) □. To obtain the entanglement of an arbitrary pure state one first has to calculate the Schmidt decomposition [47] and then, by local unitary transformation, transform the state to the form $|\psi\rangle = \alpha|00\rangle + \beta|11\rangle$. As local unitary transformations do not change the entanglement, we have therefore shown that the Bures measure of entanglement does not reduce to the entropy of entanglement for pure states.

The proof presented here can be generalized to many dimensional systems in the same fashion as the relative entropy of entanglement. Let

$$f(x) = 2 - 2F(\sigma, (1-x)\omega + x\rho) \,,$$

where $\sigma = \sum_{n_1,n_2} \sqrt{p_{n_1} p_{n_2}} |\phi_{n_1}\psi_{n_1}\rangle\langle\phi_{n_2}\psi_{n_2}|$ is the pure state for which we wish to find the Bures distance of entanglement, and

$$\omega = \sum_n p_n |\phi_n\psi_n\rangle\langle\phi_n\psi_n| \tag{4.49}$$

$$\rho = \sum_{n_1 \ldots n_4} a_{n_1} a^*_{n_2} b_{n_1} b^*_{n_2} |\phi_{n_1}\psi_{n_1}\rangle\langle\phi_{n_2}\psi_{n_2}| \,, \tag{4.50}$$

where ω is, as before, our guess for the closest state under the Bures metric (the proof also works for the modified Bures metric $2 - 2F$). F is the fidelity, which is in this case of σ being pure given by the expression:

$$F = \mathrm{Tr}\{\sigma((1-x)\omega + x\rho)\} \,. \tag{4.51}$$

We wish to show that $\frac{df}{dx}|_{x=0} \geq 0$, which would immediately prove that ω is indeed a minimum of σ. So,

$$\left.\frac{df}{dx}\right|_{x=0} = \mathrm{Tr}(\sigma\omega) - \mathrm{Tr}(\sigma\rho) \,. \tag{4.52}$$

Thus, in order to show that the above derivative is non–negative we need to show that

$$\mathrm{Tr}(\sigma\omega) \geq \mathrm{Tr}(\sigma\rho) \ . \tag{4.53}$$

A simple calculation shows that

$$\mathrm{Tr}(\sigma\omega) = \sum_n p_n^2 \tag{4.54}$$

$$\mathrm{Tr}(\sigma\rho) = \sum_{n_1 n_2} \sqrt{p_{n_1} p_{n_2}} a_{n_1} a_{n_2}^* b_{n_1} b_{n_2}^* \ . \tag{4.55}$$

But, using the well-known inequality for complex numbers [119] we obtain

$$\begin{aligned} |\sum_{n_1 n_2} \sqrt{p_{n_1} p_{n_2}} a_{n_1} a_{n_2}^* b_{n_1} b_{n_2}^*|^2 &\leq (\sum_{n_1 n_2} p_{n_1} p_{n_2} |a_{n_1}||a_{n_2}||b_{n_1}||b_{n_2}|) \\ &\leq \sum_{n_1 n_2} p_{n_1}^2 p_{n_2}^2 \sum_{n_1 n_2} |a_{n_1} a_{n_2}|^2 \sum_{n_1 n_2} |b_{n_1} b_{n_2}|^2 \\ &\leq \sum_{n_1 n_2} p_{n_1}^2 p_{n_2}^2 = (\sum_n p_n^2)^2 \ . \end{aligned} \tag{4.56}$$

Taking the square root proves inequality in eq. (4.53). Hence it follows that the above derivative is positive and that ω is therefore a minimum of σ with respect to the (modified) Bures metric.

In fact, it is now easy to see the following

Corollary. The Bures measure of entanglement for pure states is smaller than the entropy of entanglement, i.e. for any pure state σ

$$E_B(\sigma) \leq -\mathrm{Tr}\left\{\sigma_A \ln \sigma_A\right\} \ . \tag{4.57}$$

Proof. One can see quickly that for $\alpha \in [0,1]$

$$4\alpha^2(1-\alpha^2) \leq -\alpha^2 \ln \alpha^2 - (1-\alpha^2)\ln(1-\alpha^2) \tag{4.58}$$

from which the Corollary follows □.

As the Bures measure of entanglement does not satisfy condition E4, i.e. does not reduce to the entropy of entanglement for pure states one might argue that it does not provide a sensible measure of entanglement. However, it should be noted that the Bures metric immediately gives an upper bound on the following

very special purification procedure. Assume that Alice and Bob are given EPR pairs, but one pair at a time. They are allowed to perform any local operations they like, and then, upon communication, decide whether they keep the pair or discard it. Then, they are given the next EPR pair, and so on. The question is, how many pure singlet states they can possibly distill out of such a purification procedure. The answer is immediately obvious from condition E3. The best that Alice and Bob can do is to have one subensemble with pure singlets and all other subensembles with disentangled states. Then the probability to obtain a singlet is simply given by the Bures measure of entanglement for the initial ensemble. As this is smaller than the entropy of entanglement we have found the nontrivial, though not very surprising result, that this restricted purification procedure is strictly less efficient than entanglement concentration described in [110].

4.2.5.3 One Thousand and One Good Measures of Entanglement?

Fidelity and relative entropy possess various natural generalizations. First of all we can in the usual fashion define quantum Rényi overlaps as the maximum classical distinguishability of two quantum states compared in the following way:

$$F_\alpha(\sigma,\rho) = \min_{E_i} \sum_i (\mathrm{Tr}(\sigma E_i))^\alpha (\mathrm{Tr}(\rho E_i))^{1-\alpha} \,. \tag{4.59}$$

For $\alpha \to 1$ this reduces to the usual expression for fidelity considered in the previous subsection. In the similar fashion we can introduce the Rényi relative information

$$S_\alpha(\sigma,\rho) = \min_{E_i} \frac{1}{\alpha-1} \ln \left\{ \sum_i (\mathrm{Tr}(\sigma E_i)^\alpha (\mathrm{Tr}(\rho E_i))^{1-\alpha} \right\} \tag{4.60}$$

which, likewise, reduces to the von Neumann relative entropy for $\alpha \to 1$ (not really true, explain).

We now introduce a generalization of the von Neumann relative entropy, which we simply refer to as the von Neumann relative g-entropy ("g" stands for generalized). This is a direct generalization of the quantum formula and does not make

use of any classical equivalent. It is given by the following expression [122]

$$S_g(\sigma||\rho) := \mathrm{Tr}\left\{\sigma g(\frac{\rho}{\sigma} - 1)\right\} , \tag{4.61}$$

where $g(\omega) : (-1, \infty) \to \mathbf{R}$ is a strictly convex function satisfying $g(0) = 0$. For $g(\omega) = -\ln(\omega + 1)$ we recover the usual von Neumann relative entropy. There are several ways of defining the expression σ/ρ. One way is to postulate that

$$\frac{\sigma}{\rho} := e^{\ln\sigma - \ln\rho} \tag{4.62}$$

which is by the Lie-Trotter formula equivalent to $\lim_{n\to 1}(\sigma^{1/n}\rho^{-1/n})^n$ [122]. For us it is important to note that Petz [123] showed that if g is a convex operator function than the quantum relative g-entropy does not increase under completely positive, trace preserving map, i.e.

$$S_g(\Phi(\sigma)||\Phi(\rho)) \leq S_g(\sigma||\rho) \ . \tag{4.63}$$

By the same token as the fidelity, this is now satisfies the condition E3, meaning that the measure of entanglement defined as

$$E_g := \min_{\rho\in\mathcal{D}} S_g(\sigma||\rho) \tag{4.64}$$

does not increase under Φ. However not all of the quantum relative g-entropies satisfy F1, F2, F4 or F5. Those that do, immediately satisfy E1-E3, and therefore provide "good" measures of entanglement. However, a general classification of metrics according to F1-F5, or based on E1-E3 has not been attempted.

A generalization of the Bures metric is already used here in the form of the modified Bures metric. Namely, we used $2 - 2F$ as the generator for the measure of entanglement. We could, therefore, ask whether F1–F5 are satisfied by a more general quantity of the form $N(1 - F^n)$, where N is the appropriate normalization and $1 \leq n$. The crucial condition to prove is F4, since all the other are trivially satisfied. That F4 holds can easily be confirmed by inspection.

Another at first sight reasonable candidate to generate a measure of entanglement is the Hilbert–Schmidt metric. Here we have that $D(A||B) = ||A - B||^2 :=$

$\mathrm{Tr}(A-B)^2$. F1 follows from the fact that $||A-B||$ is a true metric, and F2 is obvious. However, F3 does not hold. A counter example in the 4-dimensional space follows [124]. Let A and B be 4×4 matrices defined by

$$A=\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad B=\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Then we have

$$A^\dagger A=\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

It follows that $A^\dagger A+B^\dagger B=\mathbb{1}_4$ and hence

$$T(\rho)=A\rho A^\dagger+B\rho B^\dagger,$$

where ρ is arbitrary, defines a completely positive trace preserving linear map. Let ρ and σ be density matrices defined by

$$\rho=\begin{pmatrix} 1/2 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \sigma=\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/2 \end{pmatrix}.$$

Then we have

$$(\rho-\sigma)^2=\begin{pmatrix} 1/4 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/4 & 0 \\ 0 & 0 & 0 & 1/4 \end{pmatrix}$$

and hence

$$\mathrm{Tr}[(\rho-\sigma)^2]=1.$$

On the other hand, we have

$$A\rho A^\dagger = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad B\rho B^\dagger = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$A\sigma A^\dagger = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 \end{pmatrix}, \quad B\sigma B^\dagger = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 \end{pmatrix}.$$

It follows that

$$(T(\rho) - T(\sigma))^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and hence

$$\mathrm{Tr}[(T(\rho) - T(\sigma))^2] = 2.$$

We conclude therefore

$$\mathrm{Tr}[(T(\rho) - T(\sigma))^2] > \mathrm{Tr}[(\rho - \sigma)^2].$$

Consequently we cannot prove that this is a good measure by using our sufficient conditions F1-F5. Condition E3, therefore, has to be checked using some other, to the author unknown, means. We also believe that there are numerous other nontrivial choices for $D(A||B)$ (by nontrivial we mean that the choice is not a simple scale transformation of the above candidates). Each of those generators would arise from a different physical procedure involving measurements conducted on σ and ρ^*. None of the choices could be said to be more important than any other *a priori*, but the significance of each generator would have to be seen through physical

assumptions. To illustrate this point further, let us take an extreme example. Define,

$$D(A||B) = \begin{cases} 1 & : \quad A \neq B \\ 0 & : \quad A = B \end{cases} .$$

If entanglement is calculated using this distance, then

$$E(\sigma) = \begin{cases} 1 & : \quad \sigma \notin \mathcal{D} \\ 0 & : \quad \sigma \in D \end{cases} .$$

This measure therefore tells us if a given state σ is entangled, i.e. when $E(\sigma) = 1$, or disentangled, i.e. when $E(\sigma) = 0$ (note that this measure is obtained from the generalized Bures metric $N(1 - F^n)$ when $n \to 0$). We can call it the "indicator measure" of entanglement. It should be noted that this measure trivially satisfies conditions E1-E3. This shows that there are numerous different choices for $D(A||B)$ and each is related to different physical considerations. We explain the statistical basis of the relative entropy of entanglement in section 4.4. The relative entropy of entanglement is then seen to be linked very naturally to the notion of a purification procedure. First, however, we present an efficient numerical method to obtain entanglement for arbitrary particles.

4.3 Numerics for Two Spin 1/2 Particles

In order to understand how our program for calculating the amount of entanglement works, we first need to introduce one basic definition and one important result from convex analysis [59]. From this point onwards we concentrate on the von Neumann relative entropy as a measure of entanglement although most of the considerations are of a more general nature.

Definition 2. The convex hull (co(A)) of a set A is the set of all points which can be expressed as (finite) convex combinations of points in A. In other words, $x \in \text{co}(A)$ if and only if x has an expression of the form $x = \sum_{k=1}^{K} p_k a_k$ where K is finite, $\sum_{k=1}^{K} p_k = 1$, and, for $k = 1, \ldots, K$, $p_k > 0$ and $a_k \in A$.

We immediately see that the set of disentangled states $\mathcal{D}$ is a convex hull of its pure states. This means that any state in $\mathcal{D}$ can be written as a convex combination of the form $\sum p_n|\phi_n\psi_n\rangle\langle\phi_n\psi_n|$. However, there is now a problem in the numerical determination of the measure of entanglement. We have to perform a search over the set of disentangled states in order to find that disentangled state which is closest to the state σ of which we want to know the entanglement. But how can we parametrize the disentangled states? We know that the disentangled states are of the form given by Definition 1. However, there the number of states in the convex combination is not limited. Therefore one could think that we have to look over all convex combinations with one state, then two states, then 1000 states and so forth. The next theorem, however, shows that one can put an upper limit to the number of states that are required in the convex combination. This is crucial for our minimization problem as it shows that we do not have to have an infinite number of parameters to search over.

Caratheodory's theorem. Let $A \subset \mathbf{R}^N$. Then any $x \in \mathrm{co}(A)$ has an expression of the form $x = \sum_{n=1}^{N+1} p_n a_n$ where $\sum_{n=1}^{N+1} p_n = 1$, and, for $n = 1, \ldots, N+1$, $p_n \geq 0$ and $a_n \in A$.

Proof. Suppose that $x \in \mathrm{co}(A)$ has an expression of the form $x = \sum_{k=1}^{K} p_k a_k$ where K is finite, $\sum_{k=1}^{K} p_k = 1$, and, for $k = 1, \ldots, K$, $p_k > 0$ and $a_k \in A$. Suppose that $K > N+1$ and that x cannot be expressed as a convex combination of fewer than K points. For $k = 2, \ldots, K$, the points $a_k - a_1$ are linearly dependent in $\mathbf{R}^N$, so there exists a non-trivial sequence $(\lambda_k)_{k=1}^{K} \in \mathbf{R}^K$ such that $\sum_{k=1}^{K} \lambda_k a_k = 0$ and $\sum_{k=1}^{K} \lambda_k = 0$. But then, for any m for which $\lambda_m > 0$,

$$x = \sum_{k=1}^{K} (p_k a_k - \frac{p_m}{\lambda_m} \lambda_k a_k) \; . \tag{4.65}$$

Choosing m so that p_m/λ_m is minimized over m with $\lambda_m > 0$ gives an expression for x as a convex combination of fewer than K points in A which is a clear contradiction □.

A direct consequence of Caratheodory's theorem is that any state in $\mathcal{D}$ can be

decomposed into a sum of at most $(\dim(H_1) \times \dim(H_2))^2$ products of pure states. So, for 2 spin1/2 particles there are at most 16 terms in the expansion of any disentangled state. In addition, each pure state can be described using two real numbers, so that there are altogether at most $15 + 16 \times 4 = 79$ real parameters needed to completely characterize a disentangled state in this case. In fact, this result can be improved in the case of two spin-1/2 particles, and Wootters has shown that only 4 terms are needed in the expansion of any disentangled state [102].

A random search over the 79 real parameters would still be very inefficient. However, we can now make use of another useful property of the relative entropy, which is the fact that it is convex. This means that we have to minimize a convex function over the convex set of disentangled states. It can easily be shown that any local minimum must also be a global minimum. Suppose that a convex function $f(x)$ has a minimum at x_1 so that $df(x)/dx = 0$ at x_1. Suppose also that, contrary to our claim, the function f has another minimum at x_2 such that $f(x_2) < f(x_1)$. Then since f is convex we have for any $0 \leq \lambda \leq 1$ that

$$\begin{aligned} f[x_2 + \lambda(x_1 - x_2)] - f(x_2) &= f[(1-\lambda)x_2 + \lambda x_1] - f(x_2) \\ &\leq (1-\lambda)f(x_2) + \lambda f(x_1) - f(x_2) \\ &= \lambda[f(x_1) - f(x_2)] \,. \end{aligned} \tag{4.66}$$

Dividing through by λ and taking the limit $\lambda \to 0$ gives us 0 on the left hand side, but also gives us a negative quantity on the right hand side, which is a clear contradiction. Hence we must have that $f(x_1) = f(x_2)$ and any local minimum of a convex function is therefore also a global minimum. Returning to our problem, we can perform a gradient search for the minimum (basically we calculate the gradient and then perform a step in the opposite direction and repeat this procedure until we hit the minimum). As soon as we have found any relative minimum we can stop the search, since this is also a global minimum. To make the gradient search efficient we have to chose a suitable parametrization. The parametrization that

we use has the advantage that it also provides us with another proof of Theorem 3 which states that for pure states the relative entropy of entanglement reduces to the von Neumann reduced entropy. We first explain the parametrization and then state the alternative proof for Theorem 3. The following results can easily be extended to two subsystems of arbitrary dimensions but for clarity we restrict ourselves to two spin 1/2 systems.

Our aim is to find the amount of entanglement of a state σ of two spin 1/2 states, i.e. we have to minimize $tr\{\sigma \ln \sigma - \sigma \ln \rho\}$ over all $\rho \in \mathcal{D}$. From Caratheodory's theorem we know that we only need convex combinations of at most 16 pure states ρ_k^i to represent $\rho \in \mathcal{D}$, i.e.

$$\rho = \sum_{i=1}^{16} p_i^2 \rho_1^i \otimes \rho_2^i \ . \tag{4.67}$$

(Notice that we use p_i^2 instead of p_i for convenience, so that here we require that $\sum_{i=1}^{16} p_i^2 = 1$). The parametrization we chose is now given by

$$p_i = \sin \phi_{i-1} \prod_{j=i}^{15} \cos \phi_j \quad \text{with} \quad \phi_0 = \frac{\pi}{2} \tag{4.68}$$

and

$$\begin{aligned} \rho_k^i &= |\psi_k^i\rangle\langle\psi_k^i| \\ |\psi_1^i\rangle &= \cos\alpha_i|0\rangle + \sin\alpha_i e^{i\,\eta_i}|1\rangle \\ |\psi_2^i\rangle &= \cos\beta_i|0\rangle + \sin\beta_i e^{i\,\mu_i}|1\rangle \end{aligned}$$

All angles $\alpha_i, \beta_i, \phi_i, \eta_i, \mu_i$ can have arbitrary values, but due to the periodicity only the interval $[0, 2\pi]$ is really relevant. Numerically this has the advantage that our parameter space has no edges at which problems might occur. The program for the search of the minimum is now quite straightforward. The idea is that given σ we start from a random ρ, i.e. we generate 79 random numbers. Then we compute $S(\sigma||\rho)$, as well as small variations of the 79 parameters of ρ, to obtain the approximate gradient of $S(\sigma||\rho)$ at the point ρ. We then move opposite to the gradient to obtain the next ρ. We continue this until we reach the minimum. As

explained before, a convex function over a convex set can only have a global minimum, so that the minimum value we end up with is the one and only. The method outlined above immediately generalizes to two subsystems of arbitrary dimension, however, the number of parameters rises quickly to large values which slows down the program considerably.

Before we state some numerical results we now indicate an alternative proof of Theorem 3 using Caratheodory's theorem and the parametrization given in eqs. (4.67) - (4.69). For this proof we use the fact that we can represent the logarithm of an operator ρ by

$$\ln \rho = \frac{1}{2\pi i} \oint \ln z \frac{1}{z\mathbb{1} - \rho} \tag{4.69}$$

where the path of integration encloses all eigenvalues of ρ. We can now take the partial derivative of $\ln \rho$ with respect to a parameter ϕ on which ρ might depend.

$$\frac{\partial \ln \rho}{\partial \phi} = \frac{1}{2\pi i} \oint \ln z \frac{1}{z\mathbb{1} - \rho} \frac{\partial \rho}{\partial \phi} \frac{1}{z\mathbb{1} - \rho} \ . \tag{4.70}$$

Now, we have a given pure state

$$\sigma = \alpha^2 |00\rangle\langle 00| + \alpha\sqrt{1-\alpha^2}(|00\rangle\langle 11| + |11\rangle\langle 00|) + (1-\alpha^2)|11\rangle\langle 11| \tag{4.71}$$

The suspected closest approximation to σ within the disentangled states is given by

$$\rho_{min} = \alpha^2 |00\rangle\langle 00| + (1-\alpha^2)|11\rangle\langle 11| \ . \tag{4.72}$$

If we want to represent ρ_{min} using the parametrization given in eqs. (4.67) - (4.69) then we find for these parameters $\cos^2 \phi_1 = \alpha^2$; $\alpha_2 = \beta_2 = \frac{\pi}{2}$ and zero for all other parameters. Using eq. (4.70) one can now calculate all the partial derivatives of the relative entropy around the point ρ_{min}. It is easy, but rather lengthy, to check that these derivatives vanish and that therefore ρ_{min} is a relative minimum. This concludes the proof as a relative minimum of a convex function on a convex set is also a global minimum.

After this additional proof of Theorem 3 we now state some results that we have obtained or confirmed with the program that implements the gradient search.

We present four nontrivial states σ for which we can find the closest disentangled state ρ that minimize the von Neumann relative entropy thereby giving the relative entropy of entanglement. Using the same ideas as for the proof of Theorem 3 in Eq. (4.69 - 4.72) one can then prove that these are indeed the closest disentangled states.

Example 1.

$$\sigma_1 = \lambda|\Phi^+\rangle\langle\Phi^+| + (1-\lambda)|01\rangle\langle 01| \tag{4.73}$$

$$\begin{aligned}\rho_1 &= \frac{\lambda}{2}(1-\frac{\lambda}{2})|00\rangle\langle 00| + \frac{\lambda}{2}(1-\frac{\lambda}{2})\{|00\rangle\langle 11| + \text{H.C.}\} + \\ &(1-\frac{\lambda}{2})^2|01\rangle\langle 01| + \frac{\lambda^2}{4}|10\rangle\langle 10| + \frac{\lambda}{2}(1-\frac{\lambda}{2})|11\rangle\langle 11|\end{aligned} \tag{4.74}$$

$$E(\sigma_1) = (\lambda-2)\ln(1-\frac{\lambda}{2}) + (1-\lambda)\ln(1-\lambda)\ . \tag{4.75}$$

Here $|\Phi^+\rangle$ is one of the four Bell states defined by

$$|\Phi^\pm\rangle = \frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle) \tag{4.76}$$

$$|\Psi^\pm\rangle = \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle) \tag{4.77}$$

Example 2.

$$\sigma_2 = \lambda|\Phi^+\rangle\langle\Phi^+| + (1-\lambda)|00\rangle\langle 00| \tag{4.78}$$

$$\rho_2 = (1-\frac{\lambda}{2})|00\rangle\langle 00| + \frac{\lambda}{2}|11\rangle\langle 11| \tag{4.79}$$

$$\begin{aligned}E(\sigma_2) &= s_+\ln s_+ + s_-\ln s_- \\ &-(1-\frac{\lambda}{2})\ln(1-\frac{\lambda}{2}) - (1-\frac{\lambda}{2})\ln(1-\frac{\lambda}{2}),\end{aligned} \tag{4.80}$$

where

$$s_\pm = \frac{1 \pm \sqrt{1-2\lambda(1-\frac{\lambda}{2})}}{2} \tag{4.81}$$

are the eigenvalues of σ_2. One could argue that in the above two cases the following reasoning can be applied: $\sigma_{1(2)}$ is a mixture of a maximally entangled state (for which the amount of entanglement is given by $\ln 2$) and a completely disentangled state ($E = 0$). Thus one would expect a total amount of entanglement of $\lambda \ln 2$.

It is curious that this reasoning does not work for either of the two states, since, in fact, $E(\sigma_{1(2)}) \leq \lambda \ln 2$. Now, we show how to use Theorem 4 to generate more states and their minima. For pure states $\sigma^2 = \sigma$ we know the minimum ρ. Now, the state that is a convex sum of σ and ρ should also have the same minimum ρ. So, we have the following.

Example 3.

$$\sigma_3 = A|00\rangle\langle 00| + B|00\rangle\langle 11| + B^*|11\rangle\langle 00| + (1-A)|11\rangle\langle 11| \quad (4.82)$$

$$\rho_3 = A|00\rangle\langle 00| + (1-A)|11\rangle\langle 11| \quad (4.83)$$

$$E(\sigma_3) = e_+ \ln e_+ + e_- \ln e_- - A \ln A - (1-A)\ln(1-A) \ , \quad (4.84)$$

where

$$e_\pm = \frac{1 \pm \sqrt{1 - 4A(1-A) - |B|^2}}{2} \ . \quad (4.85)$$

Using Theorem 4, the amount of entanglement can be found for a number of other spin 1/2 states. Our program can also help us infer the entanglement of some other non-trivial states as the last example shows.

Example 4.

$$\sigma_4 = A|00\rangle\langle 00| + B|00\rangle\langle 11|B^*|11\rangle\langle 00| + \quad (4.86)$$

$$+(1-2A)|01\rangle\langle 01| + A|11\rangle\langle 11| \quad (4.87)$$

$$\rho_4 = C|00\rangle\langle 00| + D|00\rangle\langle 11| + D^*|11\rangle\langle 00| + E|01\rangle\langle 01| \quad (4.88)$$

$$+(1-2C-E)|10\rangle\langle 10| + C|11\rangle\langle 11| \ , \quad (4.89)$$

where

$$E = \frac{(1-2A)(1-A)^2}{(1-A)^2 - B^2} \quad (4.90)$$

$$C = 1 - A - E \quad (4.91)$$

$$D = \sqrt{E(1-E-2C)} = \frac{(1-2A)(1-A)}{(1-A)^2 - B^2} B \ . \quad (4.92)$$

It is now easy to compute the amount of entanglement from the above information.

In addition to the above described methods there is a simple way of obtaining a lower bound for the amount of entanglement for any two spin 1/2 system. Suppose that we have a certain state σ. We first find the *maximally* entangled state $|\psi\rangle$ such that the fidelity $F = \langle\psi|\sigma|\psi\rangle$ is maximized. Then we apply local unitary transformations to σ which transform $|\psi\rangle$ into the singlet state (this is, of course, always possible). Now, we apply local random rotations [100] to both particles. These will transform σ into a Werner state, where the singlet state will have a weight F (since it is invariant under rotations) and all the other three Bell states will have equal weights of $(1 - F)/3$ (since they are randomized). Since these operations are local they cannot increase the amount of entanglement, and we have that for any σ

$$E(\sigma) \geq E(W_F) = F \ln F + (1 - F) \ln(1 - F) + \ln 2 \tag{4.93}$$

where W_F is the above described Werner state (the Relative Entropy of entanglement for a general Bell diagonal state is calculated in [103]).

We note that this efficient computer search provides an alternative criterion for deciding when a given state σ of two spin 1/2 systems is disentangled, i.e. of the form given in eq. (4.4). The criterion which already exists is the one given by Peres and the Horodecki family (see the second and third references in [67]), which states that a state is disentangled iff its partial trace over either of the subsystems is a non-negative operator. This criterion is only valid for two spin 1/2, or one spin 1/2 and one spin 1 systems. In the absence of a more general analytical criterion our computational method provides a way of deciding this question. In addition we would like to point out that the program is also able to provide us with the convex decomposition of a disentangled state ρ.

At the end of this section we mention *additivity* as an important property desired from a measure of entanglement, i.e. we would like to have

$$E(\sigma_{12} \otimes \sigma_{34}) = E(\sigma_{12}) + E(\sigma_{34}) \ , \tag{4.94}$$

where systems $1+2$ and systems $3+4$ are entangled separately from each other. The exact definition of the left hand side is

$$E(\sigma_{12}\otimes\sigma_{34}) = \min_{p_i,\rho_{13},\rho_{24}} S(\sigma_{12}\otimes\sigma_{34}||\sum_i p_i\rho_{13}^i\otimes\rho_{24}^i) \ . \tag{4.95}$$

Why this form? One would originally assume that $\sigma_{12}\otimes\sigma_{34}$ should be minimized by the states of the form $(\sum_i p_i\rho_1^i\otimes\rho_2^i)\otimes(\sum_j p_j\rho_3^j\otimes\rho_4^j)$. However, Alice, who holds systems 1 and 3, and Bob, who holds systems 2 and 4, can also perform arbitrary unitary operation on their subsystems (i.e. locally). This obviously leads to the creation of entanglement between 1 and 3 and between 2 and 4 and hence the form given in eq. (4.95). Additivity is, of course, already true for the pure states, as can be seen from the proof above, when our measure reduces to the von Neumann entropy. For more general cases we were unable to provide an analytical proof, so that the above additivity property remains a conjecture. However, for two spin 1/2 systems, our program did not find any counter-example. It should be noted that it is easy to see that we have

$$E(\sigma_{12}\otimes\sigma_{34}) \leq E(\sigma_{12}) + E(\sigma_{34}) \ . \tag{4.96}$$

In the following we will assume that Eq. (4.94) holds and use it in section 4.5 to derive certain limits to the efficiency of purification procedures.

4.4 Statistical Basis of Entanglement Measure

Let us see how we can interpret the relative entropy of entanglement in the light of experiments, i.e. statistically [105]. We have shown how the notion of the Shannon relative entropy arises in classical information theory as a measure of distinguishability of two probability distributions. We now wish to generalize this idea to the quantum case, i.e. to distinguishing between two quantum states (for a discussion of distinguishability of pure quantum states see e.g. [125]). We will see that this naturally leads to the notion of the von Neumann relative entropy.

It is then straightforward to extend this concept to explain the relative entropy of entanglement .

We have seen from the theory of types that the probability of not distinguishing the distributions $P(x)$ and $Q(x)$ after n trials asymptotically converges to $e^{-nS(P(x)||Q(x))}$, where

$$S(P(x)||Q(x)) = \sum_i p_i \ln p_i - p_i \ln q_i \tag{4.97}$$

is the Shannon relative entropy. To generalize this to quantum theory, we need a means of generating probability distributions from two quantum states σ and ρ. This is accomplished by introducing a general measurement $\sum_i E_i E_i^\dagger = \mathbb{1}$. So, the probabilities are given by

$$\begin{aligned} p_i &= \mathrm{Tr}(E_i^\dagger E_i \rho) \\ q_i &= \mathrm{Tr}(E_i^\dagger E_i \sigma) \ . \end{aligned} \tag{4.98}$$

Now, we can use eq. (4.97) to distinguish between σ and ρ. The above is not the most general measurement that we can make, however. In general we have N copies of σ and ρ in the state

$$\sigma^N = \underbrace{\sigma \otimes \sigma ... \otimes \sigma}_{\text{total of N terms}} \tag{4.99}$$

$$\rho^N = \underbrace{\rho \otimes \rho ... \otimes \rho}_{\text{total of N terms}} \ . \tag{4.100}$$

We may now apply a POVM $\sum_i A_i = \mathbb{1}$ acting on σ^N and ρ^N. Consequently, we define a new type of relative entropy

$$S_N(\sigma||\rho) := \sup_{\text{A's}} \left\{ \frac{1}{N} \sum_i \mathrm{Tr} A_i \sigma^N \ln \mathrm{Tr} A_i \sigma^N - \mathrm{Tr} A_i \sigma^N \ln \mathrm{Tr} A_i \rho^N \right\} . \tag{4.101}$$

Now it can be shown that [116]

$$S(\sigma||\rho) \geq S_N \tag{4.102}$$

where, as before,

$$S(\sigma||\rho) := \mathrm{Tr}(\sigma \ln \sigma - \sigma \ln \rho) \tag{4.103}$$

is the von Neumann relative entropy [52, 103, 105, 112, 113, 116] (for the summary of the properties of the von Neumann relative entropy see [13]). Equality is achieved in eq. (4.102) iff σ and ρ commute [28]. The inequality in eq. (4.102) can be seen as a consequence of the Holevo bound for the classical capacity of QCC in Chapter 3 [28]. We already remarked there that the Holevo bound is achieved only when all the signals commute, which is in direct analogy with the eq. (4.102) when σ and ρ commute. However, for any σ and ρ it is true that [115]

$$S(\sigma||\rho) = \lim_{N\to\infty} S_N \ .$$

In fact, this limit can be achieved by projective measurements which are independent of σ [126]. It is known that if eq. (4.97) is maximized over all general measurements E, the upper bound is given by the von Neumann relative entropy (see e.g. [116]). In quantum theory we therefore state a law analogous to Sanov's theorem presented in Chapter 2 (see also [105]),

Theorem 10 (or The quantum Sanov Theorem). The probability of *not* distinguishing two quantum states (i.e. density matrices) σ and ρ after n measurements is

$$p(\rho \to \sigma) = e^{-nS(\sigma||\rho)} \ . \tag{4.104}$$

In fact, as explained before, this bound is reached asymptotically [115], and the measurements achieving this are global projectors independent of the state σ [126]. We note that the quantum Sanov Theorem was presented by Donald in [127] as a definition justified by properties uniquely characterizing the quantity $e^{-nS(\sigma||\rho)}$. The underlying intuition in the above measurement approach and Donald's approach are basically the same. Now the interpretation of the relative entropy of entanglement becomes immediately transparent as I have shown with my co-workers in [105]. The probability of mistaking an entangled state σ for a closest, disentangled state, ρ, is $\exp\{-n \times min_{\rho\in\mathcal{D}} S(\sigma,\rho)\} = e^{-nE(\sigma)}$. If the entanglement of σ is greater, than it takes fewer measurements to distinguish it from a disentangled state (or, fixing n, there is a smaller probability of confusing it with some disentangled state). Let us give

an example. Consider a state $(|00\rangle + |11\rangle)/\sqrt{2}$, known to be a maximally entangled state. The closest to it is the disentangled state $(|00\rangle\langle 00| + |11\rangle\langle 11|)/2$ [103]. To distinguish these states it is enough to perform projections on to $(|00\rangle + |11\rangle)/\sqrt{2}$. If the state that we are measuring is the above mixture, then the sequence of results (1 for a successful projection, and 0 for an unsuccessful projection) will contain on average an equal number of 0's and 1's. For this to be mistaken for the above pure state, the sequence has to contain all n 1's. The probability for that is 2^{-n}, which also comes from using eq. (4.104). If, on the other hand, we performed projections onto the pure state itself, we would then never confuse it with a mixture, and from eq. (4.104) the probability is seen to be $e^{-\infty} = 0$.

Before we apply this simple idea to obtaining an upper bound to the efficiency of any purification procedure, we would like to explain briefly how the Bures distance of entanglement arises statistically. The (modified) Bures metric is given by $D_B(\sigma||\rho) = 2 - 2F(\sigma,\rho)$, where $F(\sigma,\rho) := \left[\mathrm{Tr}\{\sqrt{\rho}\sigma\sqrt{\rho}\}^{1/2}\right]^2$ is the fidelity [128]. The (modified) Bures metric offers a very attractive and simple operational basis for the measure of entanglement in terms of general measurements [28]. It derives from the nature of fidelity as a 'measure' of distinguishability between two probability distributions $p_{1i} = \mathrm{Tr}(\sigma A_i^\dagger A_i)$ and $p_{1i} = \mathrm{Tr}(\rho A_i^\dagger A_i)$, where $\sum_i A_i^\dagger A_i = \mathbf{I}$. More precisely,

$$F(\sigma,\rho) = \min_{A_i^\dagger A_i} \sum_i \sqrt{\mathrm{Tr}(\sigma A_i^\dagger A_i)}\sqrt{\mathrm{Tr}(\rho A_i^\dagger A_i)} \tag{4.105}$$

where the minimum is taken over all possible general measurements. This possibly enables us, in principle, to determine eq. (4.15) and therefore also the degree of entanglement experimentally. We stress that in order to satisfy F4 we need to use somewhat modified Bures metric $2 - 2F(\sigma,\rho)$, i.e. without the square root of the fidelity, but this does not change the interpretation in any fundamental way.

4.5 Thermodynamics of Entanglement

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 the 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 von Neumann 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.

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, i.e. that σ is known to Alice and Bob before they start purifying. 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 [66]. At present the best that can be done is to purify a certain class of entangled states. (see e.g. [19, 110, 142]). 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. We 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. 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 $2^{-m} = e^{-m \ln 2}$ (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 $e^{-nE(\sigma)}$. But the 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

$$e^{-nE(\sigma)} \leq e^{-m \ln 2} , \tag{4.106}$$

which implies that

$$nE(\sigma) \geq m , \tag{4.107}$$

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 measure of entanglement given in eq. (4.15), when

$D(\sigma||\rho) = S(\sigma||\rho)$, can be used to provide an upper bound on the efficiency of any purification procedure. For Bell diagonal states, Rains [106] found an upper bound on distillable entanglement using completely different methods. It turns out that the bound he obtained in this case is identical to the one provided by the relative entropy of entanglement.

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 von Neumann relative entropy. Based on the upper bound considerations we can introduce the following definition.

Definition 3. A purification procedure given by a local complete positive trace preserving map $\sigma \to \sum V_i \sigma V_i^\dagger$ is defined to be *ideal* in terms of efficiency iff

$$\sum \mathrm{Tr}(\sigma_i)\ E(\sigma_i/\mathrm{Tr}(\sigma_i)) = E(\sigma) \quad , \tag{4.108}$$

where, as usual, $\sigma_i = V_i \sigma V_i^\dagger$ and $p_i = \mathrm{Tr}(V_i \sigma V_i^\dagger)$ (i.e. the ideal purification is the one where E3 is an equality rather than an inequality). Notice that we seem to have an apparent formal analogy between a purification procedure and the Carnot cycle in Thermodynamics. The Carnot cycle is the most efficient cycle in Thermodynamics (i.e. it yields the greatest "useful work to heat" ratio), since it is reversible (i.e. it conserves the thermodynamical entropy). We would now like to claim that the ideal purification procedure is the most efficient purification procedure (i.e. it yields the greatest number of singlets for a given input state), since it is reversible (i.e. it conserves entanglement, measured by the minimum of the von Neumann relative entropy over all disentangled states). Unfortunately this analogy between the Carnot cycle and purification procedures is not exact (it is only strictly true for the pure states). This is seen when we compare the entanglement of formation with the Relative Entropy of Entanglement. In Theorem 6 we have, in fact, shown that the entanglement of formation is never smaller than the relative entropy of entanglement . As an example one can consider Bell diagonal states for which we

can exactly calculate both the entanglement of formation [103, 104] and the relative entropy of entanglement [100]. It turns out that the entanglement of formation is always strictly larger than the relative entropy of entanglement except for the limiting cases of maximally entangled Bell states or of disentangled Bell diagonal states (see Fig. 4.4 for Werner states). This result leads to the following

Implication. In general, the amount of entanglement that was initially invested in creation of σ cannot all be recovered ("distilled") by local purification procedures. Therefore, the ideal purification procedure, though most efficient, is nevertheless irreversible, and some of the invested entanglement is lost in the purification process itself. This irreversibility is a consequence of the loss of certain information as can be seen from the following analysis. Suppose we start with an ensemble of N of singlets and we want to locally create any mixed state σ. The state σ can always be written as a mixture of pure states $\Psi_1, \Psi_2, ...$ with the corresponding probabilities $p_1, p_2, ...$ We now use Bennett et al's (de)purification procedure for pure states [110] (whose efficiency is governed by the von Neumann entropy). We convert the first $p_1 \times N$ singlets into the state Ψ_1, the second $p_2 \times N$ singlets into the state Ψ_2, and so on... In this way, the whole ensemble is in the state σ. But, we have additional information: we know exactly that the first $p_1 \times N$ pairs are in the state Ψ_1, second $p_2 \times N$ states are in the state Ψ_2, and so on. This is not the same as being given an initial ensemble of identically prepared pairs in the state σ with no additional information. In this, second, case we do not have the additional knowledge of the state of each of the pairs. This is why the purification without this knowledge is less efficient, and hence the relative entropy of entanglement is smaller than the entanglement of formation. An open question remains as to whether we can use some other generator, such as the Bures metric, to give an even more stringent bound on the amount of distillable entanglement.

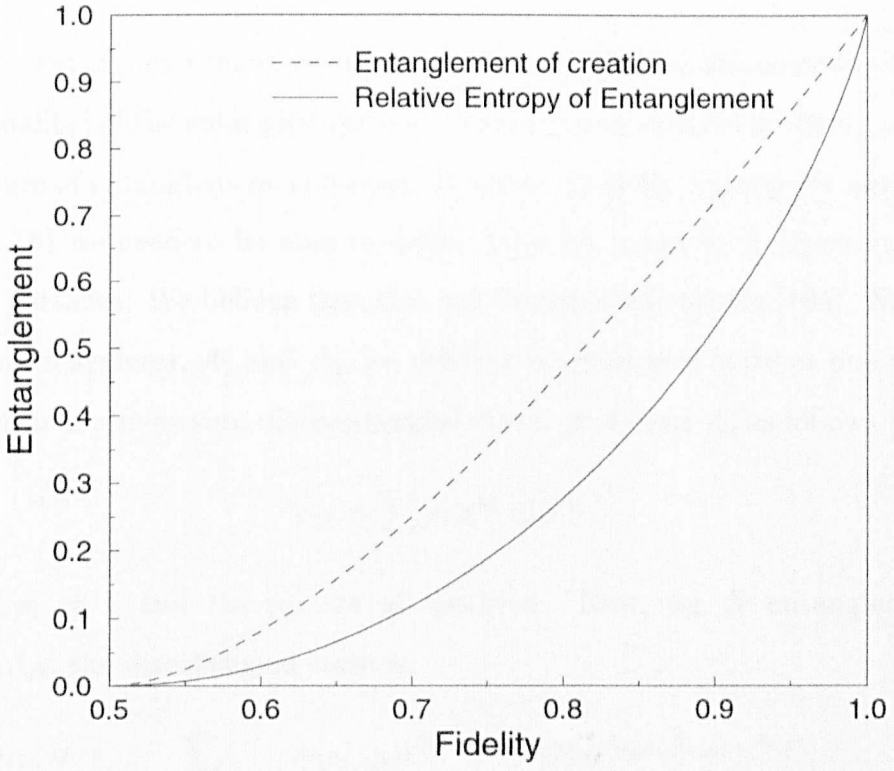


Figure 4.4: Comparison of the entanglement of formation and the Relative Entropy of Entanglement for the Werner states (these are are Bell diagonal states of the form $W = \text{diag}(F, (1-F)/3, (1-F)/3, (1-F)/3.)$ One clearly sees that the entanglement of formation is strictly larger than the relative entropy of entanglement for $0 < F < 1$.

In this section we have analysed the theoretical basis of purification procedures for two qubits. There is nothing fundamental in our treatment that limits this analysis to two qubits only. In fact, the measures of entanglement proposed here are independent of the number of systems or their dimensionality; we only need to define a distance measure and a set of disentangled states, and then check if E1-E3 hold. We turn to this subject in the next section.

4.6 More Than two Subsystems

The treatment of measures of entanglement does not refer to the number (or, indeed, dimensionality) of the entangled systems. This is a very desired property as it makes our measure of entanglement *universal.* However, in order to perform minimization in eq. (4.15) we need to be able to define what we mean by a disentangled state of say N particles. We believe that this can be done inductively [105]. Namely, for two quantum systems, A_1 and A_2, we define a disentangled state as one which can be written as a convex sum of disentangled states of A_1 and A_2 as follows [103, 105]:

$$\rho_{12} = \sum_i p_i \, \rho_i^{A_1} \otimes \rho_i^{A_2} \, , \tag{4.109}$$

where $\sum_i p_i = 1$ and the p's are all positive. Now, for N entangled systems $A_1, A_2, ...A_N$, the disentangled state is:

$$\rho_{12...N} = \sum_{\text{perm}\{i_1 i_2 ... i_N\}} r_{i_1 i_2 ... i_N} \rho^{A_{i_1} A_{i_2} ... A_{i_n}} \otimes \rho^{A_{i_{n+1}} A_{i_{n+2}} ... A_{i_N}} \, , \tag{4.110}$$

where $\sum_{\text{perm}\{i_1 i_2 ... i_N\}} r_{i_1 i_2 ... i_N} = 1$, all r's are positive and where $\sum_{\text{perm}\{i_1 i_2 ... i_N\}}$ is a sum over all possible permutations of the set of indices $\{1, 2, ..., N\}$. To clarify this let us see how this looks for 4 systems:

$$\begin{aligned} \rho_{1234} = & \sum_i \; p_i \, \rho_i^{A_1 A_2 A_3} \otimes \rho_i^{A_4} + q_i \, \rho_i^{A_1 A_2 A_4} \otimes \rho_i^{A_3} \\ & + \; r_i \, \rho_i^{A_1 A_3 A_4} \otimes \rho_i^{A_2} + s_i \, \rho_i^{A_2 A_3 A_4} \otimes \rho_i^{A_1} \\ & + \; t_i \, \rho_i^{A_1 A_2} \otimes \rho_i^{A_3 A_4} + u_i \, \rho_i^{A_1 A_3} \otimes \rho_i^{A_2 A_4} \\ & + \; v_i \, \rho_i^{A_1 A_4} \otimes \rho_i^{A_2 A_3} \end{aligned} \tag{4.111}$$

where, as usual, all the probabilities $p_i, q_i, ..., v_i$ are positive and add up to unity. The above two equations, at least in principle, define the disentangled states for any number of entangled systems. Note that this form describes a different situation from the one given in eq. (4.95) which refers to a number of pairs shared by Alice and Bob only. The above definition of a disentangled state is justified by extending the idea that local actions cannot increase the entanglement between two quantum

systems [100, 103, 105]. In the case of N particles we have N parties (Alice, Bob, Charlie, ... , Wayne) all acting locally on their systems. The general action that also includes communications can be written as [105]

$$\rho \longrightarrow \sum_{i_1,i_2,...,I_N} A_{i_1} \otimes B_{i_2} \otimes ... \otimes W_{i_N} \rho A_{i_1}^{\dagger} \otimes B_{i_2}^{\dagger} \otimes ... \otimes W_{i_N}^{\dagger} \tag{4.112}$$

and it can be easily seen that this action does not alter the form of a disentangled state in eqs. (4.110,4.111). In fact, eq. (4.110) is the most general state invariant in *form* under the transformation given by eq. (4.112). This can be suggested as a definition of a disentangled state for $N \geq 3$, i.e. it is the most general state invariant in form under local POVM and classical communications. Of course, an alternative to defining a disentangled state would be

$$\rho_{12...N} = \sum_i r_i \rho_i^{A_1} \otimes \rho_i^{A_2} \ldots \otimes \rho_i^{A_N} \quad , \tag{4.113}$$

which means that we do not allow any entanglement in any subset of the N states. This would be a disentangled state based on some local hidden variable model. Again we repeat that the particular choice of a form of disentangled states will depend on the physical background in our model and there is no absolute sense in which we can resolve this dichotomy. It should be stressed that for two particles this free choice does not exist as both pictures coincide.

4.7 Conclusions

We can look at the entanglement from two different perspectives. One insists that local actions cannot increase entanglement and do not change it if they are unitary. The other one looks at the way we can distinguish an entangled state from a disentangled one. In particular, the following question is asked: what is the probability of confusing an entangled state with a disentangled one after performing a certain number of measurements? These two, at first sight different approaches, lead to the same measure of entanglement. This results in the fact that a purification procedure can be regarded as a protocol of distinguishing an entangled state from a

disentangled set of states. From this premise we derived the upper bound on the efficiency of any purification procedure. It turns out that distil able entanglement is in general smaller than the entanglement of formation. Our entanglement measure is independent on the number of systems and their dimensionality. This suggests applying it to more than two entangled systems in order to understand multi-particle entanglement. We have shown how to compute entanglement efficiently for two spin 1/2 subsystems using computational methods. However, a closed form for the expression of this entanglement measure is desirable. An interesting problem is to specify all the states that have the same amount of entanglement. We know that all the states that are equivalent up to a local unitary transformation have the same amount of entanglement (by definition–E2). However, there are states with the same amount of entanglement but which are not equivalent up to a local unitary transformation (for example one state is pure and the other one is mixed). A question for further research is whether they are linked by a local complete measurement. Our work in addition suggest a question of finding a general local map that preserves the entanglement of a given entangled state.

Chapter 5

Quantum Error Correction

5.1 Introduction

In the previous chapter we saw how to quantify entanglement, and how to understand entanglement from the statistical point of view. Now we turn our attention to realistic situations involving entanglement manipulations. In most realistic cases entanglement is gradually lost due to the detrimental interactions with an environment. In this chapter we focus on methods of protection of quantum states in dissipative and decoherent environments. In particular, with the discovery of an algorithm to factorize a large number on a quantum computer in polynomial time instead of exponential time required by a classical computer [17], the question of how to implement such a quantum computer has received considerable attention [83]. We have already stressed that this exponential increase crucially depends on being able to maintain large entangled states for sufficiently long periods of time. However, realistic estimates soon showed that decoherence processes and spontaneous emission severely limit the bit size of the number that can be factorized by destroying entanglement [129, 130]. It has become clear that the solution to the problem does not lie in an increase in the lifetime of the transitions used in the computation. Attention has now shifted towards the inves-

tigation of methods to encode qubits such that the correction of errors due to interaction with the environment becomes possible. In a number of recent publications, possible encoding schemes have been considered and theoretical work has been undertaken to elucidate the structure of quantum error correction codes [20, 21, 22, 23, 24, 100, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140]. However, in this chapter we focus on a particular part of spontaneous emission–the conditional time evolution [141] between spontaneous emissions, which these codes do not correct perfectly. This has the effect that, for example, the encoded lower state of a qubit, which, if unencoded, is not influenced by the conditional time evolution, acquires an error due to the conditional time evolution. We then proceed to construct a code that is able to correct *one* general error and is able to correct to *all orders* the errors due to the conditional time evolution between spontaneous emissions. By one general error we mean an arbitrary one bit operation acting on a single bit of the code. The conditional time evolution, however, contains terms that act on many qubits. The code described here is the first code that has the ability to correct a special kind of error (i.e. the one due to the conditional time evolution) to *all* orders. This is an interesting feature, as one would be interested in correcting those errors which frequently occur to higher order than rare errors. In addition our code is insensitive to any possible detuning of the laser used to drive transitions in the atoms (or ions). We explain this feature in greater detail in section 5.4. The code presented here is optimal in the sense that it uses the smallest possible number of qubits required to perform its task (correcting one general error and all errors due to the conditional time evolution). Before we describe this particular code, we derive general conditions that all such codes have to satisfy in order to be able to protect quantum information.

5.2 General Conditions

We now describe an alternative way of manipulating quantum states which is best handled using the language of quantum computation. An advantage of quantum computation lies in the fact that the input can be in a coherent superposition of qubit states, which are then simultaneously processed. The practical realisation of a qubit can be constructed from any two-state quantum system e.g. a two-level atom in an ion trap, where the unitary transformations are implemented through interaction with a laser. The computation is completed by making a measurement on the output. However, a major problem is that the coherent superpositions must be maintained throughout the computation. In reality, the main source of coherence loss is due to dissipative coupling to an environment with a large number of degrees of freedom, which must be traced out of the problem. This loss is often manifested as some form of spontaneous decay, whereby quanta are randomly lost from the system. Each interaction with, and hence dissipation to, the environment can be viewed in information theoretic terms as introducing an error in the measurement of the output state. There are, however, techniques for 'correcting' errors in quantum states [20, 21, 23, 142]. The basic idea of error-correction is to introduce an excess of information, which can then be used to recover the original state after an error. These quantum error correction procedures are in themselves quantum computations, and as such also susceptible to the same errors. This imposes limits on the nature of the 'correction codes', which are explored in this section.

First we derive general conditions which a quantum error correction code has to satisfy [142] and which are, in particular, less restricting than those previously derived in [132]. We point out that our derivation is an alternative to that by Knill and Laflamme in [24], who also arrive at the same conditions, and this is the author's original contribution to the field. Assume that q qubits are encoded in terms of $n \geq q$ qubits to protect against a certain number of errors, d. We construct 2^q *code–words*, each being a superposition of states having n qubits. These code-

words must satisfy certain conditions, which are derived in this section. There are three basic errors [132] (i.e. all other errors can be written as a combination of those): amplitude, $\hat{A}$, which acts as a NOT gate; phase, $\hat{P}$, which introduces a minus sign to the upper state; and their combination, $\hat{A}\hat{P}$. A subscript shall designate the position of the error, so that $\hat{P}_{1001}$ means that the first and the fourth qubit undergo a phase error.

We consider an error to arise due to the interaction of the system with a 'reservoir' (*any* other quantum system), which then become entangled. This procedure is the most general way of representing errors, which are not restricted to discontinuous 'jump' processes, but encompass the most general type of interaction. Error correction is thus seen as a process of disentangling the system from its environment back to its original state. The operators $\hat{A}$ and $\hat{P}$ are constructed to operate only on the system, and are defined in the same way as the operators for a complete measurement described in subsection 2.4, eq. (3.28). In reality, each qubit would couple independently to its own environment, so the error on a given state could be written as a direct product of the errors on the individual qubits. A convenient error basis for a single error on a single qubit is $\{\hat{\mathbb{1}}, \hat{\sigma}_i\}$, where the $\hat{\sigma}_i$'s are the Pauli matrices. In this case, the error operators are Hermitian, and square to the identity operator, and we assume this property for convenience throughout the following analysis.

In general the initial state can be expressed as

$$|\psi_i\rangle = \sum_{k=1}^{2q} c_k |C^k\rangle |R\rangle \tag{5.1}$$

where the $|C^k\rangle$ are the code–words for the states $|k\rangle$ and $|R\rangle$ is the initial state of the environment. The state after a general error is then a superposition of all possible errors acting on the above initial state

$$|\psi_f\rangle = \sum_{\alpha\beta} \hat{A}_\alpha \hat{P}_\beta \sum_k c_k |C^k\rangle |R_{\alpha,\beta}\rangle \,, \tag{5.2}$$

where $|R_{\alpha,\beta}\rangle$ is the state of the environment. (We keep the hat notation to designate

operators in this section in order to avoid any confusion.) Note that $|R_{\alpha,\beta}\rangle$ depends only on the nature of the errors, and is *independent* of the code–words [132]. The above is, in general, not in the Schmidt form, i.e. the code–word states after the error are not necessarily orthogonal (to be shown) and neither are the states of the environment. Now, since we have no information about the environment, we must trace it out using an orthogonal basis for the environment $\{|R_n\rangle, n = 1, d\}$. The resulting state is a mixture of the form $\hat{\eta}_i = \sum_n |\psi_n\rangle\langle\psi_n|$, where

$$|\psi_n\rangle = \sum_{\alpha\beta} x_n^{\alpha\beta} \hat{A}_\alpha \hat{P}_\beta \sum_k c_k |C^k\rangle \ , \tag{5.3}$$

and $x_n^{\alpha\beta} = \langle R_n | R_{\alpha\beta}\rangle$. To detect an error, one then performs a measurement on the state $\hat{\eta}$ to determine whether it has an overlap with one of the following subspaces

$$\mathcal{H}_{\alpha\beta} = \{\hat{A}_\alpha \hat{P}_\beta |C^k\rangle, k = 1, \ldots, 2^q\} \ . \tag{5.4}$$

The initial space after the error is given by the direct sum of all the above subspaces, $\mathcal{H} = \sum_{\alpha\beta} \oplus \mathcal{H}_{\alpha\beta}$. Each time we perform an overlap and obtain a zero result, the state space $\mathcal{H}$ reduces in dimension, eliminating that subspace as containing the state after the error. Eventually, one of these overlap measurements will give a positive result which is mathematically equivalent to projecting on to the corresponding subspace. The state after this projection is then given by the mixture $\hat{\eta}_f = \sum_n |\psi_{nProj_{\alpha\beta}}\rangle\langle\psi_{nProj_{\alpha\beta}}|$, where

$$|\psi_{nProj_{\alpha\beta}}\rangle = \sum_{kl} \sum_{\gamma\delta} x_n^{\gamma\delta} \hat{A}_\alpha \hat{P}_\beta |C^k\rangle\langle C^k | \hat{P}_\beta \hat{A}_\alpha \hat{A}_\gamma \hat{P}_\delta | C^l\rangle c_l \ . \tag{5.5}$$

The successful projection will effectively take us to the state generated by a superposition of certain types of error. One might expect that to distinguish between various errors the different subspaces $\mathcal{H}_{\alpha\beta}$ would have to be orthogonal. However, we will show that this is not, in fact, necessary.

After having projected onto the subspace $\mathcal{H}_{\alpha\beta}$ we now have to correct the corresponding error by applying the operator $\hat{P}_\beta \hat{A}_\alpha$ onto $|\psi_{Proj_{\alpha\beta}}\rangle$, since $\hat{P}_\beta \hat{A}_\alpha \hat{A}_\alpha \hat{P}_\beta = \hat{\mathbb{1}}$.

In order to correct the error successfully, the resulting state has to be proportional to the initial state of code–words in $|\psi_i\rangle$. This leads to the condition

$$\sum_{kl}\sum_{\gamma\delta} x_n^{\gamma\delta}|C^k\rangle\langle C^k|\hat{P}_\beta\hat{A}_\alpha\hat{A}_\gamma\hat{P}_\delta|C^l\rangle c_l = z^{\alpha\beta n}\sum_m c_m|C^m\rangle \ . \tag{5.6}$$

where $z^{\alpha\beta n}$ is an arbitrary complex number. Now we use the fact that all code words are mutually orthogonal, *i.e.* $\langle C^k|C^l\rangle = \delta_{kl}$, to obtain that

$$\sum_{l}\sum_{\gamma\delta} c_l x_n^{\gamma\delta}\langle C^k|\hat{P}_\beta\hat{A}_\alpha\hat{A}_\gamma\hat{P}_\delta|C^l\rangle = z^{\alpha\beta n}c_k \tag{5.7}$$

for all k and arbitrary c_k. This can be written in matrix form as

$$\mathbf{F}^{\alpha\beta n}\mathbf{c} = z^{\alpha\beta n}\mathbf{c} \ , \tag{5.8}$$

where the elements of the matrix $\mathbf{F}$ are given by

$$F_{kl}^{\alpha\beta n} := \sum_{\gamma\delta} x_n^{\gamma\delta}\langle C^k|\hat{P}_\beta\hat{A}_\alpha\hat{A}_\gamma\hat{P}_\delta|C^l\rangle \ . \tag{5.9}$$

As eq. (5.8) is valid for all $\mathbf{c}$ it follows that

$$\forall \ k,l, \qquad F_{kl}^{\alpha\beta n} = z^{\alpha\beta n}\delta_{kl} \ . \tag{5.10}$$

However, we do not know the form of $x_n^{\gamma\delta}$'s as we have no information about the state of the environment. Therefore, for the above to be satisfied for *any* form of x's we need each individual term in eq. (5.9) to satisfy

$$\langle C^k|\hat{P}_\beta\hat{A}_\alpha\hat{A}_\gamma\hat{P}_\delta|C^l\rangle = y^{\alpha\beta\gamma\delta}\delta_{kl} \tag{5.11}$$

where $y^{\alpha\beta\gamma\delta}$ is *any* complex number. From eqs. (5.9,5.10,5.11) we see that the numbers x, y, and z are related through

$$\sum_{\gamma\delta} x_n^{\gamma\delta}y^{\alpha\beta\gamma\delta} = z^{\alpha\beta n} \ . \tag{5.12}$$

Eq. (5.11) is the main result in this section, and gives a general, and in fact the *only*, constraint on the construction of code–words, which may then be used

for encoding purposes. If we wish to correct for up to d errors, we have to impose a further constraint on the subscripts $\alpha, \beta, \gamma,$ and δ; namely, $\mathrm{wt}(\mathrm{supp}(\alpha) \cup \mathrm{supp}(\beta))$, $\mathrm{wt}(\mathrm{supp}(\gamma) \cup \mathrm{supp}(\delta)) \leq d$, where $\mathrm{supp}(x)$ denotes the set of locations where the n–tuple x is different from zero and $\mathrm{wt}(x)$ is the Hamming weight [36], *i.e.* the number of digits in x different from zero. This constraint on the indices of errors simply ensures that they do not contain more than d logical '1's altogether, which is, in fact, equivalent to no more than d errors occurring during the process.

We emphasise that these conditions are the most general possible, and they in particular generalise the conditions given by Ekert and Macchiavello in [132]. By substituting $z^{\alpha\beta\gamma\delta} = \delta_{\alpha\beta}\delta_{\gamma\delta}$ in eq. (5.11), we obtain the conditions

$$\langle C^k | \hat{P}_\beta \hat{A}_\alpha \hat{A}_\gamma \hat{P}_\delta | C^l \rangle = \delta_{\beta\delta}\delta_{\alpha\gamma}\delta_{kl} \tag{5.13}$$

given in [132]. These are therefore seen only as a special case of the general result in eq. (5.10). These generalized conditions (but not the Ekert–Macchiavello conditions) show the main difference between a quantum and classical error correction: it is possible for two different errors to lead to the *same* state providing that the overlap is the same for all the code–words.

Knill and Laflamme, and Bennett et al., who arrive at the same conditions as in eq. (5.11) [24, 100], give no example of a code that violates the conditions in eq. (5.13) but satisfies those of eq. (5.11). One such code, which I introduced with my co-workers in [143], will be presented in the next section (but c.f. Shor [22]). This code violates the conditions given in eq. (5.13), thereby explicitly showing that they are *not* necessary, but merely sufficient.

5.3 Error Correction in the Presence of Spontaneous Emission

We will now introduce a code that protects qubits in the presence of spontaneous emission [143]. This code is interesting for two reasons. Firstly, it corrects a non-

unitary part of the evolution (described below) to all orders, and secondly it violates the more restrictive conditions of error correction by Ekert and Macciavello in eq. (5.13). Before we present the code let us introduce mathematical formalism for treating spontaneous emission.

5.3.1 Spontaneous Emission Dynamics

In general an open system evolves according to the so called Master equation, which preserves the positivity of the density matrix. The most general form of evolution is described in the following theorem.

Theorem (Lindblad [144]). A bounded operator $\mathcal{L}$ on the set of states $\mathcal{T}(\mathcal{H})$ is the generator of norm-continuous completely positive dynamical semigroup $\{\Lambda_t : \mathcal{T}(\mathcal{H}) \to \mathcal{T}(\mathcal{H})\}$ iff it has the form

$$\mathcal{L}\rho = -i[H,\rho] + \frac{1}{2}\sum_j \left\{[V_j, \rho V_j^\dagger] + [V_j\rho, V_j^\dagger]\right\} \tag{5.14}$$

$$= -i[H,\rho] + \sum_j V_j \rho V_j^\dagger - \frac{1}{2}[\sum_j V_j^\dagger V_j, \rho] \; . \tag{5.15}$$

The most general dynamical evolution of a quantum system now looks like

$$\frac{\partial \rho}{\partial t} = \mathcal{L}\rho \; . \tag{5.16}$$

Loosely speaking, this is a continuous version of the POVM formalism introduced in Chapter 3.

In the case of spontaneous emission, it turns out that there is only one of the V operators present. We can write $V = \sqrt{\gamma}\sigma_- = 1/2\sqrt{\gamma}(\sigma_1 - i\sigma_2)$, where σ_- is the Pauli lowering operator and γ the spontaneous emission rate. Thus the evolution of the atomic density ρ in the presence of the spontaneous emission can be written as:

$$\frac{\partial \rho}{\partial t} = -i[H,\rho] + \gamma\sigma_-\rho\sigma_+ - \frac{\gamma}{2}[\sigma_+\sigma_-, \rho] \tag{5.17}$$

where the Pauli raising operator $\sigma_+ = 1/2(\sigma_1 + i\sigma_2)$ is the hermitian conjugate of σ_- and H is the two–level system Hamiltonian.

There is an interesting way of treating this equation which is mathematically very convenient for the implementation of error correction to protect against the spontaneous emission. It is usually referred to as the quantum jump approach, or for an alternative, the quantum state diffusion approach [145]. We assume that the atom is at some time t in a pure state $\rho = |\Psi(t)\rangle\langle\Psi(t)|$. Now, in the small interval Δt there are two possibilities: either the atom emits a photon or it does not emit a photon (small Δt means that no more than one emission occurs). To propagate the conditioned state vector from time t to $(t + \Delta t)$ we need to calculate the current probability of emission, given by

$$\Delta p = \gamma \Delta t \langle \Psi(t)|\hat{\sigma}_+\hat{\sigma}_-|\Psi(t)\rangle \,. \tag{5.18}$$

If the emission happens, the new state afterwards will be given by:

$$|\Psi\rangle \longrightarrow N^{-1}\hat{\sigma}_-|\Psi\rangle \tag{5.19}$$

where N^{-1} is a normalization factor. If, on the other hand the emission does not happen we propagate $|\Psi\rangle$ under the influence of the non-Hermitian effective Hamiltonian. This can easily be derived from the requirement that the resulting evolution must be given by eq. (5.17). So, after a simple calculation we obtain

$$H_{\text{eff}} = H - i\hbar\frac{\gamma}{2}\hat{\sigma}_+\hat{\sigma}_- \tag{5.20}$$

where H is the unperturbed atomic Hamiltonian. Therefore when there is no emission the evolution proceeds according to

$$|\Psi\rangle \longrightarrow N^{-1}(1 - H_{\text{eff}}\Delta t/\hbar)|\Psi\rangle \,. \tag{5.21}$$

A single trajectory of $|\Psi(t)\rangle$ evolves in a smooth evolution under H_{eff}, interrupted by the jumps (emissions). When many trajectories are averaged, the traditional exponential decay law is recovered. It is curious to note that in order to compensate for the emission, the non-emission part is non-unitary (i.e. generated by a non-Hermitian effective Hamiltonian). We now turn to analysing single error correcting codes before showing how to error correct in the presence of spontaneous emission.

5.3.2 Single error correcting codes

Several codes have been proposed to encode one qubit which can correct one general error, i.e. amplitude and phase error or a combination of both applied to the same qubit. An example [131] of such a code is one where state $|0\rangle$ is represented by

$$\begin{aligned} |0_L\rangle &= |00000\rangle + |11100\rangle - |10011\rangle - |01111\rangle \\ &\quad +|11010\rangle + |00110\rangle + |01001\rangle + |10101\rangle \end{aligned} \tag{5.22}$$

and the state $|1\rangle$ by

$$\begin{aligned} |1_L\rangle &= |11111\rangle - |00011\rangle + |01100\rangle - |10000\rangle \\ &\quad -|00101\rangle + |11001\rangle + |10110\rangle - |01010\rangle \ , \end{aligned} \tag{5.23}$$

where the subscript L indicates that the encoded state $|i_L\rangle$ differs from the initial state $|i\rangle$. As usual, we omit the obvious normalization factor in the states $|0_L\rangle$ and $|1_L\rangle$ throughout, as they are irrelevant for the present analysis. We start with a state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, this is encoded as $|\psi_L\rangle = \alpha|0_L\rangle + \beta|1_L\rangle$. If the state suffers an amplitude error A_i (which acts as a NOT operation on qubit i) or a phase error P_i (which gives the upper state of qubit i a minus sign) or the combination A_iP_i of both to the ith qubit of $|\psi_L\rangle$ it is possible to reconstruct the initial state $|\psi\rangle$. The code given in eqs. (5.22) - (5.23) has the attractive feature that it is optimal in the sense that it only requires 5 qubits which can be shown to be the minimal possible number [24]. Using ideas similar to classical error correcting codes one can estimate that if one wants to encode l qubits in terms of n qubits in such a way that one can reconstruct the state after t general errors, then the inequality

$$2^l \sum_{i=0}^{t} 3^i \binom{n}{i} \leq 2^n \tag{5.24}$$

has to be satisfied [132]. The bound eq. (5.24) is related to the sphere packing bound in classical coding theory that we introduced in Chapter 2. The reason for this is that eq. (5.24) was obtained using the assumption that different errors

lead to different mutually orthogonal error syndromes as in eq. (5.13). The factor 3^i, which distinguishes this bound from its classical counterpart, comes from the fact that there are three basic errors (the Pauli spin operators). However, we will later see that the code presented here (like the one presented in [22]), violates this assumption which indicates that it may be possible to find codes that go beyond eq. (5.24).

The code given in eqs. (5.22) - (5.23) does not correct for multiple errors. In particular, it is not able to correct to all orders for errors that arise due to the conditional time evolution between spontaneous emissions. The conditional time evolution between spontaneous emissions is unavoidable and it differs from the unit operation because the fact that no spontaneous emission has taken place provides information about the state of the system and therefore changes its wave function. The conditional time evolution of the system under the assumption that no spontaneous emission has taken place is given by the non-unitary time evolution operator $\exp\{-i\,H_{\rm eff}\,t/\hbar\}$ [141] introduced earlier. For the case that the qubits are not driven by external fields we obtain for the code given in eqs. (5.22) - (5.23) the effective Hamilton operator

$$H_{\rm eff} = \sum_{i=1}^{5} -i\,\hbar\Gamma\sigma_{11}^{(i)} \ , \tag{5.25}$$

where $\sigma_{11}^{(i)}$ is the projector $|1\rangle\langle 1|$ onto the excited state of the ith qubit leaving all other qubits unaffected. 2Γ is the Einstein coefficient of the upper level 1 of the qubits. If we apply the conditional time evolution $\exp(-i\,H_{\rm eff}t/\hbar)$ to the encoded state

$$|\psi_L\rangle = \alpha|0_L\rangle + \beta|1_L\rangle \ , \tag{5.26}$$

and subsequently apply the appropriate error correction procedure for this 5-bit code [131] we do *not* recover the original state. This becomes obvious in the special case $\Gamma t \gg 1$ in which one obtains

$$\begin{aligned} |\psi_C\rangle \ &= \ |00000\rangle + |00010\rangle + |01000\rangle - |01110\rangle \\ &\quad +|10000\rangle + |10010\rangle + |11000\rangle + |11110\rangle \ . \end{aligned} \tag{5.27}$$

This shows that this 5-bit code is not able to correct errors due to the conditional time evolution exactly. Especially striking is the effect when we assume that $\beta = 0$, i.e. we encode the (stable) ground state. The conditional time evolution then leads to no errors in the unencoded state while it changes the encoded state such that it cannot be corrected perfectly anymore. Note, however, that the error introduced by the conditional time evolution is, for short times, of fourth order. If, however, a spontaneous emission (or any other kind of error) occurs then a subsequent conditional time evolution induces contributions which after error correction lead to second order errors in the state. The code presented later in this section preserves the encoded state in both cases perfectly, i.e. to all orders.

The reason that the code eqs. (5.22)-(5.23) cannot perfectly correct errors due to the conditional time evolution derives from the fact that the words (product states) of which the code consists do not all have the same number of excited states. This leads to a difference in the rate at which the amplitude of these states decays. The amplitude of $|00000\rangle$ remains unchanged under the conditional time evolution while the amplitude of $|11100\rangle$ for example decreases at a rate $\exp(-3\Gamma t)$. This can be seen as a multiple amplitude error with which the code cannot cope. This problem is not restricted to the 5-bit code given in [131] but is present in all other previously proposed codes. It should be noted that it is *not* necessary to observe the system for these conclusions to hold. If we do not observe the system, it then has to be described by a density operator, whose time evolution follows the appropriate Bloch equations. This time evolution can in principle be decomposed into individual trajectories, each of which consists of no–jump evolutions interrupted by spontaneous emissions [141]. For each of these trajectories our considerations above hold and therefore also hold for the incoherent sum of these trajectories which make up the ensemble. Therefore our error correction code is not restricted to a particular measurement scheme such as for example the detection and reconstruction scheme discussed in [146], where it is necessary to detect individual quantum jumps. Nevertheless such a detection of individual jumps would improve the performance of

the code, as that would exclude the contribution of multiple quantum jumps with which our code cannot cope. This would enhance the importance of the conditional time evolution as a error source compared to other sources and it is here where our code is superior to previous codes.

5.3.3 Correcting spontaneous emission

The discussion of the last section shows that it is of some interest to construct a quantum error correcting code that corrects errors due to the conditional time evolution to all orders. This is possible, and in the following we present such a quantum error correcting code.

This code was constructed starting from the code (5.22)-(5.23). State $|0\rangle$ is encoded as

$$\begin{aligned} |0_L\rangle &= |00001111\rangle + |11101000\rangle - |10010110\rangle - |01110001\rangle \\ &+ |11010100\rangle + |00110011\rangle + |01001101\rangle + |10101010\rangle \ , \end{aligned} \tag{5.28}$$

while state $|1\rangle$ is encoded as

$$\begin{aligned} |1_L\rangle &= |11110000\rangle - |00010111\rangle + |01101001\rangle - |10001110\rangle \\ &- |00101011\rangle + |11001100\rangle + |10110010\rangle - |01010101\rangle \ . \end{aligned} \tag{5.29}$$

The state eq. (5.28) encoding the logical 0 was obtained in the following way. We start with state eq. (5.22) and for each word, e.g. $|11100\rangle$ we construct its bitwise inverse, i.e. $|00011\rangle$. We concatenate the two words where the second one is taken in reverse bit order to obtain $|1110011000\rangle$. This method, when applied to all words in eq. (5.22), already yields a possible code. However, it is possible to shorten the code by removing bits 5 and 6 from every word. This then yields eq. (5.28) and analogously eq. (5.29). A computer search was performed by Dr. M. B. Plenio to search for potentially shorter codes; this revealed no such codes, so we conclude that $n = 8$ qubits is the minimum number required for the task of

correcting one general error while errors due to the conditional time evolution are corrected perfectly. In the following we present some interesting properties of the code and demonstrate that it indeed has the claimed error correction properties. However, this code differs in many ways from previously proposed codes. First of all, it violates the conditions given for quantum error correcting codes in eq. (5.13) thereby explicitly showing that these conditions are overly restrictive. As these conditions were used to derive the inequality in eq. (5.24), their violation indicates that there might exist codes that require less qubits than expected from eq. (5.24). However, we did not yet succeed in constructing such a code. One should also realize that the code–words in the code eqs. (5.28)-(5.29) do *not* form a linear code as this would imply that $|00000000\rangle$ is a code–word which in turn would render impossible the task of constructing a code with code–words of equal excitation. Nevertheless, the code–words of $|0_L\rangle$ form a coset of a linear code. The coset leader is $|00001111\rangle$. This contrasts slightly with other codes such as those presented in [20, 21, 22, 131]. The code–words of the code (5.22)-(5.23) for example form a linear code. Given the initial state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, we obtain the code eqs. (5.28)-(5.29) using the network given in Fig. 5.1. To correct the error that may have appeared we first apply the encoder in the reverse direction (right to left). After the application of the decoder, the third qubit contains information about the encoded state while the remaining 7 qubits contain the error syndrome, from which one can infer the type and location of the error. We measure the qubits of the error syndrome and apply, according to the result of our measurement, a suitable unitary operation on qubit 3. We assume that after the measurement all the other qubits are reset to their ground state $|0\rangle$ so that, in principle, we can re-encode the state again using the same qubits.

In table 5.1 we give all possible outcomes of the measurement and the corresponding state of the third qubit. The necessary unitary transformation that has to be applied onto the third qubit is then obvious. Careful inspection of table 5.1 reveals that this error correction scheme has, for some errors, a slightly different

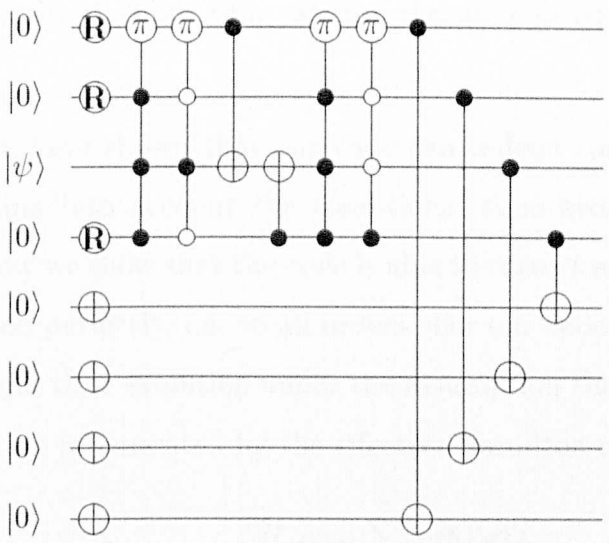


Figure 5.1: The encoding network: R describes a one bit rotation which takes $|0\rangle \to (|0\rangle + |1\rangle)/\sqrt{2}$ and $|1\rangle \to (|0\rangle - |1\rangle)/\sqrt{2}$. An encircled cross denotes a NOT operation while a dot denotes a control bit. For a filled circle the operation is carried out if the control bit is 1; for an empty circle the operation is carried out if the control bit is 0. A circle with a π represents multiplication with phase $\exp(i\pi)$. Qubit 3 is in the state $|\psi\rangle$ that we wish to encode, while all other qubits are initially in their ground state $|0\rangle$.

effect than expected. Take for example a phase errors P_1 on bit 1 and compare with the effect of a phase error P_8 on bit 8. We observe that they both lead to the same error syndrome but that the resulting state differs by a global phase -1. Therefore it is not possible to correct both states in such a way that they go over to the initial state. After the correction they differ by a global phase -1. But this also shows that the dimension of the space $\mathcal{H}_{code}$ spanned by the code together with all states that result from it by single errors is 2×21 and not as expected from eq. (5.24) 2×25. The violation of these conditions by the code eq. (5.28)-(5.29) leads to these different predictions for the dimension of $\mathcal{H}_{code}$. On the other hand it can

be checked easily that our code satisfies the more general conditions we derived in eq. (5.11).

So far we have shown that our code can indeed correct a general single error without taking into account the conditional time evolution due to spontaneous emission. Now we show that the code is able to correct errors due to the conditional time evolution perfectly, i.e. to all orders. For our code given in eqs. (5.28)-(5.29) the conditional time evolution under the assumption that no spontaneous emission has taken place is generated by the effective Hamilton operator

$$H_{eff} = \sum_{i=1}^{8} -i\hbar\Gamma\sigma_{11}^{(i)} \ . \tag{5.30}$$

If the code undergoes a conditional time evolution before it experiences an error like e.g. a spontaneous emission, it is obvious that the code eqs. (5.28)-(5.29) will work properly, as it is invariant under the conditional time evolution $\exp(-iH_{eff}t/\hbar)$. However, it is not so obvious that the code corrects general single errors that occur *before* or in between the conditional time evolution. As we do not know the time at which the general error occurs, this situation will almost certainly occur and has to be examined. If the error was a phase error, then no problem will occur, as this error does not change the excitation of the state. However, for amplitude errors or a combination of amplitude and phase errors we have to investigate the code more closely. The problem is that, for example after an amplitude error in the first qubit, we obtain

$$\begin{aligned} A_1|0_L\rangle \ = \ & \\ & |10001111\rangle - |11110001\rangle + |11001101\rangle + |10110011\rangle \\ & +|0110100\rangle - |00010110\rangle + |01010100\rangle + |00101010\rangle \ . \end{aligned} \tag{5.31}$$

Now the code words have a different degree of excitation so that their relative weights will change during the subsequent conditional time evolution. However, for $|\psi_L\rangle = \alpha|0_L\rangle + \beta|1_L\rangle$ we have the relations

Error	Error syndrome	State of qubit 3
None	0000000	$\alpha\|0\rangle + \beta\|1\rangle$
P_1	1000000	$\alpha\|0\rangle + \beta\|1\rangle$
P_2	0100000	$\alpha\|0\rangle + \beta\|1\rangle$
P_4	0010000	$\alpha\|0\rangle + \beta\|1\rangle$
A_5	0001000	$\alpha\|0\rangle + \beta\|1\rangle$
A_6	0000100	$\alpha\|0\rangle + \beta\|1\rangle$
A_7	0000010	$\alpha\|0\rangle + \beta\|1\rangle$
A_8	0000001	$\alpha\|0\rangle + \beta\|1\rangle$
P_3	1010000	$\alpha\|0\rangle - \beta\|1\rangle$
A_2	0010010	$\alpha\|0\rangle - \beta\|1\rangle$
P_6	1010000	$-\alpha\|0\rangle + \beta\|1\rangle$
A_2P_2	0110010	$-\alpha\|0\rangle + \beta\|1\rangle$
A_6P_6	1010100	$-\alpha\|0\rangle + \beta\|1\rangle$
P_5	0010000	$-\alpha\|0\rangle - \beta\|1\rangle$
P_7	0100000	$-\alpha\|0\rangle - \beta\|1\rangle$
P_8	1000000	$-\alpha\|0\rangle - \beta\|1\rangle$
A_5P_5	0011000	$-\alpha\|0\rangle - \beta\|1\rangle$
A_7P_7	0100010	$-\alpha\|0\rangle - \beta\|1\rangle$
A_8P_8	1000001	$-\alpha\|0\rangle - \beta\|1\rangle$
A_1P_1	1110001	$\beta\|0\rangle + \alpha\|1\rangle$
A_4P_4	1011000	$\beta\|0\rangle + \alpha\|1\rangle$
A_3P_3	1110100	$\beta\|0\rangle - \alpha\|1\rangle$
A_1	0110001	$-\beta\|0\rangle - \alpha\|1\rangle$
A_3	0100100	$-\beta\|0\rangle - \alpha\|1\rangle$
A_4	1001000	$-\beta\|0\rangle - \alpha\|1\rangle$

Table 5.1: We obtain an error syndrome, i.e. the state of all qubits except qubit 3, depending on the error that occurred and the place in which it occurred. P_i indicates a sign change of the upper level of qubit i, A_i an amplitude error which is given by the transformation $|0\rangle \leftrightarrow |1\rangle$. The product of both applied to the same qubit gives the third kind of error. Note that the error syndrome is not able to distinguish between P_i and P_{9-i} which leads to global phases in some of the corrected states. This table does not take into account that before and after the error a conditional time evolution takes place.

$$e^{-iH_{eff}t/\hbar}A_i|\psi_L\rangle = \frac{1}{2}e^{-3\Gamma t}\left\{(1+e^{-2\Gamma t})A_i - (1-e^{-2\Gamma t})A_iP_i\right\}|\psi_L\rangle \tag{5.32}$$

and

$$e^{-iH_{eff}t/\hbar}A_iP_i|\psi_L\rangle = \frac{1}{2}e^{-3\Gamma t}\left\{-(1-e^{-2\Gamma t})A_i + (1+e^{-2\Gamma t})A_iP_i\right\}|\psi_L\rangle \ . \tag{5.33}$$

Eq. (5.32) shows that after an amplitude error A_i on the ith qubit, the conditional time evolution transforms the state into a superposition of a state without conditional time evolution after this amplitude error, and a state without conditional time evolution obtained after a combined amplitude and phase error A_iP_i on the ith qubit. Inspecting table 1 we see that both errors A_i and A_iP_i lead to a different error syndrome. A measurement of the syndrome will then indicate one or the other error, A_i or A_iP_i, which can then be corrected. Therefore the code (5.28)-(5.29) corrects properly even if the error is followed by a conditional time evolution.

We now briefly point out another desired feature of our code. Namely, this code is insensitive to the detuning of the laser used to perform the actual quantum computation in the ion trap computer. When the laser is detuned from the $|0\rangle \rightarrow |1\rangle$ transitional frequency, then the excited atomic state gains an additional phase factor, so that the state $|\Psi(0)\rangle = a|0\rangle + b|1\rangle$ would freely evolve into the state

$$|\Psi(t)\rangle = a|0\rangle + be^{-i\Delta t}|1\rangle \ , \tag{5.34}$$

where Δ is the detuning. This, of course, is detrimental to the error correction capabilities of any code since the decoding procedure might now wrongly interpret this phase change as a phase error. However, if the code is constructed in such a way that the number of excitations is the same in each word, then this overall phase will simply factor out and not affect the calculations in any significant way.

5.4 Reliable Quantum Computation from Unreliable Components

We have seen how to protect qubits against general error, and in particular how to protect an atom against spontaneous emission. However, this protection is rather "static", i.e. our qubits are not evolving while errors occur. Suppose we would like to implement a Controlled-NOT between two qubits which can undergo an error during this operation. Is there a point to encoding these qubits in the first place, since the encoding and decoding procedures are just composed of a number of CNOTs (and other gates) which themselves can undergo errors? It appears that if we (*realistically*) allow encoding and decoding to undergo errors then there is no point is protecting gates since this action introduces even more errors. The conclusion would be that quantum error correction cannot be used in quantum computation! The same conclusion was reached in 1930s about classical computation. Then, however, von Neumann [147] showed this to be completely a erroneous conclusion and he proved that a reliable computation (classical of course, as von Neumann did not know about quantum computation) is possible from unreliable components. His argument can directly be translated into quantum computing and this gives rise to the fault tolerant quantum computation, i.e., in von Neumann's jargon, reliable quantum computation from unreliable components. We now present a sketch of this argument. This is intended only as a qualitative argument that the quantum error correction we have studied in this chapter can be applied to quantum computing in general, and no details will be given.

The idea of fault tolerant quantum computation [135] is to encode the qubits in such a way that the encoding does not introduce more errors than previously present. If the error stays at the same level we then keep performing error correction until the error has decreased in magnitude [135, 137, 148]. The present state of the art requires $5 - 10$ qubits to encode a single qubit against a single error. It is the iterative application "in depth" of the encoding that will enable us to reduce error

to an arbitrarily small level providing it is below a certain level to start with. In other words we will be encoding the encoding bits. Before we give more details let us just recapitulate the main points about a quantum computer.

An input to a quantum computer is a string of qubits. For this calculation a quantum computer is viewed as consisting of two main parts: *quantum gates* and *quantum wires*. By *basic* quantum gates we mean any set of quantum gates which can perform any desired quantum computation. A universal quantum gate is the one whose combination can be used to simulate any other quantum gate. A quantum wire is used as a representation of that part of computation of any qubit where the evolution is a simple identity operation (i.e. no gate operates on the qubit), as well as the time the qubit spends during the gate operation.

For stable quantum computation, obviously, we require that the probability of error after the fault–tolerantly encoded basic gate is of higher order (i.e. the error is smaller) than the probability of error after the unencoded gate (that is the whole point of encoding and fault–tolerant error correction!). From this we derive the bound on the size of allowed errors in the wires and in the gates. When we encode the encoding bits again, we reduce the error further and can reduce the error arbitrarily for an arbitrarily long computation. Therefore given certain initial limits on the error rate in the gates and wires we can stabilize any computation to a desirably small error rate, given an unlimited amount of time. Consider a two input two output quantum gate. The probability of having any of the three basic errors in the first as well as in the second wire is η–giving the overall first order wire error of 2η. The error in the gate itself is ϵ. We assume that the overall error of the whole basic gate is $\leq 2\eta + \epsilon$. Suppose that the basic gate is now encoded fault tolerantly against a single error of any kind, using l qubits. Then the overall second order error is at the end of the gate:

$$\eta^*(\eta, \epsilon, l) = \left(1 - \frac{l(l-1)}{2} l^4 \eta^2\right) l^2 \epsilon + \frac{l(l-1)}{2} l^4 \eta^2 (1 - l^2 \epsilon) \ , \qquad (5.35)$$

i.e. equal to having error in the wires (this time in second order) and not in the

gates plus having error in the gates and not in the wires. The term $l(l-1)/2$ comes from choosing two out of l wires to err and the factor l^4 derives from the use of l^2 gates, so that the error is transformed according to $\eta \to l^2\eta$ and is of second order. We require that the fault tolerant error correction reduces the error. Hence:

$$\left(1 - \frac{l(l-1)}{2} l^4\eta^2\right) l^2\epsilon + \frac{l(l-1)}{2} l^4\eta^2(1 - l^2\epsilon) \leq 2\eta + \epsilon \ . \tag{5.36}$$

As the RHS is $> \eta$, we simplify the above without a greater loss in generality to:

$$\left(1 - \frac{l(l-1)}{2} l^4\eta^2\right) l^2\epsilon + \frac{l(l-1)}{2} l^4\eta^2(1 - l^2\epsilon) \leq \eta \ , \tag{5.37}$$

The solutions to the equation derived from the above are:

$$\eta_{\pm} = \frac{1 \pm \sqrt{1 - 2(l^8\epsilon - 2l^{10}\epsilon^2)}}{(l^6 - 2l^8\epsilon)} \ , \tag{5.38}$$

We require that $\eta \in \mathcal{R}$ (and that $0 \leq \eta \leq 1/2$) so that we have the following two regimes of error

1. $0 < \eta < \eta_+$ and $\epsilon \leq \epsilon_-$.

2. $0 < \eta < \eta_-$ and $\epsilon \geq \epsilon_+$.

where $\epsilon_{\pm} = \frac{1}{2l^2}(1 \pm \sqrt{1 - 2l^{-6}})$. The output of the first encoded basic gate is fed into the next one (or part of the output into one next basic gate and the rest into another next basic gate). It is evident that if condition 1 holds, further encoding can only decrease the error. The residual error not taken into account is $\sim l^3(l^2\eta)^3 = l^9\eta^3$ (i.e. the second order error is not corrected by our encoding). In the worst case when $\epsilon = \epsilon_- \sim l^{-8}$ we get $\eta \sim l^{-6}$, which means that the residual uncorrected error is $\sim l^{-9}$. This error can accumulate over time if the computation is sufficiently long. However the residual error after n in depth encodings is $l^{-\mathcal{O}(n)}$, which can made be arbitrarily small using sufficiently large n. Therefore if the initial error per gate is sufficiently small, these gates can be used to perform arbitrary large quantum computations. If we need $l = 10$ qubits to fault-tolerantly encode one qubit, then the tolerant error rate is 10^{-6} which a more careful analysis shows to be correct [149].

5.5 Conclusions

We have shown how to protect quantum information against any general error. We have then applied our formalism to protecting a two level atom against the spontaneous emission. We concluded that our code is able to correct a single general error *and* in addition errors due to the conditional time evolution to *arbitrary* order. It is the first code proposed so far that can correct a general error to first order *and* a special kind of error to all orders. This is an interesting result as it shows that it is possible to correct special kinds of errors to all orders. As some errors are more frequent than others it would be in our interest to correct those errors to higher order than less frequently occurring errors. We have adapted our code to correct errors due to the conditional time evolution between spontaneous emissions. Other applications will require different adoptions. The code presented here (similar to the one given in [22]) violates the conditions for quantum codes given in eq. (5.13) which shows that these conditions are overly restrictive, as they exclude codes like the one presented here that map different errors onto the same error syndromes. This can lead to the construction of shorter quantum error correction codes than expected from the quantum sphere packing bound. We than concluded that the code for correcting for the spontaneous emission would also correct for any existing laser detuning in driving the atoms to implement quantum computations. These results may become important in different fields such as quantum computation, the distribution of entangled particles and in quantum cryptography [150, 151, 152, 153]. We also presented a heuristic argument for using error correction in quantum computing, which might play an important role in building the first actual quantum computer.

Chapter 6

Cavity QED Implementations of Purification Procedures

6.1 Introduction

In Chapters 3 and 4 we analysed purification procedures, and their efficiency limits. Chapter 5 was then devoted to exploring the idea of quantum error correction aiming at protecting quantum information in noisy environments. We now look at practical realisations of purification procedures and also examine the connections between the two (c.f. [100]). We recall that purification procedures are based on Gisin's original proposal [67] described in Chapter 3 using 'local filters' to increase correlations between two entangled quantum subsystems. Following this a number of other schemes have been designed for the purpose of local purification [19]. All of these have one idea in common: they all rely on some form of classical communication on which subsequent *post-selection* is based. This means that if we start with an ensemble of N pairs of particles in a mixed state, the final pure state will invariably have fewer particles. This was seen in Chapter 4 as a consequence of the fact that local operations (i.e. generalised filters) *cannot* increase quantum correlations. We now show that although the increase in correlations cannot be achieved, an

error correction procedure can always be applied locally, which will maintain the entanglement.

We introduce the Jaynes-Cummings model in section 6.2 [96, 154]. The cavity QED implementation of purification procedures will then be described using this model. In section 6.3 we present a simple model of atoms interacting 'locally' with two entangled cavities and give a number of feedback schemes by which the correlations *might* possibly be increased, without using any classical communication and post-selection. We show that each of these schemes fails, and we link this to the impossibility of superluminal propagation of any signal. At the end of this section we briefly show how non–local interactions can easily be used to increase correlations. Using the error correcting methods of previous chapter we then present in section 4 a simple example of how to encode two cavities against a single amplitude error on either cavity using four atoms.

6.2 Jaynes-Cummings Model

The Jaynes-Cummings model (JCM) is a fundamental model widely used in quantum optics [154]. Although at first sight is appears to be very simple, the model has been studied for more than thirty years and new, exciting and surprising discoveries are still being made. The JCM is the first fully quantized model of the interaction between a two level atom (or indeed any other two level system) and a quantized, monochromatic EM field. In order to present this model, we shall first briefly describe the process of quantization of EM field.

6.2.1 Quantization of EM Field

Let us imagine a one dimensional cavity with perfectly reflecting mirrors, placed at $z = 0$ and $z = L$, filled with a monochromatic EM field. In order to satisfy the imposed boundary conditions, i.e. the electric field has to vanish at the mirrors,

whereas the magnetic field strength reaches its maximum, we must have:

$$E_x(t) = \sqrt{\frac{2\omega^2}{\epsilon_0 L}} q(t) \sin kz \tag{6.1}$$

$$H_y(t) = \frac{\epsilon_0}{k} \sqrt{\frac{2\omega^2}{\epsilon_0 L}} \dot{q}(t) \cos kz \tag{6.2}$$

where $kL = n\pi$, n being an integer. Our choice of writing $E_x(t)$ and $H_y(t)$ in this particular form is immediately justified by observing the overall energy in the cavity, per frequency, which is given by:

$$W = \frac{1}{2} \int_0^L (\epsilon_0 E_x^2 + \mu_0 H_y^2) dz \tag{6.3}$$

Substituting the expressions for $E_x(t)$ and $H_y(t)$ and writing $p = \dot{q}$ we obtain:

$$W = \frac{1}{2}(p^2 + \omega q^2) \tag{6.4}$$

where we have used the fact that $c = 1/\sqrt{\mu_0 \epsilon_0}$ and $\omega = kc$. This is a familiar form of the energy of a unit mass harmonic oscillator at frequency ω. It strongly suggests that the EM field should be quantized via canonically conjugate variables q and p (a more rigorous analysis shows that this is a correct way of proceeding). It will, therefore, exibit all the properties of the quantized mechanical harmonic oscillator (QMHO). In particular, the energy will be quantized with the eigenvalues having the form

$$E_n = \hbar\omega(n + \frac{1}{2}), \quad n = 0, 1, 2, \ldots \tag{6.5}$$

Invoking raising and lowering operators, $\hat{a}^\dagger$ and $\hat{a}$ we can write:

$$q \to \hat{q} = \sqrt{\frac{\hbar}{2\omega}} (\hat{a} + \hat{a}^\dagger) \tag{6.6}$$

$$p \to \hat{p} = -i \sqrt{\frac{\hbar\omega}{2}} (\hat{a} - \hat{a}^\dagger) \tag{6.7}$$

Therefore, the quantized electric field, in a one dimensional cavity can be written using $\hat{a}^\dagger$ and $\hat{a}$ as

$$\hat{E}_x(t) = \sqrt{\frac{\hbar\omega}{\epsilon_0 L}} \vec{x} (\hat{a} + \hat{a}^\dagger) \sin kz \tag{6.8}$$

where $\vec{x}$ is a unit vector in the x direction, and $\sqrt{\frac{\hbar\omega}{\epsilon_0 L}}$ is the so called electric field per photon. We note in passing that since $\hat{q}$ and $\hat{p}$ obey the Heisenberg commutation relations, i.e. $[\hat{q}, \hat{p}] = -i$, so will the quantized E and H. Thus, similar uncertainty relations to the ones obeyed by $\hat{q}$ and $\hat{p}$ are also obeyed by the quantized E and H fields. In what follows we deal exclusively with quantum systems, so that we omit hats which are superfluous.

6.2.2 Spin–Boson Interaction Dynamics

The time development of an isolated quantum system is completely determined once we specify the corresponding Hamiltonian. Let us, therefore, represent the states of the atom by vectors $|g\rangle$ (ground) and $|e\rangle$ (excited). The field is at the same time represented by the energy eigenvectors of a QMHO, i.e. the number states $|n\rangle$. Invoking the well-known action of raising and lowering operators:

$$a^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle \tag{6.9}$$

$$a|n\rangle = \sqrt{n}|n-1\rangle \tag{6.10}$$

and taking the energy of the ground state to be zero, and that of the excited state to be $\hbar\omega_0$, we can write the Hamiltonian of the composite atom-field system as:

$$H = \hbar\omega_0|e\rangle\langle e| + a^\dagger a\hbar\omega + V \tag{6.11}$$

where we omitted the 'vacuum field' zero-point energy-term from the field. The interacting Hamiltonian, V, is given by the dipole-field interaction energy:

$$V = -e\vec{r}\vec{E} = -exE_x \tag{6.12}$$

where e is the electric unit charge and is not to be confused with the excited state, $|e\rangle$. Taking the Rotating Wave Approximation, i.e. ignoring the rapidly varying terms we get the following Hamiltonian:

$$H = \hbar\omega_0|e\rangle\langle e| + a^\dagger a\hbar\omega + \hbar\lambda(a^\dagger\sigma_- + a\sigma_+) \tag{6.13}$$

known as the Jaynes-Cummings Hamiltonian. Here

$$\lambda = -\frac{\langle e|ex|g\rangle\sqrt{\frac{\hbar\omega}{\epsilon_0 L}}\sin kz}{\hbar} . \tag{6.14}$$

Note that the interacting term has a very simple and natural interpretation: the first interaction term indicates that the atom is de-excited and the field receives this quantum of excitation, while the second term is the exact reverse. In general, after some time t, the atom and the field are in an entangled state of the form:

$$|\Psi(t)\rangle = \sum_{n=0}^{\infty}(a_n(t)|g\rangle \otimes |n\rangle + b_n(t)|e\rangle \otimes |n\rangle) . \tag{6.15}$$

Solving Schrödinger's equation we obtain

$$a_n(t) = ac_n \cos\lambda\sqrt{n}t - ibc_{n-1}\sin\lambda\sqrt{n}t \tag{6.16}$$

$$b_n(t) = bc_n \cos\lambda\sqrt{n+1}t - iac_{n+1}\sin\lambda\sqrt{n+1}t \tag{6.17}$$

where the atom and the field are initially in a disentangled state of the form

$$|\Psi(t=0)\rangle = (a|g\rangle + b|e\rangle) \otimes \sum_{n=0}^{\infty} c_n|n\rangle . \tag{6.18}$$

We will use the evolution equations eq. (6.17) in the following section. We now turn to describing atom-cavity models used to implement the local concentration of entanglement.

6.3 Atom–Cavity Models

We present here a simple model which aims to *increase* the amount of entanglement between two entangled subsystems. The model we present employs a technique of performing 'local' complete measurements. By this, we mean that when the two quantum systems are entangled we perform complete measurements on either subsystem separately, while not interacting directly with the other subsystem. We may regard this result to be counter-intuitive — it does not seem at first sight possible that purely local operations could increase the non-local quantum features. There

have been many schemes devised whereby correlations can be increased by local measurements on an ensemble of systems combined with *classical communication*, followed by a procedure of *post-selection*. Indeed, the model presented here can also be adapted readily to represent such a scheme. However, we verify that by local measurement alone, and without post-selection based upon classical communication, the correlations do not increase. This, in fact, verifies general results presented in Chapters 2,3 and 4.

The models used to demonstrate this are of the cavity QED type, and are both easy to understand physically and simple to analyse analytically. Our model is also easier to implement physically than the ones based on EPR–photon pairs [19], since it only requires a single pair of entangled cavities. A good outline of cavity QED is given in [155]. We consider two optical cavities, the field states of which are entangled number states (for simplicity)

$$|\Psi\rangle_{AB} = \alpha|n\rangle_A|m\rangle_B + \beta|n'\rangle_A|m'\rangle_B, \tag{6.19}$$

where the subscripts 'A' and 'B' refer to the two cavities, and without loss of generality, we assume that $|\alpha| > |\beta|$. This is a pure state but it is *not* maximally entangled. The aim is to produce the state:

$$|\Phi\rangle_{AB} = \frac{1}{\sqrt{2}}\left(|n\rangle_A|m\rangle_B + |n'\rangle_A|m'\rangle_B\right), \tag{6.20}$$

i.e. we have made $\alpha{=}\beta{=}\frac{1}{\sqrt{2}}$, which is maximally entangled.

Two-level atoms are sent, one at a time, through cavity A and interact with that individual cavity field, via the Jaynes-Cummings Hamiltonian [154], for a predetermined time period. After each atom passes through the cavity, a measurement is made which projects the atomic state into either the ground state or the excited state. Due to the entanglement developed between the atom and the field in cavity A during the interaction, this measurement also collapses the joint cavity A – cavity B field state into a different superposition, one with either the same number of photons in cavity A, or with one extra photon respectively. By successively sending

atoms through the cavity for interaction periods determined from the state of the previously measured atom, a *feedback* mechanism can be set up whereby one might expect to optimise the probability of achieving the state defined in eq. (6.20). Similar schemes have been used on single cavities for quantum state-engineering [156].

We also consider extensions to this procedure. Firstly, we mention procedures for interacting locally with both cavities, the qualitative results of which are the same. And secondly, we give two examples of non-local interactions, which give quite different results to the above local procedures.

6.3.1 Cavity Models With Local Feedback

The first model involves sending atoms through cavity A only, a schematic of which is given in Fig. 6.1; we assume the initial joint cavity field state is given by (6.19). The first atom is in the excited state, and so the initial atom-field state is

$$|\Psi_i\rangle = (\alpha|n\rangle_A|m\rangle_B + \beta|n'\rangle_A|m'\rangle_B) \otimes |e\rangle_A. \tag{6.21}$$

After interaction for a time t_1, determined from the atomic time of flight, the joint atom-field state becomes

$$\begin{aligned} |\Psi_f\rangle &= (\alpha a_n(t_1)|n\rangle_A|m\rangle_B + \beta a_{n'}(t_1)|n'\rangle_A|m'\rangle_B) \otimes |e\rangle_A \\ &+ (\alpha b_n(t_1)|n+1\rangle_A|m\rangle_B + \beta b_{n'}(t_1)|n'+1\rangle_A|m'\rangle_B) \otimes |g\rangle_A \end{aligned} \tag{6.22}$$

where the coefficients are given by $a_n(t_1) = \cos\left(\frac{R_n t_1}{2}\right)$, $a_{n'}(t_1) = \cos\left(\frac{R_{n'} t_1}{2}\right)$, $b_n(t_1) = -i\sin\left(\frac{R_n t_1}{2}\right)$, $b_{n'}(t_1) = -i\sin\left(\frac{R_{n'} t_1}{2}\right)$, and $R_i = 2\lambda\sqrt{i+1} = R_o\sqrt{i+1}$.

We now arrange that the velocity of the atom, and hence the interaction time with the field, is such that

$$\alpha a_n(t_1) = \beta a_{n'}(t_1)\,, \tag{6.23}$$

in which case the joint atom-field state becomes

$$|\Psi_f\rangle = \alpha a_n(t_1)\left(|n\rangle_A|m\rangle_B + |n'\rangle_A|m'\rangle_B\right) \otimes |e\rangle_A$$

$$+ \left(\alpha b_n(t_1)|n+1\rangle_A|m\rangle_B + \beta b_{n'}(t_1)|n'+1\rangle_A|m'\rangle_B\right) \otimes |g\rangle_A \quad . \tag{6.24}$$

From this we see that if we measure the atom in the excited state, the resulting cavity field state is maximally entangled. The probability of measuring the excited atomic state is

$$P_1(e) = 2|\alpha a_n(t_1)|^2 \, . \tag{6.25}$$

If we were to prepare a whole ensemble of cavities in precisely the same initial state in eq. (6.21), then after measurement on all of the ensemble members, we would have prepared approximately $(100 \times P_1(e))\%$ of the cavities in the maximally entangled state in eq. (6.20). We can discard all the cavities for which we measured the atom in the ground state, and we will have a whole sub-ensemble of cavities for which the entanglement has increased. This is the post-selection procedure mentioned earlier, and always requires that measurements on the whole ensemble be 'thrown away' in order to increase the entanglement of a sub-ensemble.

What we wish to do here is to increase the entanglement on an *individual pair* of entangled cavities. Instead of performing one measurement on an ensemble of cavities, we keep performing a number of measurements on this single pair until we achieve our aim. When the atom is measured in the excited state, we are there. If the outcome of the atomic state measurement was $|g\rangle_A$, the final cavity field state would be the corresponding field state in eq. (6.24), which is still entangled, but not maximally so. We can now use this field state as a *new* initial entangled cavity field. In this way, we would hope that it is just a matter of sending through 'enough' atoms until the desired state is reached.

Since the field state corresponding to a ground state measurement involves the $(n+1)$ Fock state, sending through another excited atom allows the possibility of

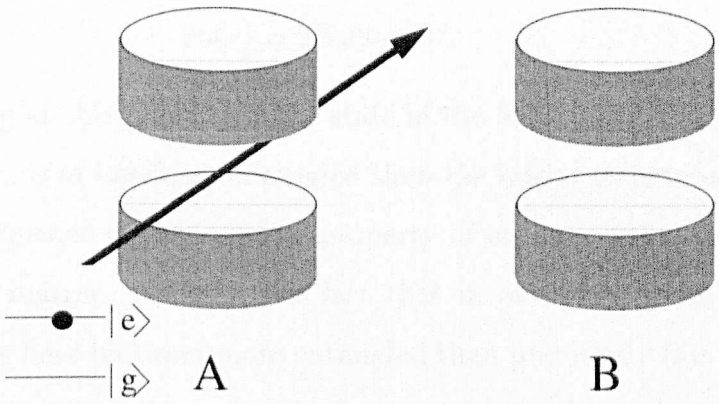


Figure 6.1: The experimental setup for local interactions: two cavities are initially entangled in a state of the form $|\Psi\rangle_{AB} = \alpha|n\rangle_A|m\rangle_B + \beta|n'\rangle_A|m'\rangle_B$, and atoms are sent through cavity A only.

generating an $(n+2)$ Fock state, which takes us further away from the initial state in eq. (6.21). We thus send through a ground-state atom, which can remove the extra photon.

Using, therefore, this as the starting field-state, we define the 'new' α and β as

$$\alpha' = \frac{\alpha b_n(t_1)}{\sqrt{(\alpha b_n(t_1))^2 + (\beta b_{n'}(t_1))^2}}, \qquad \beta' = \frac{\beta b_{n'}(t_1)}{\sqrt{(\alpha b_n(t_1))^2 + (\beta b_{n'}(t_1))^2}}. \tag{6.26}$$

and the joint atom-field state after sending through a ground state atom for time t_2, such that $b_n(t_2)\alpha' = b_{n'}(t_2)\beta'$, becomes

$$\begin{aligned} |\Psi_f\rangle &= (a_n(t_2)\alpha'|n+1\rangle_A|m\rangle_B + a_{n'}(t_2)\beta'|n'+1\rangle_A|m'\rangle_B) \otimes |g\rangle_A \\ &+ b_n(t_2)\alpha'\left(|n\rangle_A|m\rangle_B + |n'\rangle_A|m'\rangle_B\right) \otimes |e\rangle_A . \end{aligned} \tag{6.27}$$

As before, if the atom is measured in the excited state, then the cavities are left in the maximally entangled field state, once normalised, as desired. The probability

for this measurement is

$$P_2(e) = 2|b_n(t_2)\alpha'|^2. \tag{6.28}$$

It is worth noting at this point that the state of the field, after measuring a ground state of the atom, is in itself *less* entangled than the initial state in eq. (6.19). This is a direct consequence of the concave property of entropy when applied to either reduced density matrix. Namely, the fact that in one case, when registering an excited atom, the field becomes more entangled than previously (i.e. the entropy of either reduced system is greater after the interaction), implies that the entanglement of the other field state, when we register a ground atom, is 'smaller' than previously (i.e. the entropy is smaller than before the interaction). This can be quantified as follows. Let the reduced field state after the interaction be

$$\hat{\rho}'_A = \mathrm{p}\hat{\rho}'_{A1} + (1 - \mathrm{p})\hat{\rho}'_{A2} \tag{6.29}$$

where $\hat{\rho}'_A$ is the reduced density matrix for cavity A formed from eq. (6.24), and $\hat{\rho}'_{A1}$, $\hat{\rho}'_{A2}$ are the parts of $\hat{\rho}'_A$ corresponding to the measurement of an excited or ground state atom respectively. Now using the concave property eq. (3.19) we see that

$$S(\hat{\rho}_A) = S(\hat{\rho}'_A) \geq \mathrm{p}S(\hat{\rho}'_{A1}) + (1 - \mathrm{p})S(\hat{\rho}'_{A2}) \tag{6.30}$$

where $\hat{\rho}_A$ is the reduced density matrix for cavity A before the interaction. The first equality follows from the fact that the reduced density matrix does not change during this interaction, which can readily be derived for this example, and is shown generally in the next section. It follows that

$$S(\hat{\rho}_A) \geq \mathrm{p}(S(\hat{\rho}_A) + \Delta) + (1 - \mathrm{p})S(\hat{\rho}'_{A2}) \tag{6.31}$$

where Δ is the amount by which the entropy (and hence entanglement) of the reduced subsystem is constructed to increase upon measurement of $|e\rangle$, by arranging atomic interaction times. So,

$$S(\hat{\rho}_A) - S(\hat{\rho}'_{A2}) \geq \frac{\mathrm{p}\Delta}{1 - \mathrm{p}} \geq 0 \quad . \tag{6.32}$$

Hence, it is immediately seen that

$$S(\hat{\rho}_A) \geq S(\hat{\rho}'_{A2}) \tag{6.33}$$

and the result is proven.

A small amount of simple algebra applied to eq. (6.23) shows that whatever the initial values of α and β, the ratio

$$\frac{\min(\alpha, \beta)}{\max(\alpha, \beta)} \tag{6.34}$$

always *decreases* unless $n = n'$, i.e. the cavities are not entangled in the first place (a ratio equal to unity implies maximal entanglement). We thus have that $|\alpha'| > |\alpha|$ and $|\beta'| < |\beta|$. It is readily seen from this, and the fact that $|a_i(t)| < 1$ and $|b_i(t)| < 1$, that

$$P_2(e)_{\max} = 2|\beta'|^2 \qquad < \qquad P_1(e)_{\max} = 2|\beta|^2 \ . \tag{6.35}$$

Thus, there are two effects each time an atom is sent through the cavity — the first is that the probability of detecting an atom in the excited state, and hence collapsing the field state to the maximally entangled form, on average decreases with each atom that goes through; and the second is that the field-state if the atom is measured in the ground state becomes successively more disentangled. The effect is to make it successively more likely that the field will become completely disentangled, rather than completely entangled, which was the original aim. This can be seen mathematically by adding up the probabilities of detecting an atom in the excited state after sending through exactly N atoms. If the probability of detection in state $|e\rangle$ after the i-th atom is a_i, and the corresponding probability for $|g\rangle$ is b_i, then the probability of detection in $|e\rangle$ after N-atoms is

$$\begin{aligned} & a_0 + b_0 a_1 + b_0 b_1 a_2 + b_0 b_1 b_2 a_3 + b_0 b_1 b_2 b_3 a_4 + \ldots + b_0 \ldots b_{N-1} a_N \\ = \ & (1 - b_0) + b_0(1 - b_1) + b_0 b_1 (1 - b_2) + b_0 b_1 b_2 (1 - b_3) + \ldots + b_0 \ldots b_{N-1}(1 - b_N) \\ = \ & 1 - \prod_{i=0}^{N} b_i \ . \end{aligned} \tag{6.36}$$

The above product term is always less than unity since each and every b_i is individually less than unity, and similarly is always positive since all b_i are individually positive, so the probability of detection of $|e\rangle$ after N-atoms is less than unity. In the limit of $N \to \infty$, it can be verified by a computer program that the above product always tends to the value of $2|\beta|^2$. This result has the following consequence. In the limit $N \to \infty$ we either register a maximally entangled state or a completely disentangled state. However we could arrange the atom–cavity interaction time to be such that this happens when the first atom goes through the cavity. In this case it can be easily shown that the probability for the maximally entangled state to be registered (i.e. measuring the excited atomic state) is exactly $2|\beta|^2$. Thus, no matter how many atoms we send through the cavity (one or infinitely many), the highest probability of reaching the maximally entangled state is *always* less than unity. We thus see that this scheme *cannot* increase correlations between two entangled systems. We also note that the efficiency of $2|\beta|^2$ is much greater than the one given in [19]. In fact, if we confine ourselves to operation on single pairs at a time, then the scheme presented here is the optimal one, in the sense that it achieves the highest possible entanglement at the end [157].

We note also that we do not have to aim to achieve maximum entanglement for the particular initial state given by eq. (6.21). We could continue to send, for example, excited atoms and simply hope to achieve increased entanglement for *any* state. However, the same arguments given above also show that we cannot increase the entanglement of *both* field states corresponding to the two atomic measurement outcomes, as eq. (6.33) shows.

We should note that if it was possible to increase entanglement by the above local scheme, we would have a means of superluminal communication. Namely, the sender of the message could change the entanglement by operating locally on his cavity which could then be detected on the other end by the receiver in possession of the other cavity. The communication would then proceed as follows: two participants would initially share a number of not maximally entangled cavities. Then,

if the sender does nothing on one of his cavities, this could represent 'logical zero', whereas if the sender maximally entangled the cavities this would represent 'logical one'. After sharing the entangled sets of cavities, the two participants could travel spatially as far away from each other as desired. In this way, they would be able to communicate, through the above binary code, at a speed effectively instantaneously (only governed by the time to actually prepare the binary states, and to measure them at the other end). Therefore, we see that the impossibility of locally increasing the correlations is closely related to Einstein's principle of causality. This is a curious consequence of quantum mechanics, the postulates of which contain no reference to special relativity. Indeed, this could be turned upside down, and viewed as one reason why the above (or any similar) scheme would not work.

We thus find that the above scheme cannot increase correlations by local actions on one cavity alone. We might expect to compensate for this by sending independent atoms through both cavities, and arranging a feedback mechanism based upon classically communicating the knowledge of each state to the other side. In this way, we approach more closely the scheme of classical communication with post selection [67], but hope to replace the post-selection procedure with that of sending through multiple atoms until we achieve success. We would also expect to avoid superluminal communications since the method inherently involves classical communication between the two observers. The analysis for this problem is very similar to that given above for the one-atom model, except that there is much more freedom to choose which state to measure and how to optimise it. Following through a similar reasoning as in the single atom model, it is readily deduced that there is no way in this scheme to increase correlations. There are numerous variations on this above scheme: maximising the probability of detection in $|e\rangle_A|e\rangle_B$, minimising the rate of change of α and β, and so on, but the basic fact that the probability is never identically unity for any number of atoms remains the same. These results are in complete agreement with our condition E3 in Chapter 4, that the amount of entanglement (or, more precisely the average amount of entanglement) cannot

increase under local measurements aided by classical communications. It should also be clear that "feedback" schemes are of no help here. This is because if Alice and Bob perform local general measurements and then, upon communicating, perform another set of local general measurements, the combination of the two is yet another local general measurement and hence cannot increase the amount of entanglement. Namely,

$$\begin{aligned} & \sum_j C_j \otimes D_j \Big\{ \sum_i A_i \otimes B_i \; \rho_{AB} \; A_i^\dagger \otimes B_i^\dagger \Big\} C_j^\dagger \otimes D_j^\dagger = \\ = & \sum_{ij} (C_j A_i) \otimes (D_j B_i) \; \rho_{AB} \; (C_j A_i)^\dagger \otimes (D_j B_i)^\dagger \; . \end{aligned} \tag{6.37}$$

So, now we confirmed that entanglement cannot be increased locally. This means that if Alice and Bob start with a completely disentangled (separable) state, they cannot create entanglement locally. To create entanglement we need some no-local action, which is what we present next.

6.3.2 Increasing Entanglement Non-Locally

We now present two simple examples showing how a *nonlocal* operation *can* increase and, in fact, create correlations and entanglement. The procedures described here can be used to prepare initially entangled states.

6.3.2.1 Method 1

Suppose that the two cavities, A and B, start disentangled in the state:

$$|\phi_{cav}\rangle_{AB} = \frac{1}{\sqrt{2}} |0\rangle_A |0\rangle_B \; . \tag{6.38}$$

Let us send an entangled atomic pair through the cavities, each atom going through one cavity only, with the initial atomic state:

$$|\phi_{atom}\rangle_{AB} = \frac{1}{\sqrt{2}} (|e\rangle_A |g\rangle_B + |g\rangle_A |e\rangle_B) \; . \tag{6.39}$$

After the interaction for the same time t the joint state will be:

$$|\psi_{joint}\rangle_{AB} = -b_o^2(t)\,|g\rangle_A\,|g\rangle_B \left\{\frac{1}{\sqrt{2}}(|0\rangle_A|1\rangle_B + |1\rangle_A|0\rangle_B)\right\} + a_o^2(t)\left\{\frac{1}{\sqrt{2}}(|e\rangle_A|g\rangle_B + |g\rangle_A|e\rangle_B)\right\}|0\rangle_A\,|0\rangle_B\,. \quad (6.40)$$

Therefore, by simply setting $b_o(t) = 0$ we end up with certainty in the maximally entangled field state. Hence nonlocal interactions can, as expected, increase and create correlations and entanglement. The difference between this scheme and the previous two is that entanglement is being transferred to the cavities, from the atoms. This allows the cavity entanglement to 'increase', but at the expense of the entanglement of the atoms.

6.3.2.2 Method 2

This method involves only one atom, first interacting with one cavity and then with the other. This type of "entanglement generation" has been analysed in a number of other places [158]. Let the initial state of 'atom+fields' be:

$$|e\rangle|0\rangle_A|0\rangle_B\,. \quad (6.41)$$

After interaction between the atom and the cavity A for time t_1 the state is

$$(a_o(t_1)|e\rangle|0\rangle_A + b_o(t_1)|g\rangle|1\rangle)|0\rangle_B\,. \quad (6.42)$$

The atom now interacts with the cavity B for time t_2 after which the final state is

$$a_o(t_1)a_o(t_2)|e\rangle|0\rangle_A|0\rangle_B + a_o(t_1)b_o(t_2)|g\rangle|0\rangle_A|1\rangle_B + b_o(t_1)|g\rangle|1\rangle_A|0\rangle_B\,. \quad (6.43)$$

Choosing $a_o(t_2) = 0$ and $a_o(t_1) = b_o(t_2) = \frac{1}{\sqrt{2}}$ the above reduces to:

$$|g\rangle\frac{1}{\sqrt{2}}(|0\rangle_A|1\rangle_B + |1\rangle_A|0\rangle_B)\,, \quad (6.44)$$

which is the desired, maximally entangled state of the field. Thus, the method achieves an entangled cavity state by creating an entangled atom-cavity state, and transferring this to the two cavities alone.

6.4 Local Error Correction Preserves Correlations

We now show that although entanglement cannot be increased or created locally, still local error correction can be used to preserve an initially present amount of entanglement. We first describe a general method and then implement it using cavity QED techniques.

6.4.1 Theoretical Considerations

Imagine two initially entangled quantum systems A and B distributed between two spatially separated parties. Let, for the sake of simplicity, both A and B be two spin–$\frac{1}{2}$ particles in the initial EPR–like state

$$|\psi_{A+B}\rangle = \alpha|0\rangle_A|1\rangle_B + \beta|1\rangle_A|0\rangle_B \tag{6.45}$$

where the first ket describes the system A and the second the system B. Let both A's particles be encoded locally (i.e. adding locally a certain number of auxiliary qubits and performing local unitary transformations to encode) in order to protect their own qubit against a desired number of errors. We suppose that they both use the same coding, with the code–words denoted by $|C^0\rangle$ and $|C^1\rangle$. After the encoding, the state is therefore

$$|\Psi_{A+B}\rangle = \alpha|C^0\rangle_A|C^1\rangle_B + \beta|C^1\rangle_A|C^0\rangle_B \ . \tag{6.46}$$

Notice that the entanglement between the systems A and B is not changed by the encoding procedure, since local unitary operations do not change the spectrum of the reduced density matrices.

Let this state now be corrupted by errors, $\hat{E}$, which are local in nature, after which we perform the projections in eq. (5.4) of chapter 5 to obtain

$$|\Psi_{A+B}\rangle' = \alpha\,\hat{E}_i\,|C^0\rangle_A\,\hat{E}_j\,|C^1\rangle_B + \beta\,\hat{E}_i|C^1\rangle_A\,\hat{E}_j\,|C^0\rangle_B \ . \tag{6.47}$$

We wish to show that the error does not change the value of the entanglement. For

this we compute A's reduced density matrix:

$$\begin{aligned}\hat{\rho}_A &= \mathrm{Tr}_B(|\Psi_{A+B}\rangle'\langle\Psi_{A+B}|') \\ &= \langle C^0|_B(|\Psi_{A+B}\rangle'\langle\Psi_{A+B}|')|C^0\rangle_B + \langle C^1|_B(|\Psi_{A+B}\rangle'\langle\Psi_{A+B}|')|C^1\rangle_B \\ &= \langle C^0|_B\,\hat{E}_j|C^0\rangle_B\left\{\,|\alpha|^2\hat{E}_i\,|C^0\rangle_A\langle C^0|_A\,\hat{E}_i + |\beta|^2\,\hat{E}_i\,|C^1\rangle_A\langle C^1|_A\,\hat{E}_i\right\} \quad (6.48)\end{aligned}$$

which obviously has the same entropy as the original state in eqs. (6.45,6.46) and eq. (6.47). In the above derivation we used the relations in eq. (5.11) such that

$$\langle C^0|_B\hat{E}_i\,\hat{E}_j|C^0\rangle_B = \langle C^1|_B\hat{E}_i\,\hat{E}_j|C^1\rangle_B \quad (6.49)$$

$$\langle C^0|_B\hat{E}_i\,\hat{E}_j|C^1\rangle_B = 0\,. \quad (6.50)$$

Thus the entropy of the reduced density matrices of the initial pair of encoded systems, and of the systems after undergoing errors are both the same, indicating that the correlations and thus the entanglement do not change during the above described process. By a process of introducing more local degrees of freedom into the problem, we are able to maintain non-local quantum correlations. So, in fact, this process does also involve discarding information, but is different to the post selection previously described. This is so because all the error correcting particles are introduced locally, and do not form a part of the original ensemble.

6.4.2 Example With Cavities

We now present a simple example of how to locally preserve entanglement between two cavities in the state

$$\alpha|0\rangle_A|1\rangle_B + \beta|1\rangle_A|0\rangle_B \quad (6.51)$$

against a single amplitude error (action of $\hat{\sigma}_x$ Pauli operator) on either cavity. For this purpose we *locally* introduce a pair of atoms to each cavity, all of which are in the ground state. These atoms interact identically with their respective cavities. We also allow errors to happen to the atoms, as long as there is no more than one error on either side, A or B. We would like to implement the following interaction

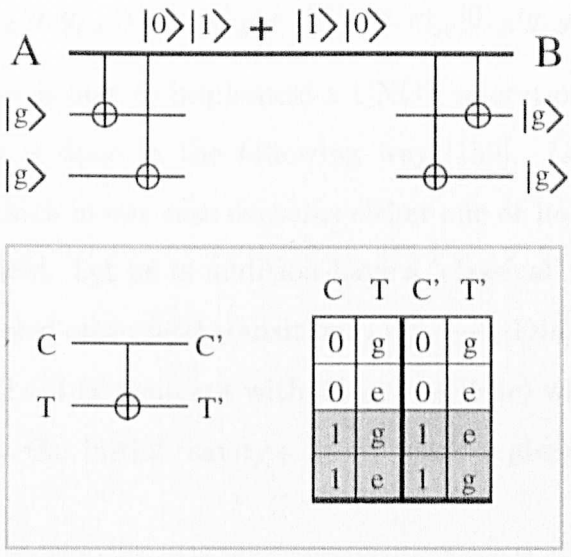


Figure 6.2: The encoding network for protecting against amplitude errors is shown in the upper diagram: the encircled cross denotes a NOT operation while a dot denotes a control bit, together making a CNOT operation. The atoms are initially in their ground states, and the order in which the gates are executed is irrelevant. The lower diagram gives a truth table for the CNOT operation; here, 'C' and 'T' represent *control* and *target* bits respectively.

in order to encode the state against an amplitude error [21] (four additional atoms for each cavity are needed to correct against a general type of single error [131])

$$|0\rangle|g\rangle_1|g\rangle_2 \longrightarrow |0\rangle|g\rangle_1|g\rangle_2 \tag{6.52}$$

$$|1\rangle|g\rangle_1|g\rangle_2 \longrightarrow |1\rangle|e\rangle_1|e\rangle_2 \,. \tag{6.53}$$

This is, in fact, an action of two CNOTs, with the control bit being the state of the cavity and the target bits being the atoms 1 and 2. We therefore perform identical interactions on both cavities and their atoms. This is shown schematically in Fig. 6.2. The state of the whole system ('2 cavities + 4 atoms') will be after the encoding

procedure,

$$\alpha|0\rangle_A|g,g\rangle_A|1\rangle_B|e,e\rangle_B + \beta|1\rangle_A|e,e\rangle_A|0\rangle_B|g,g\rangle_B \ . \tag{6.54}$$

So all we need to know is how to implement a CNOT operation between the cavity and one atom. This is done in the following way [159]. Let the atom be sent through the cavity, which in our case contains either one or no photons, interacting resonantly with the field. Let us in addition have a 'classical' light source (a laser) resonant with the dressed atom-field transition $|1\rangle|g\rangle \longrightarrow |1\rangle|e\rangle$. Due to the vacuum Rabi splitting this will not be resonant with $|0\rangle|g\rangle \longrightarrow |0\rangle|e\rangle$ which is precisely what we need. In this way the initial 'cavity+ atom' state undergoes evolution of the form

$$(\alpha|0\rangle + \beta|1\rangle)|g\rangle \longrightarrow \alpha|0\rangle|g\rangle + \beta|1\rangle|e\rangle \tag{6.55}$$

which is a CNOT gate. By repeated action of this gate we can create the state in eq. (6.54). Then if a single amplitude error occurs on either side (e.g. a spontaneous decay of the field) we can correct it by applying a unitary operation to the cavities to restore the original state, depending on the state of the four atoms [21].

Let us give a simple example of how this would work. Suppose that only the cavity A, after encoding, undergoes an amplitude error resulting in, after a small rearrangement, the joint 'cavities + atoms+environment' state of the form (eq. (5.2))

$$\begin{aligned} &(\alpha|0\rangle_A|1\rangle_B|g,g\rangle_A|e,e\rangle_B + \beta|1\rangle_A|0\rangle_B|e,e\rangle_A|g,g\rangle_B)|R_0\rangle \\ + \ &(\alpha|1\rangle_A|1\rangle_B|g,g\rangle_A|e,e\rangle_B + \beta|0\rangle_A|0\rangle_B|e,e\rangle_A|g,g\rangle_B)|R_1\rangle \ . \end{aligned} \tag{6.56}$$

To recover the original state we first have to decode the above state. This is just the inverse of encoding, i.e. we apply two CNOTs as described above, resulting in the state

$$\begin{aligned} &(\alpha|0\rangle_A|1\rangle_B + \beta|1\rangle_A|0\rangle_B)|g,g\rangle_A|g,g\rangle_B)|R_0\rangle \\ + \ &(\alpha|1\rangle_A|1\rangle_B + \beta|0\rangle_A|0\rangle_B)|e,e\rangle_A|g,g\rangle_B)|R_1\rangle \ . \end{aligned} \tag{6.57}$$

That decoding is the inverse transformation of encoding can be seen from the fact that if there was no error than we would just obtain the original state at the end, where both the atoms would be in the ground state. Otherwise we would obtain one of the other three possibilities for the state of the two atoms (ground-excited, excited-ground or excited-excited). In this second step we can make a measurement on the atoms and depending on the outcome apply an appropriate unitary transformation to the cavities. In this case we only have to consider cavity A: if both of the atoms are in the ground state then we do nothing because the joint-cavity state remains unchanged, whereas if both of the atoms are excited we apply a NOT operation to cavity A. This we do in a fashion similar to performing CNOT. We could, for example, send an excited atom through the cavity and tune the external laser to the dressed transition $|0\rangle|e\rangle \longleftrightarrow |1\rangle|e\rangle$. In this way we recover the state in eq. (6.51). We emphasise that the form in eq. (6.57) is incomplete since the terms arising from all the other amplitude errors are missing (corresponding to the cavity B and the atoms); however, it can easily be checked that the above scheme would also accommodate for this.

6.5 Conclusion

In this chapter we presented simple models to demonstrate that correlations cannot be increased by any form of local complete measurement. The consequence of this is that any purification procedure has to represent a post-selection of the original ensemble to be purified. Classical communication is an essential precursor to the post-selection procedure — we cannot post-select without classical communication, but the post-selection procedure is necessary to prepare the maximally entangled subset. We then showed that we can locally 'protect' the entanglement by standard quantum error correction schemes, such that the correlations (and therefore the entanglement) are preserved under any type of complete measurement, which can be viewed as an error in this context. We presented a simple example of how to

encode two cavities against a single amplitude error. Thus, local error correction can protect nonlocal features of entangled quantum systems, which otherwise cannot be increased by any type of local actions which exclude classical communication and post-selection. We would like to stress that as far as the implementations of quantum gates are concerned apart from cavity QED there are other possibilities, most notably a linear ion trap described before. However, the formalism describing the linear ion trap quantum computation is exactly that of the Jaynes-Cummings model used throughout this chapter. Therefore all the practical schemes that we have presented regarding purification procedures and error correction have a more general character and can immediately "translated" into a linear ion-trap quantum computer.

Chapter 7

Conclusions

In the last chapter we present a very brief summary of the main results of this thesis. We also introduce a number of open questions in quantum information theory related to this work, which will be investigated in the future.

7.1 Summary of the Thesis

In this thesis we presented basis of entanglement quantification in two or more entangled quantum subsystems. The central idea involves quantifying the amount of entanglement in a given state by calculating its distance to the set of disentangled states, i.e. to a closest disentangled state. There are two quantities to be specified in this definition. One is the distance measure to be used, and the other one is the form of a disentangled state–in other words, what we mean by "disentangled". We have seen that the distance measure very much depends on the physical way of distinguishing quantum states. We have presented three conditions E1-E3 that are based on physical reasoning and which any measure of entanglement has to obey. We have seen that there is an infinite number of measures of entanglement satisfying these conditions. However, if measurements are performed on an ensemble in a given quantum state, and we wish to distinguish it from another state, then asymptotically the quantum relative entropy will be the "right" quantity to use. In this case we

saw that the amount of entanglement is defined as

$$E(\sigma) = \min_{\rho \in \mathcal{D}} S(\sigma||\rho) \,, \tag{7.1}$$

where D is a set of all disentangled states. Otherwise S can be replaced by any other distance measure such that E satisfies E1-E3. For two subsystems we define this to be composed of all ρ of the form

$$\rho_{12} = \sum p_i \rho_1^i \otimes \rho_2^i \,. \tag{7.2}$$

This definition is intuitively attractive and indeed these states are the only ones from which we cannot distill any entanglement by local operations and classical communication. Moreover, for pure entangled states we recover the von Neumann entropy of the reduced subsystems, which is a good measure of entanglement for pure states from the Schmidt decomposition procedure and classical data and quantum entanglement compression point of view. For more than two subsystems there is an ambiguity in what we call a disentangled state. We have suggested two basic forms, but we believe that there is no unique way of doing this and the definition should depend on operational procedure used for quantification. We also presented an argument for using the above measure as the ultimate efficiency of purification procedures for two qubits. We saw that if entanglement as defined above is additive, i.e.

$$E(\sigma_1 \otimes \sigma_2) = E(\sigma_1) + E(\sigma_2) \,, \tag{7.3}$$

then from an initial ensemble of N pairs in state σ, we can distill M singlets such that

$$NE(\sigma) \geq M \ln 2 \,. \tag{7.4}$$

We have found no counter example to additivity using numerical methods for minimizing the von Neumann relative entropy. We hope that future work will find a purification protocol which will asymptotically achieve the equality in the above inequality in eq. (7.4). For pure states Bennett et al [110] have already presented such an asymptotic protocol, which we reviewed in Chapter 4. One of the most attractive

features of this way of quantifying entanglement is that it can directly be generalized to any number of subsystems of any dimensionality. The real impediment to progress here is the fact that any concrete calculation becomes exponentially more difficult with the increase in dimensions or number of systems in spite of the Caratheodory theorem.

We have then explained the basics of quantum error correction. We have derived the condition that code-words $|C^k\rangle$ have to satisfy in the presence of errors. The conditions derived are very simple and read

$$\langle C^k|\hat{P}_\beta\hat{A}_\alpha\hat{A}_\gamma\hat{P}_\delta|C^l\rangle = y^{\alpha\beta\gamma\delta}\delta_{kl}\ , \tag{7.5}$$

where P's are phase errors A's are amplitude errors, whose subscripts indicate the position of the error and $y^{\alpha\beta\gamma\delta}$ is *any* complex number. The physical intuition behind them is clear: different codewords should be transformed into mutually orthogonal states after errors have happened, since then we can distinguish them with certainty and correct the errors. Curiously, the same code-words do not have to go into orthogonal states after error, and can be corrected as long as all the code-words have the same overlap with the errors. This feature is a purely "quantum effect" and does not exist in classical error correction. We then applied this formalism to protecting information written into an atom against spontaneous emission into vacuum. We have shown that when spontaneous emission is viewed as a combination of two basic errors–jump and no-jump error, then the no jump part can be corrected to all orders of magnitude. Correcting for the spontaneous emission is very important in the ion-trap quantum computation, since spontaneous emission will be the ultimate obstacle to successful information processing if information is encoded into atoms.

We have then used a cavity QED example to show how two entangled cavities can be manipulated in order to create, and distill entanglement. This also provided a tool for encoding each of the two entangled cavities with atoms to protect entanglement locally. This is an interesting result: although entanglement (non-locality) cannot be increased by local operations and classical communication, nevertheless,

local encoding can preserve entanglement (i.e. non-locality).

7.2 Further Work

We have uncovered the number of interesting questions during the course of this research. It fact, it could be said that a number of questions initiated by this work by far outstrips the number of problems solved. We feel that in order to put the results of this thesis into the right perspective it is necessary to summarize future research prospects in quantum information theory.

- *Uniqueness of the entanglement measure.* The conditions E1-E3 presented in Chapter 4 do not lead to a single measure of entanglement as shown by presenting a number of examples of measures which satisfy them. An open question is what conditions should be added to E1-E3 in order to single out the relative entropy of entanglement as a unique measure of the amount of entanglement. It is reasonable to ask for the following additional properties:

  1. $E(\sigma)$ reduces to the von Neumann entropy for pure states.
  2. $E(\sigma)$ is continuous.
  3. $E(\sigma)$ is additive, i.e. $E(\sigma_1 \otimes \sigma_2) = E(\sigma_1) + E(\sigma_2)$.

  We proved condition 1 in Chapter 4, and Donald has proven condition 2 [160]. Condition 3 has only been confirmed by numerical calculations and no counter-example has been found. We conjecture that the only measure of entanglement that satisfies E1-E3 and together with the above conditions $1-3$ is given by

  $$E(\sigma) := \min_{\rho \in \mathcal{D}} S(\sigma||\rho) \tag{7.6}$$

  where $S(\sigma||\rho)$ is the quantum relative entropy.

- *Closed form of entanglement.* At present, calculating the Relative Entropy of entanglement involves minimization of the von Neumann relative entropy. We

would like to find a closed form for this expression so that given a bipartite qubit state we can calculate its entanglement without need for computation minimization methods.

- *Channel Capacities.* We have explained in Chapter 3 that there are two different aspects of a quantum channel: it can either be used for classical communication, in which case the Holevo bound provides the value of its capacity, or it can be used for transmitting an unknown quantum state from Alice to Bob. This second aspect is directly related to distributing entanglement and teleporting quantum states and can be introduced as follows. Alice and Bob wish to share a certain number of entangled pairs in order to ensure a perfect teleportation of an unknown state. Initially, Alice prepares an entangled state of two subsystems $|\Psi_{AB}\rangle$ and sends the subsystem B to Bob through a noisy quantum channel described by a complete measurement $\sum_i E^{\dagger}_{(B)i} E_{(B)i} = \mathbb{1}$, where B indicates that the domain of action is the subsystem B only. Let the state of $A + B$ after the action of the channel be given by ρ_{AB}. Then the quantum capacity of this channel will be understood as the largest amount of entanglement left after the transaction. It would follow that if 1. the above was the best way of transmitting an unknown state, and 2. the relative entropy of entanglement is indeed an achievable upper bound for purification, then the relative entropy of entanglement maximized over all the entangled input states would indeed be the quantum capacity of a quantum communication channel. This question is of a general importance and needs further investigation.

- *Local extraction of information.* In Chapter 4 we linked the idea of distinguishability of quantum states to the amount of entanglement. We emphasised that the asymptotic distinguishability is governed by the von Neumann relative entropy. It is known that this result is achieved by performing, in general, a non-local measurement. An open question is whether the same can

be achieved locally.

- *Purification of more than 2 particles.* We have shown that our way of quantifying entanglement can naturally be extended to more than two entangled subsystems. Further investigation into this area would help us understand the notion of a disentangled state and might be useful in practical quantum cryptographic protocols.

- *Efficient Purification procedures.* At present there exists no purification procedure for mixed bipartite states that achieves the upper bound given by the relative entropy of entanglement. Search into this question would be worthwhile because of possible benefits in efficient quantum cryptography and communication in general.

There is a number of other interesting and related questions to investigate, and the above are only the immediate next possibilities. Throughout this thesis we have emphasised the relationship between quantum mechanics and information theory. We have seen the implications that physics has for information theory in the sense that the information processing efficiency depends on whether the processing is based on classical or quantum laws. On the other hand, the ideas from information theory have been very useful in understanding the concept of correlations in quantum theory, and, in particular in this thesis, have directly provided a basis for understanding purely quantum correlations, i.e. entanglement. In the long run, we might hope to put quantum mechanics entirely on information–theoretic footing. This implies writing down axioms of information processing whose consequences would result in quantum mechanical laws. This information-theoretic way of interpreting physics might elucidate further the structure and character of quantum mechanical laws and perhaps resolve the current mysteries encompassing the measurement problem and the arrow of time. This is a good task for the next century, and a positive note on which to end this thesis.

Bibliography

[1] C. E. Shannon and W. Weaver, "The Mathematical Theory of Communication", (University of Illinois Press, Urbana, IL, 1949).

[2] L. Brillouin, "Science and Information Theory", (Academic Press, New York, 1956).

[3] S. Kullback, "Information Theory and Statistics", (John Wiley and Sons, New York, 1959).

[4] T. M. Cover and J.A. Thomas, "Elements of Information Theory" (A Wiley-Interscience Publication, 1991).

[5] B. Schumacher, Phys. Rev. A **51**, 2738 (1995).

[6] E. Schrödinger, Naturwissenschaften **23**, 807, 823, 844 (1935).

[7] A. Einstein, B. Podolsky and N. Rosen, Phys. Rev. **47**, 777 (1935).

[8] J. Bell, "Speakable and Unspeakable in Quantum Mechanics", (Cambridge Univ. Press, Cambridge, 1987).

[9] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **A 23**, 880 (1969).

[10] J. P. Gordon, *"Noise at Optical Frequences; Information Theory*, Quantum Electronics and Coherent Light, Proc. Int. School Phys. "Enrico Fermi, Course XXXI, ed. P. A. Miles, p.p. 156 (Academic Press, New York, 1964).

[11] A. S. Holevo, *Problemy Peredachi Informatsii*, **9**, 3 (1973) [A. S. Kholevo, Problems of Information Transmission, **9**, 177 (1973)]; B. Schumacher and

M. D. Westmoreland, Phys. Rev. A, **56**, 131 (1997); P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland and W. K. Wootters, Phys. Rev. A **54**, 1896 (1996); for the continuous case see H. P. Yuen and M. Ozawa, Phys. Rev. Lett. **70**, 363 (1993).

[12] R. S. Ingarden, Rep. Math. Phys. **10**, 43 (1976).

[13] M. Ohya, Rep. Math. Phys. **27**, 19 (1989).

[14] D. Deutsch, Proc. R. Soc. Lond. A **400**, 97 (1985).

[15] D. Deutsch, Proc. R. Soc. Lond. A **425**, 73 (1989).

[16] D. Deutsch and R. Jozsa, Proc. R. Soc. Lond. A **439**, 553 (1992); E. Bernstein and U. Vazirani, in *Proc. 25th ACM Symposium on the Theory of Computation*, 11 (1993); D.S. Simon, *Proceedings of the 35th Annual Symposium on the Foundations of Computer Science*, edited by S. Goldwasser (IEEE Computer Society Press, Los Alamitos, CA), 16 (1994);

[17] P. W. Shor. *In Proc. 35th Annual Symposium on Foundations of Computer Science*, ed. S. Goldwasser. (IEEE Computer Society Press, Nov. 1994) pp. 124-134.

[18] A. S. Holevo, Probl. Predachi Inform. **15**, 3 (1979), english translation: Problems of Inform. Transm., **15**, 247 (1979).

[19] D. Deutsch, A. Ekert, R. Jozsa, C. Macchiavello, S. Popescu, and A. Sanpera Phys. Rev. Lett. **77**, 2818 (1996).

[20] A. M. Steane, Phys. Rev. Lett. **77**, 793 (1996).

[21] A. Steane, Proc. Roy. Soc. London A **452**, 2551 (1996).

[22] P. W. Shor, Phys. Rev. A **52**, 2493 (1995).

[23] A. R. Calderbank and P.W. Shor, Phys. Rev. A **54**, 1098 (1996).

[24] E. Knill and R. Laflamme, Phys. Rev. A **55**, 900 (1997).

[25] P. Cochrane, "Tips for Time Travellers", (Orion Business, London, 1997).

[26] S. Kullback and R.A. Leibler, Ann. Math. Stat. **22**, 79 (1951).

[27] F. M. Reza, "An introduction to Information Theory", (New York: Dover, 1994).

[28] C. A. Fuchs, "Distinguishability and Accessible Information in Quantum Theory", PhD thesis, The University of New Mexico, Albuquerque, NM, 1996 (lanl e-print server: quant-ph/9601020).

[29] H. Everett, III, "The Theory of The Universal Wavefunction ", in "The Many–Worlds Interpretation of Quantum Mechanics" edited by B. DeWitt and N. Graham (Princeton University Press, 1973).

[30] A. N. Kolmogorov, "Foundations of The Probability Theory", (Chelsea pub. company, New york, 1950).

[31] J. von Neumann, "Mathematische Grundlagen der Quantenmechanic" (Springer, Berlin, 1932; English Translation, Princeton University Press, Princeton, 1955).

[32] G. W. Mackey, "Mathematical Foundations of Quantum Mechanics", (New York, W. A. Benjamin Inc., 1963).

[33] A. S. Holevo, "Probabilistic and Statistical Aspects of Quantum Theory", (North-Holland Publishing Company, Amsterdam, 1982).

[34] J. E. Cohen, Y. Inzsa, G. Rautu, M. B. Ruskai, E. Seneta and G. Zbaganu, Lin. Alg. Its Appl. **179**, 211 (1993).

[35] A. I. Khinchin, "Mathematical Foundations of Information Thoery", (Dover Publications, New York, 1957).

[36] V. Pless, "Introduction to the Theory of Error-Correcting Codes", (John Wiley & Sons 1982).

[37] R. Hill, "A First Course in Coding Theory", (Oxford University Press, Oxford, 1986).

[38] P. A. P. Moran, Proc. Cambridge Philos. Soc. **57**, 833 (1961).

[39] E. Bennati and H. Narhoffer, Lett. Math. Phys. **15**, 325 (1988).

[40] F. P. Kelly, "Reversibility and Stochastic Networks", (John Wiley and Sons Ltd., The University Press (Belfast) Ltd, 1979).

[41] C.-K. Peng, J. Mietus, J. M. Hausdorff, S. Havlin, H. E. Stanley and A. L. Goldberg, Phys. Rev. Lett. **70**, 1343 (1993).

[42] C. Tsallis, S. V. F. Levy, A. M. C. Souza and R. Maynard, Phys. Rev. Lett. **75**, 3589 (1995); for a popular account see C. Tsallis, Phys. World **10**, no. 7, p. 22 (1997).

[43] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).

[44] Z. Daróczy, Information and Control **16**, 36 (1970).

[45] I. Csiszár and J. Körner, "Coding Theorems for Discrete Memoryless Systems", (Academic Press, New York, 1981).

[46] I. N. Sanov, Mat. Sbornik (Moscow) **42**, 11 (1957).

[47] The original reference is E. Schmidt, "Zur Theorie der linearen und nicht linearen Integralgleichungen", Math. Annalen **63**, 433 (1907), in the context of quantum theory see H. Everett III, in *The Many-World Interpretation of Quantum Mechanics*, ed. B.S. DeWitt and N. Graham (Princeton University Press, Princeton, 1973) p. 3, and H. Everett III, Rev. Mod. Phys. **29** 454 (1957). A graduate level textbook by A. Peres, "Quantum Theory: Concepts and Methods", (Kluwer, Dordrecht, 1993), Chapt. 5 includes a brief description of the Schmidt decomposition; for quantum optical applications see A. K. Ekert and P. L. Knight, Am. J. Phys. **63**, 415 (1995).

[48] L.P. Hughston, R. Josza, and W. Wootters, Phys. Lett. **A 183**, 14 (1996).

[49] A. Peres, "Higher order Schmidt Decompositions", lanl-gov e-print server no. 9504006, 1995.

[50] M. Ohya and D. Petz, "Quantum Entropy and Its Use", Texts and Monographs in Physics, (Berlin: Springer-Verlag, 1993).

[51] R. S. Ingarden, A. Kossakowski and M. Ohya, "Information Dynamics and Open Systems - Classical and Quantum Approach", (Kluwer Academic Publishers, Dordrecht, 1997).

[52] A. Wehrl, Rev. Mod. Phys. **50**, 221 (1978).

[53] H. Araki and E. H. Lieb, Comm. Math. Phys., **18**, 160 (1970).

[54] S. J. D. Phoenix and P. L. Knight, Ann. Phys., **186**, 381 (1988); S. J. D. Phoenix and P. L. Knight, Phys. Rev. A **44**, 6023 (1991).

[55] H. Umegaki, Kodai Math. Sem. Rep. **14**, 59 (1962).

[56] M. D. Choi, Lin. Algebra and Appl. **10**, 285 (1975).

[57] E. B. Davies, "Quantum Theory of Open Systems", (Academic Press, London, 1976).

[58] B. Schumacher, Phys. Rev. A **54**, 2614 (1996).

[59] T. Rockafeller, "Convex Analysis" (Princeton University Press, New Jersey, 1970).

[60] M. H. Partovi, Phys. Lett. A **137**, 445 (1989), and the references therein.

[61] M. Redhead, "Incompleteness, Nonlocality and Realism", (Clarendon Press, Oxford, 1987).

[62] N. Gisin, Phys. Lett. A **143**, 1 (1990).

[63] R. Horodecki, P. Horodecki, and M. Horodecki, Phys. Lett. **A 200**, 340 (1995).

[64] N. Gisin, Phys. Lett. A **154**, 201 (1991).

[65] J. Baez, Lett. Math. Phys. **13**, 135 (1987).

[66] M. Horodecki, P. Horodecki and R. Horodecki, Phys. Rev. Lett. **78**, 574 (1997).

[67] N. Gisin, Phys. Lett. **A 210**, 151, (1996), and references therein; A. Peres, Phys. Rev. A **54**, 2685 (1996).

[68] A. Feinstein, "Foundations of Information Theory", (McGrow-Hill Company, Inc. New York, 1958).

[69] C. H. Bennett, C. A. Fuchs and J. A. Smolin, "Entnaglement-Enhanced Classical Communication on a Noisy Quantum Channel", eds. O. Hirota, A. S. Holevo and C. M. Caves (3rd International Workshop on Quantum Communication and Measurement, 1997).

[70] M. Sasaki, K. Kato, M. Izutsu, O. Hirota, "A simple quantum channel having superadditivity of channel capacity", lanl gov e-print server quant-ph/9705043 (1997).

[71] S. Lloyd, Phys. Rev. A **55**, 1613 (1997); B. Schumacher and M. Nielsen, Phys. Rev. A **54**, 2629 (1996).

[72] C. Bennett, C. Brassard, C. Crepeau, R. Jozsa, A. Peres and W. K. Wootters, Phys. Rev. Lett. **70**, 1895 (1993).

[73] D. Deutsch, "The Fabric of Reality", (Viking–Penguin Publishers, London, 1997).

[74] R. Landauer, IBM J. Res. Dev. **5**, 183 (1961); C.H Bennett, IBM J. Res. Dev. **32**, 16 (1988); T. Toffoli, Math. Systems Theory **14**, 13 (1981).

[75] C. H. Bennett, SIAM J. Comput. **18(4)**, 766 (1989); R. Y. Levine and A. T. Sherman, SIAM J. Comput. **19(4)**, 673 (1990).

[76] V. Vedral, A. Barenco and A. Ekert, Phys. Rev. A **54**, 147 (1996).

[77] D. Deutsch, A. Barenco and A. Ekert, Proc. R. Soc. Lond. A **449** 669 (1995).

[78] S. Lloyd, Phys. Rev. Lett. **75**, 346 (1995).

[79] A. Barenco, C. H. Bennet, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, H. Weinfurter, Phys. Rev. A, **52**, 3457 (1995).

[80] D. E. Knuth, "The Art of Computer Programming, Volume 2: Seminumerical Algorithms", (Addison-Wesley, New York, 1981).

[81] A. Ekert and R. Jozsa, Rev. Mod. Phys. **68**, 733 (1996).

[82] A. V. Aho, J. E. Hopcroft and J. D. Ullman, "Data Structures and Algorithms", (Addison–Wesley, 1983).

[83] J. I. Cirac and P. Zoller, Phys. Rev. Lett. **74**, 4091 (1995).

[84] D. G. Cory, M. D. Price, T. F. Havel, e-print quant-ph/9709001 (1997).

[85] N. A. Gershenfeld and I. L. Chuang, Science **275**, 350 (1997).

[86] E. Knill, I. Chuang, R. Laflamme, e-print quant-ph/9706053 (1997).

[87] R. Laflamme, E. Knill, W. H. Zurek, P. Catasti, S. V. S. Mariappan, e-print quant-ph/9709025 (1997).

[88] W. Paul, Rev. Mod. Phys. **62**, 531 (1990).

[89] I. Waki, S. Kassner, G. Birkl, and H. Walther, Phys. Rev. Lett. **68**, 2007 (1992).

[90] C. S. Adams and E. Riis, Prog. Quant. Electr. **21**, 1 (1997).

[91] W. M. Itano, J. C. Bergquist, J. J. Bollinger, and D. J. Wineland, Phys. Scripta **T59**, 106 (1995).

[92] C. Monroe, D. M. Meekhof, B. E. King, W. M. Itano, D. J. Wineland, Phys. Rev. Lett., **75**, 4714, 1995.

[93] C. Cohen-Tannoudji, J. Dupont-Roc, and G. Grynberg, "Atom-Photon Interactions: Basic Processes and Applications", (John-Wiley & Sons, Inc. New York, 1992).

[94] W. Vogel and R. L. de Matos Filho, Phys. Rev. A **52**, 4214 (1996).

[95] J. Steinbach, J. Twamley, P. L. Knight, Phys. Rev. A **56**, 4815 (1997) and references therein.

[96] E. T. Jaynes and F. W. Cummings, Proc. IEEE **51**, 89 (1963).

[97] C. A. Blockley, D. F. Walls and H. Risken, Europhys. Lett. **17**, 509 (1992).

[98] C. Monroe, D. M. Meekhof, B. E. King, W. M. Itano, and D. J. Wineland, Phys. Rev. Lett. **75**, 4714 (1995).

[99] A. K. Ekert, D.Phil Thesis, (Clarendon Laboratory, Oxford, 1991).

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

[101] S. Hill and W. K. Wootters, Phys. Rev. Lett. **78**, 5022 (1997).

[102] W. K. Wootters, lanl-gov e-print quant-ph/9709029, (1997).

[103] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight, Phys. Rev. Lett. **78**, 2275 (1997).

[104] V. Vedral and M. B. Plenio, Phys. Rev. A, **57**, 1619 (1998).

[105] V. Vedral, M. B. Plenio, K. Jacobs, and P. L. Knight, Phys. Rev. A **56**, 4452 (1997).

[106] E. Rains, *Entanglement purification via separable superoperators*, lanl e-print quant-ph/9707002, (1997).

[107] M. Horodecki, P. Horodecki, R. Horodecki, *Mixed state entnaglement and distillation: is there a bound entnaglement in nature*, lanl gov e-print server quant-ph/9801069 (1998).

[108] A. Peres, Phys. Rev. Lett. **77**, 1413 (1996).

[109] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Lett. A **223**, 1 (1996).

[110] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Phys. Rev. A **53**, 2046 (1996).

[111] R. Jozsa and B. Schumacher, J. Mod. Opt. **41**, 2343 (1994).

[112] G. Lindblad, Comm. Math. Phys. **40**, 147 (1975).

[113] G. Lindblad, Comm. Math. Phys. **39**, 111 (1974).

[114] M. Reed and B. Simon, "Methods of Modern Mathematical Physics–Functional Analysis", (Academic Press, New York, 1980).

[115] F. Hiai and D. Petz, Comm. Math. Phys. **143**, 99 (1991).

[116] M. J. Donald, Comm. Math. Phys. **105**, 13 (1986); M.J. Donald, Math. Proc. Camb. Phil. Soc., **101**, 363 (1987).

[117] E. H. Lieb and M. B. Ruskai, Phys. Rev. Lett. **30**, 434 (1973); E. H. Lieb and M. B. Ruskai, J. Math. Phys. **14**, 1938 (1973).

[118] M.J. Donald, private communication.

[119] G. H. Hardy, J. E. Littlewood and G. Pólya, "Inequalities", (Cambridge: Cambridge University Press, Second ed., 1952).

[120] D. Bures, Trans. Am. Math. Soc. **135**, 199 (1969); see also A. Uhlmann, Rep. Math. Phys. **9**, 273 (1976); *ibid* **24**, 229 (1986).

[121] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher, Phys. Rev. Lett. **76**, 2818 (1996).

[122] M. B. Ruskai, Rev. Math. Phys. **6**, 1147 (1994).

[123] D. Petz, Rep. Math. Phys. **23**, 56 (1986).

[124] M. Ozawa, private communication.

[125] W. K. Wootters, Phys. Rev. D **23**, 357 (1981).

[126] M. Hayashi, *Asymptotic Attainment for Quantum Relative Entropy*, lanl e-print server quant-ph/9704040 (1997).

[127] M. J. Donald, Found. Phys. **22**, 1111 (1992).

[128] R. Josza, J. Mod. Opt. **41**, 2315 (1994).

[129] M. B. Plenio and P. L. Knight, Phys. Rev. A **53**, 2986 (1996).

[130] M. B. Plenio and P. L. Knight, Proceedings of the 2nd International Symposium on Fundamental Problems in Quantum Physics, 1996, edited by M. Ferrero and A. Van der Merwe (Kluwer, Dordrecht).

[131] R. Laflamme, C. Miguel, J. P. Paz, and W. H. Zurek, Phys. Rev. Lett. **77**, 198 (1996).

[132] A. Ekert and C. Macchiavello, Phys. Rev. Lett. **77**, 2585 (1996).

[133] A. Barenco, A. Berthiaume, D. Deutsch, A. Ekert, R. Josza, and C. Macchiavello, SIAM J. Comp. **26**, 1514 (1997).

[134] A. R. Calderbank, E. M. Rains, P. W. Shor, and N. J. A. Sloane, Phys. Rev. Lett. **78**, 405 (1997).

[135] P. W. Shor, *Fault-Tolerant Quantum Computation*, lanl e-print quant-ph/9605011 (1996).

[136] A. M. Steane, Phys. Rev. A **54**, 4741 (1996).

[137] D. P. DiVincenzo and P.W. Shor, Phys. Rev. Lett. **77**, 3260 (1996).

[138] D. Gottesman, *Pasting Quantum Codes*, lanl e-print quant-ph/9607027 (1996).

[139] D. Gottesman, Phys. Rev. A **54**, 1862 (1996).

[140] R. Cleve and D. Gottesmann, Phys. Rev. A **56**, 76 (1997).

[141] J. Dalibard, Y. Castin, and K. Mølmer, Phys. Rev. Lett. **68**, 580 (1992); G. C. Hegerfeldt and T. S. Wilser, *Proceedings of the II. International Wigner Symposium, Goslar 1991*, H. D. Doebner, W. Scherer, and F. Schroeck, Eds., World Scientific, Singapore 1992; H. J. Carmichael, "An Open Systems Approach to Quantum Optics". Lecture Notes In Physics, (Springer, Berlin 1993).

[142] V. Vedral, M. A. Rippin and M. B. Plenio, J. Mod. Opt. **44**, 2185 (1997).

[143] M. B. Plenio, V. Vedral, P.L. Knight, Phys. Rev. A **55**, 67 (1997).

[144] G. Lindblad, Comm. Math. Phys. **48**, 119 (1976).

[145] M. B. Plenio and P. L. Knight, Rev. Mod. Phys. (1998).

[146] H. Mabuchi and P. Zoller, Phys. Rev. Lett. **76**, 3108 (1996).

[147] J. von Neumann, "Probabilistic logics and the synthesis of reliable organisms from unreliable components", in Automata Studies, eds. C. E. Shannon and J. McCarthy, (Princeton University Press, Princeton, 1955).

[148] M. B. Plenio, V. Vedral, P.L. Knight, Phys. Rev. A **55**, 4593 (1997).

[149] E. Knill, R. Laflamme, and W. Zurek, preprint quant-ph/9610011 (1996).

[150] C. H. Bennett and G. Brassard, in *Proceedings if IEEE Conference on Computers, Systems and Signal Processing*, 175 (1984).

[151] A. Ekert, Phys. Rev. Lett. **67**, 661 (1991).

[152] R. J. Hughes, D.M. Alde, P. Dyer, G.G. Luther, G.L. Morgan and M. Schauer, Cont. Physics **36**, 149 (1995).

[153] S. J. D. Phoenix and P. D. Townsend, Cont. Physics **36**, 165 (1995).

[154] B. W. Shore, P. L. Knight, J. Mod. Opt. **40**, 1195, (1993), and references therein.

[155] S. Haroche, "Cavity Quantum Electrodynamics", Les Houches, Session LIII, 1990, eds. J. Dalibard, J.M. Raimond, J. Zinn-Justin, (Elsevier Science publishers B.V. 1992).

[156] B. Garraway, B. Sherman, H. Moya-Cessa, P. L. Knight, G. Kurizki, Phys. Rev. A, **49**, 535, (1994).

[157] H.-K. Lo and S. Popescu, *Concentrating entnaglement by local actions–beyond mean values*, lanl gov e-print server quant-ph/9707038 (1997).

[158] J. D. Berger, H. Giessen, P. Meystre, T. Nelson, D. Haycock, S. Hamman, Phys. Rev. A, **51**, 2482, (1995) ; C. C. Gerry, Phys. Rev. A, **53**, 2857, (1996).

[159] J. M. Raimond, "Basics of Cavity Quantum Electrodynamics", published in "Quantum Optics of Confined Systems" eds. M. Ducloy and D. Bloch (Kluwer Academic Publishers, 1996).

[160] M. J. Donald, private communication.

LIST OF PUBLICATIONS

1.1 Authors : V. Vedral, A. Barenco and A. Ekert
Title : *Quantum Networks for Elementary Arithmetic Operations*
Journal : Phys. Rev. A **54**, 147 (1996)

1.2 Authors : M. Plenio, V. Vedral and P. Knight
Title : *Computers and communication in the quantum world.*
Journal : Phys. World **9**, 19 (1996)

1.3 Authors : M. B. Plenio, V. Vedral and P. L. Knight
Title : *Quantum error correction in the presence of spontaneous emission.*
Journal : Phys. Rev. A **55**, 67 (1997)

1.4 Authors : V. Vedral, M. B. Plenio, M. A. Rippin and P.L. Knight
Title : *Quantifying entanglement*
Journal : Phys. Rev. Lett. **78**, 2275 (1997)

1.5 Authors : V. Buzek, V. Vedral, M. B. Plenio, P. L. Knight and M. Hillery
Title : *Broadcasting of entanglement via local copying*
Journal : Phys. Rev. A **55**, 3327 (1997)

1.6 Authors : M. B. Plenio, V. Vedral and P. L. Knight
Title : *A fault-tolerant error correction network.*
Journal : Phys. Rev. A **55**, 4593 (1997)

1.7 Authors : V. Vedral, M. A. Rippin and M. B. Plenio
Title : *Quantum correlations, local interactions and error correction*
Journal : J. Mod. Opt. **44**, 2185 (1997)

1.8 Authors : K. Jacobs, P. L. Knight and V. Vedral
Title : *Determining the state of a single cavity mode from photon statistics*
Journal : J. Mod. Opt. **44**, 2427 (1997)

1.9 Authors : P.L. Knight, M.B. Plenio and V. Vedral
Title : *Decoherence and Quantum Error Correction*
Journal : Phil. Trans. Roy. Soc. London **335**, 2381 (1997)

1.10 Authors : V. Vedral, M. B. Plenio, K. Jacobs and P.L. Knight
Title : *Statistical Inference, Distinguishability of Quantum States, And Quantum Entanglement*
Journal : Phys. Rev. A **56**, 4452 (1997)

1.11 Authors : V. Vedral and M. B. Plenio
Title : *Entanglement Measures and Purification Procedures*
Journal : Phys. Rev. A **57**, 1619 (1998)

1.12 Authors : S. Bose, V. Vedral, and P.L. Knight
Title : *Multiparticle Schemes for Entanglement Swapping*
Journal : Phys. Rev. A **57**, 822 (1998)

1.13 Authors : M. Murao, M. B. Plenio, S. Popescu,
V. Vedral and P.L. Knight
Title : *Multiparticle Entanglement Purification Protocols*
Journal : submitted to Phys. Rev. Lett. (1997)

1.14 Authors : S. Bose, P. L. Knight, M. Murao,
M. B. Plenio and V. Vedral
Title : *Implementations of Quantum Logic: Fundamental and Practical Limits*
Journal : to appear in Phil. Trans. Roy. Soc. London, (1998)

1.15 Authors : V. Vedral and M. B. Plenio
Title : *Basics of Quantum Computing*
Journal : to appear in Prog. Q. Elect., (1998)

1.16 Authors : M. B. Plenio and V. Vedral
Title : *Entanglement in Quantum Information Theory*
Journal : to appear in Contemp. Phys., (1998)