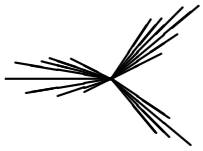

Instituto de Física Teórica
Universidade Estadual Paulista

MASTER DISSERTATION IFT–D.008/2014

A Study of the Dynamics of Quantum Correlations

Patrice Audibert Camati

Advisor

Dr. Gastão Inácio Krein

Feb/2014

Acknowledgements

I, firstly, would like to thank the National Council for Scientific and Technological Development (CNPq) for having granted me with a scholarship along the past two years. I thank my advisor for have accepted me as his student and for have allowed me to pursue my studies in a field which is not one from his expertise. I also must thank some colleagues of mine who provided me with insightful discussions about physics as well as taught me a bit of their respective fields. They are: George De Conto Santos, Guilherme Franzmann, Henrique dos Santos Flores, Lucas Lolli Savi and Marcio Woitek Junior.

"The scientist does not study nature because it is useful; he studies it because he delights in it, and he delights in it because it is beautiful. If nature were not beautiful, it would not be worth knowing, and if nature were not worth knowing, life would not be worth living."

- HENRI POINCARÉ

"The book of nature is written in the language of mathematics."

- GALILEU GALILEI

"If people do not believe that mathematics is simple, it is only because they do not realize how complicated life is."

- JOHN VON NEUMANN

"Science is much more than a body of knowledge. It is a way of thinking."

-CARL SAGAN

Abstract

In this work, we present an introduction to the quantum information theory. Before dealing with quantum information we develop some necessary tools, the classical information theory and some important concepts in quantum dynamics. We define measures of quantum correlations, concurrence and quantum discord, and study their dynamical evolution when the coupling between the system and environment is considered random. We conclude that a random environment imposes permanent loss of correlations to the system.

Key words: Entanglement; quantum discord; random coupling constant.

Areas: Quantum information theory, quantum correlations.

Resumo

Neste trabalho, apresentamos uma introdução à teoria da informação quântica. Antes de lidarmos com a informação quântica propriamente dita, desenvolvemos alguns assuntos preliminares, a teoria clássica da informação e conceitos relevantes em dinâmica de sistemas quânticos. Depois, definimos medidas de correlações quânticas, concorrência e discórdia quântica, e estudamos sua evolução dinâmica quando o acoplamento do sistema com o ambiente é modelado como aleatório. Concluímos que um ambiente aleatório impõe uma perda permanente nas correlações do sistema.

Palavras-chave: Emaranhamento, discórdia quântica, acoplamento randômico.

Campos: Teoria quântica da informação ; correlações quânticas.

Table of Contents

Chapter 1

Introduction

Since the advent of quantum mechanics there has been a huge progress in technology. Quantum mechanics underpins our modern society in many ways, from all kinds of electronic devices to how some economic sectors are structured. One of the new paradigms of technological development is the idea of construct a quantum computer. A computer which exploits in a fundamental way the unique quantum behavior. All computers use quantum mechanics in the sense of using quantum mechanical devices such as the transistor. Nevertheless, when it comes to store and process the information, our current computers use binary language, for example, either with or without electrical current flowing through the transistors. It is precisely this point that quantum computers intend to extend, by exploiting the superposition property of quantum systems. This story of quantum computers brings us to another field greatly developed in the past century, the theory of communication. We will say a few words about information theory in the next chapter. For the moment, let us comment briefly some chain of events that were important for the realization that taking into account quantum mechanics in the way we transmit information can provide more efficient algorithms to some tasks.

Entanglement has been a major concept in modern physics. It is accepted that the first discussion of this property in a scientific article was in the seminal paper written by Einstein, Podolsky and Rosen (also known as EPR article) in 1935 [1]. They questioned if the quantum theory was complete or not. As they were prone to the incompleteness, because in their work this was a consequence if the locality of quantum theory is assumed, they suggested that it might exist hidden variables which would make the quantum theory complete. This means that they have provided an example that seemed to show that quantum mechanics does not respect

locality. The next important step in this story may be considered the work of John S. Bell published in 1964 [2]. Bell set up the discussion to the experimental level. He derived an inequality (called Bell's inequality) which the supposed (local) hidden variable theory should satisfy. Even though it might have been unclear up to now, all these discussions were raised because since the beginning, with the EPR article, they were dealing with entangled states, those states that seemed to violate locality. Nobody was able to grasp what was really going on at the time. In 1972 John Clauser and Stuart Freedman carried out the first experimental test that Bell's inequality (and also the Clauser-Horne-Shimony-Holt inequality*) were, in fact, violated by quantum mechanics. Since then, there has been a myriad of experimental confirmation of this feature. The violation of Bell-like inequalities by quantum mechanics means that there is something fundamentally different that cannot be substituted by the local hidden variable theories, which are the most general notion which preserves both locality and reality (as defined by EPR). It was then that a careful study of such states (those which violate the inequalities) began to grow.

In the early 1980s arose the interest in whether it might be possible to use such quantum effects to signal faster than light. Wootters, Zurek and Dieks [3,4] showed that this is not possible and it turns out that this is prevented by the no-cloning theorem, which says that an arbitrary unknown quantum state cannot be perfectly copied (we will see this in more detail later). This was the first result in quantum mechanics which had a striking distinct characteristic when contrasted with classical information, where bits can be easily copied. In fact, the copying availability is everywhere used in error-correction schemes in classical communication. After that, the field of quantum information began to grow increasingly faster. Right after the discovery of the no-cloning theorem, Bennett and Brassard [5] discovered the first quantum cryptographic protocol, in 1984, representing the beginning of interest in using quantum systems for security of communication. This naturally leads to the will to use quantum mechanics to explore other communication concepts, such as codification, data compression, transmission of information and computation.

With the development of these studies entanglement have found its place. Until recently, it was believed that entanglement was a necessary resource for a quantum

*The CHSH inequality is a Bell-like inequality whose violation by quantum mechanics also means that quantum mechanics cannot be described by local hidden variable theories.

protocol be more efficient, if that is even the case, than a similar classical protocol which solves the same task. Some quantum protocols will be discussed later in this work, some of which are the quantum teleportation and the super dense coding. Nevertheless, there has been an increasing interest in what is called mixed-state computation. In this computation there is no entanglement, or a negligible amount of entanglement such that it is not the resource that promote the efficiency of the algorithm. These resources are still somewhat fundamentally quantum, although there is much more research ahead for clarification of this issue. One example of another resource, other than entanglement, is quantum discord, which we are also going to discuss and show an efficient quantum algorithm to solve a task that there is no known classical algorithm that does so. These quantum correlations are indeed playing the role of potential resources for speedup computation, justifying the interest on studying them.

This work has two aims. First, it is structured to provide an introduction to quantum information theory, beginning with a review of the major concepts of classical information theory and quantum mechanics, although there were some issues that were just commented and an appropriate explanation was left behind. Second, it serves to present an analysis about the behavior of some measures of quantum correlations when the coupling constant between system and environment is random.

This work is organized as follows. In Ch. 2 we introduce the classical information theory. We begin with a discussion about the concept of information and, from that, we introduce and study the measure of information of a informational source, the Shannon entropy. This comprises the Secs. 2.1, 2.2 and 2.3. Then we develop the other measures of information for two or more sources, the mutual information, conditional entropy and relative entropy in Sec. 2.4. In Ch. 3 we review the basic formalism and dynamics of quantum systems, Secs. 3.1 and 3.2, as well as we provide an introduction to more advanced topics such as open system dynamics, Sec. 3.3, and measurement dynamics, in Sec. 3.4. In Sec. 3.5 we present a general description of the quantum dynamics through the so-called quantum operations. Finally, in Sec. 3.6, we discuss two related phenomena, decoherence and environment-induced superselection, using the tools developed in the previous sections. With the basic tools of classical information theory and quantum mechanics we focus on the quantum

information theory in Ch. 4. We begin introducing the von Neumann entropy, the Shannon-like measure of information for quantum systems, and discuss the information at the two basic interfaces of communication, the preparation of quantum systems and their measurement, thus, comprising Secs. 4.1 and 4.2. In Sec. 4.3 we discuss correlations in quantum mechanics and then, in Sec. 4.4, we develop some measures of correlations such as entropy of entanglement, for pure states, concurrence and quantum discord for mixed states. We end the chapter, in Sec. 4.5, with some notable applications in quantum information theory, the no-cloning theorem, quantum teleportation, entanglement swapping, quantum dense coding and mixed-state quantum computation. In Ch. 5 we study the dynamics of quantum correlations when the interaction between system and environment is governed by a random coupling constant. There, we discuss three situations. The first where the coupling constants are actually constant, the homogeneous environment, and then two random interactions, one of which with a uniform distribution and the other with a normal distribution. Finally, we end this work with the conclusions in Ch. 6.

Chapter 2

Classical Information Theory

Information theory is a broad field of science which comprises applied mathematics, electrical engineering and computer science [6]. The landmark work that may be considered as the starting point of the information theory is the article entitled "A Mathematical Theory of Communication", written by Claude E. Shannon in 1948 [8]. Shannon was a researcher at the Math Center at the Bell Laboratories, a division created by the American Telephone and Telegraph (AT&T) specifically for conducting research and development of communication systems. From a fairly reasonable set of properties that an appropriate measure of information should have, he arrived at a function, which he himself recognized as being similar to the statistical entropy, that properly measures information, choice and uncertainty. This function is the so-called Shannon entropy.

Information theory is not concerned with the meaning of the information sent, but with practical issues such as how to send efficiently and reliably a signal through a noisy channel such that it can be recovered faithfully. The development of this field enabled the invention of the Internet, which certainly is one of the technological achievements that most shaped, and still is powerfully capable of shaping, our modern society.

The aim of this chapter is to present an introduction to the classical information theory, going through its fundamental concepts, all of which will be of key importance when we turn into the quantum information theory. We will not be concerned with actual applications.

The structure of this chapter is the following. In Sec. 2.1 we will discuss the concept of information, making the meaning of the word clear when we use it. Following this, we will reach at an objective meaning of information, thus enabling

us to quantify it and use it operationally. Encouraged by this meaning, in Section 2.2, we will study a measure of information supported by it. In Section 2.3 we will explore the fundamental concept of information theory, the Shannon entropy. Instead of being concerned with the actual derivation given by Shannon, we will present it as a definition and work with it in some examples to understand its meaning. In Section 2.4 we will begin to explore the correlations between two or more messages. In order to do so we will use other entropies, namely, the joint entropy, the conditional entropy, the mutual information and the relative entropy. These entropies rest upon the Shannon entropy and are as important as the Shannon entropy itself. The quantum analogous of some of these entropies will give us a measure of quantum correlations.

The presentation is intended to be pedagogical, as a solid understanding of the concepts of information theory in the classical case is important when one goes to the quantum case.

This chapter is heavily based on the Ref. [7].

2.1 What is information?

If we wonder about the concept of information for while it is not difficult to realize that it has many facets. In fact, it is not such a rare insight to realize that define any abstract noun as precise as possible is by no means an easy task. Semantics, the field of Linguistics that studies the meaning of words, phrases, et cetera, is a very interesting subject. Studying how we acquire the meaning of words shows us that from our early days we learn based on our experience. Our brain is biologically adapted to learn languages. We observe the behavior of those around us, listen to those mysterious sounds they make and by simply being there our brain is able to connect those sounds with the present situation and, even though it is an extremely complex process, after two or three years we are able to produce, identify precisely and understand those sounds and their meanings [9]. Based on this phenomenon the present author believes the best way to define a concept, mainly an abstract one, is heuristically. We will go through a couple of situations where the idea of information is applied and then reflect about its different meanings.

Suppose Alice and Bob want to secretly communicate with each other. They are

neighbors and live side by side in such a way that from Alice's bedroom window she can see Bob's bedroom window and vice versa. They have established a secret code using a lantern. If Alice flashes once, it means that her parents have left home. If Alice flashes twice, it means that her parents have decided not to go out anymore. Without minding why they would like to communicate with each other, this situation brings us the opportunity to explore the concept of information. We say that after Alice flashed once or twice, Bob has received information from her. So, one property of informatioin is that it can be transmitted. The code they established must be known *a priori* to both Alice and Bob. If Bob didn't know about the code, he would have not been able to read the message properly. He simply would not have understood it. The abstract message, either parents at home or not, had to be coded in a physical system. In the present case, the flashes of light. Moreover, if it was not a problem, Alice could have just yelled at him. The very same information would have been transmitted through another physical system, namely, the sound waves in the air. Thus, the same information can be stored in different physical systems.

Suppose now that Alice and Bob have met in the street near their houses and started a conversation. Alice tells Bob that in a street nearby there was a car which crashed into a post. We say naturally that Alice had this information and gave it to Bob. So, now Bob has this information too. Alice neither forgot nor lost this information in any way, both of them have the information now. Therefore, one additional and very important property of information is that it can be copied. Consider now that they have not seen each other for, say, a week. When they meet again Alice tells Bob the very same story. Alice didn't forget the information about the accident; instead, she simply has forgotten she told Bob. But Bob already knew the story and it was not a new information for him. This is a subtle and fundamental point now. Suppose that, in this situation, Alice did not tell him the very same story, instead she told a couple of details she haven't told him before, and therefore he did not know about. We would say that he acquired a little amount of information but not as much as he would have if she had told him from the very beginning, in the first encounter, the standard story plus the details. Of course at the end we would agree that he had the same amount of information about the story. The subtlety here is that we implicitly are measuring the amount of information. One could

even say that he had been listening inattentively or uninterestedly until the moment when Alice told him something new, unexpected. He, then, might have turned his attention on her to listen carefully the details which were new to him. This is the key. We measure information as how unexpected, uncertain, unpredictable something is. The more uncertain, unexpected, unpredictable something is the more information we will gain when we acquire that knowledge. We will make these concepts more precise in the sections to follow.

With this sense, we have a proper lead to an objective (and hence scientifically useful) definition of information because in order to use it we must have an operational definition. Something we can manipulate and calculate mathematically.

To summarize this section, we concluded that information is something that can be coded, transmitted, copied and stored in different ways. The scientific concept which we will develop from now on will not be concerned with the meaning of the words or symbols, because this is a subjective property from which we can not make science. Instead, we will measure information based on unexpectancy, uncertainty and unpredictability.

2.2 Measuring Information

The focus of this section is to develop a measure of the amount of information. We concluded the last section realizing that we measure information qualitatively relating the knowledge of some event as how much unexpected or unpredictable it is. The most unexpected the event is the more surprised we will be while acquiring that knowledge and consequently more information we gain.

Suppose that Alice plays the lotto and she is searching for the information of the lotto's results. When she gets the information she reads it and it says: The winning combination is 5, 38, 4, 22. What is important in this sentence is the set of numbers. After she compared these numbers with the ones in her tickets the two important events are: W = she has won or L = she has lost. There is a probability associated with each of these events. From experience we know that the odds of winning are quite small, and the odds of losing are, therefore, substantially high.

We may assume, for instance, that

$$p(W) = 0.000001, \tag{2.1}$$

$$p(L) = 0.999999. \tag{2.2}$$

Now the question is, how much information does Alice get when actually knowing the lotto result? As we know from our experience that the odds of winning are small, so does Alice. Therefore she has not much expectation of winning, she will not learn much upon realizing that she has, indeed, lost. But on the contrary case, if she wins, it would be quite a surprise for her (as it would be for us). There is a lot of information in the winning event, to the same extent there is little information in the losing event. Now we have to find a proper mathematical function to measure the information. This function must obey some conditions. This function, call it $I(x)$, should approach zero for events closed to absolute certainty, and infinity for events reaching the impossibility. It must also satisfy that $I(x_1) > I(x_2)$ if the two events satisfy $p(x_1) < p(x_2)$, meaning that the less likely event is associated with the greater information. An information measure that satisfies all these requirements is the following

$$I(x) = -K \log [p(x)], \tag{2.3}$$

where $\log(x)$ is the base two logarithm and K is a dimensionless constant which we will set up to 1 for the moment and later we shall see why this is appropriate. In our lotto example we have

$$I(W) = -\log [0.000001] = 19.9315686, \tag{2.4}$$

$$I(L) = -\log [0.999999] = 1.0 \times 10^{-6}. \tag{2.5}$$

As expected by construction, the winning event has more information because it is more unexpected.

By convention we set the symbol $\log(x)$ to stand for base two logarithm. We could have used any basis for the logarithmic function, though. This would just have changed the scale of the measure or, in other words, its units. This measure is dimensionless and the unit we use is the bit*. But why would we use the bit instead of other units? The smallest nontrivial set of events from which we may speak of information is the set composed of two symbols, or events. For example, a coin. If we toss

*If we have chosen a three based logarithm it would be a trit, a four based a quad and so on.

an unbiased coin there is a probability of one half to get either heads or tails. Hence $p\,(\text{heads}) = p\,(\text{tails}) = 1/2$. The information associated with each of these events is $I\,(\text{heads}) = I\,(\text{tails}) = 1$ bit. Therefore what a bit means is the exact amount of information necessary to describe the outcome of a coin tossing. This is why we had set up, by convention, the constant K in Eq. (2.3) to 1: to match up the scale of the measure with the simplest case. The logarithmic function can again be seen to be appropriate as it fits our intuition that if we had two coins the amount of information needed to describe the outcome of both coins would be doubled. For instance, we would have $p\,(\text{heads,heads}) = p\,(\text{heads,tails}) = p\,(\text{tails,heads}) = p\,(\text{tails,tails}) = 1/4$ and therefore $I\,(\text{heads,heads}) = I\,(\text{heads,tails}) = I\,(\text{tails,heads}) = I\,(\text{tails,tails}) = 2$ bits.

Let us now apply this measure in another situation. Consider an unbiased dice. The possible outcomes of rolling a dice are $X = \{1, 2, 3, 4, 5, 6\}$. The probability of each event is $p\,(x = 1) = \cdots = p\,(x = 6) = 1/6$. The amount of information in each outcome is, therefore,

$$I\,(X) = -\log\,(1/6) \approx 2.584 \text{ bits}. \tag{2.6}$$

The above result may seem confusing at first sight. How do we interpret a real amount of information? The meaning of this result is that we need a string of at least 3 symbols to code all possible outcomes unambiguously. Using a string of length 3 we can code the outcomes as: $1 = 001$, $2 = 010$, $3 = 011$, $4 = 100$, $5 = 101$ and $6 = 110$. Note that we are also left with two possible blocks: 000 and 111. Even though the coding is arbitrary, for any coding of length three, there will remain two blocks unused.

There is an alternative interpretation for the amount of information measured in bits: an operationally one. The logical truth values YES and NO can be thought of as a binary alphabet. This alphabet may be interpreted as the answer (outcome) of a YES-or-NO question. How many questions do we need to ask in order to know the outcome of a coin tossing? One question may be: is the result 0? If yes, then we already know the result. If no, by complementarity we know that the result was 1[†]. Therefore, there exists exactly one YES-or-NO question such that when answered

[†]We could have asked about the outcome 1, i.e., it might exist more than one appropriate question.

we know the result unambiguously. The existence of exactly one question is directly associated with the amount of information which is, in this case, 1 bit. In the dice rolling we have 6 equiprobable outcomes. How many questions do we need to ask in order to know the outcome of a dice rolling? The first question which comes to mind is, like before: is the result 3 (or any other)? If the answer is yes then we know the result. If it is no we don't. This is a difference between the coin tossing and the dice rolling. In the coin tossing example, there are only two possible outcomes. So, one such question is enough to know the outcome with certainty. But this is not the case in the dice rolling because it does not have only 1 bit of information. Only with questions of the kind "is the outcome of the dice rolling x ?" we would need six questions to know the outcome with certainty. Let us try to find a small set of questions which unambiguously gives the result when answered. Consider the three following questions:

1. Is the result x even? If YES, then $x \in \{2,4,6\} = U_1$, if NO, then $x \in \{1,3,5\} = U_2$.

2. Is the result x strictly greater than 3? If YES, then $x \in \{4,5,6\} = V_1$, if NO, then $x \in \{1,2,3\} = V_2$.

3. Is the result x divisible by 3? If YES, then $x \in \{3,6\} = W_1$, if NO, then $x \in \{1,2,4,6\} = W_2$.

When these questions are answered they yield precisely the value of x:

- $\{1\} = U_2 \cap V_2 \cap W_2$,i.e., 1 = NO/NO/NO;
- $\{2\} = U_1 \cap V_2 \cap W_2$,i.e., 1 = YES/NO/NO;
- $\{3\} = U_2 \cap V_2 \cap W_1$,i.e., 1 = NO/NO/YES;
- $\{4\} = U_1 \cap V_1 \cap W_2$,i.e., 1 = YES/YES/NO;
- $\{5\} = U_2 \cap V_1 \cap W_2$,i.e., 1 = NO/YES/NO;
- $\{6\} = U_2 \cap V_2 \cap W_2$,i.e., 1 = YES/YES/YES.

Note again that there are two combinations that are not used, YES/NO/YES ($U_1 \cap V_2 \cap W_1 = \emptyset$) and NO/YES/YES ($U_2 \cap V_1 \cap W_1 = \emptyset$). So we used six combinations

of the binary symbols YES and NO out of eight possible ones. This is why the informational content of the dice rolling is between 2 and 3. If the source was given by eight equiprobable symbols $Y = \{1, 2, 3, 4, 5, 6, 7, 8\} = X \cup \{7, 8\}$ we would have 3 bits of information because each symbol would have probability $1/8$. What does the difference $I(Y) - I(X) = 0.416$ mean? Well, it is exactly the amount of information which we gain from source Y when it is no longer equiprobable because its two last symbols have now (from the point of view of the source X) zero probability of occurrence. This excess of probability are redistributed among the other six symbols to form a new source of equiprobable six symbols, namely, X. Each of these six symbols will have their probabilities increased by $8/6$ which in terms of information means $I(\Delta Y) = -\log(8/6) = -0.416$. After this re-arrangement the new source has information $I(X) = I(Y) + I(\Delta Y) = 3 - 0.416 = 2.584$ bits. This supports the interpretation of information as being related to certainty. If there are two symbols that do not show up anymore then we are more certain about the possible outcomes (because we have eliminated two possibilities) and therefore less information is associated.

Summarizing this section, we have seen that the simplest alphabet is composed by the binary digits (bits) 0 and 1 (or YES and NO, or whichever). The measure of information $I(x)$ is dimensionless and its common unit is the bit. The amount of information of a given event, as measured by $I(x)$, is a positive real number. In terms of coding the information in bits, the shortest string of bits to code an event, or a message, of information $I(x)$ is given by the smallest n such that $n \geq I(x)$.

2.3 Shannon Entropy

We have seen how to measure the amount of information of some events. Even though it may seem that this is the central concept in information theory the fact is that it is not. The central concept is entropy, more precisely the Shannon entropy. In his landmark paper [8], Shannon was interested mainly in two situations. Given a source of information, what is the efficient way of encoding this information in order to transmit it through a channel. The second was, given a noisy channel (one that may produce error in the transmission), what is the largest rate of information that can be sent through the channel such that the information can be recovered

reliably after the transmission. These motivated him to seek an improved measure of information. From a small set of hypotheses that this measure should obey he deduced mathematically a function which he called entropy. We are not going to discuss the hypotheses made by Shannon or how he arrived at the expression of this entropy. Instead, we are going to define it and apply it to some examples in order to grasp its meaning because this is what is important when we turn our attention to quantum information theory in Ch. 4.

It turns out that the Shannon entropy (also called source entropy*) is given by the average of the information measure we have been studying so far, i.e.,

$$H(X) = \langle I(x) \rangle = -\sum_{x \in X} p(x) \log p(x) . \tag{2.7}$$

The entropy of a source is, therefore, the average amount of information per source symbol. If the source is equiprobable then the source entropy reduces to the information, i.e.,

$$H(X) = -\sum_{x \in X} p(x) \log p(x) = -\sum_{i=1}^{N} \frac{1}{N} \log \frac{1}{N} = \log N. \tag{2.8}$$

This means that we have been working with the Shannon entropy all along but in a special case. If we use the base two logarithm then the unit of the entropy is bit/symbol.

The entropy of a dice roll is

$$H(X) = \log 6 = 2.584 \text{ bit/symbols}, \tag{2.9}$$

which is the same as the amount of information found in Eq. (2.6) because of the equiprobability of the events, as stated above. Let us consider that we throw two dices in a row and we want to study the sum of the two outcomes. Hence the outcomes of this new source are $X = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. Table 2.1 accounts for the probabilities of each of these events. With the data of the table we can evaluate the source entropy, which is found to be $H(X) = 3.274$ bits/symbol. This means that, on average, each event can be described through a number of bits between 3 and 4, which is expected since we have 11 symbols to represent and 4 bits $\leftrightarrow 2^4 = 16 > 11 > 8 = 2^3 \leftrightarrow 3$ bits. As physically there is no such thing of

*The concept of a source is going to be addressed later on. For now it is enough to say that a source is a set of events (or symbols) with a probability distribution for them.

Sum of the dices	Probability $p(x)$	$-p(x)\log p(x)$
$2 = 1 + 1$	0.027 777 778	0.143 609 03
$3 = 1 + 2 = 2 + 1$	0.055 555 556	0.231 662 5
$4 = 2 + 2 = 3 + 1 = 1 + 3$	0.083 333 333	0.298 746 88
$5 = 4 + 1 = 1 + 4 = 3 + 2 = 2 + 3$	0.111 111 111	0.352 213 89
$6 = 5 + 1 = 1 + 5 = 4 + 2 = 2 + 4 = 3 + 3$	0.138 888 889	0.395 555 13
$7 = 6 + 1 = 1 + 6 = 5 + 2 = 2 + 5 = 4 + 3 = 3 + 4$	0.166 666 667	0.430 827 08
$8 = 6 + 2 = 2 + 6 = 5 + 3 = 3 + 5 = 4 + 4$	0.138 888 889	0.395 555 13
$9 = 6 + 3 = 3 + 6 = 5 + 4 = 4 + 5$	0.111 111 111	0.352 213 89
$10 = 6 + 4 = 4 + 6 = 5 + 5$	0.083 333 333	0.298 746 88
$11 = 6 + 5 = 5 + 6$	0.055 555 556	0.231 662 5
$12 = 6 + 6$	0.027 777 778	0.143 609 03

Table 2.1: Probabilities of the summation of the outcomes of rolling two dices. This table was trascripted from Ref. [7].

rational binary symbols we need four bits to code the outcomes of this source. As in the case considered in Sec. 2.2 we will be left with possible coding blocks unused whichever coding we perform.

There is a nice operational interpretation of the Shannon entropy which we will address in the following [10], but a small reflection at this point is worthy.

As we have discussed in Sec. 2.1, for all practical purposes we don't have interest in the meaning of the messages. We want to study the properties of messages in an objective way. In fact, we have already been doing this when we treated the coin tossing and the dice rolling. Let us think about the kinds of messages we are used to in daily life. Texts in any human language are a complex structure made of sentences, words, letters, syntactic rules, et cetera. Furthermore, exactly because of this complex structure, the probability of a given letter occur within a string of letters (and spaces) depends on the word it finds itself in. Therefore, in order to measure the information that a new letter would have if it filled in the next position in the string we would need to consider too many joint and conditional probabilities. Let us forget that these complications exist and study a relatively simple source and see what we can learn from it.

Suppose we have an alphabet with N letters (or, abstractly, a set with N symbols). Let us consider a machine that prints out letters in a sheet of paper in a

sequential manner. The probability p_i of the letter x_i occur in a given position of the sequence is completely uncorrelated with what has already been written or will be written. This machine is called a signal source without memory and the set $\{x_i, p_i\}$ is called signal ensemble[†] (see Fig. 2.1). From now on a message is going to be a definite sequence of n letters generated by this kind of machine.

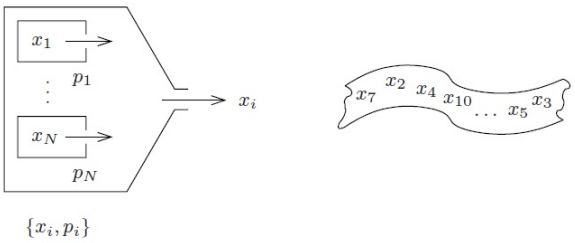

Figure 2.1: Representation of the signal source with signal ensemble $\{x_i, p_i\}$. This figure was copied from Ref. [10]

Alice uses this printing machine to send messages to Bob. She doesn't send the message as a whole but letter by letter. Therefore, Bob does not know what letter may come next but he does know the signal ensemble. Remember our discussion in Sec. 2.1, Bob did know the code.

Consider that Alice prints out a sequence of n letters. If the alphabet contains N different letters, then there are N^n such sequences. The relative frequency of a letter x_i to appear in these sequences is of the order np_i. With a large enough sequence the frequency of a letter x_i may be considered $n_i = np_i$. A sequence such as $x_i x_i \ldots x_i$ is not excluded but has a probability p_i^n, which is very small since $0 \leq p_i \leq 1$ and n is large. So, when n is sufficiently large, Bob may assume that he will only receive the sequences which contain the letters x_i's with the frequencies n_i's. These sequences are called typical sequences. A fairly reasonable question is: how many of these sequences are there? Well, there are $n!$ ways to arrange the n letters. Permutations of the same letter in a sequence does not change it. As there are n_i occurrences of the letter x_i we need to remove the $n_i!$ possible permutations. Hence the number of typical sequences is

$$Z_n = \frac{n!}{n_1! n_2! \ldots n_N!}. \tag{2.10}$$

[†]This set is equivalently represented by a random variable X.

Using the Stirling's formula $\log(n!) \approx n\log n - n$, which is valid when n is large, we obtain for the logarithm of Z_n

$$\log Z_n \approx n\log n - n - \sum_{i=1}^{N}(n_i \log n_i - n_i) = -n\sum_{i=1}^{N} p_i \log p_i. \tag{2.11}$$

If we divide this by n and take the limit $n \to \infty$ we obtain the Shannon entropy

$$H(X) := \lim_{n\to\infty} \frac{1}{n}\log Z_n = -\sum_{i=1}^{N} p_i \log p_i. \tag{2.12}$$

Or equivalently

$$Z_n = 2^{nH(X)}(\text{asymptotically}). \tag{2.13}$$

Since many (few) possibilities reflect a large (small) measure of *a priori* uncertainty for Bob, $H(X)$ from Eq. (2.12) is a measure of the mean *a priori* uncertainty of a character which is received by Bob. Then H is at the same time the average information which Bob receives per transmitted character. This result is a simplified way of stating the Shannon's source coding theorem.

As we said previously the simplest source is composed by two bits. Let us calculate the Shannon entropy for this source. The Shannon entropy is simply

$$H(X) = -p\log p - (1-p)\log(1-p), \tag{2.14}$$

where p is the probability to get the bit 0, for example. This function is also called binary entropy function and it is plotted in Fig. 2.2.

We see that the Shannon entropy has a maximum, when the two bits have the same probability of occurrence, namely, 1/2, and it has two minima, when one of the bits has certainty of occurrence. The maximum situation is the one which we are more uncertain about the next symbol given by the source. Exactly because there is the same probability to occur either 0 or 1, there is no way to discriminate. As the probability of one symbol is increasing, the entropy is reduced because we know that one symbol occurs more frequently than the other. The extreme case is when the source just prints out one symbol. In this case we are completely certain of the next symbol and therefore the entropy reaches zero.

In this section we studied the central concept of information theory, the Shannon entropy. The Shannon entropy is interpreted as the average amount of information per symbol associated with a given source. Next we are going to extend this notion and define entropies that enable the study of correlation between two or more sources.

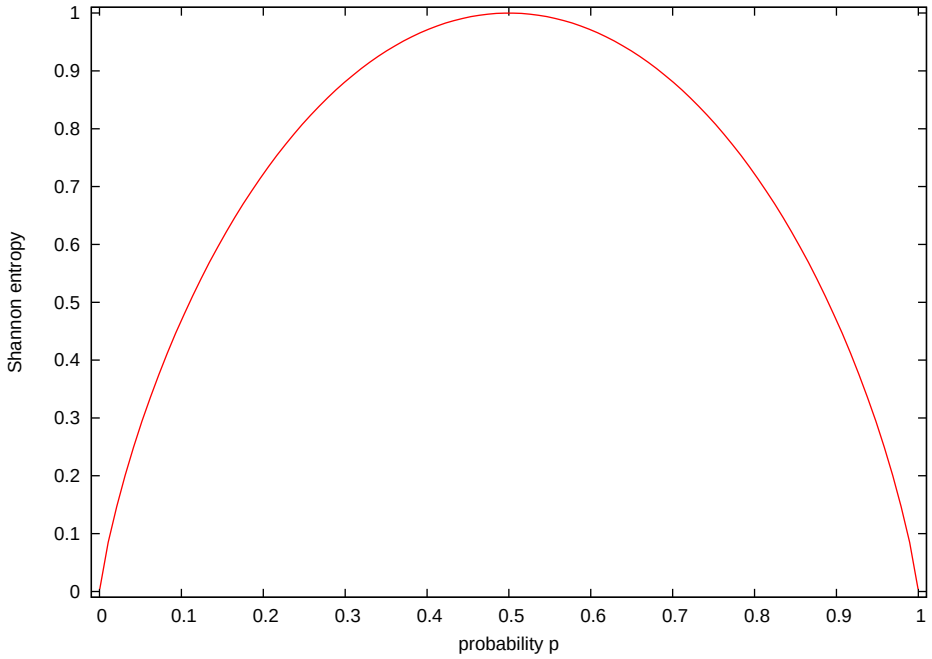

Figure 2.2: Shannon entropy for a 1-bit source.

2.4 Mutual Information and Other Entropies

We have been discussing about events, messages, random variables, information and entropy. Both information and entropy are quantities which depend on one message (or one random variable). In this chapter we shall deal with other entropies, all of which tell us about the correlations between two (or more) messages. We shall answer questions like: How much information do these two messages share in common? How much information do I gain reading this message given I've read the other? As we shall see in Ch. 4 these entropies play a central role in quantum information theory. We will consider a concrete example and define these entropies as we go through.

Consider two event sources related with a public company [7]. The first event source is about three possible conclusions for the quarterly sales report, namely $x \in X = \{\text{good}, \text{same}, \text{bad}\} = \{x_1, x_2, x_3\}$, meaning that the results are in excess of the predictions (good), or on target (same), or under target (bad). The second event source is the company's stock value, as reflected by the stock exchange, with $y \in Y = \{\text{up}, \text{steady}, \text{down}\} = \{y_1, y_2, y_3\}$. Then assume that there exists some form of correlation between the company's result ($x \in X$) and its stock value ($y \in Y$). A

		y_1 (up)	y_2 (steady)	y_3 (down)				
$p(x_i, y_j)$	p	0.300	0.500	0.200	$\sum =$			
x_1 (good)	**0.160**	0.075	0.075	0.010	**0.160**			
x_2 (same)	**0.750**	0.210	0.400	0.140	**0.750**			
x_3 (bad)	**0.090**	0.015	0.025	0.050	**0.090**			
	$\sum =$	**0.300**	**0.500**	**0.200**	**1.000**			
		y_1 (up)	$p(y_1, x_i)$	y_2 (steady)	$p(y_2, x_i)$	y_3 (down)	$p(y_3, x_i)$	
$p(y_j, x_i)$	p	0.300	$= p(y_1\|x_i)p(x_i)$	0.500	$= p(y_2\|x_i)p(x_i)$	0.200	$= p(y_3\|x_i)p(x_i)$	
x_1 (good)	**0.160**	0.469	0.075	0.469	0.075	0.063	0.010	
x_2 (same)	**0.750**	0.280	0.210	0.533	0.400	0.187	0.140	
x_3 (bad)	**0.090**	0.167	0.015	0.278	0.025	0.556	0.050	
	$\sum =$		**0.300**		**0.500**		**0.200**	**1.000**
		x_1 (good)	$p(x_1, y_j)$	x_2 (same)	$p(x_2, y_j)$	x_3 (bad)	$p(x_3, y_j)$	
$p(x_i, y_j)$	p	0.160	$= p(x_1\|y_j)p(y_j)$	0.750	$= p(x_2\|y_j)p(y_j)$	0.090	$= p(x_3\|y_j)p(y_j)$	
y_1 (up)	**0.300**	0.250	0.075	0.700	0.210	0.050	0.015	
y_2 (steady)	**0.500**	0.150	0.210	0.800	0.400	0.050	0.025	
y_3 (down)	**0.200**	0.010	0.015	0.700	0.100	0.250	0.050	
	$\sum =$		**0.300**		**0.750**		**0.090**	**1.000**

Table 2.2: Numerical example of the quarterly sales report and the company's stock value sources. This table was trascripted from Ref. [7].

numerical example of joint and conditional probability data, $p(x_i, y_j)$, $p(x_i|y_j)$ and $p(y_j|x_i)$ is shown in Table 2.2.

The first group of numerical data, shown at the top of Table 2.2, corresponds to the joint probabilities $p(x_i, y_j)$. Summing up the data by rows (i) or by columns (j) yields the probabilities $p(x_i)$ or $p(y_j)$, respectively. The double checksum (bottom and right) which yields unity through summing by row or by column is also shown for consistency. The two other groups of numerical data in Table 2.2 correspond to the conditional probabilities $p(x_i|y_j)$ and $p(y_j|x_i)$. These are calculated through Bayes' rule $p(x_i|y_j) = p(x_i, y_j)/p(y_j)$ and $p(y_j|x_i) = p(x_i, y_j)/p(x_i)$. The intermediate columns providing the data $p(x_i|y_j)\,p(y_j) = p(y_j|x_i)\,p(x_i) = p(x_i, y_j)$ and their checksums by column are shown in the table for consistency.

We define the joint entropy $H(X, Y)$ associated with the joint distribution $p(x, y)$,with $x \in X$ and $y \in Y$, as

$$H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x, y). \tag{2.15}$$

It is easily recognizable to be a natural extension of the Shannon entropy for a joint

probability distribution. This entropy represents the average information derived from joint events occurring from two sources X and Y. The unit of $H(X,Y)$ is bit/symbol.

We then define the conditional entropy $H(X|Y)$ through

$$H(X|Y) = -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(x|y)\,. \tag{2.16}$$

The conditional entropy $H(X|Y)$ corresponds to the average information conveyed by the conditional probability distribution $p(x|y)$. Put simply, $H(X|Y)$ represents the information we learn from source X given the information we have from source Y. Its unit is also bit/symbol.

Note that in the joint and conditional entropy definitions the two-dimensional averaging over the event space $\{X,Y\}$ consistently involves the joint distribution $p(x,y)$.

The conditional entropy $H(Y|X)$ is defined in the same way, namely,

$$H(Y|X) = -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(y|x)\,. \tag{2.17}$$

As the conditional probabilities $p(x|y)$ and $p(y|x)$ are in general different, the two conditional entropies will also be different.

If the two sources are statistically independent then $p(x,y) = p(x)\,p(y)$. Consequently, this implies that

$$\begin{aligned} H(X,Y) &= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(x,y) \\ &= -\sum_{x\in X} p(x)\log p(x) - \sum_{y\in Y} p(y)\log p(y) = H(X) + H(Y)\,. \end{aligned} \tag{2.18}$$

Moreover $p(x|y) = p(x)$, then

$$\begin{aligned} H(X|Y) &= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(x|y) \\ &= -\sum_{x\in X} p(x)\log p(x)\left(\sum_{y\in Y} p(y)\right) = H(X) \end{aligned} \tag{2.19}$$

and

$$\begin{aligned} H(Y|X) &= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(y|x) \\ &= -\sum_{y\in Y} p(y)\log p(y)\left(\sum_{x\in X} p(x)\right) = H(Y)\,. \end{aligned} \tag{2.20}$$

These results match with our intuition that if the two sources are statistically independent the knowledge of one of them does not provide any advance knowledge about the other.

In the general case where the sources are not necessarily statistically independent we find a relation between the joint and conditional entropies given by

$$\begin{aligned} H(X,Y) &= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(x,y) \\ &= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(x|y) - \sum_{y\in Y} p(y)\log p(y) \\ &= H(X|Y) + H(Y), \end{aligned} \tag{2.21}$$

or equivalently,

$$\begin{aligned} H(X,Y) &= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(x,y) \\ &= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(y|x) - \sum_{x\in X} p(x)\log p(x) \\ &= H(Y|X) + H(X). \end{aligned} \tag{2.22}$$

These expressions may be easier to memorize under the forms

$$H(X|Y) = H(X,Y) - H(Y), \tag{2.23}$$

$$H(Y|X) = H(X,Y) - H(X), \tag{2.24}$$

which state that, given a source X (Y), any advance knowledge from the other source Y (X) reduces the joint entropy $H(X,Y)$ by the net amount $H(Y)$ $(H(X))$, respectively. It will be shown in Sec. 2.4.1 that all these quantities are positive. In other words, the prior information one may gain from a given source is made at the expense of the information available from the other source, unless the two are statistically independent.

We can illustrate the above properties through our stock-exchange probability distributions' data from Table 2.2. We want, here, to determine how the average information from the company's sales, $H(X)$, is affected by that concerning the stocks, $H(Y)$, and vice versa. The computations of $H(X)$, $H(Y)$, $H(X,Y)$, $H(X|Y)$ and $H(Y|X)$ are detailed in Fig. 2.3. It is seen from the table that the results and stock entropies compute to $H(X) = 1.046$ and $H(Y) = 1.485$, respectively (for

$u_i = p(x_i)$
$v_j = p(y_j)$

	u_i	$-u_i \log u_i$		v_j	$-v_j \log v_j$
$\mathbf{x_1}$**(good)**	0.160	0.423	$\mathbf{y_1}$**(up)**	0.300	0.521
$\mathbf{x_2}$**(same)**	0.750	0.311	$\mathbf{y_2}$**(steady)**	0.500	0.500
$\mathbf{x_3}$**(bad)**	0.090	0.313	$\mathbf{y_3}$**(down)**	0.200	0.464
	$H(X) =$	1.0469		$H(Y) =$	1.4855
$H(X) +$	$H(Y) =$	2.5324			

$u_{ij} = p(x_i, y_j)$

	u_{i1}	$-u_{i1} \log u_{i1}$	u_{i2}	$-u_{i2} \log u_{i2}$	u_{i3}	$-u_{i3} \log u_{i3}$
$\mathbf{x_1}$	0.075	0.280	0.075	0.280	0.010	0.066
$\mathbf{x_2}$	0.210	0.473	0.400	0.529	0.140	0.397
$\mathbf{x_3}$	0.015	0.091	0.025	0.133	0.050	0.216
	$\sum =$	0.844	$\sum =$	0.942	$\sum =$	0.680
$H(X,Y) =$	2.4657					

$v_{ji} = p(y_j|x_i)$

	v_{1i}	u_{i1}	$-u_{i1} \log v_{1i}$	v_{2i}	u_{i2}	$-u_{i2} \log v_{2i}$	v_{3i}	u_{i3}	$-u_{i3} \log v_{3i}$
$\mathbf{x_1}$	0.469	0.075	0.082	0.469	0.075	0.082	0.063	0.010	0.040
$\mathbf{x_2}$	0.280	0.210	0.386	0.533	0.400	0.363	0.187	0.140	0.339
$\mathbf{x_3}$	0.167	0.015	0.039	0.278	0.025	0.046	0.556	0.050	0.042
		$\sum =$	0.506		$\sum =$	0.491		$\sum =$	0.421
$H(Y\|X) =$	1.4188								

$w_{ij} = p(x_i|y_j)$

	w_{1j}	u_{1j}	$-u_{i1} \log w_{1j}$	w_{2j}	u_{2j}	$-u_{i2} \log w_{12}$	w_{3j}	u_{3j}	$-u_{31} \log w_{3j}$
$\mathbf{y_1}$	0.250	0.075	0.150	0.700	0.210	0.108	0.050	0.015	0.065
$\mathbf{y_2}$	0.150	0.075	0.205	0.800	0.400	0.129	0.050	0.025	0.108
$\mathbf{y_3}$	0.050	0.010	0.043	0.700	0.140	0.072	0.250	0.050	0.100
		$\sum =$	0.398		$\sum =$	0.309		$\sum =$	0.421

$H(X\|Y) =$	0.9802	
$H(X,Y) =$	H$(Y\|X) +$	$H(X)$
$2.4657 =$	$1.4188+$	1.0469
$=$	$H(X\|Y) +$	$H(Y)$
$=$	$0.9802+$	1.4855

qwe

Table 2.3: Entropies for stock-exchange example. This table was trascripted from Ref. [7].

easier reading, we omit here the bit/symbol units). The joint entropy is found to be $H(X,Y) = 2.466$, which is lower than the sum $H(X) + H(Y) = 2.532$. This proves that the two sources are not independent, namely, that they have some information in common. We find that the conditional entropies satisfy

$$H(Y|X) = 1.418 < H(Y) = 1.485, \tag{2.25}$$

$$H(X|Y) = 0.980 < H(X) = 1.046, \tag{2.26}$$

or equivalently, using four decimal places, for accuracy (see Fig. 2.3)

$$H(Y) - H(Y|X) = 1.4855 - 1.4188 = 0.0667, \tag{2.27}$$

$$H(X) - H(X|Y) = 1.0469 - 0.9802 = 0.0667. \tag{2.28}$$

These two results mean that the prior knowledge of the company's stocks contain an average of 0.0667 bit/symbol of information on the company's quarterly result, and vice versa. As we shall see below, the two differences above are always equal and they are called mutual information. Simply put, the mutual information is the average information that two sources share in common.

We can define the mutual information of two sources X and Y as the bit/symbol quantity

$$H(X;Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)}. \tag{2.29}$$

We may note the absence of a minus sign in the above definition, unlike in $H(X,Y)$, $H(Y|X)$ and $H(X|Y)$. Also note the ";" separation, which distinguishes mutual information from joint entropy $H(X,Y)$. Mutual information is also often referred to in the literature as $I(X;Y)$ instead of $H(X;Y)$.

Since the logarithm argument is unity when the two sources are statistically independent, we immediately observe that the mutual information is equal to zero in this case. This reflects the fact that independent sources do not have any information in common.

It is pertinent, now, to derive some formulas which relate these different entropies. Using the properties of the logarithmic function, the Bayes' rule and the relations $\sum_x p(x,y) = p(y)$ and $\sum_y p(x,y) = p(x)$ for the joint probability distri-

bution, the mutual information can be written in the following ways

$$\begin{aligned} H(X;Y) &= \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)\,p(y)} \\ &= -\sum_{x\in X} p(x)\log p(x) + \sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(x|y) \\ &= H(X) - H(X|Y), \end{aligned} \tag{2.30}$$

$$\begin{aligned} H(X;Y) &= \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)\,p(y)} \\ &= -\sum_{y\in Y} p(y)\log p(y) + \sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(y|x) \\ &= H(Y) - H(Y|X) \end{aligned} \tag{2.31}$$

and

$$H(X;Y) = H(X) + H(Y) - H(X,Y), \tag{2.32}$$

where in the last formula one can use either Eq. (2.30) or Eq. (2.31) together with Eqs. (2.23). The first two equalities above confirm the observation derived from our previous numerical example. They can be interpreted according to the following statement: mutual information is the reduction of uncertainty in X that we get from the knowledge of Y (and vice versa).

The last equality, as rewritten under the form

$$H(X,Y) = H(X) + H(Y) - H(X;Y), \tag{2.33}$$

shows that the joint entropy of two sources is generally less than the sum of the source entropies. The difference is the mutual information that the sources have in common, which reduces the net uncertainty or joint entropy.

Finally, we note from the Eqs. (2.30), (2.31) and (2.32) that the mutual information is symmetrical with respect to its arguments, namely, $H(X;Y) = H(Y;X)$, as is expected from its very meaning.

The different entropies introduced up to this point may seem a bit abstract and their different relations apparently is not very practical to memorize. But the situation becomes different after we draw an analogy with the property of ensembles.

Consider, indeed, two ensembles, called A and B. The two ensembles may be united to form a whole, which is denoted as $F = A \cup B$ (A union B). The two

ensembles may or may not have elements in common. The set of common elements is called $G = A \cap B$ (A intersection B). The same definitions of union and intersection apply to any three ensembles A, B, and C. Fig. 2.3 shows the Venn diagram representations of such ensemble combinations. In the case of two ensembles, there exist four subset possibilities, as defined by their elements's properties:

- Elements common to A or B: $A \cup B$,
- Elements common to A and B: $A \cap B$,
- Elements from A and not B: $A \cap \neg B$,
- Elements from B and not A: $B \cap \neg A$.

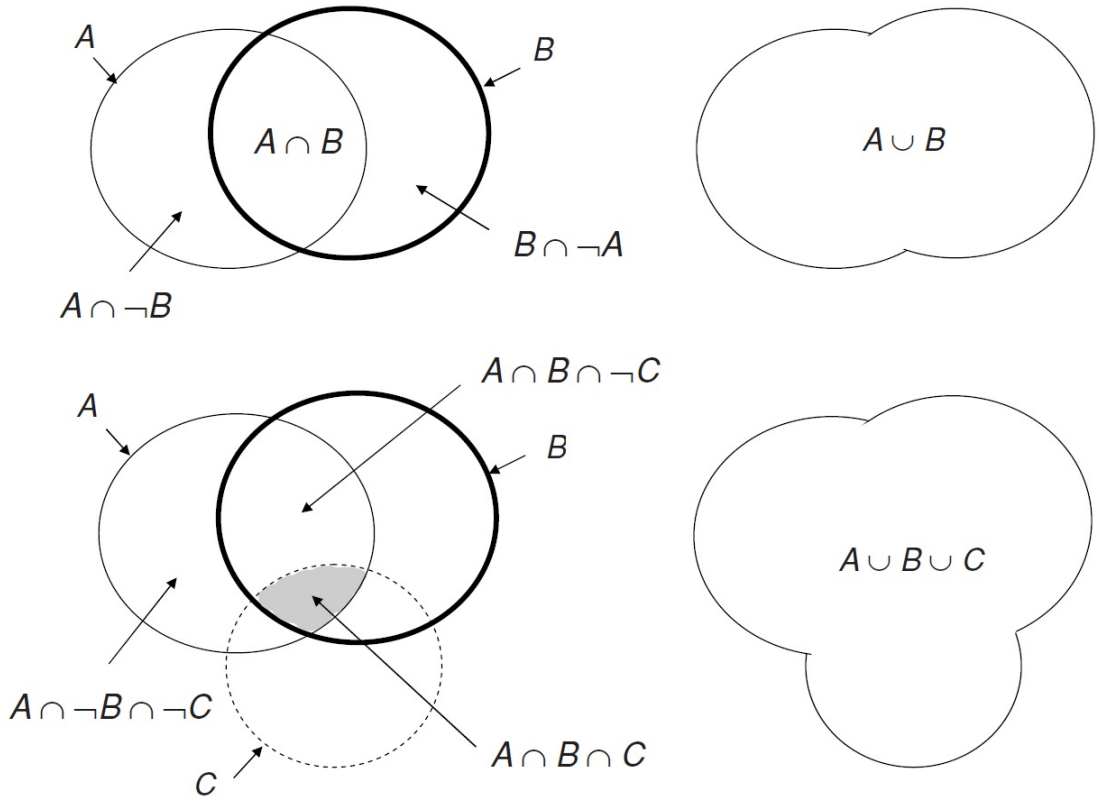

Figure 2.3: Venn diagram representation of two (top) and three (bottom) ensembles.This figure has been taken from Ref. [7].

In the conventional notations shown at right, we see that the symbol $\cup$ stands for a logical OR, the symbol $\cap$ stands for a logical AND, and the symbol $\neg$ stands for a

logical NO. These three different symbols, which are also called Boolean operators*, make it possible to perform various mathematical computations in the field called Boolean logic. In the case of three ensembles A, B, C, we observe that there exist many more subset possibilities (e.g., $A \cap B \cap C$, $A \cap B \cap \neg C$). The interest of the above visual description with the Venn diagrams is the straightforward correspondence with the various entropies that have been introduced. Based on Eqs. (2.18), (2.19), (2.20), (2.23), (2.30), (2.31) and (2.32) we can draw a Venn diagram which respects them and see that the following equivalences hold (compare Fig. 2.3 with Fig. 2.4):

$$\begin{cases} H(X,Y) \leftrightarrow H(X \cup Y) \\ H(X;Y) \leftrightarrow H(X \cap Y) \\ H(X|Y) \leftrightarrow H(X \cap \neg Y) \\ H(Y|X) \leftrightarrow H(\neg X \cap Y) \end{cases} \tag{2.34}$$

The first equivalence in Eq. (2.34) means that the joint entropy of two sources is the entropy of the source defined by their combined events. The second equivalence in Eq. (2.34) means that the mutual information of two sources is the entropy of the source containing the events they have in common. The last two equivalences in Eq. (2.34) provide the relationship with the conditional entropy. For the case of the conditional entropy of a source X given the information on a source Y, the conditional entropy is given by the contributions of all the events belonging to X but not to Y. Fig. 2.4 illustrates all the above logical equivalences through Venn diagrams, using up to three sources. Considering the two-source case, we can immediately visualize from Fig. 2.4 to which subsets the differences $H(X) - H(X;Y)$ and $H(Y) - H(X;Y)$ actually correspond. Given the identities listed in Eq. (2.34), we can call these two subsets $H(X|Y)$ and $H(Y|X)$, respectively, which proves the previous point. We also observe from the Venn diagram that $H(X|Y) \leq H(X)$ and $H(Y|X) \leq H(Y)$, with equality if the sources are independent.

The above property can be summarized by the statement according to which conditioning reduces entropy. A formal demonstration, using the concept of "relative entropy" is provided later.

The three-source case, as illustrated in Fig. 2.4, is somewhat more tricky because

*These are not exactly the symbols which usually represent the Boolean operators but this is not relevant to us.

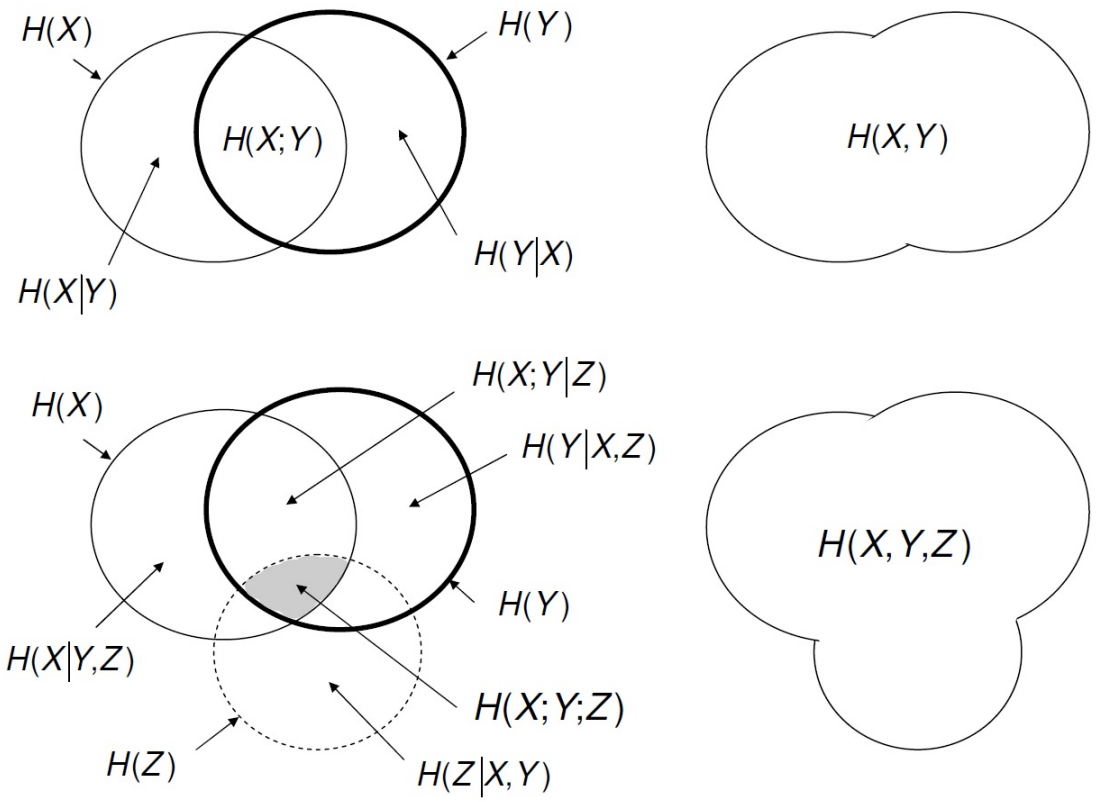

Figure 2.4: Venn diagram representation for the different entropies.This figure has been taken from Ref. [7].

it generates more complex entropy definitions with three arguments X, Y, and Z. Conceptually, defining joint or conditional entropies and mutual information with three (or more) sources is not that difficult. Considering the joint probability $p(x, y, z)$ for the three sources, we can indeed generalize the previous two-source definitions according to the following:

$$H(X, Y, Z) = -\sum_{x \in X} \sum_{y \in Y} \sum_{z \in Z} p(x, y, z) \log p(x, y, z), \tag{2.35}$$

$$H(X; Y; Z) = +\sum_{x \in X} \sum_{y \in Y} \sum_{z \in Z} p(x, y, z) \log \frac{p(x, y, z)}{p(x) p(y) p(z)}, \tag{2.36}$$

$$H(Z|X, Y) = -\sum_{x \in X} \sum_{y \in Y} \sum_{z \in Z} p(x, y, z) \log p(z|x, y), \tag{2.37}$$

$$H(X, Y|Z) = -\sum_{x \in X} \sum_{y \in Y} \sum_{z \in Z} p(x, y, z) \log p(x, y|z). \tag{2.38}$$

These four definitions correspond to the joint entropy of the three sources X, Y, Z (Eq. (2.35)), the mutual information of the three sources X, Y, Z (Eq.(2.36)), the

entropy of source Z given the known entropy of X, Y (Eq. (2.37)), and the joint entropy of X, Y given the known entropy of Z (Eq. (2.38)). The last two definitions are seen to involve conditional probabilities of higher orders, namely, $p(z|x,y)$ and $p(x,y|z)$, which are easily determined from the generalization of Bayes' rule [†]. Other entropies of the type $H(X;Y|Z)$ and $H(X|Y;Z)$ are more tricky to determine from the above definitions. But we can resort in all confidence to the equivalence relations and the corresponding two-source or three-source Venn diagrams shown in Fig. 2.4. Indeed, a straightforward observation of the diagrams leads to the following correspondences:

$$H(X;Y|Z) = H(X;Y) - H(Z), \tag{2.39}$$

$$H(X|Y;Z) = H(X) - H(Y;Z). \tag{2.40}$$

Finally, the Venn diagrams (with the help of Eq. (2.34)) make it possible to establish the following properties for $H(X,Y|Z)$ and $H(X|Y,Z)$. The first is

$$H(X,Y|Z) = H(X|Z) + H(Y|X,Z), \tag{2.41}$$

which is easy to memorize if a condition $|Z$ is applied to both sides of the definition of joint entropy, $H(X,Y) = H(X) + H(Y|X)$. The second is

$$H(X,Y|Z) = H(Y|Z) + H(X|Y,Z), \tag{2.42}$$

comes from the permutation in Eq. (2.41) of the sources X, Y, since the joint entropy $H(X,Y)$ is symmetrical with respect to its arguments. The lesson learned from using Venn diagrams is that there is, in fact, little to memorize, as long as we are allowed to make drawings. The only general rule to remember is:

$H(U|Z)$ is equal to the entropy $H(U)$ defined by the source U (for instance, $U = X,Y$ or $U = X;Y$) minus the entropy $H(Z)$ defined by the source Z, the reverse being true for $H(Z|U)$. But the use of Venn diagrams require us not to forget the unique correspondence between the ensemble or Boolean operators ($\cup$ $\cap$ $\neg$) and the separators (, ; |) in the entropy-function arguments.

[†]As we have

$$p(x,y,z) = p(z|x,y)\,p(x)\,p(y) \rightarrow p(z|x,y) = p(x,y,z) \,/\, [p(x)\,p(y)]$$

and

$$p(x,y,z) = p(x,y|z)\,p(z) \rightarrow p(x,y|z) = p(x,y,z) \,/ p(z).$$

2.4.1 Relative Entropy

Relative entropy is our last word on the different kinds of (classical) entropy in this work. It is related to the concept of distance between two probability distributions. It is not strictly a distance because it is not symmetrical, for instance[‡]. The actual usefulness of this entropy can only be appreciated when one studies advanced topics in information theory. For us it is going to be useful as a preparation for the quantum relative entropy in Ch. 4 and to prove some inequalities about the other entropies.

We shall introduce the relative entropy between two probability distribution functions, which is also called the Kullback-Leibler (KL) divergence. Consider two probability distribution functions, $p(x)$ and $q(x)$, where the argument x belongs to a single source X. The relative entropy is denoted by $D[p(x)||q(x)]$ and is defined as follows:

$$D[p(x)||q(x)] = \left\langle \log \frac{p(x)}{q(x)} \right\rangle_p = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}. \tag{2.43}$$

To show that the relative entropy is nonnegative we will need the following inequality for logarithmic functions $\log(x)\ln(2) = \ln(x) \leq x - 1$, or in a more convenient way

$$-\log(x) \geq \frac{1-x}{\ln 2}. \tag{2.44}$$

Hence it follows

$$\begin{aligned} H(p||q) &= -\sum_{x \in X} p(x) \log \frac{q(x)}{p(x)} \\ &\geq \frac{1}{\ln 2} \sum_{x \in X} p(x) \left(1 - \frac{q(x)}{p(x)}\right) = \frac{1}{\ln 2} \sum_{x \in X} (p(x) - q(x)) = 0. \end{aligned} \tag{2.45}$$

This inequality is called Gibbs' inequality. The equality only holds when the two distributions are the same, i.e., $p(x) = q(x)$.

An important particular case occurs when the distribution $q(x)$ is a uniform distribution. If the source X has N events, the uniform probability distribution is

[‡]It also does not satisfy the triangle inequality. But it does satisfy the non-negativity as we shall show.

thus given by $q(x) = 1/N$. Replacing this definition in Eq. (2.43) yields

$$\begin{aligned} D[p(x)||q(x)] &= \sum_{x\in X} p(x) \log \frac{p(x)}{1/N} \\ &= \log N \sum_{x\in X} p(x) + \sum_{x\in X} p(x) \log p(x) = \log N - H(X). \end{aligned} \tag{2.46}$$

Since the distance $D[p||q]$ is always nonnegative, it follows from the above that $H(X) \leq \log N$. This result shows that the entropy of a source X with N elements has $\log N$ for its upper bound, which (in the absence of any other constraint) represents the entropy maximum. Assume next that p and q are joint distributions of two variables x, y. Similarly to the definition in Eq. (2.43), the relative entropy between the two joint distributions is

$$D[p(x,y)||q(x,y)] = \left\langle \log \frac{p(x,y)}{q(x,y)} \right\rangle_p = \sum_{x\in X} \sum_{y\in Y} p(x,y) \log \frac{p(x,y)}{q(x,y)}. \tag{2.47}$$

The relative entropy is also related to the mutual information. Indeed, recalling the definition of mutual information (cf. Eq. (2.29)) we get

$$H(X;Y) = \sum_{x\in X} \sum_{y\in Y} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)} = D[p(x,y)||p(x)\,p(y)], \tag{2.48}$$

which shows that the mutual information between two sources X, Y is the relative entropy (or KL divergence) between the joint distribution $p(x,y)$ and the distribution product $p(x)\,p(y)$. Since the relative entropy is always nonnegative, it follows that mutual information is always nonnegative. Indeed, recalling the Eqs. (2.30) and (2.31) together with the Gibbs' inequality yields

$$H(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \geq 0, \tag{2.49}$$

which thus implies the two inequalities

$$H(X|Y) \leq H(X), \tag{2.50}$$

$$H(Y|X) \leq H(Y). \tag{2.51}$$

The above result can be summarized under the fundamental conclusion, which has already been established: conditioning reduces entropy. Thus, given two sources X and Y, the information we obtain from source X, given the prior knowledge of the information from Y, is less than or equal to that available from X alone, meaning that entropy has been reduced by the fact of conditioning. The strict inequality

applies in the case where the two sources have nonzero mutual information. If the two sources are disjoint, or made of independent events, then the equality applies, and conditioning from Y has no effect on the information of X.

Even though we are not going to reach the actual interpretation of the relative entropy we are going to state it [11]: The relative entropy $D\left[p\left(x\right)||q\left(x\right)\right]$ is a measure of the information lost when the probability distribution q is used to approximate the distribution p.

Chapter 3

Quantum Dynamics

Quantum mechanics is one of the greatest theoretical developments ever made by human beings. It dictates how our subworld behaves. By subworld we mean the world beneath our world, the world of invisible (to the naked eye) entities, the world which goes beyond our senses. The study of this invisible subworld has brought almost every kind of technology we use today and any which might be in ongoing process of development.

Since the beginning of quantum mechanics until nowadays there are some researchers who are interested in its interpretation. There are many interpretations that have been put forth, the most famous being the Copenhagen interpretation formulated by Bohr and Heisenberg. Among these interpretations there is one that can be viewed as minimal, in the sense that it provides a concrete-based interpretation for the mathematical objects of the theory. This is the statistical (or ensemble) interpretation [12] which takes the Born's rule to its fullest extent. The wavefunction is viewed as a mathematical tool to describe statistically the possible outcomes of an identically prepared ensemble of quantum systems. This interpretation is older than of what we may call today the operational interpretations. An operational interpretation attempts to interpret the properties of quantum systems and mathematical objects of the theory in terms of their practical results, in the same way as in the statistical interpretation. The statistical interpretation touches basically the interpretational problem of the wavefunction but, on the other hand, does not say a word with respect to other fundamental concepts of the theory such as entanglement. From this foundational point of view, quantum information theory sheds light on new ways to understand such concepts. Quantum entanglement is seen as a resource which enables some efficient algorithms to solve tasks in a quantum

computer.

In Ref. [13] one may find a recent poll carried out in a 2011 conference on quantum foundations presenting the viewpoint of the participants about foundational issues in quantum theory. In this poll, it can be seen how quantum informational concepts merged with quantum mechanics. Some questions asked to the participants that manifest this are: "What about quantum information?", whose most voted answer was "It's a breath of fresh air for quantum foundations"; "When will we have a working and useful quantum computer?", whose most voted answer was "In 10 to 25 years"; "What interpretation of quantum states do you prefer?", where the epistemic/informational interpretation is one that received considerable votes; and finally "What is your favorite interpretation of quantum mechanics", with the information-based/information-theoretical as the second most voted, behind only by the traditional Copenhagen interpretation.

The aim of this chapter is to provide a unified way to describe all possible quantum dynamics; from time evolution to measurement. In Sec. 3.1, we will present the dynamics of closed systems. We will study the pure and mixed states and then as an example the dynamics of a qubit. In Sec. 3.2, we will set up the mathematical language to describe systems of many particles (subsystems), introducing fundamental concepts such as entanglement and partial trace. In Sec. 3.3, we will study, in a fairly general way, the dynamics of open quantum systems. We will study the properties of general dynamical maps and derive a differential equation for the density operator [10,14]. In Sec. 3.4, we will study the generalized description of measurement dynamics. Following this, in Sec. 3.5, we will conclude our purpose showing an unified picture of quantum dynamics in terms of quantum operations. As an application of the formalism developed in the previous sections, we will address the phenomenon of decoherence in Sec. 3.6 and see that it is intimately related with another phenomenon called environment-induced superselection. Most of this chapter was based on Ref. [10].

3.1 Dynamics of Closed Quantum Systems

In order to discuss the dynamics of closed quantum systems we have to know what a closed system is. For the moment we consider that the universe is everything that

exists * and we can divide it into two distinct parts, the system and the environment. The system is that part of the universe in which we are interested to describe. For example, a collection of atoms or molecules. The environment is all that is in the universe and which doesn't fit into the system. Being established these concepts, a closed system is a system that, for all practical purposes, we can neglect all spoiler effects of the environment over the system. So for example, we are interested in studying the kinematics of a marble rolling freely † in a flat table, say, to verify Newton's laws. In this example the system would be the marble. The environment, consisted of the table and everything else, does affect the system dynamics, namely, the table exerts a normal force which balances the weight of the marble. But this effect is necessary to maintain the marble's movement constrained to the plane, which is the region where we are interest to study the movement. Therefore this influence does not spoil the experiment. On the other hand, the friction with the local atmosphere or with the table could spoil our attempt to verify Newton's law, as they would affect the movement in the plane. If these effects can be made negligible by an appropriate experimental setup then the marble is considered a closed system and we will verify Newton's laws. Let us begin to study the quantum dynamics of such a systems.

3.1.1 Dynamics of Pure and Mixed States

When we begin to learn the mathematical formulation of quantum mechanics we put the Schrödinger equation as a postulate. The Schrödinger equation is a differential equation which gives the dynamics of a state vector $|\psi_0\rangle$. The equation is

$$\begin{cases} i\hbar \frac{d}{dt} |\psi(t)\rangle = H(t) |\psi(t)\rangle & , \\ |\psi(0)\rangle = |\psi_0\rangle & . \end{cases} \tag{3.1}$$

For the sake of completeness let us establish a common language and notation at this point. The symbol i stands for the imaginary unit and $\hbar$ for the Planck's constant. The state vector $|\psi(t)\rangle$ stands for the mathematical representation, within the theory, of a pure state. A state vector is a vector in some Hilbert space $\mathcal{H}$,

*We are not concerned with philosophical problems which may be raised with these concepts but just setting up a common conceptual ground to understand the concept of closed systems.

†Freely means that we could set up an experiment such that the dynamical friction coefficient between the marble and the table is negligible.

also hereafter referred to as state space[‡]. The object $H(t)$ is the Hamiltonian of the system. It provides the structure of the dynamical evolution of a quantum system. The Hamiltonian is an operator over the state space $\mathcal{H}$. It lives in another Hilbert space denoted by $\mathcal{L}(\mathcal{H})$, the space of linear operators over $\mathcal{H}$. All state vectors must be normalized in order to preserve the probabilistic characteristics of quantum mechanics. The normalization condition is mathematically stated as $\langle \psi, \psi \rangle = \langle \psi \,|\, \psi \rangle = 1$, where $\langle \cdot \,|\, \cdot \rangle$ is the inner product of the Hilbert space $\mathcal{H}$. As we shall see later on the Schrödinger equation preserves the normalization condition as the time goes by.

From the Schrödinger equation we can deduce an algebraic equation for the time evolution of the state vector. To deduce this equation we first rewrite the Schrödinger equation in its integral form,

$$|\psi(t)\rangle = |\psi_0\rangle + \frac{1}{i\hbar}\int_0^t dt' H(t')\,|\psi(t')\rangle\,. \tag{3.2}$$

Now we substitute this equation on itself, leading us to

$$\begin{aligned} |\psi(t)\rangle = &|\psi_0\rangle + \frac{1}{i\hbar}\int_0^t dt_1 H(t_1)\,|\psi_0\rangle \\ &+ \left(\frac{1}{i\hbar}\right)^2 \int_0^t dt_1 H(t_1) \int_0^{t_1} dt_2 H(t_2)\,|\psi(t_2)\rangle\,. \end{aligned} \tag{3.3}$$

Observe the time ordering $t \geq t_1 \geq t_2 \geq 0$ in the second integral. The integrals $\int_0^t dt_1 \int_0^{t_1} dt_2$ can be written as $\frac{1}{2}\int_0^t dt_1 \int_0^t dt_2$. With this relation the above equation can be rewritten as

$$\begin{aligned} |\psi(t)\rangle = &\left[1 + \left(-\frac{i}{\hbar}\right)\int_0^t dt_1 H(t_1) + \left(-\frac{i}{\hbar}\right)^2 \frac{1}{2!}\int_0^t dt_1 \int_0^t dt_2 H(t_1) H(t_2)\right] |\psi_0\rangle \\ &+ O\left(\frac{1}{\hbar^3}\right). \end{aligned} \tag{3.4}$$

Performing some more steps one can convince itself that we can write Eq. (3.2) algebraically as

$$|\psi(t)\rangle = U(t)\,|\psi_0\rangle\,, \tag{3.5}$$

where $U(t)$ is the time evolution operator

$$U(t) = T_{\leftarrow}\left\{\exp\left(-\frac{i}{\hbar}\int_0^t dt' H(t')\right)\right\}. \tag{3.6}$$

[‡]We will tacitly assume throughout this work that the state space is finite dimensional.

The exponential function of an operator must be interpreted, by definition, as its Taylor's series. The symbol $T_{\leftarrow}\{\dots\}$ denotes the antichronological-ordering, i.e., the operators inside it must be ordered in the following way: $T_{\leftarrow}\{H(t_1)\,H(t_2)\cdots H(t_n)\}$ $= H(t_1)\,H(t_2)\cdots H(t_n)$ for $t_1 \geq t_2 \geq \cdots \geq t_n$.

In general, the Hamiltonian of the system does not depend explicitly on time. When this is the case the expression for the time evolution operator reduces to the usual form

$$U(t) = \exp\left(-\frac{itH}{\hbar}\right). \tag{3.7}$$

One can see easily that the time evolution operator is unitary, i.e., $U(t)\,U^{\dagger}(t) = U^{\dagger}(t)\,U(t) = I$, where I is the identity operator in $\mathcal{L}(\mathcal{H})$. Physically the time evolution operator being unitary means that the norm of the state vector doesn't change with time. We can see this simply calculating the inner product of the evolved state vector: $\langle\psi(t)\,|\,\psi(t)\rangle = \langle\psi_0|\,U^{\dagger}(t)\,U(t)\,|\psi_0\rangle = \langle\psi_0\,|\,\psi_0\rangle$. As this algebraic approach for the time evolution of the state vector is completely equivalent to that of the Schrödinger equation it follows that the Schrödinger equation also preserves the normalization condition. We refer to this evolution to be a continuous unitary evolution (or dynamics). This contrasts with the other possible evolution to which a state vector may be subjected, the measurement. Within a dynamical point of view, the measurement is a discontinuous evolution in which the state vector, at a particular instant of time, suddenly changes to another. Moreover, the measurement dynamics is non-unitary because once a state has changed there is no way to know which original state it was, there is no inverse. This time evolution is referred in the literature as discontinuous non-unitary evolution (dynamics) or measurement dynamics of the state vector. How can it exist, in such a fundamental theory like quantum mechanics, two diametrically opposed dynamics? One that is continuous and unitary and another which is discontinuous and non-unitary[§]? For some researchers this may be enough to claim that there must exist some other quantum theory, or some underlying theory.

The usual measurement studied in quantum mechanics is the projective measurement or von Neumann measurement. It is mathematically characterized by a set of mutually orthogonal projection operators. Given an observable this set of projection

[§]The non-unitarity is not the problem, in fact. As we shall see, open systems undergo non-unitary evolutions.

operators is constructed with its eigenvectors[¶]. Consider some observable A and its set of orthogonal eigenvectors $\{|a_k\rangle\}$. From them we construct the orthogonal set of projection operators $\{P_k = |a_k\rangle\langle a_k|\}$. They satisfy the properties of orthogonality and idempotency through $P_k P_l = P_k \delta_{kl}$ and completeness, $\sum_k P_k = I$. The probability of obtaining an outcome a_k is given by the Born's rule

$$p(a_k) = \langle P_k \rangle_{|\psi\rangle} = \langle\psi| (|a_k\rangle\langle a_k|) |\psi\rangle = |\langle a_k | \psi\rangle|^2 . \tag{3.8}$$

After the measurement the system is found to be in the state

$$|\tilde{\psi'}\rangle = P_k |\psi\rangle , \tag{3.9}$$

which is a non normalized state[‖]. Normalizing it we obtain

$$|\psi'\rangle = \frac{1}{\sqrt{p(a_k)}} P_k |\psi\rangle . \tag{3.10}$$

We see that this state vector is normalized after a simple calculation,

$$\begin{aligned} \langle\psi' | \psi'\rangle =& \frac{1}{p(a_k)} \langle\psi| P_k^2 |\psi\rangle = \frac{1}{p(a_k)} \langle\psi| P_k |\psi\rangle \\ =& \frac{1}{p(a_k)} \langle P_k \rangle_{|\psi\rangle} = \frac{1}{p(a_k)} p(a_k) = 1. \end{aligned} \tag{3.11}$$

So far we have studied the two standard dynamics, the unitary and a particular measurement dynamics of the state vector. Let us turn now to another possible representation of a pure state. This representation is going to be of fundamental importance for what follows. It will enable us to treat the dynamics of open systems.

Consider a state vector $|\psi_0\rangle \in \mathcal{H}$ associated with a pure state. From it we can build a projection operator called the (pure) density operator $\rho = |\psi_0\rangle\langle\psi_0|$. Using Eq. (3.5) it is straightforward to find the time evolution of the density operator to be

$$\rho(t) = U(t) \rho(0) U^\dagger(t) . \tag{3.12}$$

Suppose $\{|a_k\rangle\}$ is an arbitrary orthonormal basis for $\mathcal{H}$. Then the normalization condition is translated to

$$\langle\psi_0 | \psi_0\rangle = \langle\psi_0| \left(\sum_k |a_k\rangle\langle a_k| \right) |\psi_0\rangle = \sum_k \langle a_k | \psi_0\rangle\langle\psi_0 | a_k\rangle = \mathrm{Tr}(\rho) = 1. \tag{3.13}$$

[¶]An observable is a self-adjoint operator over $\mathcal{H}$. The possible outcomes one may find measuring this observable are its eigenvalues.

[‖]The tilde symbol ˜ will always stand for a non normalized state.

As the density operator is a projection operator it is obvious that

$$\mathrm{Tr}\left(\rho^2\right) = 1. \tag{3.14}$$

Let us find what is the equation that gives the dynamics of the density operator. From Eq. (3.12) take the time derivative to obtain

$$\frac{d}{dt}\rho\left(t\right) = \left(\frac{d}{dt}U\left(t\right)\right)\rho\left(0\right)U^{\dagger}\left(t\right) + U\left(t\right)\rho\left(0\right)\left(\frac{d}{dt}U^{\dagger}\left(t\right)\right), \tag{3.15}$$

which can be further evaluated to, by means of Eq. (3.7),

$$\begin{aligned}\frac{d}{dt}\rho\left(t\right) &= \left(-\frac{i}{\hbar}HU\left(t\right)\right)\rho\left(0\right)U^{\dagger}\left(t\right) + U\left(t\right)\rho\left(0\right)\left(U^{\dagger}\left(t\right)\frac{i}{\hbar}H\right) \\ &= -\frac{i}{\hbar}\left[H,\rho\left(t\right)\right].\end{aligned} \tag{3.16}$$

This equation is called the Liouville-von Neumann equation. Plus some initial condition $\rho\left(0\right) = \rho_0$ it gives the time evolution of the density operator, $\rho\left(t\right)$. Up to this point, these two different ways of describing a pure state are completely equivalent, as it should be. In this way, a equation with this structure provides a unitary evolution. But this latter representation, namely in terms of the density operators, gives us the opportunity to extend our description of physical systems. Let us pause for a moment and briefly discuss an operational view of quantum mechanics [15].

A quantum system in a given quantum state may be always thought to be the result of a preparation procedure. In other words, any quantum state is, at least in principle, the result of some preparation procedure. A preparation procedure is a sequence of steps performed in the laboratory such that at the end of the process a quantum system emerges and the experiment indeed begins. In this way, we say that the quantum system was prepared in the state, say, ρ. Consider the following realistic situation. Our experimental device prepares a pure state with a reliability of 90%, i.e., in 10% of the preparations it prepares some other pure state. The question is: what is the appropriate object in the theory that describes this particular situation? This preparation can be thought to be as composed by the weighted distribution of the preparations of the two pure states. Each one of these pure states are described by the density operators ρ_1 and ρ_2, respectively. Therefore the appropriate weighted distribution is given by

$$\rho = p_1\rho_1 + p_2\rho_2, \tag{3.17}$$

where $p_1 = 0.9$ and $p_2 = 0.1$. This means exactly that in 90% of the cases we have one pure state, ρ_1, and in 10% the other, ρ_2, and we are taking this into account in ρ when describing the statistical predictions of this peculiar preparation. This object still is an operator over $\mathcal{H}$ and furthermore its trace is one,

$$\mathrm{Tr}\,(\rho) = p_1 \mathrm{Tr}\,(\rho_1) + p_2 \mathrm{Tr}\,(\rho_2) = p_1 + p_2 = 1, \tag{3.18}$$

where we used the fact that both ρ_1 and ρ_2 are pure density operators. Thus, this operator guarantees the normalization condition and it describes a possible physical preparation. In other words, this operator is also a density operator and it is called a mixed density operator. We say that this is a mixed density operator because this preparation involves a statistical mixture between two pure states with probability distribution $\{p_k\}$. Notwithstanding this argumentation, in 1957 Gleason mathematically showed that for any probability measure μ on a state space of dimension at least 3, i.e., any way to define probabilities mathematically, implies the existence of a trace class-one positive semidefinite operator** ρ (our density operator) such that

$$\mu\,(A) = \mathrm{Tr}\,[\rho P_A]\,. \tag{3.19}$$

A is a closed subset of the state space and P_A is its orthogonal projection. This is exactly the Born's rule for the density operators, as we shall see in a moment [16]. This means that the density operator is the most general object compatible with a probability structure over Hilbert spaces. Observe that in this way we eliminate altogether the description involving vectors in $\mathcal{H}$ because both observables and states are described by operators over $\mathcal{H}$. All dynamics and predictions are mathematically stated in the space of operators over $\mathcal{H}$. Even though this will not be important for this work, it is, nevertheless, noteworthy that the observables and density operators do not live in the same Hilbert space. This is so because the definition of a Hilbert space involves an inner product and the space of density operators and observables possess different inner products.

Considering what we have seen, the most general density operator is given by

$$\rho = \sum_k p_k \rho_k, \tag{3.20}$$

**A positive semidefinite operator is an operator that for any vector of the Hilbert space its average value is non-negative, $\langle\rho\rangle_{|\psi\rangle} \geq 0 \;\forall\, |\psi\rangle \in \mathcal{H}$. In other words all its eigenvalues are non-negative and therefore positivity implies self-adjointness.

where p_k is the probability of the system to be in the pure state associated with ρ_k[††] and moreover $\sum_k p_k = 1$ to maintain the normalization condition. In other words, quantum mechanics says that any preparation procedure leads to a state described by a density operator of the above form. The natural question which may be asked here is: is there a mathematical criterion to distinguish a pure from a mixed density operator? The answer is yes. As the density operator is self-adjoint we can always diagonalize it and rewrite it as $\rho = \sum_k p'_k \left|k\right\rangle \left\langle k\right|$, where p'_k is a new probability distribution. Take the trace of the squared density operator to give us

$$\begin{aligned}\mathrm{Tr}\left(\rho^2\right) =& \mathrm{Tr}\left(\sum_{k,l} p'_k p'_l \left|k\right\rangle \left\langle k \,|\, l\right\rangle \left\langle l\right|\right) = \sum_{k,l} p_k p_l \delta_{kl} \mathrm{Tr}\left(\left|k\right\rangle \left\langle l\right|\right) \\ =& \sum_k p_k^2 \mathrm{Tr}\left(\left|k\right\rangle \left\langle k\right|\right) = \sum_k p_k^2 \leq 1. \end{aligned} \tag{3.21}$$

We used the fact that all the probabilities are smaller than one. If the density operator is associated with a pure state, the inequality is indeed an equality, $\mathrm{Tr}\left(\rho^2\right) = 1$, as we have already concluded previously (cf. Eq. (3.14)). If this is not the case, then necessarily the strict inequality holds, i.e., for mixed density operators $\mathrm{Tr}\left(\rho^2\right) < 1$. In fact this quantity has a name. It is the purity of the density operator. Just to formalize, the purity is defined as

$$\xi = \mathrm{Tr}\left(\rho^2\right) \tag{3.22}$$

and from this we can assert that

$$\begin{cases} \xi = 1 & \text{, for pure density operators and} \\ \xi < 1 & \text{, for mixed density operators.} \end{cases} \tag{3.23}$$

We shall see why this quantity is important when we study how to measure the entanglement between two pure states. The purity measures the degree of mixedness of a quantum state. Consider the density operator of a two-level system given by $\rho_1 = \frac{1}{2}\left|0\right\rangle \left\langle 0\right| + \frac{1}{2}\left|1\right\rangle \left\langle 1\right|$. We are equally unacquainted with which state we are dealing, because they have exactly the same probability to be prepared. On the other hand, if the mixed state is given by $\rho_2 = \frac{1}{4}\left|0\right\rangle \left\langle 0\right| + \frac{3}{4}\left|1\right\rangle \left\langle 1\right|$, then we are more confident that the state which we may be dealing with is, in this case, the state $\left|1\right\rangle$. In this way we say that the density operator ρ_1 is more mixed than ρ_2. Furthermore

[††]The density operators ρ_k may also be mixed density operators, even though we can always write them in terms of pure density operators.

ρ_1 is a maximally mixed density operator because its distribution is uniform and this gives complete ignorance about which pure state we are dealing with. The purity of these states are

$$\xi_1 = \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{2} \tag{3.24}$$

and

$$\xi_2 = \left(\frac{1}{4}\right)^2 + \left(\frac{3}{4}\right)^2 = \frac{5}{8}, \tag{3.25}$$

respectively. The state ρ_2 is less mixed because its purity is close to the purity of pure states, $\xi_1 \leq \xi_2 \leq 1$. For an N-dimensional Hilbert space the maximally mixed state is given by the equiprobable distribution of a basis of pure states and therefore the purity is bounded as $1/N \leq \xi \leq 1$.

The last question to be addressed in this section is related to the mathematical formulation to study the time evolution of these newly introduced mixed density operators. Answering in advance, it will be the same used previously, but let us check this out.

We want to find the proper differential equation which describes the time evolution of the most general density operator (3.20). Let $\rho_k(0)$ be the initial (pure) density operator of each of the states which compose this mixed preparation. Therefore, the initial mixed density operator is $\rho(0) = \sum_k p_k \rho_k(0)$. We know from Eq. (3.12) how each of the ρ_k's evolve. Thus

$$\begin{aligned}\rho(t) =& \sum_k p_k \left(U(t)\,\rho_k(0)\,U^\dagger(t)\right) \\ =& U(t)\left(\sum_k p_k \rho_k(0)\right) U^\dagger(t) = U(t)\,\rho(0)\,U^\dagger(t).\end{aligned} \tag{3.26}$$

As this expression is formally the same as Eq. (3.12), it follows that the derivation for its time evolution is the same. So, the time evolution of any density operator is given by

$$\frac{d}{dt}\rho(t) = -\frac{i}{\hbar}\left[H, \rho(t)\right]. \tag{3.27}$$

Therefore, this equation describes the most general dynamics of a closed system.

The measurement dynamics is expressed in this approach as follows. The pro-

bability to find an outcome a_k of an observable A is given by the Born's rule

$$\begin{aligned} p(a_k) =& \mathrm{Tr}\left[\rho P_k\right] = \sum_k p_k \mathrm{Tr}\left[\rho_k \left|a_k\right\rangle \left\langle a_k\right|\right] = \sum_k p_k \left\langle a_k \right| \left.\psi_k\right\rangle \left\langle \psi_k \right| \left. a_k\right\rangle \\ =& \sum_k p_k \left|\left\langle a_k \right| \left.\psi_k\right\rangle\right|^2 = \sum_k p_k p_k\left(a_k\right), \end{aligned} \tag{3.28}$$

which reflects that the probability is the weighted sum of the probabilities to measure the outcome a_k with respect to the pure states ρ_k. The normalized post-measurement state is given by

$$\begin{aligned} \rho' =& \frac{1}{p\left(a_k\right)} P_k \rho P_k = \frac{1}{p\left(a_k\right)} \left|a_k\right\rangle \left\langle a_k\right| \rho \left|a_k\right\rangle \left\langle a_k\right| \\ =& \frac{1}{p\left(a_k\right)} p\left(a_k\right) \left|a_k\right\rangle \left\langle a_k\right| = \left|a_k\right\rangle \left\langle a_k\right|. \end{aligned} \tag{3.29}$$

In this section we described the dynamics of pure and mixed states, their unitary and measurement dynamics, in terms of projective measurements. We will see other measurements in Sec. (3.4). In the next section we will study the dynamics of the system that is the building block of quantum information, the qubit.

3.1.2 Dynamics of One Qubit

In this section we will apply the concepts developed in the previous section to a simple quantum system. The simplest non-trivial quantum mechanical system is a system composed of just one isolated spin-1/2 particle. Its state space is a two-dimensional Hilbert space $\mathcal{H}_2$. A single spin-1/2 is just one physical system which has such a state space. A system which can be described by the Hilbert space $\mathcal{H}_2$ is called a qubit. In the next chapter, when we address the quantum information theory, we shall see why this system is important. Briefly stated, it plays the role somehow analogous to the binary alphabet in classical information theory. There may be several physical systems which can be said to be a qubit. Some examples are:

- Atoms, ions, molecules which under specific conditions of the experimental setup can be described effectively by a two-dimensional Hilbert space. One standard example is that of atoms interacting resonantly with the radiation field (Jaynes-Cummings model [17]);

- Spin-1/2 particles;

- Polarization states of single photons;
- Quantum dots.

With that being said, let us turn to the theoretical description of a qubit. The most general qubit state is

$$|\psi\rangle = c_0 |0\rangle + c_1 |1\rangle , \tag{3.30}$$

where $\{|0\rangle , |1\rangle\}$ are usually taken to be the eigenstates of the Pauli matrix σ_3, corresponding to the eigenvalues $\{+1, -1\}$, respectively. This basis is referred to as the computational basis or the standard basis. Moreover, the coefficients must satisfy the normalization condition $|c_0|^2 + |c_1|^2 = 1$. We saw in the previous section that a more general description of quantum systems is in terms of density operators. Therefore let us study the operators over the state space of a qubit.

An operator which acts on $\mathcal{H}_2$ is represented by a two by two matrix. Therefore, the space of operator over $\mathcal{H}_2$, represented by $\mathcal{L}(\mathcal{H}_2)$, is the space of all two by two matrices. A basis in this space is given by $\{I, \sigma_1, \sigma_2, \sigma_3\}$, where I is the identity matrix and the σ_k are the three Pauli matrices. This space of operators can be a Hilbert space itself if endowed with the Hilbert-Schmidt inner product

$$\langle A, B\rangle = \mathrm{Tr}\left(A^\dagger B\right) \ , \forall A, B \in \mathcal{L}(\mathcal{H}_2) . \tag{3.31}$$

Regarding this inner product the previous basis can become an orthonormal basis if taken to be $\left\{\frac{1}{\sqrt{2}} I, \frac{1}{\sqrt{2}}\sigma_k\right\}$. A general operator $A \in \mathcal{L}(\mathcal{H}_2)$ can, then, be written as

$$\begin{aligned} A =& \left\langle \frac{1}{\sqrt{2}} I, A \right\rangle \frac{1}{\sqrt{2}} I + \sum_{k=1}^{3} \left\langle \frac{1}{\sqrt{2}}\sigma_k, A \right\rangle \frac{1}{\sqrt{2}}\sigma_k \\ =& \frac{1}{2}\mathrm{Tr}[A]\, I + \frac{1}{2}\sum_{k=1}^{3} \mathrm{Tr}[A\sigma_k]\, \sigma_k . \end{aligned} \tag{3.32}$$

The density operators representing one qubit are nothing but operators in $\mathcal{L}(\mathcal{H}_2)$ with unit trace. Using the operator expansion above the most general density operator is given by

$$\rho = \frac{1}{2}\left(I + \sum_{k=1}^{3} \mathrm{Tr}[\rho\sigma_k]\, \sigma_k\right) := \frac{1}{2}\left(I + \vec{r}\cdot\vec{\sigma}\right), \tag{3.33}$$

where

$$\vec{r} := \mathrm{Tr}[\rho\vec{\sigma}] \tag{3.34}$$

is a three dimensional real vector. With Eq. (3.33), and the property of Pauli matrices $(\vec{\sigma}\cdot\vec{a})\left(\vec{\sigma}\cdot\vec{b}\right)=\left(\vec{a}\cdot\vec{b}\right)I+i\vec{\sigma}\cdot\left(\vec{a}\times\vec{b}\right)$, we obtain

$$\begin{aligned}\mathrm{Tr}\left[\rho^2\right]=&\frac{1}{4}\mathrm{Tr}\left[I+2\vec{r}\cdot\vec{\sigma}+\left(\vec{r}\cdot\vec{\sigma}\right)\left(\vec{r}\cdot\vec{\sigma}\right)\right]=\frac{1}{4}\mathrm{Tr}\left[I+2\vec{r}\cdot\vec{\sigma}+\left(\vec{r}\cdot\vec{r}\right)I+i\vec{\sigma}\cdot\left(\vec{r}\times\vec{r}\right)\right]\\=&\frac{1}{4}\left(1+|r|^2\right)\mathrm{Tr}\left[I\right]+2\vec{\sigma}\cdot\mathrm{Tr}\left[\vec{\sigma}\right]=\frac{1}{2}\left(1+|\vec{r}|^2\right).\end{aligned}\tag{3.35}$$

Consider, for example, a pure density operator $\rho=|\psi\rangle\langle\psi|$. As we saw in the previous section, this implies that $\mathrm{Tr}\left[\rho^2\right]=1$ and, therefore, the above equation implies $|\vec{r}|=1$. Now let us pause for a moment to parametrize the state vector of Eq. (3.30) in a more convenient way.

The constants c_0 and c_1 are two complex numbers, which means that the state vector would need four real numbers to be specified uniquely. If we write both complex numbers in their polar representation, $c_0=|c_0|\,e^{i\varphi_1}$ and $c_1=|c_1|\,e^{i\varphi_2}$, the state vector may be written as

$$|\psi\rangle=|c_0|\,e^{i\varphi_0}\,|0\rangle+|c_1|\,e^{i\varphi_1}\,|1\rangle=e^{i\varphi_0}\left(|c_0|\,|0\rangle+|c_1|\,e^{i\varphi}\,|1\rangle\right),\tag{3.36}$$

where $\varphi=\varphi_1-\varphi_0$, and consider that $\varphi\in[0,2\pi)$ as the complex exponential is 2π-periodic. As the global phase can be completely forgotten because physically we cannot measure it, we are left with three real parameters. The normalization condition suggests to parametrize the two real numbers $|c_0|$ and $|c_1|$ as $\cos(\theta/2)$ and $\sin(\theta/2)$, respectively, with $\theta\in[0,\pi]$. Thus

$$|\psi\rangle=\cos\left(\frac{\theta}{2}\right)|0\rangle+\sin\left(\frac{\theta}{2}\right)e^{i\varphi}\,|1\rangle.\tag{3.37}$$

In conclusion, we need two real numbers θ and φ taking values in the intervals above to represent an arbitrary pure qubit state. This parametrization is the same parametrization of a point in a sphere centered at the origin of a three-dimensional real space.

We have seen that an arbitrary qubit state is parametrized as a point in a three-dimensional sphere and, at the same time, represented by a vector in a three-dimensional space with constant norm, i.e., a sphere. Therefore, all pure density operators lie in a sphere of radius one, called the Bloch sphere (see Fig. 3.1). The vector $\vec{r}$ is called the Bloch vector. If ρ were a mixed density operator, then $\mathrm{Tr}\left[\rho^2\right]<1$ and consequently $|r|<1$. Therefore, a mixed density operator lies inside

the Bloch sphere. Thus, the Bloch sphere is a convex set[‡‡] which represents all pure and mixed states of a qubit, the pure states are represented by the boundary of the sphere and the mixed states are represented by all the interior of the sphere.

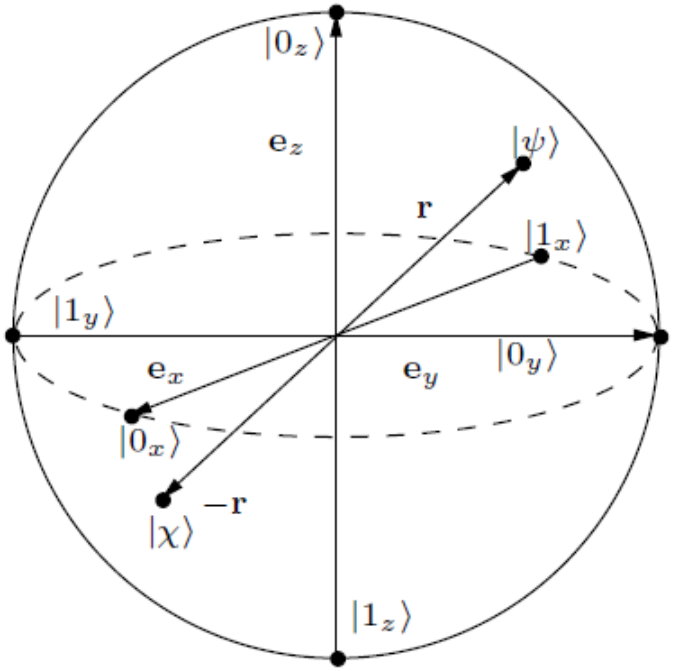

Figure 3.1: The Bloch sphere with the vectors associated to the eigenvalues of the Pauli operators. Also two arbitrary orthonormal qubit states $|\psi\rangle$ and $|\xi\rangle$, showing that they lie in opposite points on the sphere's surface. This figure was copied from Ref. [10]

Let us visualize the measurement and unitary dynamics of a qubit in a pure state in the Bloch sphere. For an arbitrary pure state, $\vec{r} = \mathrm{Tr}\left[\rho\vec{\sigma}\right]$ can be written as

$$\vec{r} = \mathrm{Tr}\left[|\psi\rangle\langle\psi|\,\vec{\sigma}\right] = \langle\psi|\,\vec{\sigma}\,|\psi\rangle\,. \tag{3.38}$$

Therefore $\vec{r}\cdot\vec{r} = 1 = \langle\psi|\,\vec{r}\cdot\vec{\sigma}\,|\psi\rangle$ implies that $\vec{r}\cdot\vec{\sigma}\,|\psi\rangle = |\psi\rangle$. So, $|\psi\rangle$ is an eigenvector with eigenvalue $+1$. As $\vec{r}\cdot\vec{\sigma}$ is a self-adjoint matrix with zero trace, it must have another eigenvector with eigenvalue -1, $\vec{r}\cdot\vec{\sigma}\,|\chi\rangle = -\,|\chi\rangle$. Consider the inner product

$$\langle\psi\,|\,\chi\rangle = \langle\psi|\,\vec{r}\cdot\vec{\sigma}\,|\chi\rangle\,, \tag{3.39}$$

[‡‡]Simply stated a convex set is a set such that every two elements in it can be connected by a straight line which is also inside the set. All polygons are convex but a boomerang is not.

using that $\langle\psi| = \langle\psi|\,\vec{r}\cdot\vec{\sigma}$. Considering that $|\chi\rangle$ is the eigenvector of $\vec{r}\cdot\vec{\sigma}$ with negative eigenvalue, one has

$$\langle\psi\,|\,\chi\rangle = -\langle\psi\,|\,\chi\rangle \Longrightarrow \langle\psi\,|\,\chi\rangle = 0. \tag{3.40}$$

Moreover,

$$\langle\chi|\,\vec{r}\cdot\vec{\sigma}\,|\chi\rangle = \vec{r}_\psi\cdot\langle\chi|\,\vec{\sigma}\,|\chi\rangle = \vec{r}_\psi\cdot\vec{r}_\chi = -1 \Longrightarrow \vec{r}_\psi = -\vec{r}_\chi. \tag{3.41}$$

Therefore, any two orthogonal vectors on the Bloch sphere are necessarily diametrically opposed (cf. Fig. (3.1)).

Any observable has two eigenvectors which are opposed in the Bloch sphere as described above. We set the Cartesian coordinates such that the observable eigenvectors are $|0\rangle$ and $|1\rangle$. An arbitrary qubit state can be expanded in this basis as

$$|\psi\rangle = c_0\,|0\rangle + c_1\,|1\rangle\,. \tag{3.42}$$

The r_z projection of the Bloch vector is

$$r_z = \cos(\theta) = \cos^2\left(\frac{\theta}{2}\right) - \sin^2\left(\frac{\theta}{2}\right) = |c_0|^2 - |c_1|^2\,. \tag{3.43}$$

Since $|c_0|^2 + |c_1|^2 = 1$, it follows that

$$|c_0|^2 = \frac{1+r_z}{2}, \tag{3.44}$$

and

$$|c_1|^2 = \frac{1-r_z}{2}. \tag{3.45}$$

Therefore, the probability to find the outcome associated with one of the eigenvectors of the observable in question is proportional to the projection of the Bloch vector associated with the qubit onto the axis defined by the two eigenvectors.

To visualize the unitary dynamics in the Bloch sphere, we will need a parametrization for an arbitrary unitary matrix $U \in SU(2)$. It can be shown that a parametrization of U is given by

$$U = e^{i\kappa}\begin{pmatrix} e^{-i\lambda/2}\cos(\mu/2)\,e^{-i\nu/2} & -e^{-i\lambda/2}\sin(\mu/2)\,e^{i\nu/2} \\ e^{i\lambda/2}\sin(\mu/2)\,e^{-i\nu/2} & e^{i\lambda/2}\cos(\mu/2)\,e^{i\nu/2} \end{pmatrix} \tag{3.46}$$

or, as an operator product,

$$U = e^{i\kappa}e^{-\frac{i}{2}\lambda\sigma_3}e^{-\frac{i}{2}\mu\sigma_2}e^{-\frac{i}{2}\nu\sigma_3}. \tag{3.47}$$

We can define a new unitary operator $\hat{U} := e^{-i\kappa}U$ without the global phase. Decomposing $\hat{U}$ in terms of the basis $\left\{\frac{1}{\sqrt{2}}I, \frac{1}{\sqrt{2}}\sigma_k\right\}$ according to Eq. (3.32) we find

$$\hat{U} = v_0 I - i\vec{v}\cdot\vec{\sigma}, \tag{3.48}$$

where

$$v_0 = \frac{1}{2}\mathrm{Tr}\left[\hat{U}\right] \text{ and } \vec{v} = \frac{i}{2}\mathrm{Tr}\left[\hat{U}\vec{\sigma}\right]. \tag{3.49}$$

Observing Eq. (3.46), we can read off the following properties

$$\hat{U}_{00} = \hat{U}^*_{11}, \tag{3.50}$$

$$\hat{U}_{10} = -\hat{U}^*_{01}, \tag{3.51}$$

$$\hat{U}_{00}\hat{U}_{11} - \hat{U}_{01}\hat{U}_{10} = 1. \tag{3.52}$$

With these properties one can show that v_0, v_1, v_2, $v_3 \in R$. The unitarity condition implies that

$$v_0^2 + v^2 = 1. \tag{3.53}$$

We can define

$$v_0 := \cos\left(\frac{\phi}{2}\right), \ \vec{v} := \sin\left(\frac{\phi}{2}\right)\vec{e}, \tag{3.54}$$

with $\vec{e}$ being a unit vector in R^3. For a given $\hat{U}$, the quantities ϕ and $\vec{e}$ are determined by Eq. (3.48). With this parametrization we can write

$$\hat{U} = e^{-i\frac{\phi}{2}\vec{e}\cdot\vec{\sigma}}. \tag{3.55}$$

Therefore, every unitary transformation over a qubit can be written uniquely in this form up to a global phase $e^{i\kappa}$.

Finally, we want to explain the effect of $\hat{U}$ on the Bloch vector. Without any loss of generality we can choose $\vec{e}$ to be the direction of $\vec{e}_z$. Hence, Eq. (3.55) reduces to

$$\hat{U} = e^{-i\frac{\phi}{2}}\left|0\right\rangle\left\langle 0\right| + e^{+i\frac{\phi}{2}}\left|1\right\rangle\left\langle 1\right|. \tag{3.56}$$

Using the parametrization of Eq. (3.37) we find that

$$\hat{U}\left|\psi\right\rangle = \cos\left(\frac{\theta}{2}\right)\left|0\right\rangle + \sin\left(\frac{\theta}{2}\right)e^{i\varphi+\phi}\left|1\right\rangle. \tag{3.57}$$

Therefore, the effect of a unitary operator U on the Bloch vector of some state vector $\left|\psi\right\rangle$ is to rotate it by an angle ϕ over the axis $\vec{e}$.

In this section we have seen the measurement and unitary dynamics of a qubit. In the next section we will begin to consider quantum systems composed of many subsystems.

3.2 Composite Quantum Systems

In this section we will set up the mathematical ground to study quantum systems composed of two or more subsystems. The idea of subsystems is well posed in classical mechanics where we deal with distinguishable objects. For instance, the solar system can be treated as being composed by nine subsystems, one star (the Sun) and eight planets. As they are macroscopical objects they are easily distinguishable. But now imagine a fluid and a small ball within it. The ball is easily seen to be a subsystem but the individual molecules whose compound the fluid are not easily distinguishable (by the naked eye) and therefore not easily classifiable as subsystems. In classical physics there is the idea that in principle we could paint these individual molecules with different colors and therefore be able to distinguish them somewhat easily. As we go more deeply into matter we learn that the elementary particles are not distinguishable*. Then, how can we treat them as subsystems if we cannot even distinguish them? From the point of view of quantum mechanics, each particle is described by its own state space. Therefore, the state space of a system of particles will be given by an appropriate composition of the Hilbert spaces of the individual particles (also called factor spaces). A closely related issue is the notion of localization. In classical mechanics it is completely natural to talk about subsystems because each subsystem is well localized in space. When we measure one of two subsystems which are separated far apart we know that this measurement does not disturb the other subsystem (besides negligible effects such as the gravitational force). But in quantum mechanics this is not the case because quantum correlations come in. We will define two important concepts in this section, entanglement and the partial trace map.

As we said in the beginning of this section, the Hilbert space which describes the properties of a system of particles is going to be an appropriate composition of the Hilbert spaces of each individual particles. Two possible constructions are the direct sum and the tensor product of the Hilbert spaces. Suppose that we have two two-level systems; for example, two spins. Each spin is described by a Hilbert space which is two-dimensional. If we construct a Hilbert space which is the direct sum

*At least two elementary particles relatively close to each other or which had interacted in the past. Of course an electron created on Earth and another created in the Sun may be distinguished.

of the two Hilbert spaces we will end with a space four-dimensional[†] $(2+2)$. If we construct a Hilbert space which is the tensor product of the two Hilbert spaces we will end with a space which is also four-dimensional[‡] (2×2). The conclusion is that two spins are not enough (at least in this superficial discussion) to discriminate between the two possible constructions. So let us consider three spins, instead. In this case if we take the direct sum we obtain a six-dimensional space $(2+2+2)$ and if we take the tensor product we obtain an eight-dimensional space $(2\times 2\times 2)$. Now we have to think what properties we expect the composed space to have. Physically, a system of three spins is expected to have the following degrees of freedom: $|\text{up},\text{up},\text{up}\rangle$, $|\text{up},\text{up},\text{down}\rangle$, $|\text{up},\text{down},\text{up}\rangle$, $|\text{down},\text{up},\text{up}\rangle$, $|\text{up},\text{down},\text{down}\rangle$, $|\text{down},\text{up},\text{down}\rangle$, $|\text{down},\text{down},\text{up}\rangle$ and $|\text{down},\text{down},\text{down}\rangle$. Hence 8 degrees of freedom. Therefore the appropriate way to construct the Hilbert space of three spins is taking the tensor product of each of the Hilbert spaces of the individual systems. That being said, we conclude that the right way to construct the Hilbert space of a system of particles is to take the tensor product of the Hilbert spaces of each of the particles, because this is the way to account correctly for the degrees of freedom of the composite system.

First let us establish notations, conventions and definitions which we shall use throughout this dissertation. As we have seen the (tensor) product of Hilbert spaces is the appropriate tool to construct the Hilbert space of many-body quantum systems. Consider that $\mathcal{H}^A$ is the state space associated with the subsystem S^A and $\mathcal{H}^B$ is the state space associated with the system S^B. The product Hilbert space is denoted by $\mathcal{H}^{AB}=\mathcal{H}^A\otimes\mathcal{H}^B$. If $\left|\varphi^A\right\rangle\in\mathcal{H}^A$ and $\left|\chi^B\right\rangle\in\mathcal{H}^B$, there is a product vector in $\mathcal{H}^{AB}$ which can be written in the following way

$$\left|\varphi^A\right\rangle\otimes\left|\chi^B\right\rangle=\left|\varphi^A\right\rangle\left|\chi^B\right\rangle=\left|\varphi^A,\chi^B\right\rangle=\left|\varphi,\chi\right\rangle. \tag{3.58}$$

The map $\otimes:\mathcal{H}^A\times\mathcal{H}^B\to\mathcal{H}^{AB}$ is bilinear, i.e.,

$$\left|\varphi^A\right\rangle\otimes\left(\lambda_1\left|\chi_1^B\right\rangle+\lambda_2\left|\chi_2^B\right\rangle\right)=\lambda_1\left|\varphi^A\right\rangle\otimes\left|\chi_1^B\right\rangle+\lambda_2\left|\varphi^A\right\rangle\otimes\left|\chi_2^B\right\rangle \tag{3.59}$$

and

$$\left(\lambda_1\left|\varphi_1^A\right\rangle+\lambda_2\left|\varphi_2^A\right\rangle\right)\otimes\left|\chi^B\right\rangle=\lambda_1\left|\varphi_1^A\right\rangle\otimes\left|\chi^B\right\rangle+\lambda_2\left|\varphi_2^A\right\rangle\otimes\left|\chi^B\right\rangle. \tag{3.60}$$

[†]The dimension of the direct sum space is the sum of the dimensions of its factor spaces.

[‡]The dimension of the tensor product space is the product of the dimensions of its factor spaces.

If $\{|n^A\rangle\}$ is a basis of $\mathcal{H}^A$ and $\{|i^B\rangle\}$ is a basis of $\mathcal{H}^B$, then $\{|n^A, i^B\rangle\}$ is a basis for the product Hilbert space $\mathcal{H}^{AB}$. From this we conclude that $\dim \mathcal{H}^{AB} = (\dim \mathcal{H}^A)\cdot(\dim \mathcal{H}^B)$. Moreover, as a direct consequence, an arbitrary vector $|\psi^{AB}\rangle$ in $\mathcal{H}^{AB}$ can be expanded in this basis

$$|\psi^{AB}\rangle = \sum_{n,i} \alpha_{ni} |n^A, i^B\rangle . \tag{3.61}$$

These same properties hold for a product Hilbert space constructed from a finite number of Hilbert spaces, larger than 2.

The linear functionals over the product Hilbert space are easily generalized from the Dirac notation. To represent a linear functional one may write any of the forms

$$\langle\varphi^A| \otimes \langle\chi^B| = \langle\varphi^A| \langle\chi^B| = \langle\varphi^A, \chi^B| = \langle\varphi, \chi| . \tag{3.62}$$

The inner product in $\mathcal{H}^{AB}$ is generalized to be

$$\langle\varphi^A, \chi^B \,|\, \xi^A, \zeta^B\rangle = \langle\varphi^A \,|\, \xi^A\rangle \langle\chi^B \,|\, \zeta^B\rangle . \tag{3.63}$$

Now one can verify that $\{|n^A, i^B\rangle\}$ is an orthonormal basis in $\mathcal{H}^{AB}$ provided $\{|n^A\rangle\}$ and $\{|i^B\rangle\}$ are orthonormal bases in $\mathcal{H}^A$ and $\mathcal{H}^B$, respectively. One can see this just using the definition given above for the inner product,

$$\langle n^A, i^B \,|\, n'^A, i'^B\rangle = \langle n^A \,|\, n'^A\rangle \langle i^B \,|\, i'^B\rangle = \delta_{nn'}\delta_{ii'}. \tag{3.64}$$

We can already define mathematically what an entangled state is. Pick a state of the product Hilbert space of two subsystems S^A and S^B, say, $|\psi^{AB}\rangle$. If in some cleverly chosen basis this state can be written as a product state, i.e., $|\psi^{AB}\rangle = |\varphi^A\rangle \otimes |\chi^B\rangle$, then we say that this state is not entangled. If there is no basis such that this decomposition is possible then we say that the state $|\psi^{AB}\rangle$ is entangled. This tells us that an entangled state is a synonym for superposition of product states. The most famous entangled states of two qubits are the so-called Bell states. They are

$$\begin{cases} |\Phi^{AB}_{\pm}\rangle = & \frac{1}{\sqrt{2}} \left(|0^A, 0^B\rangle \pm |1^A, 1^B\rangle\right) , \\ |\Psi^{AB}_{\pm}\rangle = & \frac{1}{\sqrt{2}} \left(|0^A, 1^B\rangle \pm |1^A, 0^B\rangle\right) . \end{cases} \tag{3.65}$$

As we shall see in Ch. 4, these states are maximally entangled. Moreover, they form an orthonormal basis in $\mathcal{H}^{AB}$. Considering the computational basis to be the basis

of the spin operator in the z direction, one may recognize the singlet state of two spins to be $\left|\Psi_{-}^{AB}\right\rangle$.

In Sec. 3.1.1 we realized that the general description of quantum systems must be in terms of density operators because state vectors do not account for a description of mixtures. Therefore we turn our attention to the description of linear operators over a product Hilbert space $\mathcal{H}^{AB}$. We will represent a product operator by the following notation

$$C^A \otimes D^B = C^A D^B. \tag{3.66}$$

The action of such an operator on a state is given by

$$\left[C^A \otimes D^B\right] \left|\varphi^A, \chi^B\right\rangle = \left(C^A \left|\varphi^A\right\rangle\right) \otimes \left(D^B \left|\chi^B\right\rangle\right) = \left|C^A \varphi^A, D^B \chi^B\right\rangle. \tag{3.67}$$

The operator is linear, meaning that

$$\left[C^A \otimes D^B\right] \left(\sum_{n,i} \alpha_{ni} \left|n^A, i^B\right\rangle\right) = \sum_{n,i} \alpha_{ni} \left|C^A n^A, D^B i^B\right\rangle. \tag{3.68}$$

One important concept is that of an extended operator (or subsystem operator, or local operator). If C^A is an operator of the system S^A and we embed it in a larger system we begin to interpret S^A as a subsystem. The operator must be appropriately extended to the larger space of operators. This is done by inserting as many identities as we need. For example, in a tripartite system S^{ABC} the operator C^A must become

$$\hat{C}^{ABC} = \hat{C}^A = C^A \otimes I^B \otimes I^C. \tag{3.69}$$

Two extended operators associated with different factor Hilbert spaces commute as they obey the relation

$$\hat{C}^{AB} \hat{D}^{AB} = \left(C^A \otimes I^B\right) \left(I^A \otimes D^B\right) = C^A I^A \otimes I^B D^B = C^A \otimes D^B. \tag{3.70}$$

The identity in the product Hilbert space may be expanded in terms of an orthonormal basis as

$$\begin{aligned} I^{AB} &= \sum_{n,i} \left|n^A, i^B\right\rangle \left\langle n^A, i^B\right| \\ &= \left(\sum_n \left|n^A\right\rangle \left\langle n^A\right|\right) \otimes \left(\sum_i \left|i^B\right\rangle \left\langle i^B\right|\right) = I^A \otimes I^B. \end{aligned} \tag{3.71}$$

Therefore, in the same way that we can decompose an arbitrary vector in a basis of states we can decompose an arbitrary operator in a basis of operators as

$$\begin{aligned} Z^{AB} =& I^{AB} Z^{AB} I^{AB} \\ =& \sum_{n,m} \sum_{i,j} \left\langle n^A, i^B \right| Z^{AB} \left| m^A, j^B \right\rangle \left(\left| n^A \right\rangle \left\langle m^A \right| \right) \otimes \left(\left| i^B \right\rangle \left\langle j^B \right| \right), \end{aligned} \tag{3.72}$$

where the coefficients of this expansion are the matrix elements of Z^{AB}.

The trace is defined in the usual way as

$$\mathrm{Tr}\left(Z^{AB}\right) = \mathrm{Tr}_{AB}\left(Z^{AB}\right) = \sum_{n,i} \left\langle n^A, i^B \right| Z^{AB} \left| n^A, i^B \right\rangle. \tag{3.73}$$

For product operators it follows that

$$\begin{aligned} \mathrm{Tr}_{AB}\left(C^A \otimes D^B\right) =& \sum_{n,i} \left\langle n^A, i^B \right| \left(C^A D^B\right) \left| n^A, i^B \right\rangle \\ =& \left(\sum_n \left\langle n^A \right| C^A \left| n^A \right\rangle \right) \left(\sum_i \left\langle i^B \right| D^B \left| i^B \right\rangle \right) \\ =& \mathrm{Tr}_A\left(C^A\right) \mathrm{Tr}_B\left(D^B\right). \end{aligned} \tag{3.74}$$

The next concept is of fundamental importance. It is the partial trace. Partial trace is a map from the composite state space to a set of factor spaces. We only take the trace on one or more subsystems. For example,

$$\mathrm{Tr}_B\left(Z^{AB}\right) = \sum_i \left\langle i^B \right| Z^{AB} \left| i^B \right\rangle. \tag{3.75}$$

The overall trace may be seen as a series of partial traces

$$\mathrm{Tr}_{AB}\left(Z^{AB}\right) = \mathrm{Tr}_A\left[\mathrm{Tr}_B\left(Z^{AB}\right)\right] = \mathrm{Tr}_B\left[\mathrm{Tr}_A\left(Z^{AB}\right)\right]. \tag{3.76}$$

At the beginning of this section we discussed composite systems and how describe them in quantum mechanics. The conclusion was that we construct the state space of the system of particles in a bottom-up fashion, i.e., the state space of the whole system is constructed putting together, in an appropriate way, the state space which would describe the individual particles. This leaves room for a natural description of a subsystem taking into account the influence of the disconsidered part. The strategy is simply to use the concept of partial trace described above to trace out the degrees of freedom we are not mainly interested in the density operator of the composite system. The dynamics of subsystems will be addressed in the next section. Here we

will discuss some elementary properties and comparisons of the subsystems and the whole system.

We shall consider a system composed of two subsystems S^A and S^B, but the conclusions can be easily generalized for an arbitrary number of subsystems. Suppose ρ^{AB} is a density operator which describes all the statistical properties of the system S^{AB}. We define the reduced density operator for the subsystem S^A to be

$$\rho^A := \mathrm{Tr}_B \left[\rho^{AB}\right]. \tag{3.77}$$

And analogously for the subsystem S^B. A physical operation made by, say, Alice on her subsystem is represented by a local operator (cf. Eq. (3.69)) C^A. If $\{c_n\}$ is the set of eigenvalues of the observable C^A then the probability to find an outcome is

$$p\left(c_n\right) = \mathrm{Tr}_{AB}\left[\rho^{AB}\hat{P}_n^A\right] = \mathrm{Tr}_A\left[\rho^A P_n^A\right], \tag{3.78}$$

where $P_n^A = \left|c_n^A\right\rangle\left\langle c_n^A\right|$ is a local projection operator. Moreover, the average value of the observable is

$$\left\langle C^A\right\rangle_{\rho^A} = \mathrm{Tr}_{AB}\left[\rho^{AB}\hat{C}^A\right] = \mathrm{Tr}_A\left[\rho^A C^A\right]. \tag{3.79}$$

These two relations show that the reduced density operator is the appropriate object to treat the statistical properties of the subsystems, as they satisfy the expected properties of general density operators.

Let us evaluate the reduced density operators for some pure state. Consider the Bell state $\left|\Psi_-^{AB}\right\rangle = \frac{1}{\sqrt{2}}\left(\left|0^A,1^B\right\rangle - \left|1^A,0^B\right\rangle\right)$. Its density operator is

$$\begin{aligned}\rho^{AB} = \left|\Psi_-^{AB}\right\rangle\left\langle\Psi_-^{AB}\right| = \frac{1}{2}\Big(&\left|0^A,1^B\right\rangle\left\langle 0^A,1^B\right| - \left|0^A,1^B\right\rangle\left\langle 1^A,0^B\right| \\ &- \left|1^A,0^B\right\rangle\left\langle 0^A,1^B\right| + \left|1^A,0^B\right\rangle\left\langle 1^A,0^B\right|\Big).\end{aligned} \tag{3.80}$$

The reduced density operator for the subsystem S^A is

$$\rho^A = \mathrm{Tr}_B\left[\rho^{AB}\right] = \left\langle 0^B\right|\rho^{AB}\left|0^B\right\rangle + \left\langle 1^B\right|\rho^{AB}\left|1^B\right\rangle \tag{3.81}$$

$$= \frac{1}{2}\left(\left|0^A\right\rangle\left\langle 0^A\right| + \left|1^A\right\rangle\left\langle 1^A\right|\right) = \frac{1}{2}I^A. \tag{3.82}$$

And the reduced density operator for the subsystem S^B is

$$\rho^B = \mathrm{Tr}_A\left[\rho^{AB}\right] = \left\langle 0^A\right|\rho^{AB}\left|0^A\right\rangle + \left\langle 1^A\right|\rho^{AB}\left|1^A\right\rangle \tag{3.83}$$

$$= \frac{1}{2}\left(\left|0^B\right\rangle\left\langle 0^B\right| + \left|1^B\right\rangle\left\langle 1^B\right|\right) = \frac{1}{2}I^B. \tag{3.84}$$

The conclusion of these results is that locally, i.e., for both Alice and Bob's laboratories, their subsystems exhibit the statistical behavior, under local operations, of mixed states. One can show that all four Bell states lead to this same result. As we have said before the Bell states are maximally entangled. These results show that the reduced density operators of a maximally entangled pure bipartite state are maximally mixed. This is not a coincidence as we shall see in Ch. 4. In fact, this provides a clue to measure entanglement of a pure bipartite state, namely, as how mixed their reduced density operators are. This will be made mathematically precise in Sec. (4.4.1) in terms of a quantum entropy.

3.3 Dynamics of Open Quantum Systems

We have described the dynamics of closed systems in Sec. 3.1. In this section we will turn our attention to the problem of open dynamics. Recalling the distinction of system and environment made in the beginning of the Sec. 3.1, we supposed that the universe is everything that exists. As we also have seen, the environment may interact and yet the system will be isolated, since the environment does not spoil what we are trying to describe or study. In order to describe the open dynamics we need a description, within the theory, of the environment. Therefore it would be unrealistic to consider the universe as everything because we could not ever account mathematically all different physical scales contained in the environment. The environment is, thus, composed by some other systems which influence the system. In the example of the marble rolling in a plane, the environment is the flat surface and the local atmosphere, for instance. The purpose of the theory of open systems is to describe the behavior of the systems when their environment affects them. So we have to consider, at some degree of precision, the dynamics of the environment. The classical example of the description of an open system dynamics is the Brownian motion. One paradigmatic example of the Brownian motion consists of a very tiny particle yet macroscopic, a grain of pollen, representing the system and a fluid that surrounds the particle, which represents the environment. Macroscopically the Brownian particle seems to just jiggle around. But the fact is that there is a precise mathematical description of this phenomenon in terms of

stochastic processes. We are able to know how the average velocity changes with the temperature of the environment, for example.

From the point of view of quantum mechanics we have a composite system. One subsystem is called system . This is the subsystem which we actually are interested to describe the behavior. The other subsystem is called the environment and this is the subsystem which is going to affect the dynamics of the system. Therefore the state space of the composite system is given by $\mathcal{H}^{SE} = \mathcal{H}^{S} \otimes \mathcal{H}^{E}$. The dynamics of the system plus environment (henceforth referred to S^{SE}) seen as a closed system is given by the Eq. (3.12)

$$\rho^{SE}(t) = U(t)\,\rho^{SE}(0)\,U^{\dagger}(t), \tag{3.85}$$

with

$$U(t) = \exp\left(-\frac{itH^{SE}}{\hbar}\right). \tag{3.86}$$

We consider the most general time-independent Hamiltonian

$$H^{SE} = H^{S} \otimes I^{E} + I^{S} \otimes H^{E} + H^{I}. \tag{3.87}$$

If we trace out the environment we obtain the algebraic time evolution equation for the reduced density operator of the system

$$\rho^{S}(t) = \mathrm{Tr}_{E}\left[U(t)\,\rho^{SE}(0)\,U^{\dagger}(t)\right]. \tag{3.88}$$

We will consider that at the initial time the system and environment are in a separable state, i.e., $\rho^{SE}(0) = \rho^{S}(0) \otimes \rho^{E}$. In fact it is a fairly reasonable hypothesis that $\rho^{SE}(t) \approx \rho^{S}(t) \otimes \rho^{E}$, i.e., that the system-environment state remains approximately separable for all times and that the state of the environment does not change. This hypothesis is called Born approximation and it says that the system-environment coupling is sufficiently weak and the environment is reasonably large such that changes of the density operator of the environment are negligible. It is an *a posteriori* verification that such hypotheses are extremely compatible with most open systems. Nowadays, with the increasing control on quantum systems, it is known that there exist various important physical systems which do not respect these hypotheses, though. One example is low-temperature solid-state systems. Understand the dynamics when the initial state is not separable is still an open problem. Thus,

$$\rho^{S}(t) = \mathrm{Tr}_{E}\left[U(t)\,\rho^{S}(0)\,\rho^{E}U^{\dagger}(t)\right]. \tag{3.89}$$

The expression above suggests the definition of a dynamical map, $\Phi(t,0)$, that maps the initial state into the state at time t,

$$\begin{aligned} \Phi(t,0) : S\left(\mathcal{H}^S\right) &\to S\left(\mathcal{H}^S\right), \\ \rho^S(t) = \Phi(t,0)\,\rho^S(0) &= \Phi_t\left(\rho^S(0)\right). \end{aligned} \tag{3.90}$$

$S(\mathcal{H})$ is defined as the space of density operator associated with the Hilbert space $\mathcal{H}$.

As this dynamical map maps density operators into density operators, it is also called a superoperator. Considering the existence of this map for every $t \geq 0$ and maintaining ρ^E fixed, we obtain a family of dynamical maps $\{\Phi_t | t \geq 0\}$. These dynamical maps are linear, trace-preserving and completely positive. We will go through each of these properties. Linearity is a natural and necessary property for dynamical maps as it implies that the time evolution of mixed density operator is the statistical mixture of the time evolution of its pure states,

$$\begin{aligned} \Phi_t\left(p_1\rho_1^S + p_2\rho_2^S\right) =& \mathrm{Tr}_E\left[U(t)\left(p_1\rho_1^S + p_2\rho_2^S\right)\rho^E U^\dagger(t)\right] \\ =& p_1\mathrm{Tr}_E\left[U(t)\,\rho_1^S\rho^E U^\dagger(t)\right] + p_2\mathrm{Tr}_E\left[U(t)\,\rho_2^S\rho^E U^\dagger(t)\right] \\ =& p_1\Phi_t\left(\rho_1^S\right) + p_2\Phi_t\left(\rho_2^S\right). \end{aligned} \tag{3.91}$$

The dynamical map is trace-preserving meaning that, assuming ρ^E normalized,

$$\begin{aligned} \mathrm{Tr}_S\left[\rho^S(t)\right] =& \mathrm{Tr}_S\left[\Phi_t\left(\rho^S(0)\right)\right] = \mathrm{Tr}_S\left[\mathrm{Tr}_E\left\{U(t)\,\rho^S(0)\,\rho^E U^\dagger(t)\right\}\right] \\ =& \mathrm{Tr}_{SE}\left[\rho^S(0)\,\rho^E\right] = \mathrm{Tr}_S\left[\rho^S(0)\right]. \end{aligned} \tag{3.92}$$

This property is expected because it says that the local dynamics preserves the probability characteristics of quantum mechanics. We emphasize that this does not mean that the dynamical map is unitary. Usually the dynamical map is not unitary because it has no inverse. This can be easily understood if we observe that the partial trace of different density operators can give the same reduced density operator, therefore having no inverse.

The last property will not be fully demonstrated since it is equal to the second part of the proof presented in App. B. We will only explain the terms used there and point out the only subtle difference.

An operator A is called positive (semi-definite) if its average value with respect to any state is always non-negative, $\langle A\rangle_\psi \geq 0$ for every $|\psi\rangle$. This also means

that its eigenvalues are non-negative. Therefore, positive operators are a subclass of self-adjoint operators, namely, those with non-negative eigenvalues. A positive map is a map which preserves positivity, i.e., maps positive operators into positive operators. It is natural to understand that the dynamical maps should be positive maps because, after all, for any time t we expect to have a well-defined density operator irrespective the detail of interactions or initial conditions. Consider an operator A and the mapped operator $\Phi_t(A)$. Then

$$[\Phi_t(A)]^\dagger = \left[\mathrm{Tr}_E\left\{U(t) A\rho^E U^\dagger(t)\right\}\right]^\dagger = \mathrm{Tr}_E\left\{U(t) A^\dagger\rho^E U^\dagger(t)\right\} = \Phi_t\left(A^\dagger\right). \quad (3.93)$$

Hence if A is self-adjoint then $\Phi_t(A)$ is also self-adjoint. As self-adjointness is a more general property than positiveness we can restrict the conclusion to if A is positive implies $\Phi_t(A)$ being positive and therefore Φ_t is a positive map. Consider that we couple the system S^A under this evolution with other arbitrary system S^B and consider the transformation $\Phi_t^A \otimes I^B$. Complete positivity means that whichever system S^B coupled to S^A the map $\Phi_t^A \otimes I^B$ must be a positive map, now over $S\left(\mathcal{H}^A \otimes \mathcal{H}^B\right)$. Physically, this means that we cannot exclude that the system S^A under consideration is open and hence is a subsystem of a larger system S^{AB}. In this case, via the effect Φ_t^A on S^A, the composite system S^{AB} could be influenced and therefore the extended dynamical map should transform an initial composite density operator (positive) ρ^{AB} to another ρ'^{AB}. The proof of the complete positivity of $\Phi_t(A)$ is similar to the second part of the theorem in App. B. The only difference is that there we are assuming the map $\mathcal{E}^A(\rho) \otimes I^B$ to be positive, which is exactly what we want to show. Examining carefully, the assumption is only used to write the ensemble decomposition of C^{AB} (cf. Eq. (B.13)). But this is guaranteed by the fact that the map $\Phi_t^A \otimes I^B$ preserves the self-adjointness because both Φ_t^A and I^B preserve. Therefore, the demonstration is the same, showing that the dynamical map admits a Kraus decomposition and so it is complete positive.

If we now consider that the family of dynamical maps $\{\Phi_t | t \geq 0\}$ satisfy the additional property

$$\Phi(t,0) \circ \Phi(s,0) = \Phi(t+s,0) \;\; \forall t, s \geq 0, \quad (3.94)$$

where $\circ$ stands for the composition of maps, this set of dynamical maps acquires the property of a semigroup. Being a semigroup means that there exists a generator $\mathcal{L}$

such that

$$\Phi(t,0) = e^{\mathcal{L}t}, \tag{3.95}$$

where $\mathcal{L}$: $S(\mathcal{H}) \to S(\mathcal{H})$. We can write

$$\rho^S(t) = \Phi(t,0)\,\rho^S(0) = e^{\mathcal{L}t}\rho^S(0). \tag{3.96}$$

Differentiating with respect to time we obtain

$$\frac{d}{dt}\rho^S(t) = \mathcal{L}\left[\rho^{\mathcal{S}}(t)\right]. \tag{3.97}$$

The complete positivity of the semigroup leads to important statements on the general structure of the generator. The famous Gorini-Kossakowski-Sudarshan-Lindblad theorem [18,19] states that $\mathcal{L}$ is the generator of a semigroup of completely positive quantum dynamical maps if and only if

$$\mathcal{L}\left[\rho^S\right] = -i\left[H^S, \rho^S\right] + \sum_i \gamma_i \left[A_i \rho^S A_i^\dagger - \frac{1}{2}\left\{A_i^\dagger A_i, \rho^S\right\}\right], \tag{3.98}$$

where H^S is an effective Hamiltonian (i.e., it need not coincide with the microscopic Hamiltonian in the original H^{SE}), A_i are arbitrary system operators, often called Lindblad operators, describing the various decay modes of the system and γ_i are the corresponding decay rates. We will present a rather simple demonstration of the theorem [10].

Suppose that

$$\rho(t+dt) = \rho(t) + O(dt), \tag{3.99}$$

with dt very small. This hypothesis is connected with the Markov approximation, where the dynamics at an instant of time depends only on the previous instant of time, in our case infinitely close. Using the Kraus representation (cf. App. B) we can write

$$\rho(t+dt) = \sum_i K_i(dt)\,\rho(t)\,K_i^\dagger(dt). \tag{3.100}$$

From our Markovian hypothesis one of the Kraus operators, say, the operator K_1 must be of the form

$$K_1 = I + O(dt) = I + (R - iH)\,dt, \tag{3.101}$$

where we used the fact that any linear operator may be written as a complex combination of two self-adjoint operators. Thus, $R^\dagger = R$ and $H^\dagger = H$. The remaining

Kraus operators must be of order $\sqrt{dt}$. Hence consider

$$K_i = L_i\sqrt{dt} \tag{3.102}$$

for $i \neq 1$. The completeness relation gives us

$$\begin{aligned}\sum_i K_i^\dagger K_i =& I = (I + Rdt + iHdt)(I + Rdt - iHdt) + \sum_{i>1} L_i^\dagger L_i \\ =& I + \left(2R + \sum_{i>1} L_i^\dagger L_i\right) dt + O\left(dt^2\right)\end{aligned} \tag{3.103}$$

From this equation we obtain

$$R = -\frac{1}{2}\sum_i L_i^\dagger L_i + O\left(dt^2\right). \tag{3.104}$$

From the Kraus decomposition in Eq. (3.100) we find

$$\begin{aligned}\rho(t+dt) =& (I + Rdt + iHdt)\,\rho(t)\,(I + Rdt + iHdt) + \sum_{i>1} L_i\rho(t)\,L_i^\dagger dt \\ =& \rho(t) - i[H, \rho(t)]\,dt + \sum_i L_i\rho(t)\,L_i^\dagger dt + \{R, \rho\}\,dt.\end{aligned} \tag{3.105}$$

With R given by the formula above we can write

$$\frac{\rho(t+dt) - \rho(t)}{dt} = -i[H, \rho(t)] + \sum_i \left(L_i\rho(t)\,L_i^\dagger - \frac{1}{2}\left\{L_i^\dagger L_i, \rho(t)\right\}\right). \tag{3.106}$$

Taking the limit $dt \to 0$ we obtain the differential equation

$$\frac{d\rho}{dt} = -i[H, \rho(t)] + \sum_i \left(L_i\rho(t)\,L_i^\dagger - \frac{1}{2}\left\{L_i^\dagger L_i, \rho(t)\right\}\right) = \mathcal{L}(\rho), \tag{3.107}$$

which proves the theorem. We can make the following identification between the operators: $L_i = A_i\sqrt{\gamma_i}$.

In this section we have studied the properties of dynamical maps, which govern the continuous dynamics of open quantum systems. In the next section we will see in more detail the measurement dynamics of a quantum system.

3.4 Measurement Dynamics

Measurement is a fundamental concept in science. It alludes to the idea of experimentation. A measurement is performed by a physical system which is composed of particles and therefore obeys, in principle, the same physics we are describing the

system. In classical mechanics reigns the idea that a measurement does not disturb the system. Even without quantum mechanics we could think that this could reasonably be not true for the fundamental particles of our world. This is so because in order to measure some quantity of some system we necessarily have to interact with it. If nothing interacted with the system we would not perceive the existence of it and the concept would not be scientific. As in the microscopic world we are dealing with the fundamental objects of nature it is natural that a measurement will in general disturb the state of the system. We are not invoking any quantum mechanical uncertainty relation but nevertheless arguing that this could be expected. Quantum mechanics validates this intuition. But this is not the unique difference of measurement in quantum mechanics. There are also non intuitive characteristics, one of which is the fact that the statistical properties of a system are affected by the measurement of another distant system. This is related with entanglement and will be discussed in the next chapter. In this section we will study the measurement dynamics of quantum states. We divide the measurements into two main categories, the selective and non-selective measurements. Within these categories there exist different measurements such as the projective measurements, already studied in Sec. (3.1), and generalized measurements, which can describe the imperfection of the measurement apparatus.

We will exemplify the selective and non-selective measurements using the double-slit experiment. The experimental apparatus consists of a beam, which is able to emit one particle at each time: a plate, which contains two rectangular slits in the middle, and an observation screen behind the slits to detect the position of the particle. Besides these apparatuses, which belong to the preparation procedure, we will, moreover, consider four distinguished preparation procedures through the following criteria. If the second slit is closed (cover it with any material which prohibits the passage of the particles) then we have a preparation procedure $\hat{S}_1$. If we close the slit one we have the preparation procedure $\hat{S}_2$. The third preparation procedure consists in flipping a coin and if we get heads then we close the first slit, if we get tails we close the second slit. This preparation procedure is represented by $\hat{S}_m\,(1,2)^*$. The fourth preparation procedure is just leave the two slits opened and it

*The index m stands for mixture as the distribution is a weighted sum of the distribution associated with $\hat{S}_1$ and $\hat{S}_2$.

is represented by $\hat{S}_p\,(1,2)^{\dagger}$. The distribution patterns on the screen are respectively, a gaussian behind slit one, a gaussian behind slit two, the addition of two gaussians, one of which behind slit one and the other behind slit two, and an interference pattern. All these preparation procedures have a density operator associated with them which describe these statistical properties within the theory. We will turn to the mathematical representation in a moment.

Consider now that we want to know which slit the particle has passed by. In order to do so we use a light beam which can be scattered by the particles directly behind only one of the two slits. In this way we observe that a light flash, corresponding to a scattering event, is always seen behind only one of the two slits. This reproduces the measurement as desired. Target the light beam, say, to behind slit one and consider just the preparations $\hat{S}_m\,(1,2)$ and $\hat{S}_p\,(1,2)$. Considering what has been said, if we now keep only the data associated with the particles which have passed the slit one, we know this because we are measuring it, we will end up with a distribution equal to that of $\hat{S}_1$ irrespective of the preparations being $\hat{S}_m\,(1,2)$ or $\hat{S}_p\,(1,2)$. We have thus performed a selective measurement because we indeed selected one path rather than another. But this is a pure state and therefore a selective measurement always results in a pure state. This pure state is, in general, different from the pure state associated with the preparation procedure, in our example $\hat{S}_p\,(1,2)$. If we keep track of the whole data, still measuring which particles passed slit one, we obtain a distribution which is equal to $\hat{S}_m\,(1,2)$. This second case we did not select the post measured state but instead we measured and continued considering the whole data. This measurement is called a non-selective measurement. As we can infer from this experimentation, a non-selective measurement always results in a mixed state. Let us represent this mathematically.

The ket associated with "passing through slit one" is denoted by $|1\rangle$ and the ket associated with "passing through slit two" is denoted by $|2\rangle$. The four preparation procedures give rise to the four density operators

$$\rho_1 = |1\rangle\,\langle 1| \tag{3.108}$$

$$\rho_2 = |2\rangle\,\langle 2| \tag{3.109}$$

[†]The index p stands for pure.

$$\rho_{12}^{m} = \frac{1}{2}\left(\rho_1 + \rho_2\right) \tag{3.110}$$

$$\rho_{12}^{p} = \left|\psi_{12}\right\rangle \left\langle\psi_{12}\right|, \tag{3.111}$$

where $\left|\psi_{12}\right\rangle = \frac{1}{\sqrt{2}}\left(\left|1\right\rangle + \left|2\right\rangle\right)$. The selective measurement presented in this example is represented by a projective measurementAs we discussed briefly previously (cf. Sec. (3.1)) the projective measurements are characterized by a complete set of mutually orthogonal states. In this case $\{P_k = \left|k\right\rangle \left\langle k\right|, k \in \{1,2\}\}$. Our selective measurement is such that it selects the particles which pass the first slit, therefore the final normalized state is

$$\rho' = \frac{1}{p\left(1\right)} P_1 \rho P_1. \tag{3.112}$$

The two (recall we are only considering $\hat{S}_m\left(1,2\right)$and $\hat{S}_p\left(1,2\right)$) possible post selected states are

$$\rho_{12}^{m'} = \rho_{12}^{p'} = \left|1\right\rangle \left\langle 1\right|, \tag{3.113}$$

which are pure states.

As the non-selective measurement does not select a particular state, but at the same time distinguishes the particles which pass through the slits, the final density operator is

$$\rho' = P_1 \rho P_1 + P_2 \rho P_2. \tag{3.114}$$

In our cases the non-selective measurement results in the following density operators

$$\rho_{12}^{m'} = \rho_{12}^{p'} = \frac{1}{2}\left|1\right\rangle \left\langle 1\right| + \frac{1}{2}\left|2\right\rangle \left\langle 2\right|, \tag{3.115}$$

which are statistical mixtures.

We will study now generalized measurements and at the final of the section provide an unified picture for any measurement on quantum states.

There are some discontinuous dynamics that are not associated with a projective measurement and yet may occur. Three examples are the following. Suppose we have a composite system with two subsystems, and they interact in such a way that the unitary evolution U^{AB} is able to produce an entangled state, say, at a time T. If a local selective measurement is performed on the subsystem S^B, then the reduced state of the subsystem S^A is also going to evolve discontinuously. Regarding just the different possible states of the subsystem S^A after the measurement, the set of operators which would describe the different probabilities for S^A to end in such

states do not form a mutually orthogonal set of projectors. Thus, a projective measurement on S^B does not induce a discontinuous evolution on S^A which can be described by projection operators. Another example is the case where, for some reason, the measurement apparatus is noisy, i.e., with a probability p it actually measures the system and with probability $1-p$ it simply gives a random state completely uncorrelated with the system. It describes a reasonably realistic situation where the measurement apparatus possess a probability of error. This discontinuous dynamics does not fit in the contest of projective measurements and we need a generalized notion to describe it. The last example is the case where, because of the conditions of the experiment, the measurement apparatus is not able to reliably distinguish which state is being measured and therefore it is also not possible to fit this situation in a projective measurement, as the projective measurement always distinguishes between two orthogonal states. Let us study the third situation.

We will base our discussion on the Stern-Gerlach experiment. The experimental setup consists of a source of spin-1/2 particles, an inhomogeneous magnetic field $\vec{B}$ and a screen. The arrangement is such that the particles acquire a trajectory in the z_+ direction of the screen if they have the z-component of spin given by $+1/2$ and acquire a direction in the z_- direction if they have the z-component of spin given by $-1/2$. At the screen what we see are two gaussian distributions, one in the z_+ region and another in the z_- region. Usually we consider that the two gaussian distributions on the screen do not overlap considerably, by putting the screen sufficiently distant from the magnetic field, in order to know that the particles which arrive at z_+ are indeed the particles with spin polarization $+1/2$ and equally for the other polarization. Consider now a non-optimal Stern-Gerlach experiment. In this case the two gaussian distributions do overlap in a way that we cannot quite differentiate if the particle which arrived at the z_+ region is necessarily a particle with polarization $+1/2$. We will model this in the following way.

The $+1/2$ and $-1/2$ spin polarizations in the z-direction are going to be represented by the kets $|\uparrow\rangle$ and $|\downarrow\rangle$, respectively. The detector is modeled quantum mechanically and its states are $|+\rangle$ and $|-\rangle$. The $|+\rangle$ state characterizes the detector being triggered at the z_+ region and the $|-\rangle$ being triggered at the z_- region. In the optimal situation the evolution can be modeled by a unitary operator which

completely characterizes the preparation state, i.e.,

$$|\uparrow\rangle |i\rangle \xrightarrow{U} |\uparrow\rangle |+\rangle \tag{3.116}$$

$$|\downarrow\rangle |i\rangle \xrightarrow{U} |\downarrow\rangle |-\rangle , \tag{3.117}$$

where $|i\rangle$ is some initial state of the detector. Therefore, an arbitrary spin state $|\varphi\rangle = c_\uparrow |\uparrow\rangle + c_\downarrow |\downarrow\rangle$ ends in the pre-measured state

$$|\varphi\rangle |i\rangle \xrightarrow{U} c_\uparrow |\uparrow\rangle |+\rangle + c_\downarrow |\downarrow\rangle |-\rangle . \tag{3.118}$$

A projective measurement on the basis $\{|+\rangle , |-\rangle\}$ of the detector gives $p_+ = |c_\uparrow|^2$ as the probability for the detector be triggered in the z_+ region and $p_- = |c_\downarrow|^2$ as the probability for the detector be triggered in the z_- region. The state final state of the spin is $|\uparrow\rangle$ and $|\downarrow\rangle$, respectively. Therefore, in the optimal experiment a projective measurement of the detector leads to an indirect projective measurement of the spin polarization.

In a real Stern-Gerlach experiment the gaussians may overlap and thus the $|+\rangle$ state of the detector can be triggered by the $|\downarrow\rangle$ spin state. This situation can be modeled by the following unitary evolution

$$|\uparrow\rangle |i\rangle \xrightarrow{U} |\uparrow\rangle \left(\sqrt{1-p_0}\, |+\rangle + \sqrt{p_0}\, |-\rangle\right), \tag{3.119}$$

$$|\downarrow\rangle |i\rangle \xrightarrow{U} |\downarrow\rangle \left(\sqrt{p_1}\, |+\rangle + \sqrt{1-p_1}\, |-\rangle\right), \tag{3.120}$$

where p_0 (p_1) is the probability of the detector being triggered by the state $|\downarrow\rangle$ ($|\uparrow\rangle$). In the optimal case $p_0 = p_1 = 0$. A general spin state undergoing this evolution results in

$$\begin{aligned} |\varphi\rangle |i\rangle \xrightarrow{U} &\left(\sqrt{1-p_0} c_\uparrow |\uparrow\rangle + \sqrt{p_1} c_\downarrow |\downarrow\rangle\right) |+\rangle \\ &+ \left(\sqrt{p_0} c_\uparrow |\uparrow\rangle + \sqrt{1-p_1} c_\downarrow |\downarrow\rangle\right) |-\rangle . \end{aligned} \tag{3.121}$$

Define the measurement operators

$$M_+ := \sqrt{1-p_0}\, |\uparrow\rangle \langle\uparrow| + \sqrt{p_1}\, |\downarrow\rangle \langle\downarrow| , \tag{3.122}$$

$$M_- := \sqrt{p_0}\, |\uparrow\rangle \langle\uparrow| + \sqrt{1-p_1}\, |\downarrow\rangle \langle\downarrow| . \tag{3.123}$$

They obey the completeness relation

$$M_+^\dagger M_+ + M_-^\dagger M_- = I. \tag{3.124}$$

The state after the interaction is then

$$|\psi\rangle = M_+ |\varphi\rangle |+\rangle + M_- |\varphi\rangle |-\rangle . \tag{3.125}$$

Measuring the detector in the $\{|+\rangle, |-\rangle\}$ basis, the system undergoes a discontinuous dynamics ending in the normalized state As a generalization of the projective measurement (cf. Eq. (3.8))

$$|\varphi\rangle \longrightarrow |\varphi_+\rangle = M_+ |\varphi\rangle \cdot \frac{1}{\text{Norm}}, \tag{3.126}$$

with probability

$$p_+ = \text{Tr}\left[(I \otimes |+\rangle \langle +|) |\psi\rangle \langle \psi|\right] = \langle \varphi| M_+^\dagger M_+ |\varphi\rangle = (1 - p_0) |c_\uparrow|^2 + p_1 |c_\downarrow|^2 , \tag{3.127}$$

or the state

$$|\varphi\rangle \longrightarrow |\varphi_-\rangle = M_- |\varphi\rangle \cdot \frac{1}{\text{Norm}}, \tag{3.128}$$

with probability

$$p_- = \text{Tr}\left[(I \otimes |-\rangle \langle -|) |\psi\rangle \langle \psi|\right] = \langle \varphi| M_-^\dagger M_- |\varphi\rangle = p_0 |c_\uparrow|^2 + (1 - p_1) |c_\downarrow|^2 . \tag{3.129}$$

The probabilities $p_\pm$ are thus the probabilities of the detector being triggered in the z_+, z_- region, respectively. From the point of view of the system, it undergoes a discontinuous dynamics governed by the two measurement operators, which are not orthogonal, and the probability of ending in the state $|\varphi_+\rangle$ ($|\varphi_-\rangle$) is p_+ (p_-).

From this example we see that there are situations that the common projective measurement does not account for. A generalized selective measurement is defined by a set of linear operators $\{M_k\}$ called measurement operators. The two conditions over these operators are that $M_k^\dagger M_k$ is a positive operator to assure that the probabilities are positive and

$$\sum_k M_k^\dagger M_k = I, \tag{3.130}$$

to ensure the normalization of probability. The probability of the outcome k is given by

$$p(k) = \langle \psi| M_k^\dagger M_k |\psi\rangle , \tag{3.131}$$

and the normalized state vector after the meaurement is given by

$$|\psi'\rangle = \frac{1}{\sqrt{p(k)}} M_k |\psi\rangle . \tag{3.132}$$

In terms of density operators, these relations change to

$$p\left(k\right) = \mathrm{Tr}\left[M_k \rho M_k^\dagger\right], \tag{3.133}$$

and

$$\rho_k' = \frac{1}{p(k)} M_k \rho M_k^\dagger. \tag{3.134}$$

And finally, if we perform a non-selective measurement, the post measurement state is going to be

$$\rho_{n.s.}' = \sum_k M_k \rho M_k^\dagger. \tag{3.135}$$

Again, it is a sum of selective measurements, i.e., we measure but do not select and therefore one must account for all possibilities.

For future references we define here what is a Bell measurement. As any orthogonal basis defines a projective measurement, just by defining the projection operators to be the projections over each state of the basis, a Bell measurement is a projective measurement on two subsystems, S^A and S^B, given by the four projection operators $\left\{\left|\Psi_\pm^{AB}\right\rangle, \left|\Phi_\pm^{AB}\right\rangle\right\}$, where the two-qubit states are the four Bell states (cf. Eq. 3.65). This measurement plays an important role in quantum protocols in the same amount entangled states do. This is because this measurement projects the state vector onto one out of four maximal entangled states. We will use this measurement in Sec. (4.5) for the quantum teleportation and entanglement swapping protocols.

This comprises the most general measurement dynamics of a quantum system. From this perspective, the projective measurements are just a particular case when the measurement operators satisfy the mutually orthogonal relation. In the next section we will put together in a same framework the generalized measurements and the continuous dynamics of quantum systems.

3.5 Generalized Quantum Dynamics

In the previous sections we have studied the most general dynamical evolution and the most general measurement dynamics of a quantum state. The properties derived in Sec. (3.3) for the dynamical maps are quite general and the only thing to change in order to encompass the generalized measurements is to change the trace-preserving condition to a trace-non-increasing condition, as the measurement

usually reduces to a pure state which is not normalized. A quantum operation is a mapping $\mathcal{E} : S(\mathcal{H}) \to S(\mathcal{H})$ such that

- $\mathcal{E}$ is a linear map;
- $\mathcal{E}$ is trace-non-increasing, i.e., $\mathrm{Tr}[\mathcal{E}(\rho)] \leq \mathrm{Tr}[\rho]$ (a contraction);
- $\mathcal{E}$ is a completely positive map.

In this way we accomplish the purpose of this chapter which was to provide a unified picture of the most general dynamics of a quantum system. There are only two final remarks. One is that in order to work always with a normalized density operator one must normalize it after the quantum operation by

$$\rho' = \frac{\mathcal{E}(\rho)}{\mathrm{Tr}[\mathcal{E}(\rho)]}. \tag{3.136}$$

The other is that, again, we call attention to the fact that we did not consider initial correlations between the system and environment in the dynamical maps studied.

In App. B we prove a representation for quantum operations which is called the Kraus representation. A map $\mathcal{E}$ is a quantum operation if and only if there exists a set of operators, called Kraus operators, such that

$$\mathcal{E}(\rho) = \sum_i K_i^\dagger \rho K_i, \tag{3.137}$$

where K_i are linear operators which fulfill the condition

$$\sum_i K_i^\dagger K_i \leq I. \tag{3.138}$$

The equality applies for trace-preserving quantum operations. We will apply this representation in a simple case which is going to be useful in the next section.

We will find the Kraus operators, i.e. the quantum operation, for a system S^A which undergoes a dynamics with another system S^B through the unitary operator U^{AB}. Suppose the initial state is separable, i.e., $\rho^{AB} = \rho^A \otimes \left|i^B\right\rangle \left\langle i^B\right|$. This state undergoes a unitary evolution to another state ρ'^{AB}. This induces a mapping $\rho^A \longrightarrow \tilde{\rho}'^A$ which is the quantum operation we are interested in. The final state is

$$\rho'^{AB} = U^{AB} \left|i^B\right\rangle \rho^A \left\langle i^B\right| U^{AB\dagger}. \tag{3.139}$$

The final state for the system S^A is given by

$$\rho'^A = \mathrm{Tr}_B\left[\rho'^{AB}\right] = \sum_n \left\langle e_n^B\right| U^{AB} \left|i^B\right\rangle \rho^A \left\langle i^B\right| U^{AB\dagger} \left|e_n^B\right\rangle, \tag{3.140}$$

where $\left\{\left|e_n^B\right\rangle\right\}$ is an orthonormal basis of $\mathcal{H}^B$. Define

$$K_n^A := \left\langle e_n^B\right| U^{AB} \left|i^B\right\rangle, \tag{3.141}$$

and

$$K_n^{A\dagger} := \left\langle i^B\right| \left(U^{AB}\right)^\dagger \left|e_n^B\right\rangle. \tag{3.142}$$

From the unitarity of U^{AB}

$$\sum_n K_n^{A\dagger} K_n^A = I^A. \tag{3.143}$$

We can rewrite Eq. (3.140) as

$$\rho'^A = \sum_n K_n^A \rho^A K_n^{A\dagger}, \tag{3.144}$$

and therefore we found the Kraus operators.

In this section we presented the general picture of quantum dynamics and its relation with Kraus representation. In the last section we will use some tools we have discussed to study the phenomenon of decoherence and environment-induced superselection, both of which are intimately related.

3.6 Decoherence and Environment-Induced Superselection

Decoherence is the phenomenon by which several superposition states are supressed dynamically when interacting with an environment. This means that several initial pure states are transformed into (possibly different) mixed states, losing the coherence between the relative phases they had at the beginning. In general, we cannot control all environment degrees of freedom and thus decoherence is seen as a ubiquitous process outside the laboratory. For quantum computers, decoherence may be a difficult problem to deal with since many processing protocols are based on maximally entanglement states, which are pure states. In 1980s Zurek introduced the concepts of pointer basis [20] and environment-induced superselection [21], where the environment selects dynamically a basis of states of the system which do not suffer decoherence. Decoherence occurs in any superposition of such states. From this point of view, decoherence arises simply because the interaction

of environmental particles with the system creates correlations between the system and the environment, and this process tends to transform the original pure quantum state into a statistical mixture. This correlation created with the environmental particles carries information (coherence) away and consequently spoils the capacity of the system to be used to perform the desirable tasks. In this section we will present a simple model for decoherence. This model is very simple but captures the underlying idea of the decoherence process.

Consider a scatterer system S^A on which another quantum system S^B is scattered. The scatterer center is supposed to have two orthonormal states $\left|0^A\right\rangle$ and $\left|1^A\right\rangle$ which are not modified by the scattering (stable states). The scattered system S^B arrives with the state $\left|i^B\right\rangle$ and is transformed asymptotically into the state $\left|0^B\right\rangle$ ($\left|1^B\right\rangle$) if scattered by the state $\left|0^A\right\rangle$ ($\left|1^A\right\rangle$). The three states are supposed to for an orthonormal basis in $\mathcal{H}_3^B$ as they are all mutually distinguishable in the laboratory. One can consider $\left|0^A\right\rangle$ and $\left|1^A\right\rangle$ to be two energy levels and $\left|i^B\right\rangle$, $\left|0^B\right\rangle$ and $\left|1^B\right\rangle$ to be three momentum states.

The scattering process is a unitary process of the composite system S^{AB}. The unitary operator has the following effect on initial states

$$\hat{U}^{AB}\left|0^A, i^B\right\rangle = \sqrt{1-p}\left|0^A, i^B\right\rangle + \sqrt{p}\left|0^A, 0^B\right\rangle, \tag{3.145}$$

$$\hat{U}^{AB}\left|1^A, i^B\right\rangle = \sqrt{1-p}\left|1^A, i^B\right\rangle + \sqrt{p}\left|1^A, 1^B\right\rangle, \tag{3.146}$$

where p is the scattering probability. Based in this process we can write a unitary operator responsable for the scattering process as

$$\begin{aligned} U^{AB} =& \sqrt{1-p}\left(\left|0^A, i^B\right\rangle\left\langle 0^A, i^B\right| + \left|1^A, i^B\right\rangle\left\langle 1^A, i^B\right|\right) \\ &+ \sqrt{p}\left(\left|0^A, 0^B\right\rangle\left\langle 0^A, i^B\right| + \left|1^A, 1^B\right\rangle\left\langle 1^A, i^B\right| + H.C.\right). \end{aligned} \tag{3.147}$$

Using Eqs. (3.141) and (3.142) we find the Kraus operators of the quantum operation that acts on the subsystem S^A to be

$$K_i^A = \left\langle i^B\right| U^{AB}\left|i^B\right\rangle = \sqrt{1-p}I^A = \sqrt{1-p}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{3.148}$$

$$K_0^A = \left\langle 0^B\right| U^{AB}\left|i^B\right\rangle = \sqrt{p}\left|0^A\right\rangle\left\langle 0^A\right| = \sqrt{p}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \tag{3.149}$$

$$K_1^A = \left\langle 1^B \right| U^{AB} \left| i^B \right\rangle = \sqrt{p} \left| 1^A \right\rangle \left\langle 1^A \right| = \sqrt{p} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.150}$$

The condition

$$K_i^{A\dagger} K_i^A + K_0^{A\dagger} K_0^A + K_1^{A\dagger} K_1^A = I^A, \tag{3.151}$$

for the Kraus operators can be easily seen to be fulfilled. If the system S^A was in the state ρ^A before the scattering then the quantum operation which gives the state ρ'^A after the scattering is given by

$$\rho'^A = \mathcal{E}^A\big(\rho^A\big) = K_i^A \rho^A K_i^{A\dagger} + K_{\mathrm{o}}^A \rho^A K_{\mathrm{o}}^{A\dagger} + K_{\mathrm{i}}^A \rho^A K_{\mathrm{i}}^{A\dagger}, \tag{3.152}$$

or, in the computational basis

$$\rho'^A = \begin{pmatrix} \rho_{00}^A & (1-p)\,\rho_{01}^A \\ (1-p)\,\rho_{10}^A & \rho_{11}^A \end{pmatrix}. \tag{3.153}$$

If scattering occurs with certainty ($p = 1$), the non-diagonal elements of the density operator of S^A vanish in the computational basis. For $0 \leq p < 1$, the non-diagonal elements decrease by the factor $(1 - p)$. Scattering causes decoherence.

In the limiting case of perfect scattering, one scattering process produces an entangled state,

$$\left(c_0 \left|0^A\right\rangle + c_1 \left|1^A\right\rangle\right) \left|i^B\right\rangle \overset{U(p=1)}{\longrightarrow} c_0 \left|0^A, 0^B\right\rangle + c_1 \left|1^A, 1^B\right\rangle, \tag{3.154}$$

and therefore the previous pure (coherent in the computational basis) state of S^A becomes a statistical mixture under any local operations performed on it.

We will also use this toy model of decoherence to discuss the environment-induced superselection [22,23] or simply einselection.

At the beginning of quantum mechanics or the old quantum theory, the attempts to understand quantum phenomena were mainly based on classical concepts. For instance, the planetary model of the atom put forward by Rutherford [24]. Not long after that, physicists realized that a quantum theory could not be built upon classical concepts. De Broglie's idea of wave-particle duality and matter (phase) wave set the stage for the development of a wave mechanics by Schrödinger. At this time the new quantum theory was born. A theory which barely hinged on our classical reality or perception of the world. A theory which still today is capable

of striking physicists' mind. The reductionist idea which revolve around physics (or even science in general) asserts that a phenomenon can be fully understood if we understand the parts contained in it. Within this philosophy a natural query is: How can we understand the classical world (physics) from quantum mechanics? The point we will address here is an attempt to understand why we don't see macroscopic superpositions of position eigenstates. Even though we are motivating einselection from a philosophical perspective this phenomenon is important for controlling the dynamics of quantum systems [25]. We will see that decoherence and extracting information about the system are synonyms.

The quantum operation of scattering used above is associated with a one-event scattering. In the practical world a reasonably large molecule is embedded in an environment full of particles which scatter it off. In general there are a lot of particles colliding randomly and we would need a huge state space for the system-environment. Instead, we will consider that one particle scatters at each time and this situation simplifies the discussion because it is enough to consider just a composition of quantum operations. The qualitative information is maintained, though. Suppose N particles scatter the system and the scattering probability is given by p. Therefore, with Eq. (3.153) the final density operator of the system after N scatterings is

$$\rho'^{A} = \begin{pmatrix} \rho_{00}^{A} & (1-p)^{N}\rho_{01}^{A} \\ (1-p)^{N}\rho_{10}^{A} & \rho_{11}^{A} \end{pmatrix}, \tag{3.155}$$

where ρ^A is an arbitrary initial density operator. We have seen in Sec. (3.1) that the purity measures the mixedness of a quantum state, $\xi = 1$ for a pure state and $\xi \to 0$ for a complete mixture. In our situation the purity of the final state is given by

$$\xi_f = (1-p)^{2N}\xi_i, \tag{3.156}$$

where ξ_i is the purity of the initial state. In addition suppose the initial state is a pure (coherent) state. The first thing to note is that for a sufficient large number of scattering particles N, the diagonal terms of the density operator of the system get arbitrarily small for $p \neq 0$. From Eq. (3.156) this means that the system becomes more mixed and in the limit of $N \to \infty$ it becomes maximally mixed. Another point is that it does get mixed in this particular basis, namely $\left\{\left|0^A\right\rangle, \left|1^A\right\rangle\right\}$. Let us understand physically why this happens.

The manner which we defined the interaction between the two systems, the system S^B, after the scattering, may carry information about the state of S^A. If S^A was in the state $\left|0^A\right\rangle$ with probability p the final state is $\left|0^A\right\rangle \otimes \left|0^B\right\rangle$ and therefore by measuring the system S^B we are able to know the state of the system S^A. The same happens if $\left|1^A\right\rangle$ was the initial state of S^A. In this sense the system S^B monitors the states $\left|0^A\right\rangle$ and $\left|1^A\right\rangle$ of S^A. From the point of view of a multi-particle scattering, the environment, composed of N particles, contains information about the state of S^A. On the other hand, from the point of view of the whole system S^{AB} they may get in a pure entangled state of the form given in Eq. (3.154). Summarily if the system S^A is either in $\left|0^A\right\rangle$ or $\left|1^A\right\rangle$ with or without scattering the final state is separable. If scattering happens, the system S^B (environment) monitors the state of S^A and we can obtain this information by measuring it. On the other hand, if S^A is in a superposition state then the whole system S^{AB} may get entangled and the system S^A gets mixed. Thus, the interpretation of the first observation is that the more mixed the system becomes (decoherence process) the more entangled it is with its environment, and consequently this means that the environment carries information about the system. In other words, our loss of local information about the system (information which may be acquired by local operations on S^A) is due to the fact that the environment carried away this information. This may be good or bad. In practice the environment is composed by an enormous amount of particles which we are not able to control and this leads to the loss of coherence of the quantum system under study. On the other hand, under a controlled way we could exploit this to obtain information about the system, implementing interactions with a few environment subsystems.

The second observation is that the statistical mixture ends up in a combination of $\left\{\left|0^A\right\rangle, \left|1^A\right\rangle\right\}$ and not other states. This is due to the specific form of the interaction. The way we modeled the interaction the environment carries information about the system being in these states, and not in others. The form of the interaction defines this preferred states (or preferred basis). In other words, the decoherence occurs in this basis, i.e., the superposition of this basis is suppressed. As a consequence there are a few superpositions which are allowed, these being any superposition which

results in either the state $\left|0^A\right\rangle$ or $\left|1^A\right\rangle$. For instance, the superpositions given by

$$\frac{1}{\sqrt{2}}\left(\left|0_x^A\right\rangle \pm \left|1_x^A\right\rangle\right), \tag{3.157}$$

where $\left|0_x^A\right\rangle$ and $\left|1_x^A\right\rangle$ are the eigenvectors of the operator σ_x^A. These two superpositions remain intact because a direct calculation of them gives the two preferred states $\left|0^A\right\rangle$ and $\left|1^A\right\rangle$, respectively. The conclusion here is that not all quantum superpositions are treated equally by decoherence. The superpositions which remain allowed are exactly those which give as a result the preferred states. Physically, the interaction of the system with the environment selects a set of preferred states which are not affected by decoherence but nonetheless the superpositions of these states are suppressed. This set of states which are robust against decoherence are defined by the interaction. This interaction is such that the environment can be seen to be monitoring the preferred states. As the environment (plus interaction) selects these robust states this phenomenon is called environment-induced superselection.

With this phenomenon in mind we can understand, by extrapolation, why we don't see superposition of macroscopic objects. In general the classical interactions, gravitational and electromagnetic, are position dependent. This means that the environmental particles possess information of the position of the system particles, or that the environment monitors the position of the system. Therefore the superposition states of the system are the preferred states and any superposition of them are rapidly suppressed by decoherence. A more realistic model using actual scattering theory can be found in chapter 3 of the reference [26].

Chapter 4

Quantum Information Theory

Quantum information is a relatively recent field of science. Its main scope is generally the same as the classical information but now taking into account, in a fundamental way, the properties of quantum mechanics. Our current computers are fundamentally based on transistors. As of 2012, the highest transistor count in a commercially available CPU is over 2.5 billion transistors [27]. A transistor is a semiconductor device, and hence, it is based on quantum mechanics. But this does not necessarily mean that quantum mechanics plays a role in the processing of information in a computer. In fact, nowadays, it does not. The way the transistors are used they process information in a purely classical manner, i.e., in terms of two binary digits 0 and 1. On the other hand, quantum information uses a qubit as its fundamental system to process information. This means that 0 and 1 are now two orthogonal states of a qubit, namely, $|0\rangle$ and $|1\rangle$. But in quantum mechanics these are just two out of infinitely many different possible states of a qubit. It is precisely this feature that quantum information intends to explore. For this purpose quantum information is intimately connected with quantum computation. At the end of this chapter we shall give some applications of quantum information theory.

At the same time not only quantum information is contributing to potentially new applications but it has also been contributing for the foundations of quantum mechanics itself. Quantum information has provided a new language to talk about quantum mechanics and therefore to understand other fields of physics. As the development of quantum information has shown the importance of entanglement, physicists have been studying this property of quantum systems in different contexts. These studies show how broadly applicable the concept of entanglement can be. Entanglement is a resource which can help in such tasks as the reduction of

classical communication complexity and clock synchronization. It has also given new insights for understanding many physical phenomena including super-radiance, disordered systems and emergence of classicality. As a final example, it was also used to characterize phase transitions: divergence of correlations at critical point is always accompanied by divergence of a suitably defined entanglement length - see [28] and references therein.

This chapter is organized as follows. In Sec. 4.1 we define the von Neumann entropy and show its relationship with the Shannon entropy. In Sec. 4.2 we evaluate and discuss the von Neumann entropy of two important steps in the processing of information, the preparation, which is related with the coding, and the measurement, which is related with the readout of the information. In Sec. 4.3 we discuss when a quantum system is correlated or not as well as if its correlations are classical or quantum. In Sec. 4.4 we study different measures of correlations, such as entropy of entanglement, concurrence and quantum discord. In Sec. 4.5 we show some interesting applications and phenomena using the concepts developed in the present chapter, some of which are the quantum dense coding and quantum teleportation. This chapter is heavily based on Ref. [10].

4.1 The von Neumann Entropy

In Ch. 2 we studied the classical theory of information and we saw that the concept of entropy has a close relationship with the amount of information. With the example in Sec. 2.3 we have seen that the Shannon entropy is a measure of the average information per source symbol, or that the minimum number of bits needed to characterize a given source is given by the nearest integer greater than the Shannon entropy. In this section we will study the counterpart of this entropy in quantum mechanics.

As we have seen, a classical signal source generates the letters of an alphabet, one after another. This signal source can be described by an ensemble $\{x_{i,}, p_i\}$ with $i = 1, \ldots, N$. Suppose we want to send a message through a quantum channel, i.e., the system in which we will encode the symbols of our source is a quantum system, e.g., atoms of the same type with spin 1/2 , or photons. Therefore, these quantum systems take on the role of carriers of the letters of the message. For each letter

x_i, an apparatus with the index i prepares a quantum system in the signal state $|\psi_i\rangle$ and transmits it. The relationship between the letter x_i and the state $|\psi_i\rangle$ is unambiguous. The entire setup is called the quantum signal source. By means of the preparation procedure, the classical information is thus coded in terms of pure quantum states. This procedure produces a statistical mixture of signal states with the density operator

$$\rho = \sum_{i=1}^{N} p_i \left|\psi_i\right\rangle \left\langle\psi_i\right| \tag{4.1}$$

which acts over a Hilbert space $\mathcal{H}_d$ of dimension d. The associated ensemble is the quantum signal ensemble $\{|\psi_i\rangle, p_i\}$ (cf. Fig. 4.1 and compare with Fig. 2.1).

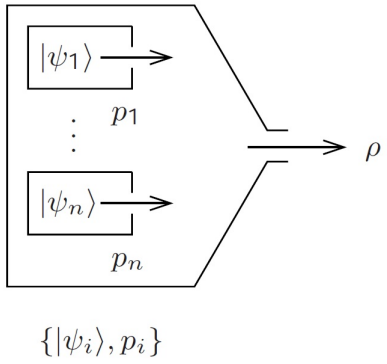

Figure 4.1: Representation of the quantum signal source with signal ensemble $\{|\psi_i\rangle, p_i\}$. This figure was copied from Ref. [10].

It is important that we do not require, in general, that the normalized state vectors $|\psi_i\rangle$ to be orthogonal and, furthermore, the dimension d need not be the same as N. The quantum channel should be free of disturbances and isolated from outside influences, making the quantum signal ensemble to remain unchanged during the transmission.

At a given moment there would be an attempt to read out the information which was originally input into the signal source via projective measurements. For this purpose, a detector which measures some observable D is used. The orthonormal eigenstates of the observable D are

$$D\left|d_m\right\rangle = d_m \left|d_m\right\rangle, \tag{4.2}$$

and they form an orthonormal basis of $\mathcal{H}_d$. The associated eigenvalues d_m are assumed to be nondegenerate. Then the correspondence between the measured

values d_m and the states $|d_m\rangle$ after the measurement procedure is unambiguous. The occurrence probability of a measured value d_m in a measurement performed on ρ is denoted by $p(d_m)$. Transmission of a signal via a quantum channel reflects the underlying scheme of quantum theory, i.e., at the beginning stands the preparation of a state and at the end, a measurement. The input consists of a sequence of classical signals with the Shannon entropy $H(\tilde{p})$, and the output is a sequence of measured values with the Shannon entropy $H(\tilde{p}(d))$.

To complete the description, consider the spectral decomposition of the density operator ρ,

$$\rho = \sum_{m=1}^{d} \lambda_m |m\rangle \langle m| \ , \ \langle m | m\rangle = \delta_{mm'} \tag{4.3}$$

with the eigenvectors $|m\rangle$ and the eigenvalues λ_m. The $\{|m\rangle, m = 1, \ldots, d\}$ form an orthonormal basis of $\mathcal{H}_d$.

We first consider a special situation, in which the classical information that is being input can again be read out without losses. The quantum system is, to this end, chosen so that the dimension d of $\mathcal{H}_d$ is the same as the number N of characters in the classical signal ensemble. By a suitable choice of the preparation procedure, corresponding to the character x_i, an eigenstate $|d_i\rangle$ of some observable D is generated (i.e. $|\psi_i\rangle = |d_i\rangle$)

$$\rho = \sum_{i}^{N} p_i |d_i\rangle \langle d_i| = \sum_{i}^{N} \lambda_i |i\rangle \langle i| . \tag{4.4}$$

In this case, we thus have $p_i = \lambda_i$ and $|d_i\rangle = |i\rangle$. The quantum signal source becomes a quasi-classical source due to the distinguishability of the signal states. Subsequently, the observable D is measured. The occurrence of the value d_i gives a unique indication of the original input of the signal character x_i, owing to their distinguishability. All of the probability distributions involved are the same, $p(d_i) = p_i = \lambda_i$. Correspondingly, we obtain for Shannon's entropy of the signal ensembles and of the ensemble of the measured values the value $H(\tilde{p}) = H(\tilde{p}(d))$.

The unique relation between the ensembles $\{x_i, p_i\}$, $\{|\psi_i\rangle, p_i\}$ and $\{d_i, p(d_i)\}$ in this particular quasi-classical situation and the corresponding agreement of the three probability distributions suggest that we associate to the statistical mixture, with the density operator ρ of Eq. (4.4), a quantum entropy $S(\tilde{\lambda})$ which has the

same value as Shannon's entropy, $S(\tilde{\lambda}) = H(\tilde{p})$, given by

$$S(\tilde{\lambda}) = -\sum_{i=1}^{d} \lambda_i \log \lambda_i \geq 0. \tag{4.5}$$

Using the spectral decomposition of ρ in Eq. (4.3), $S(\tilde{\lambda})$ can be written as a function of the density operator ρ

$$S(\rho) := S(\tilde{\lambda}) = -\mathrm{Tr}\left[\rho \log \rho\right]. \tag{4.6}$$

This quantum entropy $S(\rho)$ is also called the von Neumann entropy of the density operator ρ. The unit of this entropy is a quantum bit by symbol or a qubit.

Since $S(\rho)$ is unambiguously determined when ρ is fixed, we can generalize from this special procedure described above and associate formally a von Neumann entropy $S(\rho)$ as in Eq. (4.6) to every density operator ρ and thus to every quantum state, even in physical situations in which there is no signal transmission or processing. $S(\rho)$ characterizes a density operator ρ independently of how the corresponding state was prepared physically. ρ can also be a reduced density operator, which describes the state of a subsystem of a multipartite system. A state ρ with a spectral decomposition (4.3) cannot be distinguished from a statistical mixture of the states of the eigenbasis with the ensemble $\{|m\rangle, \lambda_m\}$. The state ρ can thus be completely simulated in this way. With this statistical mixture, if it is generated as a signal source, the classical information $H(\tilde{\lambda}) = S(\rho)$ will, on the average, be transmitted per signal state. This route of producing a physical situation in which the Shannon and von Neumann entropies are equal and then extends the von Neumann entropy as a measure of information in the quantum realm is one possible procedure. There is another suggestion that the von Neumann entropy is really the quantum analogue of the Shannon entropy. This work was carried out by Schumacher in 1995 [29]. We will not go through his derivation, but its main conclusion is that: "The von Neumann entropy $S(\rho)$ of a signal ensemble of pure states can be interpreted as the number of qubits per signal necessary to transpose it with near-perfect fidelity." The signal ensemble referred in the quotation is the same notion discussed in this section. Transposition is a process that maps the quantum state of a system S^A to the system S^B, at the cost of destroying the original state in S^A, i.e.,

$$\left|\psi^A\right\rangle \otimes \left|0^B\right\rangle \rightarrow \left|0^A\right\rangle \otimes \left|\psi^B\right\rangle. \tag{4.7}$$

Fidelity is a measure which indicates how distinguishable two quantum states are. The greater the fidelity the more similar the two quantum states are. For pure states the fidelity is given simply by $F\left(|\psi\rangle, |\phi\rangle\right) = |\langle\psi \,|\, \phi\rangle|$. An ideal encoder-decoder system should provide the maximum fidelity, meaning that the original state at the sender (source) is the same state recovered at the destination. In the same way as the Shannon entropy represents the minimum amount of binary digits necessary to code a source such that error at the decoder is arbitrarily small (source coding theorem), the Schumacher's theorem states that the von Neumann entropy gives the minimum amount of two-state systems (physical qubits) necessary to encode a quantum signal (density operator) such that the recovered quantum signal is almost exactly to the original after the transmission (transposition).

Being established the relationship between the von Neuman entropy and the Shannon entropy let us explore some properties of this quantum entropy. If ρ is a pure density operator, then it is a projection operator $\rho = |\psi\rangle\langle\psi|$. Therefore its eigenvalues are 0 and 1. Hence the von Neumann entropy is $S(\rho) = 0$ (we define $0 \log 0 = 0$). As the eigenvalues of the density operator are non-negative, the von Neumann entropy is also non-negative. The minimum value being zero and as it was shown above it occurs for every pure quantum state. What is the maximum value that $S(\rho)$ can take? In order to determine it we need to extremize the von Neumann entropy using the Lagrange multiplier method, as the values of λ_i's of the spectral decomposition of ρ satisfy the constraint $\sum_k \lambda_k = 1$. We need to extremize the function $F = S + \alpha\left(1 - \sum_k \lambda_k\right)$. Hence

$$\frac{\partial F}{\partial \lambda_i} = \frac{\partial S}{\partial \lambda_i} - \alpha = -\sum_k \frac{\partial \lambda_k}{\partial \lambda_i}\left[1 + \log \lambda_k\right] - \alpha = 0. \tag{4.8}$$

Isolating α,

$$\alpha = -\left(1 + \log \lambda_i\right), \tag{4.9}$$

and then λ_i gives us

$$\lambda_i = 2^{-(1+\alpha)}. \tag{4.10}$$

Applying the constraint on λ_i's we obtain

$$\sum_k \lambda_k = \sum_k^N 2^{-(1+\alpha)} = N 2^{-(1+\alpha)} = N\lambda_k = 1 \Rightarrow \lambda_k = \frac{1}{N}. \tag{4.11}$$

This means that the extremum value of $S(\rho)$ is achieved by the maximally mixed state in $\mathcal{H}_N$. Substituting this in the von Neumann entropy we obtain $S(\rho) = \log N$.

As this is greater than zero this extremum is a maximum. The von Neumann entropy for a density operator over a finite dimensional Hilbert space $\mathcal{H}_N$ is bounded as

$$0 \leq S(\rho) \leq \log N. \tag{4.12}$$

Moreover, zero is achieved only for pure states and $\log N$ for maximally mixed states. Therefore, the von Neumann entropy can be used to measure the mixedness of density operators.

To derive further properties we need to define the quantum relative entropy. Recalling the classical definition in Eq. (2.43) the quantum analogue is simply extended to be

$$S(\rho||\sigma) := \mathrm{Tr}[\rho \log \rho] - \mathrm{Tr}[\rho \log \sigma] = -S(\rho) - \mathrm{Tr}[\rho \log \sigma]. \tag{4.13}$$

As we derived the Gibbs' inequality in the classical case, here the analogue is called the Klein's inequality. To prove it consider the decomposition of ρ and σ to be

$$\rho = \sum_{m=1}^{d} \lambda_m |\phi_m\rangle \langle\phi_m|, \ \sigma = \sum_{m=1}^{d} \kappa_m |\xi_m\rangle \langle\xi_m|. \tag{4.14}$$

Then,

$$\begin{aligned} S(\rho||\sigma) &= \sum_m \lambda_m \log \lambda_m - \sum_m \langle\phi_m| (\rho \log \sigma) |\phi_m\rangle \\ &= \sum_m \lambda_m \left(\log \lambda_m - \sum_m \langle\phi_m| \log \sigma |\phi_m\rangle \right). \end{aligned} \tag{4.15}$$

Evaluating the second part,

$$\begin{aligned} \langle\phi_m| \log \sigma |\phi_m\rangle &= \langle\phi_m| \log \sigma \left(\sum_{m'} |\xi_{m'}\rangle \langle\xi_{m'}| \right) |\phi_m\rangle \\ &= \sum_{m'} \log \kappa_{m'} \langle\phi_m | \xi_{m'}\rangle \langle\xi_{m'} | \phi_m\rangle = \sum_{m'} P_{mm'} \log \kappa_{m'}, \end{aligned} \tag{4.16}$$

where $P_{mm'} \geq 0$ and $\sum_m P_{mm'} = \sum_{m'} P_{mm'} = 1$. Therefore we can write

$$S(\rho||\sigma) = \sum_m \lambda_m \left(\log \lambda_m - \sum_{m'} P_{mm'} \log \kappa_{m'} \right). \tag{4.17}$$

The logarithm function is a concave function $\left(f''(x) < 0\right)$, thus $\sum_{m'} P_{mm'} \log \kappa_{m'} \leq \log\left(\sum_{m'} P_{mm'} \kappa_{m'}\right)$ or $-\sum_{m'} P_{mm'} \log \kappa_{m'} \geq -\log(\mu_m)$, thus

$$S(\rho||\sigma) = \sum_m \lambda_m \log\left(\frac{\lambda_m}{\mu_m}\right) = D[\lambda||\mu] \geq 0, \tag{4.18}$$

where one recognizes the classical relative entropy.

With Klein's inequality one may easily show that the von Neumann entropy is subadditive. Consider $\rho = \rho^{AB}$ and $\sigma = \rho^A \otimes \rho^B$, therefore

$$\begin{aligned} S(\rho||\sigma) =& - S\left(\rho^{AB}\right) - \mathrm{Tr}_{AB}\left[\rho^{AB}\log\left(\rho^A \otimes \rho^B\right)\right] \\ =& - S\left(\rho^{AB}\right) - \mathrm{Tr}_{AB}\left[\rho^{AB}\log\left(\rho^A\right)\right] - \mathrm{Tr}_{AB}\left[\rho^{AB}\log\left(\rho^B\right)\right] \\ =& - S\left(\rho^{AB}\right) + S\left(\rho^A\right) + S\left(\rho^B\right) \geq 0, \end{aligned} \tag{4.19}$$

or,

$$S\left(\rho^{AB}\right) \leq S\left(\rho^A\right) + S\left(\rho^B\right). \tag{4.20}$$

This means that the information, or uncertainty, contained in a composite state cannot be greater than the addition uncertainty of the subsystems. This already contrasts with the classical case where the joint entropy is always greater or equal to the Shannon entropy of the independent sources. This means that, unlike the classical case, if the subsystems are not independent the information of knowing the whole is reduced when compared with the addition of the information of the individual parts. As we will see later on, this is related to the possibility of storing information nonlocally in a quantum state, via quantum correlations such as entanglement.

Concavity can also be easily proved using Klein's inequality. Consider two density operators ρ_1 and ρ_2. Moreover, consider a convex combination of them to be a new density operator $\rho = \lambda\rho_1 + (1-\lambda)\rho_2$, with $0 < \lambda < 1$. By Klein's inequality

$$\lambda S(\rho_1||\rho) + (1-\lambda) S(\rho_2||\rho) \geq 0. \tag{4.21}$$

From Eq. (4.13) we can expand this expression as

$$\begin{aligned} &\lambda\left(-S(\rho_1) - \mathrm{Tr}\left[\rho_1\log\rho\right]\right) + (1-\lambda)\left(-S(\rho_2) - \mathrm{Tr}\left[\rho_2\log\rho\right]\right) \\ =& - \lambda S(\rho_1) - (1-\lambda) S(\rho_2) - \mathrm{Tr}\left[\left(\lambda\rho_1 + (1-\lambda)\rho_2\right)\log\rho\right] \geq 0, \end{aligned} \tag{4.22}$$

or equivalently,

$$\begin{aligned} &-\lambda S(\rho_1) - (1-\lambda) S(\rho_2) + S(\rho) \geq 0 \\ \Rightarrow& S\left(\lambda\rho_1 + (1-\lambda)\rho_2\right) \geq \lambda\rho_1 + (1-\lambda)\rho_2. \end{aligned} \tag{4.23}$$

This means that given two states and their informational content, a statistical mixture built with them always gives a state with increased uncertainty. The equality only holds if $\lambda = 0$ or $\lambda = 1$.

As the von Neumann entropy depends on the eigenvalues of the density operator ρ it is invariant under unitary transformation, i.e.,

$$S\left(U\rho U^{\dagger}\right) = S\left(\rho\right). \tag{4.24}$$

This means that our information about the state does not change under unitary dynamic evolution. In the same spirit it changes with a measurement dynamics as we shall see now.

4.2 Entropy at the Interfaces: Preparation and Measurement

When we structured the relationship between Shannon and von Neumann entropies in the previous section we had to suppose that both the preparation and measurement procedures were particularly special. In this section we will discuss the behavior of the entropy at these two interfaces. The measurement is related with how we read out information, and the preparation says how we code information. These two situations do not play a fundamental role in classical information theory but in the quantum case, as we may have non-orthogonal states, this may lead to fundamental losses of information. We will begin studying the measurement process.

4.2.1 The Entropy of Projective Measurements

We will describe the production of entropy through a projective measurement. In a quantum channel the states are described by a density operator, say ρ, and therefore possess a quantum entropy $S\left(\rho\right)$. A non-selective measurement, associated with a decoding observable D leads to a probability distribution $\{p\left(d_m\right)\}$ of the possible values which can be measured, $\{d_m\}$. After a non-selective measurement the post-measured state is given by

$$\rho' = \sum_{m=1}^{d} p\left(d_m\right) \left|d_m\right\rangle \left\langle d_m\right| = \sum_{m=1}^{d} P_m \rho P_m. \tag{4.25}$$

In turn, this new state has a quantum entropy $S\left(\rho'\right)$. This also consitutes a Shannon entropy of a signal ensemble $\{d_m, p\left(d_m\right)\}$, denoted as $H\left(\tilde{p}\left(d_m\right)\right)$. As the set of projectors of a projetive measurements are mutually orthogonal, the Shannon and

von Neumann entropies are equal, i.e.,

$$S(\rho') = H(\tilde{p}(d_m)). \tag{4.26}$$

The non-selective measurement through projective measurement generates a new quasi-classical signal. We want to compare the entropy after and before the measurement, $S(\rho')$ and $S(\rho)$. Klein's inequality permits us to do so.

We start with

$$0 \leq S(\rho||\rho') = -S(\rho) - \mathrm{Tr}[\rho \log \rho'], \tag{4.27}$$

and consider the second term more carefully. We have

$$\mathrm{Tr}[\rho \log \rho'] = \mathrm{Tr}\left[\left(\sum_l P_l\right) \rho \log \rho'\right] = \mathrm{Tr}\left[\sum_l P_l \rho \log(\rho') P_l\right], \tag{4.28}$$

where we used the properties of the projection operators and permuted the terms within the trace. From Eq. (4.25) we see that $P_l \rho' = P_l \rho P_l = \rho' P_l$ holds. Therefore, P_l also commutes with the operator function $\log \rho'$. Using this we find

$$\mathrm{Tr}[\rho \log \rho'] = \mathrm{Tr}\left[\sum_m P_m \rho P_m \log \rho'\right] = \mathrm{Tr}[\rho' \log \rho'] = -S(\rho'). \tag{4.29}$$

Thus, after inserting this into Eq. (4.27), we obtain the overall result

$$S(\rho') \geq S(\rho). \tag{4.30}$$

In a non-selective projective measurement, the von Neumann entropy of the state ρ' after the measurement is the same as the von Neumann entropy of the state ρ before the measurement if and only if the measurement takes place in the eigenbasis of ρ; otherwise, it is greater. A non-selective measurement, therefore, transforms the system into a new signal ensemble with a, in general, greater entropy and in this manner it destroys information. The a priori uncertainty has increased as a result of the non-selective measurement.

This can be verified with a very simple example. The pure state

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \tag{4.31}$$

has a vanishing entropy. Non-selective measurement in the computational basis leads to the maximally mixed state

$$\rho = \frac{1}{2} I, \tag{4.32}$$

with maximum entropy, $S(\rho) = 1$.

4.2.2 The entropy of preparation

In quantum coding the classical signal source with the ensemble $\{x_i, p_i\}$ is coded into the quantum ensemble $\{|\psi_i\rangle, p_i\}$, described by the density operator ρ. The maximum value of the quantum entropy is given by Eq. (4.12), i.e., it is bounded by the logarithm of the dimension of the Hilbert space of the quantum system we are using to code. The maximum value of the Shannon entropy $H(\tilde{p})$ is proportional to the number N of characters in the classical alphabet x_i through the same functional form $\log N$. In the quantum case N is the number of pure states which we are going to code the letters, i.e., $|\psi_i\rangle \leftrightarrow x_i$. Since the quantum states $|\psi_i\rangle$ need not be orthogonal the dimension d of the Hilbert space of the quantum channel may be smaller than the number N of states used. As we are coding classical distinguishable letters in non-orthogonal states, which possess redundant information, it is expected that we are losing information. Therefore we expect a relation of the form

$$H(\tilde{p}) \geq S(\rho). \tag{4.33}$$

The equality only holds if the states $|\psi_i\rangle$ are mutually orthogonal. This inequality is demonstrated in App C.

Therefore, if the states are not orthogonal, there is no decoding observable with which the full information content of the coded classical message could be read out again. ρ transmits less information than was contained in the original classical signal. Even with an optimally-matched final measurement, the information can no longer be completely recalled.

4.3 Correlations

We live in a, fairly reasonable to be considered, causal world. Causality implies correlation but the reverse is not true. Following the Macmillan Dictionary [30] the definition of correlation is: "a connection or relationship between two or more things that is not caused by chance". If A causes B then they are correlated in the sense that if A happens then we know that B is going to happen too. On the other hand, if we know that the two events A and B have a probability to occur together, this does not imply causality neither from A to B nor from B to A. In general, everyday life events are so complex that we certainly are disregarding many

potentially influential factors. What we can say is that two events are correlated, or not. Correlation is a statistical concept. Two events are correlated if their outcomes (or the appropriate notion for them) are statistically dependent, or in other words, they are not statistically independent. From probability theory there are at least two ways of classifying independence events. If the joint probability distribution is separable, $p(x,y) = p(x)\,p(y)$, or if the conditional probability is independent of the previous event, $p(x|y) = p(x)$ and $p(y|x) = p(y)$. These two conditions are synonyms as joint and conditional probabilities are connected by Bayes' rule $p(x,y) = p(x|y)\,p(y) = p(y|x)\,p(x)$. Using two random variables we can also state the independence through the Shannon entropy studied in Ch. 2. Two random variables are uncorrelated if and only if the mutual information is zero. With this criterion we may begin to interpret the correlation as a consequence of information sharing between the two random variables. In the quantum realm the probability measure is associated with the density operator of the quantum state, i.e., the density operator of a quantum system is the object that we use as a probability distribution to calculate probabilities of outcomes and averages. What we are going to show now is the condition over a density operator of two systems S^A and S^B, ρ^{AB}, that makes them uncorrelated.

Within a fully probabilistic perspective, the joint probability distribution of two outcomes a and b associated with two observables A and B, where A is an observable of the system S^A and B is an observable of the system S^B, is given by

$$p(a,b) = \mathrm{Tr}_{AB}\left[\rho^{AB}\,|a\rangle\langle a| \otimes |b\rangle\langle b|\right]. \tag{4.34}$$

Based on what we stated before the two systems are statistically independent, or uncorrelated, if and only if $p(a,b) = p(a)\,p(b)$, where $p(a) = \mathrm{Tr}_A\left[\rho^A\,|a\rangle\langle a|\right]$ and $p(b) = \mathrm{Tr}_B\left[\rho^B\,|b\rangle\langle b|\right]$. As usual, ρ^A and ρ^B stand for the reduced density operators of the system S^A and S^B. The only way to satisfy this condition is if the composite state is a product state, i.e.,

$$\rho^{AB} = \rho^A \otimes \rho^B. \tag{4.35}$$

We see that the separability condition is transfered from the probabilities to the density operators.

Entanglement is one type of quantum correlation whereas correlations may appear as classical or quantum, local or nonlocal. Usually, local and nonlocal correla-

tions are measured by Bell-like inequalities and classical and quantum correlations are measured by an appropriate measure which is based on a notion of nonclassicality. It is known that not all entangled states violate the Bell inequalities, meaning that quantum correlations may exhibit local or nonlocal character [31]. This reveals the complex structure of correlations. In this work we will only focus on the classical-quantum character of correlations.

When considering just pure states a relevant question for this matter is whether the state is entangled or not. We have already defined an entangled state as a superposition of product states (cf. Sec. (3.2)). A non-entangled state is necessarily a product state which is nothing but a separable state. So separability is a reasonably good criterion to characterize entangled states. A state is either entangled or separable. Even though the state given by Eq. (4.35) is separable the class of separable states for density operator is more general than that. An arbitrary density operator is called to be separable if and only if

$$\rho^{AB} = \sum_i p_i \rho_i^A \otimes \rho_i^B, \tag{4.36}$$

where $\{p_i\}$ is a normalized probability distribution, i.e., $p_i \geq 0$ and $\sum_i p_i = 1$. We shall see that there is a definite way to prepare a state like this and we will only use classical concepts for this, and therefore, interpretationally, we should not expect any entanglement. Mathematically stated a separable state is a convex combination of product states.

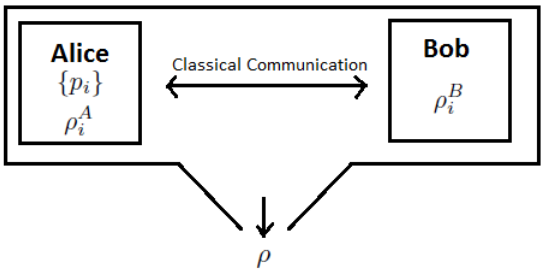

Figure 4.2: Preparation of a separable state from local operations and classical communication (LOCC).

Consider Fig. 4.2, Alice has access to a classical probability distribution $\{p_i\}$ and she has already established with Bob that if the outcome is i she prepares

the state ρ_i^A in the system S^A and Bob prepares the state ρ_i^B in the system S^B. The states ρ_i^A and ρ_i^B are in general different. Then, whenever an outcome of the probability distribution results, she uses a classical channel to communicate to Bob which outcome occurred so that he can prepare the corresponding state. After preparing both states they send both systems to a third party where someone can perform measurements with both systems. The correct density operator for this third person to use in order to describe the outcomes of measurements is the separable state given by Eq. (4.36). Therefore, any separable state can, in principle, be prepared by local operations and classical communication (LOCC). A natural question is: how can we know if a given state is separable or entangled? For pure states the strategy can be reduced to finding a basis such that the product nature of the state is explicitly shown. For an arbitrary density operator one should seek that it can (or can not) be written in the form given by Eq. (4.36). This is a non-trivial task which has been solved explicitly only for a few cases. This problem is called the separability criterion problem and we will not be concerned with it in this work. Rather, we just say for completeness that there are some methods of identifying the separability of a density operator and some of them are: the Peres-Horodecki (positive partial transpose, PPT) criterion, which says that the partial transposition* is a necessary condition for separability [32] and in some particular cases even sufficient [33], the reduction criterion and the majorization criterion [34].

There is at least another way to look at correlations which is through the correlations of measurements. If we are given, sequentially, a pair of boxes with one ball in each of them and we are asked to "measure" the color of them, after a sufficiently large number of measurements we would be able to infer the probability of ball one be blue when ball two is red, et cetera. This set of probabilities of joint measurements gives us the information of the correlation between the two balls. Let us interpret this with some examples.

Suppose the balls can be in any of two states, blue or red. As these states are macroscopically distinguishable, it is possible to represent them quantum mechanically as two orthogonal states $|B_i\rangle$ and $|R_i\rangle$, where i is the labeling of the boxes.

*Partial transposition of some party, say Alice, is the transposition of the indices of that party in the density operator. With a density operator of two parties the partial transposition maps $\rho_{m\mu,n\nu} \to \left(\rho^{T_A}\right)_{n\mu,m\nu}$.

Therefore, if the ensemble of boxes are described by a density operator

$$\rho = |B_1\rangle \langle B_1| \otimes |R_2\rangle \langle R_2| \tag{4.37}$$

we will always find ball one to be blue and ball two to be red. It seems that this state is perfectly correlated but in fact this high precision means that they are on their own and therefore independent. Now suppose the state

$$\rho = \frac{1}{3} |B_1\rangle \langle B_1| \otimes |R_2\rangle \langle R_2| + \frac{2}{3} |R_1\rangle \langle R_1| \otimes |B_2\rangle \langle B_2| . \tag{4.38}$$

This state tells us that there is a probability of one-third to find the combination B_1R_2 and two-thirds to find the R_1B_2. This state is a convex-sum of product states and therefore is a separable state. As this state is completely classical, despite our Dirac notation, it is said to be classically correlated. As this state is separable, one might be tempted to think that all separable states are classically correlated, but this is not true. There are notions of classicality, such as that provided by quantum discord, that show that separable states can be quantum correlated even though we presented the class of separable states as being reasonably connected with classical preparation ways, classical communication and local measurements. Quantum correlations can arise in separable states because the states being prepared by LOCC may not be distinguishable, a property exclusivelly quantum.

These two states (blue and red) were based on classical ideas but to deepen the discussion we will inevitably have to consider really quantum states, exploiting the non-distinguishablity. The two balls now are replaced by two spins, blue color is the up state $|0\rangle$ and red color is the down state $|1\rangle$ in a conventionally established z-direction. Now there is the possibility to have x-polarized spin states which share information about the z-polarized ones, i.e., that are not orthogonal. We represent by $|+\rangle$ and $|-\rangle$ the x_+-polarization and x_--polarization, respectively. Consider the state

$$\rho = \frac{1}{3} |0\rangle \langle 0| \otimes |+\rangle \langle +| + \frac{2}{3} |1\rangle \langle 1| \otimes |-\rangle \langle -| . \tag{4.39}$$

Comparing the states of Eqs. (4.38) and (4.39), in the first situation, classical-like, we are dealing with only just one property of the particle (ball), its color. Whereas in the second situation, quantum-like, we have two different measurable properties, z- and x-polarized spin directions. Would we get a state like that of Eq. (4.39) if we tried to extend the classically correlated state considering two

properties of the balls? The answer is no, precisely because quantum states may not be distinguishable. Classically, if we considered another property of the ball, like dyeing a small black dot on them or not, this property will necessarily be distinguishable from the previous one. This will classically always happen because of the way nature is at that level. Being always distinguishable may be roughly stated as, the ball may always possess the two properties at the same time. This is sharply contrasted with quantum mechanics because, in our example, a particle cannot have well-defined polarizations in the two directions. This gives rise to the non-orthogonality and indistinguishability. The state given by Eq. (4.39) surely possesses correlations but are neither classical nor entanglement. These correlations are called quantum correlations.

The last situation occurs when we have entangled states. Consider the singlet state $\left|\Psi_{-}^{AB}\right\rangle = \frac{1}{\sqrt{2}}\left(|01\rangle - |10\rangle\right)$. Its density operator is given by

$$\rho = \frac{1}{2}\left[|01\rangle\langle 01| - |01\rangle\langle 10| - |10\rangle\langle 01| + |10\rangle\langle 10|\right]. \tag{4.40}$$

This state is also possess quantum correlations but from a particular type, entanglement.

In this section we discussed the various types of correlations, classical, quantum and entanglement. In the next section our goal will be to provide tools to measure these different correlations.

4.4 Measuring Correlations

Quantum correlations have been experimentally verified in optical [35,36,37,38,39] [40] and nuclear magnetic ressonance (NMR) [42,43,44,45,46] systems. With three particles [38] and continuous degree of freedom [39]. In this section we will study some measures of correlation for bipartite quantum systems. As we emphasized previously, a pure state can be entangled or not. There is no room for classical correlations in a pure state because a pure state, as it says, is purely quantum. In order to have classical correlations we necessarily need mixed states. Therefore, we will separate the measuring of correlations into two parts. In the first part we will study how we measure the amount of entanglement in pure states and in the second part we will see how to measure quantum and classical correlations in general density operators, taking into account the mixed states. As we shall see in the next section,

knowing how to measure quantum correlations is of paramount importance in quantum information theory. It is widely accepted nowadays that for a specific protocol on a quantum computer surpass a similar protocol in a classical computer we must use quantum correlations as a resource in the computation. For instance, in a work by Brukner, *et al.* [47] they showed that: "there always exists a communication complexity problem, for which a protocol assisted by states which violate the [Bell] inequality is more efficient than any classical protocol." It still is an open problem to know if an arbitrary multipartite state is entangled, classically correlated or if it possess quantum correlations. For the bipartite case there are several measures of correlations and some of those we will study in this section.

4.4.1 Correlations in Pure States; Entanglement

As we concluded previously, a pure state which is correlated is necessarily entangled. Moreover, we also established that for a pure quantum state be entangled it must be a superposition of product states or, in other words, it must not be a product state.

One way to identify if a pure state is entangled or not is using the Schmidt decomposition (cf. App. A). The Schmidt decomposition is nothing but a theorem in linear algebra applied to the quantum states. This theorem is called Singular Value Decomposition and its demonstration can be found in the App. A. In the context of quantum mechanics it states the following. Consider two Hilbert spaces $\mathcal{H}^A$ and $\mathcal{H}^B$ with dimensions n_A and n_B, respectively, and an arbitrary composite pure state $\left|\psi^{AB}\right\rangle$. Moreover, suppose $\left\{\left|i^A\right\rangle, i \in \{1,\ldots,n_A\}\right\}$ and $\left\{\left|j^B\right\rangle, j \in \{1,\ldots,n_B\}\right\}$ are two orthonormal basis for $\mathcal{H}^A$ and $\mathcal{H}^B$, respectively. This state may be written as

$$\left|\psi^{AB}\right\rangle = \sum_{i=1}^{n_A}\sum_{j=1}^{n_B} \alpha_{ij} \left|i^A, j^B\right\rangle. \tag{4.41}$$

The Schmidt decomposition guarantees that such an arbitrary state may always be rewritten as

$$\left|\psi^{AB}\right\rangle = \sum_{n=1}^{k} \sqrt{p_n} \left|n^A, n^B\right\rangle, \tag{4.42}$$

where $k = \min(n_A, n_B)$ is the Schmidt rank, $p_n \geq 0$ for every n and $\left\{\left|n^A\right\rangle\right\}$ and $\left\{\left|n^B\right\rangle\right\}$ are orthonormalized eigenvectors of ρ^A and ρ^B, respectively. The Schmidt number is defined to be $N_S = \sum_n^k p_n^2$. We can observe directly from the Eq. (4.42)

that if the state is not entangled it means that both Schmidt rank and Schmidt number are equal to one. If the state is maximally entangled, then $p_n = 1/k$ for every n and therefore the Schmidt number assumes the value $N_S = 1/k$. The Schmidt rank can be used to identify if the state is separable and therefore not entangled. The Schmidt number can be used as a measure of entanglement, as it may vary in the range $1/k \leq N_S \leq 1$.

Comparing the Schmidt number and the purity (cf. Eq. (3.22)) for bipartite systems we see that they are the same. Therefore, the purity is also a measure of the amount of entanglement.

The von Neumann entropy of a pure state is always zero as we have already discussed in Sec. (4.1). Notwithstanding we can use it to measure the entanglement of a bipartite state. The protocol consists in evaluate the reduced density operator for the system S^A and calculate the von Neumann entropy for it. Or the same for the system S^B. As we already said, the entanglement in a system is connected with the mixedness of its subsystems. The von Neumann entropy for a bipartite system of dimension N is bounded by $0 \leq S(\rho) \leq \log N$. Zero is achieved when ρ is a pure state and the upper bound is achieved when ρ is completely mixed. The von Neumann entropy for any pure bipartite quantum state is therefore always zero, $S\left(\rho^{AB}\right) = 0$. If the state is a product state then $S\left(\rho^A\right) = S\left(\rho^B\right) = 0$. But for a maximally entangled state, such as the four Bell states, we find $S\left(\rho^A\right) = S\left(\rho^B\right) = 1$, which means that locally we lost information about the system altogether. The information about the composite system is contained exclusively in the correlations between the subsystems. Therefore we define the entropy of entanglement for a bipartite pure state as

$$E(\psi) = S\left(\rho^A\right) = S\left(\rho^B\right) = -\sum_{i=1}^{k} p_i \log p_i. \tag{4.43}$$

This entropy is bounded by $0 \leq E(\psi) \leq 1$. Observe that we claimed that $S\left(\rho^A\right) = S\left(\rho^B\right)$ in the expression above. We can see this from the Schmidt decomposition as ρ^A and ρ^B have the same eigenvalues (cf. App. A).

We see that these three measures are closely connected. The conclusion is that for pure bipartite states, entanglement is directly proportional to the mixedness of the reduced states. This means that the information about the composite system is stored in the quantum correlations, entanglement, i.e., locally we can not distinguish

a system prepared in a maximally mixed state from a system which is maximally entangled with a distant system, the last of which is sometimes referred to an improper mixture. Let us calculate the entropy of entanglement for a given state.

Consider the following state

$$\left|\psi^{AB}\right\rangle = \sqrt{\lambda}\left|00\right\rangle + \sqrt{(1-\lambda)}\left|11\right\rangle, \tag{4.44}$$

where $0 \leq \lambda \leq 1$. Observe that λ is a parameter which measures the weight of the states in the superposition. If λ is zero or one then we have a product state. If λ is $1/2$ we have a Bell state, which we are finally going to show that it is maximally entangled. Thus, by varying λ, we vary the amount of entanglement in the state.

In order to calculate the entropy of entanglement we must obtain the reduced density operator of one of the subsystems. Let us obtain the reduced state for the system S^A. The composite density operator is given by

$$\begin{aligned}\rho^{AB} =& \lambda^2 \left|00\right\rangle\left\langle 00\right| + \sqrt{\lambda\left(1-\lambda\right)}\left|00\right\rangle\left\langle 11\right| \\ &+ \sqrt{\lambda\left(1-\lambda\right)}\left|11\right\rangle\left\langle 00\right| + \left(1-\lambda\right)\left|11\right\rangle\left\langle 11\right|.\end{aligned} \tag{4.45}$$

The reduced density operator for the system S^A is

$$\begin{aligned}\rho^A = \mathrm{Tr}_B\left[\rho^{AB}\right] =& \left\langle 0^B\right|\rho^{AB}\left|0^B\right\rangle + \left\langle 1^B\right|\rho^{AB}\left|1^B\right\rangle \\ =& \lambda\left|0^A\right\rangle\left\langle 0^A\right| + \left(1-\lambda\right)\left|1^A\right\rangle\left\langle 1^A\right|.\end{aligned} \tag{4.46}$$

The reduced state is already diagonalized, therefore the entropy of entanglement is given by

$$E\left(\psi^{AB}\right) = -\lambda\log\lambda - \left(1-\lambda\right)\log\left(1-\lambda\right). \tag{4.47}$$

A plot of this quantity is presented in Fig. 4.3. As we claimed the entanglement achieves its maximum for the Bell state, when $\lambda = 1/2$, and it is zero when $\lambda = 0, 1$.

4.4.2 Correlations in Arbitrary Bipartite States; Accounting for Mixed States

In this section we will discuss other measures of correlations. We will be able to distinguish, in some measure, the quantum from the classical correlations and, moreover, see that entanglement is not the only quantum correlation that exists. When we studied the correlations of two messages in Ch. 2 we interpreted the mutual information $I\left(X;Y\right)$ as the amount of information shared by the two sources,

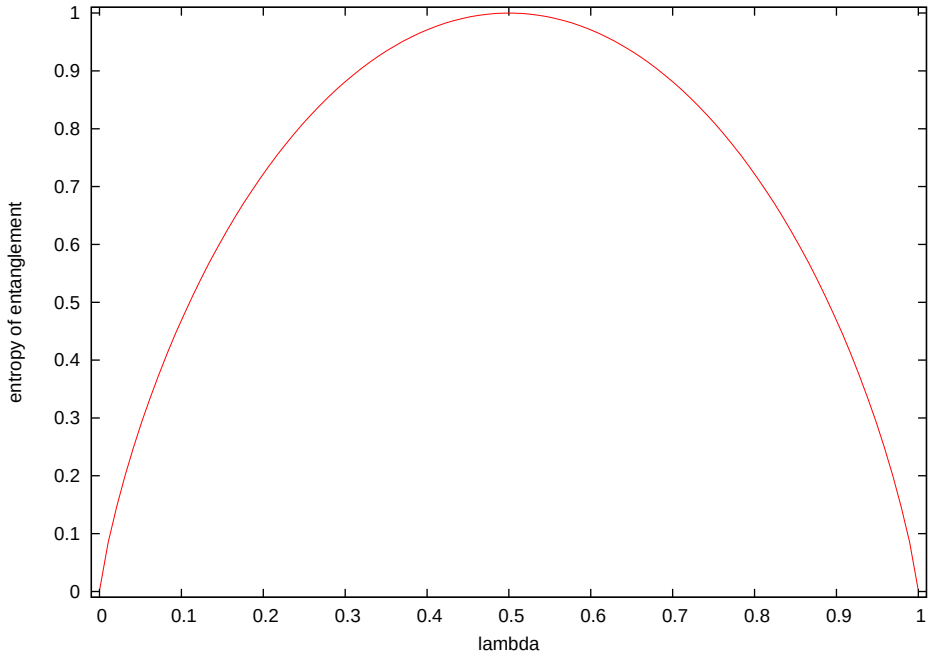

Figure 4.3: Entropy of entanglement as a function of λ.

i.e., they are correlated or not. The two sources have no information in common if and only if they are uncorrelated, $p(x,y) = p(x)\,p(y)$. Therefore, the mutual information can be seen to measure the total amount of correlation between the two sources X and Y. We will now extend the concept of mutual information to the quantum level in terms of the von Neumann entropy. This quantum mutual information is commonly interpreted to measure the total amount of correlation between two subsystems.

In order to define mutual information recall that we needed the definition of joint entropy. Given a system S^{AB} which is composed by two subsystems S^A and S^B they can be in an arbitrary state given by the density operator ρ^{AB}. The quantum analogue of the joint entropy is just the von Neumann entropy for this density operator,

$$S\left(\rho^{AB}\right) = S(AB) = -\mathrm{Tr}\left[\rho^{AB}\log\rho^{AB}\right]. \tag{4.48}$$

If the two subsystems are uncorrelated, $\rho^{AB} = \rho^A \otimes \rho^B$, then $S\left(\rho^A \otimes \rho^B\right) = S\left(\rho^A\right) + S\left(\rho^B\right)$ which is an expected property for the joint entropy to satisfy. The quantum mutual information is defined analogously to the classical counterpart as

$$I\left(\rho^{AB}\right) = I(A:B) = S\left(\rho^A\right) + S\left(\rho^B\right) - S\left(\rho^{AB}\right). \tag{4.49}$$

Let us evaluate the quantum mutual information for some states.

Firstly consider two uncorrelated qubits

$$\rho^{AB} = \frac{1}{4} I^{AB} = \left(\frac{1}{2} I^{A}\right) \otimes \left(\frac{1}{2} I^{B}\right) = \rho^{A} \otimes \rho^{B}. \tag{4.50}$$

The joint entropy is $S(AB) = 2$. The two reduced states are both maximally mixed and therefore $S(A) = S(B) = 1$. Thus the quantum mutual information is $I(A:B) = 0$. Observe that the Venn diagram interpretation given in Sec. 2.4 is valid.

In the second example consider a classically correlated state

$$\rho^{AB} = \frac{1}{2}\left(\left|0^{A}\right\rangle\left\langle 0^{A}\right| \otimes \left|1^{B}\right\rangle\left\langle 1^{B}\right| + \left|1^{A}\right\rangle\left\langle 1^{A}\right| \otimes \left|0^{B}\right\rangle\left\langle 0^{B}\right|\right). \tag{4.51}$$

In the computational basis, $\rho^{AB} = \mathrm{diag}\left(0, \frac{1}{2}, \frac{1}{2}, 0\right)$. Therefore, $S(AB) = 1$. The reduced density operators are again maximally mixed and therefore $S(A) = S(B) = 1$. The quantum mutual information takes the value $I(A:B) = 1$. For this situation the Venn diagram interpretation is still valid.

The last example assumes an entangled state

$$\rho^{AB} = \left|\Phi_{+}^{AB}\right\rangle\left\langle\Phi_{+}^{AB}\right|. \tag{4.52}$$

This state is pure, $S(AB) = 0$. The reduced states are again maximally mixed, $S(A) = S(B) = 1$. Therefore $I(A:B) = 2$. The noteworthy aspect is that with these values for the quantum entropies, the Venn diagram interpretation collapses. This is a fundamental consequence of quantum mechanics, of entanglement. Classically it is impossible for the mutual information be greater than the joint entropy, as happened here. This suggests that entanglement is a property fundamentally different from any classical counterpart. Other inequalities between entropies are violated, e.g., $H(X) \geq I(X;Y)$.

Let us now study a proposed measure of entanglement for bipartite mixtures. We have seen that the entropy of entanglement is given by $E\left(\psi^{AB}\right) = S\left(\rho^{A}\right) = S\left(\rho^{B}\right)$. So let us calculate the entropy of entanglement for an arbitrary two qubits state

$$\left|\psi^{AB}\right\rangle = a\left|00\right\rangle + b\left|01\right\rangle + c\left|10\right\rangle + d\left|11\right\rangle. \tag{4.53}$$

This provides a density operator which in the basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ can be

written as

$$\rho^{AB} = \begin{pmatrix} |a|^2 & ab^* & ac^* & ad^* \\ ba^* & |b|^2 & bc^* & bd^* \\ ca^* & cb^* & |c|^2 & cd^* \\ da^* & db^* & dc^* & |d|^2 \end{pmatrix}. \tag{4.54}$$

The reduced density operator for the system S^A can be evaluated given us

$$\rho^A = \begin{pmatrix} |a|^2 + |b|^2 & ac^* + bd^* \\ ca * + db^* & |c|^2 + |d|^2 \end{pmatrix}. \tag{4.55}$$

In order to evaluate the von Neumann entropy of this state we must diagonalize it. The eigenvalues of a two-by-two matrix A are easily shown to be

$$\lambda_\pm = \frac{\mathrm{Tr}\left[A\right] \pm \sqrt{\left(\mathrm{Tr}\left[A\right]\right)^2 - 4\mathrm{Det}\left[A\right]}}{2}. \tag{4.56}$$

The trace is trivially given by $\mathrm{Tr}\left[\rho^A\right] = 1$, by normalization. The determinant is

$$\begin{aligned} \mathrm{Det}\left[\rho^A\right] &= \left(|a|^2 + \left|b^2\right|\right)\left(|c|^2 + |d|^2\right) - \left(ac^* + bd^*\right)\left(ca^* + db^*\right) \\ &= |a|^2|c|^2 + |a|^2|d|^2 + |b|^2|c|^2 + |b|^2|d|^2 - |a|^2|c|^2 - adc^*b^* - cbd^*a^* - |b|^2|d|^2 \\ &= |a|^2|d|^2 - 2\mathrm{Re}\left[adb^*c^*\right] + |b|^2|c|^2 = (ad - bc)(ad - bc)^* = |ad - bc|^2. \end{aligned} \tag{4.57}$$

Concurrence [48] is defined as

$$C\left(\psi^{AB}\right) := 2\left|ad - bc\right|. \tag{4.58}$$

The eigenvalues of ρ^A can be written as

$$\xi_\pm = \frac{1 \pm \sqrt{1 - \left[C\left(\psi^{AB}\right)\right]^2}}{2}. \tag{4.59}$$

The entropy of entanglement is given by

$$\begin{aligned} E\left(\psi^{AB}\right) = -\xi_+ \log \xi_+ - \xi_- \log \xi_- =& -\left(\frac{1+\sqrt{1-C^2}}{2}\right)\log\left(\frac{1+\sqrt{1-C^2}}{2}\right) \\ & -\left(\frac{1-\sqrt{1-C^2}}{2}\right)\log\left(\frac{1-\sqrt{1-C^2}}{2}\right) \\ =& h\left(\frac{1+\sqrt{1-C^2}}{2}\right), \end{aligned} \tag{4.60}$$

where the newly introduced function is defined as

$$h\left(x\right) = -\left(x \log x + (1-x)\log\left(1-x\right)\right). \tag{4.61}$$

As the entropy of entanglement is a monotonic function of the concurrence, the concurrence is also used as a measure of entanglement. If $C = 0$, the state is separable and if $C = 1$, the state is maximally entangled. Until now we just found another measure for bipartite pure states. But we claimed that this measure is valid for mixtures. We will write the concurrence as a function of the density operator elements and then we will be able to extend the measure for mixtures.

The concurrence can also be written as

$$C\left(\psi^{AB}\right) = \left|\left\langle \psi^{AB} \,\middle|\, \bar{\psi}^{AB} \right\rangle\right|. \tag{4.62}$$

The state vector $\left|\bar{\psi}^{AB}\right\rangle$ is obtained from $\left|\psi^{AB}\right\rangle$, by first taking the complex conjugate of the coefficients in the computational basis, from Eq. (4.53)

$$\left|\psi^{AB}\right\rangle^* = a^* \left|00\right\rangle + b^* \left|01\right\rangle + c^* \left|10\right\rangle + d^* \left|11\right\rangle, \tag{4.63}$$

and then applying the following operator

$$\left|\bar{\psi}^{AB}\right\rangle = \sigma_2^A \otimes \sigma_2^B \left|\psi^{AB}\right\rangle^*. \tag{4.64}$$

The effect of σ_2[*] is to change the states of the computational basis and to insert the relative phase $\pm i$. In the basis $\{\left|00\right\rangle, \left|01\right\rangle, \left|10\right\rangle, \left|11\right\rangle\}$ we find

$$\sigma_2^A \otimes \sigma_2^B = \begin{pmatrix} 0 & -i\sigma_2 \\ i\sigma_2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}. \tag{4.65}$$

The resulting state vector is

$$\left|\bar{\psi}^{AB}\right\rangle = -d^* \left|00\right\rangle + c^* \left|01\right\rangle + b^* \left|10\right\rangle - a^* \left|11\right\rangle. \tag{4.66}$$

The inner product is, thus,

$$\left\langle \psi^{AB} \,\middle|\, \bar{\psi}^{AB} \right\rangle = -a^* d^* + b^* c^* + c^* b^* - d^* a^* = -2\left(a^* d^* - b^* c^*\right), \tag{4.67}$$

and finally

$$\begin{aligned} \sqrt{\left|\left\langle \psi^{AB} \,\middle|\, \bar{\psi}^{AB} \right\rangle\right|^2} &= \sqrt{4\left(a^* d^* - b^* c^*\right)\left(ad - bc\right)} \\ &= \sqrt{4\left|ad - bc\right|^2} = 2\left|ad - bc\right| = C\left(\psi^{AB}\right). \end{aligned} \tag{4.68}$$

[*] $\sigma_2 = -i\left|0\right\rangle\left\langle 1\right| + i\left|1\right\rangle\left\langle 0\right|$ is a Pauli matrix.

The next step is to write this expression in terms of density operators. Begin with

$$\begin{aligned} C^2\left(\psi^{AB}\right) =& \left|\left\langle\psi^{AB}\,\middle|\,\bar{\psi}^{AB}\right\rangle\right|^2 = \left\langle\psi^{AB}\,\middle|\,\bar{\psi}^{AB}\right\rangle\left\langle\bar{\psi}^{AB}\,\middle|\,\psi^{AB}\right\rangle \\ =& \mathrm{Tr}\left[\left(\left|\psi^{AB}\right\rangle\left\langle\psi^{AB}\right|\right)\left(\left|\bar{\psi}^{AB}\right\rangle\left\langle\bar{\psi}^{AB}\right|\right)\right] \\ =& \mathrm{Tr}\left[\rho^{AB}\bar{\rho}^{AB}\right] = \mathrm{Tr}\left[R^{AB}\right], \end{aligned} \tag{4.69}$$

where $\rho^{AB} = \left|\psi^{AB}\right\rangle\left\langle\psi^{AB}\right|$, $\bar{\rho}^{AB} = \left|\bar{\psi}^{AB}\right\rangle\left\langle\bar{\psi}^{AB}\right|$ and $R^{AB} = \rho^{AB}\bar{\rho}^{AB}$. The density operator $\bar{\rho}^{AB}$ is naturally constructed as

$$\bar{\rho}^{AB} = \left|\bar{\psi}^{AB}\right\rangle\left\langle\bar{\psi}^{AB}\right| = \left(\sigma_2^A \otimes \sigma_2^B\right)\rho^{*AB}\left(\sigma_2^A \otimes \sigma_2^B\right), \tag{4.70}$$

where ρ^{*AB} is generated by writing ρ^{AB} as a matrix in terms of the computational basis and taking the complex conjugate of the matrix elements.

For any density operator ρ^{AB} we can now find the concurrence $C\left(\rho^{AB}\right)$ using Eq. (4.69). Wootters has shown that the concurrence can also be written explicitly in the form [48]

$$C\left(\rho^{AB}\right) = \max\left\{0, \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4\right\}, \tag{4.71}$$

where λ_1, λ_2, λ_3 and λ_4 are the square roots of the eigenvalues of the matrix $R^{AB} = \rho^{AB}\bar{\rho}^{AB}$. They are non-negative real numbers and λ_1 is the largest eigenvalue. The question which remains is that if this measure of entanglement for mixtures is satisfactory physically. Wootters also showed this and we will discuss this next.

Any mixture can be written as an ensemble of pure states

$$\rho^{AB} = \sum_i p_i \left|\psi_i^{AB}\right\rangle\left\langle\psi_i^{AB}\right|. \tag{4.72}$$

The mean entanglement of the state ρ^{AB} is found as the mean value of the entropy of entanglement $E\left(\psi_i^{AB}\right)$ of the pure states, i.e. $\sum_i p_i E\left(\psi_i^{AB}\right)$. The entanglement of formation of ρ^{AB} is defined as the minimum value of the mean entanglement if one considers all the possible ensemble decompositions of ρ^{AB}, i.e.,

$$E_f\left(\rho^{AB}\right) = \min \sum_i p_i E\left(\psi_i^{AB}\right). \tag{4.73}$$

This means that if we prepare a quantum system with the density operator ρ^{AB} as a mixture, then at least the entanglement $E_f\left(\rho^{AB}\right)$ must be produced on average. Wootters also showed that the entanglement of formation, which is a physically

reasonable measure of entanglement for mixtures, is a function of concurrence as given by

$$E_f\left(\rho^{AB}\right) = h\left(\frac{1+\sqrt{1-C^2\left(\rho^{AB}\right)}}{2}\right), \tag{4.74}$$

where h is given by Eq. (4.61). As the function h is monotonic, the concurrence can also be considered a measure of entanglement for mixtures, even without a direct interpretation. As an example of the calculation we refer to Ch. 5 where we use the concurrence as a measure for the toy model studied in this work and experimental determination of concurrence can be found in Refs. [35,36,37].

We studied two measures of entanglement for arbitrary bipartite states, entanglement of formation and concurrence. Now we are going to study a measure of correlations in a density operator. This approach is fundamentally based on the classical two message entropies, mutual information, conditional entropy and joint entropy. We will begin recalling briefly some relations. For a more detailed explanation see Ch. 2.

We have seen that, in the classical theory of information, the Shannon entropy $H(X)$ is the appropriate measure of information of a random variable X. The extension to two random variables is straightforward just replacing the probabilities by the joint probability distribution. This entropy is the joint entropy $H(X,Y)$ and it measures the total amount of non redundant information in the two messages. The mutual information $I(X;Y)$ measures the amount of common information or, the total amount of correlations between the two random variables. There is also the conditional entropy $H(X|Y)$ which measures the amount of information we will receive from X given that we know Y. The mutual information is related with the other entropies through two different expressions, namely

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \tag{4.75}$$

and

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X). \tag{4.76}$$

In the context of quantum mechanics, the Shannon entropy is associated with the von Neumann entropy $S(\rho)$. The analogous of the joint entropy is just the von Neumann entropy of a bipartite system, i.e., $S\left(\rho^{AB}\right)$. This is somewhat natural because this density operator characterizes all the statistical prediction of measurements over the composite system S^{AB}. From Eq. (4.75) we define the quantum

mutual information as

$$I\left(\rho^{AB}\right) = S\left(\rho^{A}\right) + S\left(\rho^{B}\right) - S\left(\rho^{AB}\right). \tag{4.77}$$

This expression resembles directly the classical mutual information. Because of this we interpret this quantity as a measure of the total amount of correlations between S^A and S^B. Remember the beginning of this section where we explored the quantum mutual information in some examples and it is at least reasonable to interpret it like that. The turning point here is that there exists another classical way to write the mutual information. That is given by the Eq. (4.76). How can we extend this equation to the quantum realm? This is a very interesting question and it was worked out in 2001 by Ollivier and Zurek in an article entitled "Quantum Discord: A Measure of Quantum Correlations" [49]. We shall reproduce some of their calculations.

Mainly, the problem is how to write a quantum conditional entropy, as it involves the notion of a pre-acquired knowledge. To obtain any knowledge of a quantum system one must measure it. But quantum mechanically this is a strong intervention, as this changes the abruptly the state of the quantum system. Moreover, different measures may change more or less the state. In this way a quantum conditional entropy should be measurement-dependent. Suppose we perform a projective measurement on the system S^B, $\left\{\hat{\Pi}_k^B\right\}$, in a composite quantum state ρ^{AB}. If the outcome is k then the normalized post-measurement state of the system S^A is given by

$$\rho^{A}_{\Pi_k^B} = \frac{\mathrm{Tr}_B\left[\hat{\Pi}_k^B \rho^{AB} \hat{\Pi}_k^B\right]}{p_k}, \tag{4.78}$$

where the hat means the extended operator, $\hat{\Pi}_k^B = I^A \otimes \Pi_k^B$. The amount of information of the system S^A which we can extract after getting this outcome k for the system S^B is $S\left(\rho^{A}_{\Pi_k^B}\right)$. But this set of projection operators could have given any of the other results. Therefore the average information we can extract from the system S^A given that we have measured the system S^B with the set $\left\{\Pi_k^B\right\}$ of projection operators is

$$\tilde{S}\left(A|B\right)_{\left\{\Pi_k^B\right\}} = \sum_k p_k S\left(\rho^{A}_{\Pi_k^B}\right). \tag{4.79}$$

This is the interpretation for a conditional entropy. Therefore the quantum mutual information is also given by

$$J_{\left\{\Pi_k^B\right\}}\left(\rho^{AB}\right) = S\left(\rho^{A}\right) - \tilde{S}\left(A|B\right)_{\left\{\Pi_k^B\right\}}. \tag{4.80}$$

This quantity is interpreted as a measure of classical correlations as it gives the difference in uncertainty about the subsystem S^A before and after the measurement. But still, this is a measurement-dependent quantity. In principle we could measure the system S^B with any set of mutually orthogonal projection operators, as we are considering a projective measurement. Therefore, in order to have a measure of classical correlations independent of the measurement, we consider the maximization of $J_{\left\{\Pi_k^B\right\}}\left(\rho^{AB}\right)$. Maximize this quantity means to consider the largest amount of classical correlations that can be obtained through a local measurement on S^B. In spite of the fact we began with the classical mutual information which is symmetric if we exchange X and Y, this quantity is not. Thus, the total amount of correlations is given by $I\left(\rho^{AB}\right)$ and the amount of classical correlations by

$$J\left(\rho^{AB}\right) = \max_{\left\{\Pi_k^B\right\}} J_{\left\{\Pi_k^B\right\}}\left(\rho^{AB}\right) = S\left(\rho^A\right) - \min_{\left\{\Pi_k^B\right\}} \tilde{S}\left(A|B\right)_{\left\{\Pi_k^B\right\}}, \tag{4.81}$$

leading us to a measure of quantum correlations given by the difference

$$Q\left(\rho^{AB}\right) = I\left(\rho^{AB}\right) - J\left(\rho^{AB}\right). \tag{4.82}$$

The quantity $Q\left(\rho^{AB}\right)$ is called quantum discord and we will show below that it may be a resource for quantum computation.

At the same time of the quantum discord's paper, Henderson and Vedral published an article [50] arguing that the quantity given by Eq. (4.81) is a proper measure of the amount of classical correlations in a bipartite quantum system. Experimentally, quantum discord have been widelly used to measure the amount of quantum correlation in systems [41,42,43,44]. For instance, in Ref. [42] it was the first experimental evidence of nonclassicality in NMR systems. They measure the dynamics of quantum discord with a sample of ^{23}Na nuclear spin in a lyotropic liquid-crystal at room temperature ($26°C$). Also, the dynamics of quantum discord and its classical counterpart have been studied in different local environments [51] showing a rich dynamics of these correlations.

4.5 Applications of Quantum Correlations

In this section we will see some applications of the concepts developed in this chapter, those of which arised in the context of quantum information. Each sec-

tion will deal with some property or concept derived from the study of quantum correlations.

4.5.1 No-Cloning Theorem

As we shall see below, the no-cloning theorem is a direct consequence of the linearity of quantum operations [3,4], which says that an unknown quantum state cannot be cloned (copied). For practical applications the most striking consequence of this statement is that one cannot transcribe naively the classical error correction techniques for quantum computation, since they are mostly based on backing up copies of the state in the middle of a computation. For this reason we have to develop quantum error correction techniques in order to deal with quantum computations [52].

An operation U which would copy a quantum state $|\psi\rangle$ must be such that

$$\left|\psi^A\right\rangle \left|a^B\right\rangle \xrightarrow{U} \left|\psi^A\right\rangle \left|\psi^B\right\rangle, \tag{4.83}$$

where $\left|a^B\right\rangle$ is an auxiliary quantum state which would store the copied state.

First we will consider a situation in which copying is possible. Let $\{|0\rangle, |1\rangle\}$ be two states of a qubit. Consider also a unitary transformation on a system composed of two qubits given by the matrix (the basis used is $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$).

$$CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{4.84}$$

This operation is called controlled NOT gate. It is an extremely important operation in quantum computation. It is a reversible generalization of the XOR gate of classical computation*. Operationally this gate affects the qubits in the following way: $|x, y\rangle \longrightarrow |x, y \oplus x\rangle$, where the symbol $\oplus$ stands for the module-2 sum. The first qubit is called the control qubit, as it remains the same and it may or may not change the second qubit. The second qubit is called the target qubit because it is the qubit which may or may not be changed by the gate. Two of the possible

*The XOR logical gate implements an exclusive "or", i.e., a true output results if one, and only one, of the two inputs is true. This is contrasted with the standard OR gate where besides these true restults it also results in a true output when the two inputs are both true.

operations out of the four possible pairs of input qubit-qubit states can be interpreted as a coping operation, namely

$$\left|0^A, 0^B\right\rangle \overset{CNOT}{\longrightarrow} \left|0^A, 0^B\right\rangle \tag{4.85}$$

and

$$\left|1^A, 0^B\right\rangle \overset{CNOT}{\longrightarrow} \left|1^A, 1^B\right\rangle. \tag{4.86}$$

Consider now an unknown qubit state $\left|\psi^A\right\rangle = \alpha\left|0^A\right\rangle + \beta\left|1^A\right\rangle$, the CNOT gate transforms it into

$$\begin{aligned}\left|\psi^A\right\rangle\left|0^B\right\rangle &= \alpha\left|0^A, 0^B\right\rangle + \beta\left|1^A, 0^B\right\rangle \\ &\overset{CNOT}{\longrightarrow} \alpha\left|0^A, 0^B\right\rangle + \beta\left|1^A, 1^B\right\rangle \neq \left|\psi^A\right\rangle\left|\psi^B\right\rangle.\end{aligned} \tag{4.87}$$

Instead of being copied, the unknown state gets entangled with the auxiliary state.

Let us present a slightly more rigorous demonstration for the general case. A quantum system S^A is in the state $\left|\psi^A\right\rangle$. This state is supposed to be copied, in other words a second quantum system S^B, which is originally in the state $\left|i^B\right\rangle$, is supposed to be transformed into the state $\left|\psi^B\right\rangle$. Here, the initial state $\left|i^C\right\rangle$ of the copier system S^C, which completes the composite system, can itself be transformed into a new state $\left|f\left(\psi\right)^C\right\rangle$ in a manner which depends upon $\left|\psi^B\right\rangle$. The overall procedure is supposed to be universal, as with an office copying machine, i.e., a copy of an arbitrary state of S^A is supposed to be made using the same unitary transformation U of the composite system.

$\left|\varphi^A\right\rangle$ is a second state which is to be copied. Suppose there exists such a transformation U such that

$$\left|\psi^A\right\rangle\left|i^B\right\rangle\left|i^C\right\rangle \overset{U}{\longrightarrow} \left|\psi^A\right\rangle\left|\psi^B\right\rangle\left|f\left(\psi\right)^C\right\rangle, \tag{4.88}$$

$$\left|\varphi^A\right\rangle\left|i^B\right\rangle\left|i^C\right\rangle \overset{U}{\longrightarrow} \left|\varphi^A\right\rangle\left|\varphi^B\right\rangle\left|f\left(\varphi\right)^C\right\rangle. \tag{4.89}$$

The unitary transformation maintains the inner product

$$\left\langle\psi^A \,\middle|\, \varphi^A\right\rangle = \left\langle\psi^A \,\middle|\, \varphi^A\right\rangle\left\langle\psi^B \,\middle|\, \varphi^B\right\rangle\left\langle f\left(\psi\right)^C \,\middle|\, f\left(\varphi\right)^C\right\rangle. \tag{4.90}$$

If $\left|\psi^A\right\rangle$ and $\left|\varphi^A\right\rangle$ are not orthogonal, ($\left\langle\psi^A \,\middle|\, \varphi^A\right\rangle \neq 0$), it follows that

$$1 = \left\langle\psi^B \,\middle|\, \varphi^B\right\rangle\left\langle f\left(\psi\right)^C \,\middle|\, f\left(\varphi\right)^C\right\rangle. \tag{4.91}$$

Since all the states are normalized, thus $\left|\left\langle \psi^B \,\middle|\, \varphi^B \right\rangle\right| \leq 1$ and $\left|\left\langle f\left(\psi\right)^C \,\middle|\, f\left(\varphi\right)^C \right\rangle\right| \leq 1$, and hence

$$\left|\left\langle \psi^B \,\middle|\, \varphi^B \right\rangle\right| = 1, \text{ i.e., } \left|\psi^B\right\rangle = \pm\left|\varphi^B\right\rangle \tag{4.92}$$

is a necessary condition for fulfilling Eq. (4.91). Therefore, if there exists such an operation for some state $\left|\psi^A\right\rangle$, at least this operation is not able to copy any non-orthogonal state to $\left|\psi^A\right\rangle$, in other words the machine could, in principle, only copy another state $\left|\varphi^A\right\rangle$ which is orthogonal to $\left|\psi^A\right\rangle$. Recall our discussion of environment-induced superselection in Sec. 3.6, which says that if the environment monitors a set of states the superposition of such states is supressed by decoherence. Our result of the no-cloning theorem may be seen as an extreme situation of this phenomenon. By means of a possibly necessary system S^C, the system S^B monitors perfectly some states of S^A, i.e., S^B does not extract some information of the state of the system S^A, as in the einselection case, but instead all information about the state, as this is the very meaning of copying. This could be interpreted to provide a zero decoherence time, which in turn implies that there exists no instant of time such that any superposition of the states could ever be possible. The only states possible of copying must be orthogonal, as einselection also provides with the concept of preferred basis. Finally, this occurs because of entanglement. The states which the environment (in our case the copying machine) extracts information get entangled with the environment and therefore there is no possibility of obtaining a separable state in the final of the computation, which would occur if the copying would have been performed successfully.

4.5.2 Quantum Teleportation

As we have seen in the previous section an unknown quantum state cannot be copied because of entanglement. Nevertheless entanglement can be a resource to many interesting phenomena, one of which is quantum teleportation [53]. Quantum teleportation is the process to transfer a quantum state from one system to another regardless their spatial separation. The protocol to achieve this goal is the following. Alice and Bob share a Bell state $\left|\Phi_+^{AB}\right\rangle$. The subsystems are the quantum systems S^A and S^B, which are at Alice's and Bob's locations, respectively. Alice has an

additional quantum system S^C in a pure state

$$\left|\varphi^C\right\rangle = a\left|0^C\right\rangle + b\left|1^C\right\rangle, \tag{4.93}$$

which is unknown to Alice. This state $|\varphi\rangle$, not the quantum system S^C itself, is to be teleported to Bob. That means that we are seeking a procedure by means of which Bob's subsystem S^B can be prepared in the pure state $\left|\varphi^B\right\rangle$. S^B is then necessarily no longer entangled with any other system and, knowning the no-cloning theorem, the system S^C must have changed its state.

Considering all subsystems together we are dealing with a tripartite system. Let us carry out some algebraic manipulations on this tripartite state. As stated above we begin with the state

$$\left|\varphi^C\right\rangle\left|\Phi_+^{AB}\right\rangle = \frac{1}{\sqrt{2}}\left(a\left|0^C\right\rangle + b\left|1^C\right\rangle\right)\left(\left|0^A\right\rangle\left|0^B\right\rangle + \left|1^A\right\rangle\left|1^B\right\rangle\right), \tag{4.94}$$

which can be manipulated as

$$\begin{aligned}\left|\varphi^C\right\rangle\left|\Phi_+^{AB}\right\rangle =& \frac{1}{\sqrt{2}}\left(a\left|0^C\right\rangle\left|0^A\right\rangle\left|0^B\right\rangle + a\left|0^C\right\rangle\left|1^A\right\rangle\left|1^B\right\rangle\right.\\ &\left. + b\left|1^C\right\rangle\left|0^A\right\rangle\left|0^B\right\rangle + b\left|1^C\right\rangle\left|1^A\right\rangle\left|1^B\right\rangle\right)\\ =& \frac{1}{2}\left\{a\left(\left|\Phi_+^{CA}\right\rangle + \left|\Phi_-^{CA}\right\rangle\right)\left|0^B\right\rangle + a\left(\left|\Psi_+^{CA}\right\rangle + \left|\Psi_-^{CA}\right\rangle\right)\left|1^B\right\rangle\right.\\ &\left. + b\left(\left|\Psi_+^{CA}\right\rangle + \left|\Psi_-^{CA}\right\rangle\right)\left|0^B\right\rangle + b\left(\left|\Phi_+^{CA}\right\rangle + \left|\Phi_-^{CA}\right\rangle\right)\left|1^B\right\rangle\right\},\end{aligned} \tag{4.95}$$

where we have inserted some convenient zero states. Continuing with the manipulation,

$$\begin{aligned}\left|\varphi^C\right\rangle\left|\Phi_+^{AB}\right\rangle =& \frac{1}{2}\left\{\left|\Phi_+^{CA}\right\rangle\left(a\left|0^B\right\rangle + b\left|1^B\right\rangle\right) + \left|\Psi_+^{CA}\right\rangle\left(a\left|0^B\right\rangle + b\left|1^B\right\rangle\right)\right.\\ &\left. + \left|\Psi_-^{CA}\right\rangle\left(a\left|0^B\right\rangle + b\left|1^B\right\rangle\right) + \left|\Phi_-^{CA}\right\rangle\left(a\left|0^B\right\rangle + b\left|1^B\right\rangle\right)\right\}\\ =& \frac{1}{2}\left\{\left|\Phi_+^{CA}\right\rangle\left|\varphi^B\right\rangle + \left|\Psi_+^{CA}\right\rangle\sigma_1^B\left|\varphi^B\right\rangle\right.\\ &\left. + \left|\Psi_-^{CA}\right\rangle\left(-i\sigma_2^B\right)\left|\varphi^B\right\rangle + \left|\Phi_-^{CA}\right\rangle\sigma_3^B\left|\varphi^B\right\rangle\right\}.\end{aligned} \tag{4.96}$$

So far all we have done was only write mathematically the terms in the space $\mathcal{H}_2^C \otimes \mathcal{H}_2^A$ in terms of the Bell basis, which is accessible to Alice. This naturally led to a state where $\left|\varphi^B\right\rangle$ shows up explicitly, besides some Pauli operations, in $\mathcal{H}_2^B$. In the next step of the protocol Alice carries out a Bell measurement on the subsystems S^C and S^A. There are four possible results of this measurement. Alice informs

Bob of the result of her measurement by means of classical communication. The associated two bits of classical information are transmitted by Alice to Bob, e.g., by telephone. He, then, applies the corresponding unitary transformation, I^B, σ_1^B, $i\sigma_2^B$ or σ_3^B, to his subsystem in order to finally obtain the state $\left|\varphi^B\right\rangle$ on his subsystem, ending the protocol. In Fig. 4.4 below one can see a picture which represents the described protocol[†]. There is an interesting way to look at this protocol, which

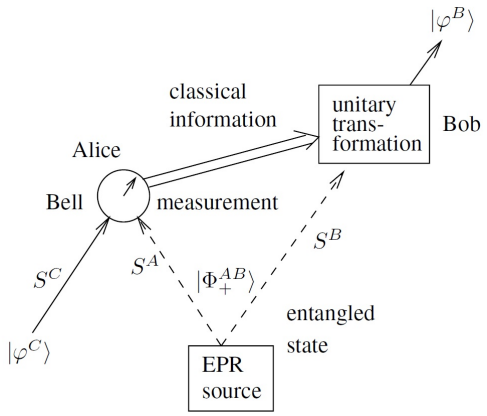

Figure 4.4: Representation of the quantum teleportation protocol. This figure was copied from Ref. [10].

is the following. At the beginning the subsystems S^A is entangled with S^B and the state which we want to teleport is in a separate subsystem S^C. This defines the preparation procedure. After the Bell measurement carried out by Alice, the final state is a composition of an entangled state of the subsystems S^A and S^C and a separate state, the one which we teleported, on the subsystem S^B. The Bell measurement can be seen to have transferred the amount of entanglement, originally in S^{AB}, to S^{AC}. In spite of that, observe that in order the protocol to work, Alice and Bob must share information classically and therefore the quantum teleportation cannot be used to transfer information faster than light. This can also be seen as a consequence of the no-cloning theorem.

At the introduction we have said that the reason why we cannot use quantum effects to send information faster than the speed of light was because of the no-cloning theorem. Let us see why. Consider the same setup of the teleportation

[†]An EPR source is simply a system that generates one of the four maximally entangled two-qubit states, the Bell states.

protocol. Alice wants to send information to Bob, say, either the bit zero or the bit one. She then measures her qubit either in the X basis or in the Z basis. Alice and Bob have conventionally agreed that a X measurement means bit zero and a Z measurement a bit one. After the measurement she uses the teleportation protocol to transfer this state to Bob's location. The task for Bob to read out the information is find a way to discover if Alice has measured in the X or Z basis. If it was possible for Bob to generate an arbitrary large number of copies of the quantum state he would be able to know which basis Alice has measured. Observe that in this scenario Alice and Bob do not need to exchange classical communication resource because Bob would have a way to know the state without Alice's help. Therefore, if this were possible they could be really space-like separated such that those copies and measurements that would be required to be performed by Bob in his laboratory would take less time than the propagation of light between them and hence, it would have been possible to send information faster than light. However, the no-cloning theorem prohibits such copy machines and therefore Alice and Bob must use classical resource in order to communicate, thus using at most a communication at the speed of light.

Several groups have verified the protocol of quantum teleportation. Nowadays, the larger distance which a qubit has been teleported was 143 Km, between the two Canary Islands of La Palma and Tenerife [54].

4.5.3 Entanglement Swapping

In general, entanglement is created through interactions which are local processes. We shall see a protocol to entangle two subsystems which have never interacted. Consider two EPR sources I and II, each produce at the same time a bipartite system S^{AB} and S^{CD} in the Bell states $\left|\Psi_{-}^{AB}\right\rangle$ and $\left|\Psi_{-}^{CD}\right\rangle$, respectively. The initial state of the protocol is, thus,

$$\left|\Psi_{-}^{AB}\right\rangle \left|\Psi_{-}^{CD}\right\rangle = \frac{1}{2}\left(\left|0^{A},1^{B}\right\rangle - \left|1^{A},0^{B}\right\rangle\right)\left(\left|0^{C},1^{D}\right\rangle - \left|1^{C},0^{D}\right\rangle\right). \tag{4.97}$$

Using the Bell bases in the spaces $\mathcal{H}^{A}\otimes\mathcal{H}^{D}$ and $\mathcal{H}^{B}\otimes\mathcal{H}^{C}$ the above state can be written as

$$\begin{aligned}\left|\Psi_{-}^{AB}\right\rangle \left|\Psi_{-}^{CD}\right\rangle =& \frac{1}{2}\left(\left|\Psi_{+}^{AD}\right\rangle \left|\Psi_{+}^{BC}\right\rangle - \left|\Psi_{-}^{AD}\right\rangle \left|\Psi_{-}^{BC}\right\rangle\right.\\ &\left. - \left|\Phi_{+}^{AD}\right\rangle \left|\Phi_{+}^{BC}\right\rangle + \left|\Phi_{-}^{AD}\right\rangle \left|\Phi_{-}^{BC}\right\rangle\right).\end{aligned} \tag{4.98}$$

Following a similar procedure of that we used in the quantum teleportation protocol would be very tricky to arrive at this expression. The best way to see that this is true is projecting the original state onto the four Bell states of the subsystem S^{AD} and take note of the resulting state, which is necessarily an state of S^{BC}, and amplitude. One can immediately see that a Bell measurement on the subsystems S^B and S^C transforms the previously non-entangled subsystems S^A and S^D into a Bell state. The Bell measurement trasfer the same amount of entanglement from S^{AB} and S^{CD} to S^{AD} and S^{BC}. The Fig. 4.5 represents this protocol.

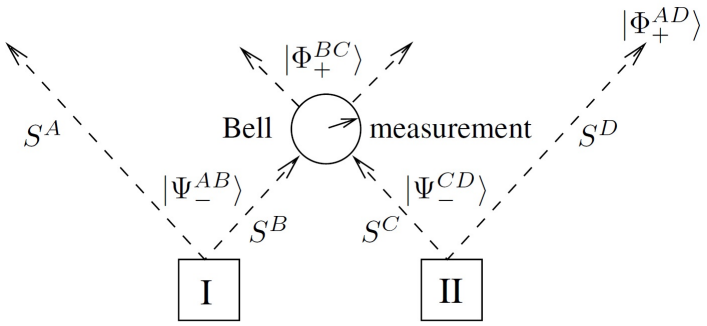

Figure 4.5: Representation of the entanglement swapping protocol. This figure was copied from Ref. [10].

A recent experimental realization of the entanglement swapping between photons can be found in Ref. [55].

4.5.4 Quantum Dense Coding

Every natural decimal number may be mapped into a string of bits. The shortest string of bits to represent a natural number is the binary representation of that number. No matter how large the natural number is, we always find that the string of bits is finite. A rational number is the ratio of two integers, therefore it is also necessary a finite string of bits to represent a rational number[‡]. But an arbitrary real number (one that is not rational) would need an infinitely long string of bits to store it, with maximum precision. A qubit needs two real numbers to define it precisely, therefore we would never be able to store an arbitrary state of a qubit in a classical computer with maximum precision. But this is not a problem as in practice all

[‡]This is so because the set of rational numbers is countable, i.e., it is isomorphic to the set of natural numbers.

experiments are bounded to the experimental precision of the equipments. Looking the other way around, in a quantum computer where the elementary units for storing information are the qubits itself, it is tempting to think that we could store an infinite amount of information as we could code a real number in it, e.g. in its relative phase, which would be analogous to store an infinitely long string of bits. Nevertheless, this is not true because we have to read out the information stored in a qubit carrying out a measurement and whichever basis we choose to measure, say, the spin of a spin-1/2 particle, there are always two orthogonal states. Therefore a qubit may transmit also 1 bit of information.

If we exploit the entanglement we can do better, send 2 bits of information transmitting just one qubit, though. This protocol is called quantum dense coding or superdense coding [56].

Suppose Alice and Bob share an entangled state, say, the Bell state $\left|\Phi_+^{AB}\right\rangle$. Alice then performs one out of the following four possible local operations on her qubit, $I^A \otimes I^B$, $\sigma_1^A \otimes I^B$, $\sigma_2^A \otimes I^B$ or $\sigma_3^A \otimes I^B$. The effect on the composite state for each of these local measurements is

$$I^A \otimes I^B \left|\Phi_+^{AB}\right\rangle = \left|\Phi_+^{AB}\right\rangle, \tag{4.99}$$

$$\sigma_1^A \otimes I^B \left|\Phi_+^{AB}\right\rangle = \left|\Psi_+^{AB}\right\rangle, \tag{4.100}$$

$$\sigma_2^A \otimes I^B \left|\Phi_+^{AB}\right\rangle = -i\left|\Psi_-^{AB}\right\rangle, \tag{4.101}$$

$$\sigma_3^A \otimes I^B \left|\Phi_+^{AB}\right\rangle = \left|\Phi_-^{AB}\right\rangle. \tag{4.102}$$

After performing the local measurement Alice sends her qubit to Bob and Bob, who has now both qubits, performs a Bell measurement on both qubits and therefore there are four possible outcomes. Alice has, thus, sent two bits of information to Bob (base two logarithm of the four possibilities).

The transmission of just one qubit, and nothing more, could never transmit two bits, therefore this is only possible to be acomplished by means of an entanglement-assisted transmission, which means that they both shared previously an entangled state.

Once more, entanglement plays a crucial role in exploring new ways of transmitting information.

4.5.5 Quantum Discord as a Resource

The aim of this section is to provide one example of a quantum protocol whose improvement over classical ones is not due to entanglement. The task is to calculate the normalized trace of an arbitrary $n \times n$ operator U. The algorithm is called deterministic quantum computation with one qubit (DQC1) [57].

Consider $n+1$ qubits where n of them are in the maximally mixed state

$$\frac{1}{2^n} I^{\otimes^n}, \tag{4.103}$$

and one qubit is in the state

$$\rho = \frac{(1-\alpha)}{2} I + \alpha \left|0\right\rangle \left\langle 0\right|. \tag{4.104}$$

The initial total state is, therefore,

$$\Omega_i = \rho \otimes \frac{1}{2^n} I^{\otimes^n}, \tag{4.105}$$

where the first qubit is the control qubit, or C-qubit, and the others are the target qubits, or T-qubits. We will let this state evolve following the circuit given by Fig. 4.6. After the Hadamard gate we have the state

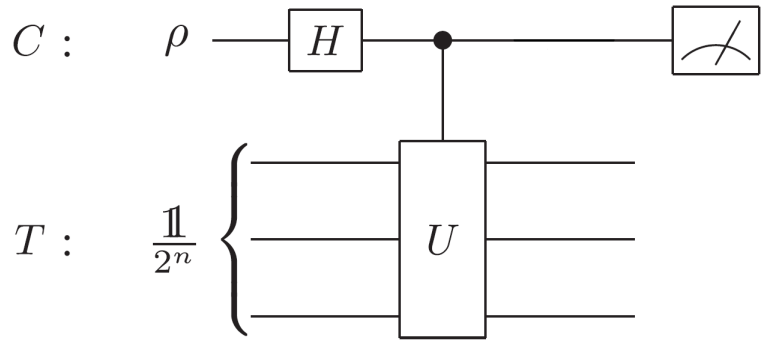

Figure 4.6: Quantum circuit of the QDC1 algorithm. Figure copied (and corrected) from Ref. [57].

$$\Omega_1 = \left[\frac{1-\alpha}{2} I + \alpha \left|+\right\rangle \left\langle +\right|\right] \otimes \frac{I^{\otimes^n}}{2^n}. \tag{4.106}$$

Opening this state explicitly gives us

$$\begin{aligned}\Omega_1 =& \left[\frac{1-\alpha}{2}\left(|0\rangle\langle 0| + |1\rangle\langle 1|\right) + \frac{\alpha}{2}\left(|0\rangle\langle 0| + |0\rangle\langle 1| + |1\rangle\langle 0| + |1\rangle\langle 1|\right)\right] \\ & \otimes \frac{I^{\otimes^n}}{2^n} \\ =& \frac{1}{2}\left[|0\rangle\langle 0| \otimes \frac{I^{\otimes^n}}{2^n} + |1\rangle\langle 1| \otimes \frac{I^{\otimes^n}}{2^n} + \alpha\,|0\rangle\langle 1| \otimes \frac{I^{\otimes^n}}{2^n} + \alpha\,|1\rangle\langle 0| \otimes \frac{I^{\otimes^n}}{2^n}\right]. \quad (4.107)\end{aligned}$$

Applying the controlled-U gate we obtain the final state

$$\Omega_f = \frac{1}{2}\left[I \otimes \frac{I^{\otimes^n}}{2^n} + \alpha\,|0\rangle\langle 1| \otimes \frac{U^\dagger}{2^n} + \alpha\,|1\rangle\langle 0| \otimes \frac{U}{2^n}\right]. \quad (4.108)$$

With this state we can evaluate the reduced density operator for the C-qubit, which leads to

$$\rho_f = \frac{1}{2}\begin{pmatrix} 1 & \frac{\alpha \mathrm{Tr}(U^\dagger)}{2^n} \\ \frac{\alpha \mathrm{Tr}(U)}{2^n} & 1 \end{pmatrix}. \quad (4.109)$$

Now evaluating the expectation value of the observables σ_1 and σ_2,

$$\mathrm{Tr}\left(\sigma_1\rho_f\right) = \frac{\alpha}{2^n}\mathrm{Re}\left[\mathrm{Tr}\left(U\right)\right]\ ,\ \mathrm{Tr}\left(\sigma_2\rho_f\right) = \frac{\alpha}{2^n}\mathrm{Im}\left[\mathrm{Tr}\left(U\right)\right], \quad (4.110)$$

gives us the estimation of the normalized trace of the unitary operator U.

It can be shown that the algorithm is efficient (scales polynomially with n) and that the best-known classical algorithm for estimating the trace of a unitary matrix requires resources that are an exponential function of n (inefficient) [57].

A study of the correlations in this computation (which we will not present but we refer to the Ref. [57] and references therein for a complete discussion) shows that there is no entanglement between C and T (this means to calculate the bipartite entanglement considering the cut $\mathcal{H}^{control} \otimes \mathcal{H}^{targets}$). An analysis of the quantum correlations as given by quantum discord shows that the discord scales similar to the efficiency of the algorithm, showing quantitatively, for the first time, that quantum correlations other than entanglement also play an important role in the speedup associated with quantum algorithms. This is very interesting since preservation of coherence of the quantum states is difficult and requires a controllable environment whereas quantum mixtures are indeed the common states at room temperature.

Though superficially presented, this algorithm posed new hopes for exploiting quantum computation at large scale [58]. This kind of computation has been referred to in the literature as mixed-state quantum computation and one reason for us to study the quantum discord in this dissertation is because of such results.

Chapter 5

The Effect of Random Coupling Constants on Quantum Correlations

We have seen that exploiting quantum correlations we can build quantum algorithms that either surpass their classical counterparts or provide ways to transmit information otherwise unimagined to be possible, e.g. through quantum teleportation. The algorithms we have discussed are ideal in the sense that they don't take into account the effect of the environment, which usually is present in nature. So, in real-world applications of quantum information it is important to take into account how the environment affects the behavior of quantum correlations.

The aim of this chapter is to study the effect on the dynamics of quantum correlations on a toy model, as measured by concurrence, quantum mutual information and quantum discord, due to a random interaction with the environment. Randomness of the coupling constants models a dynamical environment as seen by the system. The present study was inspired by the work of Castagnino *et al.* [59], where they discuss the effect of a qubit environment on decoherence of a central qubit. In this model we consider two noninteracting qubits interacting individually with a common qubit environment.

The Hamiltonian of the composite system is

$$H^{SE} = H^I = \left(\sigma_3^A + \sigma_3^B\right) \otimes \sum_{i=1}^{N} \hbar g_i \sigma_3^{(i)}. \tag{5.1}$$

Due to the fact that we are interested in a qualitative study of the effect of the interaction, exclusively, we do not consider self-energy Hamiltonian terms for the qubits.

Consider the most general initial state which has no initial system-environment

correlation and no internal environment correlations

$$\begin{aligned}|\psi(0)\rangle &= \left|\psi^S(0)\right\rangle \otimes |E(0)\rangle \\ &= \left(\sum_{i,j=0,1} \alpha_{ij} \left|i^A\right\rangle \otimes \left|j^B\right\rangle\right) \otimes \left(\bigotimes_{i=1}^{N} \left(\beta_i \left|0^i\right\rangle + \gamma_i \left|1^i\right\rangle\right)\right). \end{aligned} \tag{5.2}$$

The coefficients must satisfy $\sum_{i,j} |\alpha_{ij}|^2 = 1$ and $|\beta_i|^2 + |\gamma_i|^2 = 1$ for each i. The time evolution operator is given by

$$\begin{aligned}U(t) &= \exp\left\{-\frac{i}{\hbar} t H^I\right\} = \exp\left\{-\frac{i}{\hbar} t \left(H^A + H^B\right)\right\} \\ &= \exp\left\{-\frac{i}{\hbar} t H^A\right\} \exp\left\{-\frac{i}{\hbar} t H^B\right\}, \end{aligned} \tag{5.3}$$

where $H^{A,B} = \sigma_3^{A,B} \otimes \sum_i \hbar g_i \sigma_3^{(i)}$ and we used the fact that $\left[H^A, H^B\right] = 0$. Therefore, the composite state vector at a time t is

$$|\psi(t)\rangle = U(t)|\psi(0)\rangle = \left(\sum_{i,j=0,1} \alpha_{ij} e^{-\frac{i}{\hbar} t H^A} |i\rangle \otimes e^{-\frac{i}{\hbar} t H^B} |j\rangle\right) \otimes |E(0)\rangle. \tag{5.4}$$

This can be further evaluated to

$$\begin{aligned}|\psi(t)\rangle =& \alpha_{00} e^{-\frac{i}{\hbar} t H^A} \left|0^A\right\rangle e^{-\frac{i}{\hbar} t H^B} \left|0^B\right\rangle |E(0)\rangle + \alpha_{01} e^{-\frac{i}{\hbar} t H^A} \left|0^A\right\rangle e^{-\frac{i}{\hbar} t H^B} \left|1^B\right\rangle |E(0)\rangle \\ &+ \alpha_{10} e^{-\frac{i}{\hbar} t H^A} \left|1^A\right\rangle e^{-\frac{i}{\hbar} t H^B} \left|0^B\right\rangle |E(0)\rangle + \alpha_{11} e^{-\frac{i}{\hbar} t H^A} \left|1^A\right\rangle e^{-\frac{i}{\hbar} t H^B} \left|1^A\right\rangle |E(0)\rangle \\ =& \alpha_{00} e^{-itE} \left|0^A\right\rangle e^{-itE} \left|0^B\right\rangle |E(0)\rangle + \alpha_{01} e^{-itE} \left|0^A\right\rangle e^{+itE} \left|1^B\right\rangle |E(0)\rangle \\ &+ \alpha_{10} e^{+itE} \left|1^A\right\rangle e^{-itE} \left|0^B\right\rangle |E(0)\rangle + \alpha_{11} e^{+itE} \left|1^A\right\rangle e^{+itE} \left|1^B\right\rangle |E(0)\rangle, \end{aligned} \tag{5.5}$$

where we have defined $E = \sum_k g_k \sigma_3^{(k)}$.

Defining the environment states $|E_\pm(t)\rangle = \exp\{\mp 2itE\} |E(0)\rangle$ the final state vector is given by

$$\begin{aligned}|\psi(t)\rangle =& \alpha_{00} |00\rangle |E_+(t)\rangle + \alpha_{01} |01\rangle |E(0)\rangle \\ &+ \alpha_{10} |10\rangle |E(0)\rangle + \alpha_{11} |11\rangle |E_-(t)\rangle. \end{aligned} \tag{5.6}$$

The density operator for this state can be evaluated to be

$$\rho(t) = \begin{pmatrix} |\alpha_{00}|^2 P_{++}(t) & \alpha_{00}^* \alpha_{01} P_{+0}(t) & \alpha_{00}^* \alpha_{10} P_{+0}(t) & \alpha_{00}^* \alpha_{11} P_{+-}(t) \\ \alpha_{01}^* \alpha_{00} P_{0+}(t) & |\alpha_{01}|^2 P_{00} & \alpha_{01}^* \alpha_{10} P_{00} & \alpha_{01}^* \alpha_{11} P_{0-}(t) \\ \alpha_{10}^* \alpha_{00} P_{0+}(t) & \alpha_{10}^* \alpha_{01} P_{00} & |\alpha_{10}|^2 P_{00} & \alpha_{10}^* \alpha_{11} P_{0-}(t) \\ \alpha_{11}^* \alpha_{00} P_{-+}(t) & \alpha_{11}^* \alpha_{01} P_{-0}(t) & \alpha_{11}^* \alpha_{10} P_{-0}(t) & |\alpha_{11}|^2 P_{--}(t) \end{pmatrix}, \tag{5.7}$$

where $P_{ij} = |E_i(t)\rangle\langle E_j(t)|$, with $i,j \in \{+,-,0\}$, and $|E_0(t)\rangle = |E(0)\rangle$.

The reduced density operator for the system is obtained from

$$\rho^S(t) = \mathrm{Tr}_E(\rho(t)) \tag{5.8}$$

and gives us

$$\rho^S(t) = \begin{pmatrix} |\alpha_{00}|^2 & \alpha_{00}^*\alpha_{01}r_1(t) & \alpha_{00}^*\alpha_{10}r_1(t) & \alpha_{00}^*\alpha_{11}r_2(t) \\ \alpha_{01}^*\alpha_{00}r_1^*(t) & |\alpha_{01}|^2 & \alpha_{01}^*\alpha_{10} & \alpha_{01}^*\alpha_{11}r_3(t) \\ \alpha_{10}^*\alpha_{00}r_1^*(t) & \alpha_{10}^*\alpha_{01} & |\alpha_{10}|^2 & \alpha_{10}^*\alpha_{11}r_3(t) \\ \alpha_{11}^*\alpha_{00}r_2^*(t) & \alpha_{11}^*\alpha_{01}r_3^*(t) & \alpha_{11}^*\alpha_{10}r_3^*(t) & |\alpha_{11}|^2 \end{pmatrix}, \tag{5.9}$$

where we have defined $r_1(t) = \langle E(0)\,|\,E_+(t)\rangle$, $r_2(t) = \langle E_-(t)\,|\,E_+(t)\rangle$ and $r_3(t) = \langle E_-(t)\,|\,E(0)\rangle$.

Suppose we have the initial state for the composite system

$$|\psi(0)\rangle = (\cos(\alpha)|00\rangle + \sin(\alpha)|11\rangle) \otimes \left|\psi_0^E\right\rangle. \tag{5.10}$$

As we will see, the initial amount of entanglement, as measured by concurrence, will depend on this parameter α. If α is either 0 or $\pi/2$ the state is not entangled and if α is $\pi/4$ it is maximally entangled: it is a Bell state. For this initial state the reduced density operator of the system is given by

$$\rho^S(t) = \begin{pmatrix} \cos^2(\alpha) & 0 & 0 & \sin(\alpha)\cos(\alpha)\,r_2(t) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \sin(\alpha)\cos(\alpha)\,r_2^*(t) & 0 & 0 & \sin^2(\alpha) \end{pmatrix}. \tag{5.11}$$

Let us now calculate the function $r_2(t)$.

This function is given by

$$\begin{aligned} r_2(t) &= \langle E_-(t)\,|\,E_+(t)\rangle \\ &= \left(\bigotimes_{k=1}^{N}\left(\beta_k^* e^{-2ig_k t}\left\langle 0^k\right| + \gamma_k e^{+2ig_k t}\left\langle 1^k\right|\right)\right) \times \\ &\quad \left(\bigotimes_{l=1}^{N}\left(\beta_l e^{-2ig_l t}\left|0^l\right\rangle + \gamma_l e^{+2ig_l t}\left|1^l\right\rangle\right)\right). \end{aligned} \tag{5.12}$$

As the states are individually orthogonal the result is

$$r_2(t) = \prod_{k=1}^{N}\left[|\beta_k|^2 e^{-4ig_k t} + |\gamma_k|^2 e^{+4ig_k t}\right]. \tag{5.13}$$

As we will need later on, the norm-squared of this function is given by

$$|r_2(t)|^2 = \prod_{k=1}^{N} \left[|\beta_k|^4 + |\gamma_k|^4 + 2|\beta_k|^2 |\gamma_k|^2 \cos(8g_k t) \right]. \tag{5.14}$$

In the next two sections we will evaluate the measures of quantum correlations and present our numerical results.

5.1 Calculation of the Measures of Quantum Correlations

In this section we will calculate the concurrence, quantum mutual information and quantum discord for the state (5.11).

As we have seen in Sec. 4.4.2, in order to evaluate the concurrence we need to calculate the eigenvalues of the matrix $R = \rho^S \overline{\rho}^S$, with $\overline{\rho}^S$ given by Eq. (4.70). Thus

$$\begin{aligned} \overline{\rho}^S &= \left(\sigma_2^A \otimes \sigma_2^B\right) \rho^{*S} \left(\sigma_2^A \otimes \sigma_2^B\right) \\ &= \begin{pmatrix} \sin^2(\alpha) & 0 & 0 & \sin(\alpha)\cos(\alpha) r_2(t) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \sin(\alpha)\cos(\alpha) r_2^*(t) & 0 & 0 & \cos^2(\alpha) \end{pmatrix}. \end{aligned} \tag{5.15}$$

Hence R is given by

$$R = \begin{pmatrix} \sin^2(\alpha)\cos^2(\alpha)\left(1+|r_2(t)|^2\right) & 0 & 0 & \sin(2\alpha)\cos^2(\alpha) r_2(t) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \sin(2\alpha)\sin^2(\alpha) r_2^*(t) & 0 & 0 & \sin^2(\alpha)\cos^2(\alpha)\left(1+|r_2(t)|^2\right) \end{pmatrix}. \tag{5.16}$$

As this matrix can be put into a block diagonal form, and moreover one block is a null two-by-two matrix, two of its eigenvalues are zero. Call them $\lambda_3 = \lambda_4 = 0$. We are left with the other two eigenvalues which are eigenvalues of the matrix

$$\tilde{R} = \begin{pmatrix} \sin^2(\alpha)\cos^2(\alpha)(1+|r_2(t)|^2) & \sin(2\alpha)\cos^2(\alpha) r_2(t) \\ \sin(2\alpha)\sin^2(\alpha) r_2^*(t) & \sin^2(\alpha)\cos^2(\alpha)(1+|r_2(t)|^2) \end{pmatrix}. \tag{5.17}$$

The eigenvalues of this matrix are given by

$$\lambda_{1,2} = \frac{\mathrm{Tr}\left[\tilde{R}\right] \pm \sqrt{\left(\mathrm{Tr}\left[\tilde{R}\right]\right)^2 - 4\mathrm{Det}\left[\tilde{R}\right]}}{2}. \tag{5.18}$$

The trace is

$$\mathrm{Tr}\left[\tilde{R}\right] = \frac{\sin^2(2\alpha)}{2}\left(1+|r_2(t)|^2\right), \tag{5.19}$$

and the determinant is

$$\mathrm{Det}\left[\tilde{R}\right]=\frac{\sin^4(2\alpha)}{16}\left(1+|r_2(t)|^2\right)^2-\frac{\sin^4(2\alpha)}{4}|r_2(t)|^2. \tag{5.20}$$

The discriminant is

$$\begin{aligned}\left(\mathrm{Tr}\left[\tilde{R}\right]\right)^2-4\mathrm{Det}\left[\tilde{R}\right]=&\frac{\sin^4(2\alpha)}{4}\left(1+|r_2(t)|^2\right)^2-\frac{\sin^4(2\alpha)}{4}\left(1+|r_2(t)|^2\right)^2\\&+\sin^4(2\alpha)|r_2(t)|^2\\=&\sin^4(2\alpha)|r_2(t)|^2.\end{aligned} \tag{5.21}$$

The eigenvalues are, therefore,

$$\lambda_{1,2}=\frac{\mathrm{Tr}\left[\tilde{R}\right]}{2}\pm\frac{\sin^2(2\alpha)|r_2(t)|}{2}. \tag{5.22}$$

With these eigenvalues in hand we can calculate the concurrence as (cf. Eq. (4.69))

$$C\left(\rho^S(t)\right)=\max\{0,\lambda_1-\lambda_2-\lambda_3-\lambda_4\}=\max\left\{0,\sin^2(2\alpha)|r_2(t)|\right\}. \tag{5.23}$$

Now let us evaluate the quantum mutual information for our state. Recall that the quantum mutual information is given by

$$I\left(\rho^{AB}\right)=S\left(\rho^A\right)+S\left(\rho^B\right)-S\left(\rho^{AB}\right). \tag{5.24}$$

Thus we have to evaluate the reduced density operator of the qubits. From Eq. (5.11) the reduced density operator of the qubit A is

$$\rho^A=\mathrm{Tr}_B\left[\rho^S\right]=\begin{pmatrix}\cos^2(\alpha)&0\\0&\sin^2(\alpha)\end{pmatrix}. \tag{5.25}$$

It is already diagonal and therefore the von Neumann entropy for it is simply

$$S\left(\rho^A\right)=-\cos^2(\alpha)\log\left[\cos^2(\alpha)\right]-\sin^2(\alpha)\log\left[\sin^2(\alpha)\right]. \tag{5.26}$$

The von Neumann entropy for the qubit B is the same, by the Schmidt decomposition. In order to obtain the von Neumann entropy for the composite state we must evaluate its eigenvalues. Again, as it is a block diagonal matrix it is enough to evaluate the eigenvalues of the matrix

$$A=\begin{pmatrix}\cos^2(\alpha)&\sin(\alpha)\cos(\alpha)r_2(t)\\\sin(\alpha)\cos(\alpha)r_2^*(t)&\sin^2(\alpha)\end{pmatrix}. \tag{5.27}$$

They are easily found to be

$$\xi_{1,2} = \frac{1 \pm \sqrt{1 - \sin^2(2\alpha)\left(1 - |r_2(t)|^2\right)}}{2}. \tag{5.28}$$

Hence it follows that the joint entropy is

$$S\left(\rho^{AB}\right) = -\xi_1 \log \xi_1 - \xi_2 \log \xi_2. \tag{5.29}$$

With Eqs (5.26) and (5.29) we can easily evaluate the quantum mutual information.

The next measure of quantum correlation we are going to use is the quantum discord. Calculate the quantum discord analytically is not an easy task because it involves an optimization over measurements. In 2008 Luo [60] calculated the quantum discord analytically for a class of density operators. He began with the parametrization of two qubit states given by

$$\rho^{AB} = \frac{1}{4}\left(I^{AB} + \vec{u}\vec{\sigma}^A \otimes I^B + \vec{v} I^A \otimes \vec{\sigma}^B + \sum_{i,j} w_{ij}\sigma_i^A \otimes \sigma_j^B\right), \tag{5.30}$$

where all coefficients are real. Using the singular value decomposition (cf. App. A) he reduced this to the following state

$$\rho^{AB} = \frac{1}{4}\left(I^{AB} + \vec{a}\vec{\sigma}^A \otimes I^B + \vec{b} I^A \otimes \vec{\sigma}^B + \sum_k c_k \sigma_k^A \otimes \sigma_k^B\right). \tag{5.31}$$

As he was interested in the correlations he considered only those states whose marginals (reduced states) are maximally mixed. This is accomplished considering $\vec{a} = \vec{b} = \vec{0}$, hence

$$\rho^{AB} = \frac{1}{4}\left(I^{AB} + \sum_k c_k \sigma_k^A \otimes \sigma_k^B\right). \tag{5.32}$$

He then developed a procedure to evaluate the quantum discord. He parametrizes the von Neumann measurements and performs explicit calculations and finds an analytic result. This class of states considered by Luo is a subset of the so-called X states. An X state has the form

$$\rho_X^{AB} = \begin{pmatrix} \rho_{11} & 0 & 0 & \rho_{14} \\ 0 & \rho_{22} & \rho_{23} & 0 \\ 0 & \rho_{32} & \rho_{33} & 0 \\ \rho_{41} & 0 & 0 & \rho_{44} \end{pmatrix}. \tag{5.33}$$

Others also have calculated the quantum discord for other classes of states which are also a subset of X states. In 2010 Ali *et al.* [61] proposed a procedure to calculate the quantum discord for an arbitrary X state. They used the same parametrization of von Neumann measurements Luo had used before. Basically they have four parameters and one constraint. Shortly after that, in 2011, Lu *et al.* [62] showed that they had forgotten one constraint and therefore had found the incorrect optimal measurement to minimize the conditional entropy. In 2013, Huang [63] numerically demonstrated that the analytical result of Ali, et al. is valid with worst-case error 0.0021, in this way not invalidating the works built upon Ali *et al.* procedure. Back in 2011, Chen *et al.* [64] proposed an analytic expression to calculate the quantum discord when the matrix elements of the X state satisfy some conditions. As the state we are going to consider fits in one of the two conditions they have shown to be possible an exact calculation we will state their results (for detail we refer to the original work [64] and references therein).

If the two following properties are satisfied,

$$|\rho_{23} + \rho_{14}| \geq |\rho_{23} - \rho_{14}|\,, \tag{5.34}$$

$$(|\rho_{23}| + |\rho_{14}|)^2 \leq (\rho_{11} - \rho_{22})\,(\rho_{44} - \rho_{33})\,, \tag{5.35}$$

then the optimal set of measurement operators for the quantum discord $Q\left(\rho_X^{AB}\right)$ is $\left\{\left|0^B\right\rangle\left\langle 0^B\right|, \left|1^B\right\rangle\left\langle 1^B\right|\right\}$*.

For our state (cf. Eq. (5.11)) the first relation is

$$\sin(\alpha)\cos(\alpha)\,|r_2(t)| \geq \sin(\alpha)\cos(\alpha)\,|-r_2(t)|\,, \tag{5.36}$$

which is clearly satisfied. The second is

$$\sin^2(\alpha)\cos^2(\alpha)\,|r_2(t)|^2 \leq \left(\cos^2(\alpha)\right)\left(\sin^2(\alpha)\right), \tag{5.37}$$

which in turn implies

$$|r_2(t)|^2 \leq 1. \tag{5.38}$$

Recall the expression for this function from Eq. (5.14),

$$|r_2(t)|^2 = \prod_{k=1}^{N}\left[a_k^2 + (1-a_k)^2 + 2a_k(1-a_k)\cos(8g_k t)\right], \tag{5.39}$$

*I implicitly assumed that we are measuring the subsystem S^B for the optimization.

where we have substituted $|\beta_k|^2 = a_k$ and used the fact that the coefficients are normalized $|\beta_k|^2 + |\gamma_k|^2 = 1$. The $\cos(x)$ function is bounded on $[-1, 1]$. Therefore for the kth term of the product

$$(a_k - (1-a_k))^2 \leq a_k^2 + (1-a_k)^2 + 2a_k(1-a_k)\cos(8g_k t) \leq (a_k + (1-a_k))^2, \quad (5.40)$$

or

$$0 \leq (2a_k - 1)^2 \leq a_k^2 + (1-a_k)^2 + 2a_k(1-a_k)\cos(8g_k t) \leq 1. \quad (5.41)$$

Hence each term is bounded and consequently

$$0 \leq |r_2(t)|^2 \leq 1, \quad (5.42)$$

satisfying the condition. The quantum discord can be analitically evaluated, therefore, by the set measurement operators $\left\{M_0^B = \left|0^B\right\rangle\left\langle 0^B\right|, M_1^B = \left|1^B\right\rangle\left\langle 1^B\right|\right\}$. We have to calculate the classical correlations for this set of measurement operators. Recall that the classical correlations related with the discord are measured by

$$J\left(\rho^{AB}\right) = S\left(\rho^A\right) - \sum_k p_k S\left(\rho^A_{M_k^B}\right), \quad (5.43)$$

where we are already considering the optimal measurement operators. The probabilities are $p_k = \mathrm{Tr}_{AB}\left[\hat{M}_k^B \rho^{AB} \hat{M}_k^{B\dagger}\right]$. Let us calculate the states $\rho^A_{M_k^B}$. For $k = 0$ we have

$$\begin{aligned}\rho^{AB}_{M_0^B} =& \frac{1}{p_0} I^A \otimes M_0^B \rho^S(t) I^A \otimes M_0^{B\dagger} \\ =& \frac{1}{p_0}\left(|00\rangle\langle 00| + |10\rangle\langle 10|\right)\rho^S(t)\left(|00\rangle\langle 00| + |10\rangle\langle 10|\right). \end{aligned} \quad (5.44)$$

As our state is given by

$$\begin{aligned}\rho^S(t) =& \cos^2(\alpha)|00\rangle\langle 00| + \sin(\alpha)\cos(\alpha) r_2(t)|00\rangle\langle 11| \\ &+ \sin(\alpha)\cos(\alpha) r_2^*(t)|11\rangle\langle 00| + \sin^2(\alpha)|11\rangle\langle 11| \end{aligned} \quad (5.45)$$

we obtain

$$\rho^{AB}_{M_0^B} = \frac{1}{p_0}\cos^2(\alpha)|00\rangle\langle 00|. \quad (5.46)$$

The probability p_0 is given by

$$\begin{aligned} p_0 =& \mathrm{Tr}\left[\rho^S(t) I^A \otimes M_0^B\right] = \mathrm{Tr}\left[\rho^S(t)\left(|00\rangle\langle 00| + |10\rangle\langle 10|\right)\right] \\ =& \langle 00|\rho^S(t)|00\rangle + \langle 10|\rho^S(t)|10\rangle = \cos^2(t). \end{aligned} \quad (5.47)$$

Therefore

$$\rho^{A}_{M_0^B} = \mathrm{Tr}_B\left[\rho^{AB}_{M_0^B}\right] = \mathrm{Tr}_B\left[|00\rangle\langle 00|\right] = |0\rangle\langle 0|\,. \tag{5.48}$$

This is a pure state and has zero entropy. The second state, $k = 1$, is given by

$$\begin{aligned}\rho^{AB}_{M_1^B} =& \frac{1}{p_1} I^A \otimes M_1^B \rho^S(t)\, I^A \otimes M_1^{B\dagger} \\ =& \frac{1}{p_1}\left(|01\rangle\langle 01| + |11\rangle\langle 11|\right)\rho^S(t)\left(|01\rangle\langle 01| + |11\rangle\langle 11|\right),\end{aligned} \tag{5.49}$$

which gives us

$$\rho^{AB}_{M_1^B} = \frac{1}{p_1}\sin^2(\alpha)\,|11\rangle\langle 11|\,. \tag{5.50}$$

The probability p_1 is

$$\begin{aligned}p_1 =& \mathrm{Tr}\left[\rho^S(t)\, I^A \otimes M_1^B\right] = \mathrm{Tr}\left[\rho^S(t)\left(|01\rangle\langle 01| + |11\rangle\langle 11|\right)\right] \\ =& \langle 01|\,\rho^S(t)\,|01\rangle + \langle 11|\,\rho^S(t)\,|11\rangle = \sin^2(\alpha)\,,\end{aligned} \tag{5.51}$$

and therefore

$$\rho^{A}_{M_1^B} = \mathrm{Tr}_B\left[\rho^{AB}_{M_1^B}\right] = \mathrm{Tr}_B\left[|11\rangle\langle 11|\right] = |1\rangle\langle 1|\,. \tag{5.52}$$

This is also a pure state and hence the second term in the classical correlations is zero.

The first term is given by Eq. (5.26) and therefore

$$J\left(\rho^{AB}\right) = -\cos^2(\alpha)\log\left[\cos^2(\alpha)\right] - \sin^2(\alpha)\log\left[\sin^2(\alpha)\right]. \tag{5.53}$$

Observe that for our particular situation the classical correlations are constant through time. The quantum discord is, thus,

$$Q\left(\rho^{AB}\right) = -\cos^2(\alpha)\log\left[\cos^2(\alpha)\right] - \sin^2(\alpha)\log\left[\sin^2(\alpha)\right] + \xi_1\log\xi_1 + \xi_2\log\xi_2, \tag{5.54}$$

with $\xi_{1,2}$ given by Eq. (5.28).

5.2 Numerical Results and Discussions

In this section we will analyze the dynamics of the measures calculated in the previous section under three different environments. Before that we make some remarks.

The parameters of our model are: α, which controls the initial system qubit state, the β_i's in the initial environmental states, the coupling constants g_i's and the number of environmental qubits N.

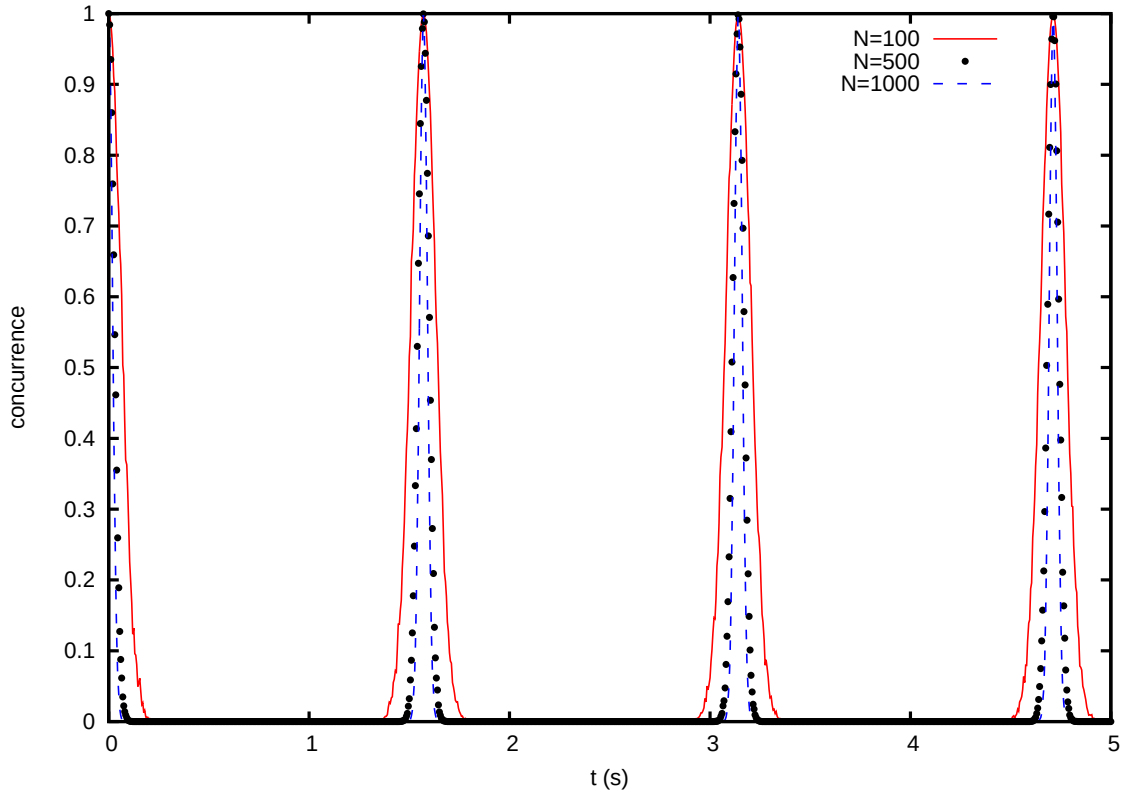

Figure 5.1: Effect of the size of a homogeneous environment on the concurrence. Here, $g = 0.5\ s^{-1}$ and $\alpha = \pi/4$.

As can be seen for the expressions determining the concurrence, mutual information and quantum discord, these quantities depend on the function $|r_2(t)|$. On the other hand this function depends on the initial states of the environmental qubits. As we want to do a more qualitative study we do not care for a particular initial state of the environmental qubits. Therefore, we select the β_k's coefficients from a probability distribution. In fact, as $|\beta_k|^2$ actually appears in the function $|r_2(t)|$ we select each of which from a uniform distribution in the unit interval $[0, 1]$. This is done for all results shown in the following. Another remark is that for our interaction and initial state the measure of classical correlations which appears in the quantum discord is time-independent, being just a function of α. Thus, the features of quantum mutual information and quantum discord become qualitatively the same. For this reason the two relevant quantities to be analyzed are the concurrence and the quantum discord. Before we consider the random coupling constants let us see what are the effects of a homogeneous environment on the quantum correlations.

By a homogeneous environment we mean simply that all coupling constants have the same value. The time-dependence of the measures of correlations is given by the behavior of the function

$$|r_2(t)|^2 = \prod_{k=1}^{N} \left[|\beta_k|^4 + |\gamma_k|^4 + 2|\beta_k|^2 |\gamma_k|^2 \cos(8g_k t)\right] = \prod_{k=1}^{N} f_k(t). \tag{5.55}$$

Each function $f_k(t)$ is periodic on its own. Since we are considering that all g_k's are equal the whole function $|r_2(t)|^2$ has the same period of the f_k's. First suppose that all coupling constants are equall to $g = 0.5$ and $\alpha = \pi/4$ (maximally entangled state). Fig. 5.1 shows the behavior of concurrence in this situation when we change the size of the environment. It can be seen that the peaks become sharply defined when N increases. The first entanglement decaying also gets faster because the more environmental qubits the more coherence is stolen by the environment. This behavior also occurs for the quantum mutual information and quantum discord.

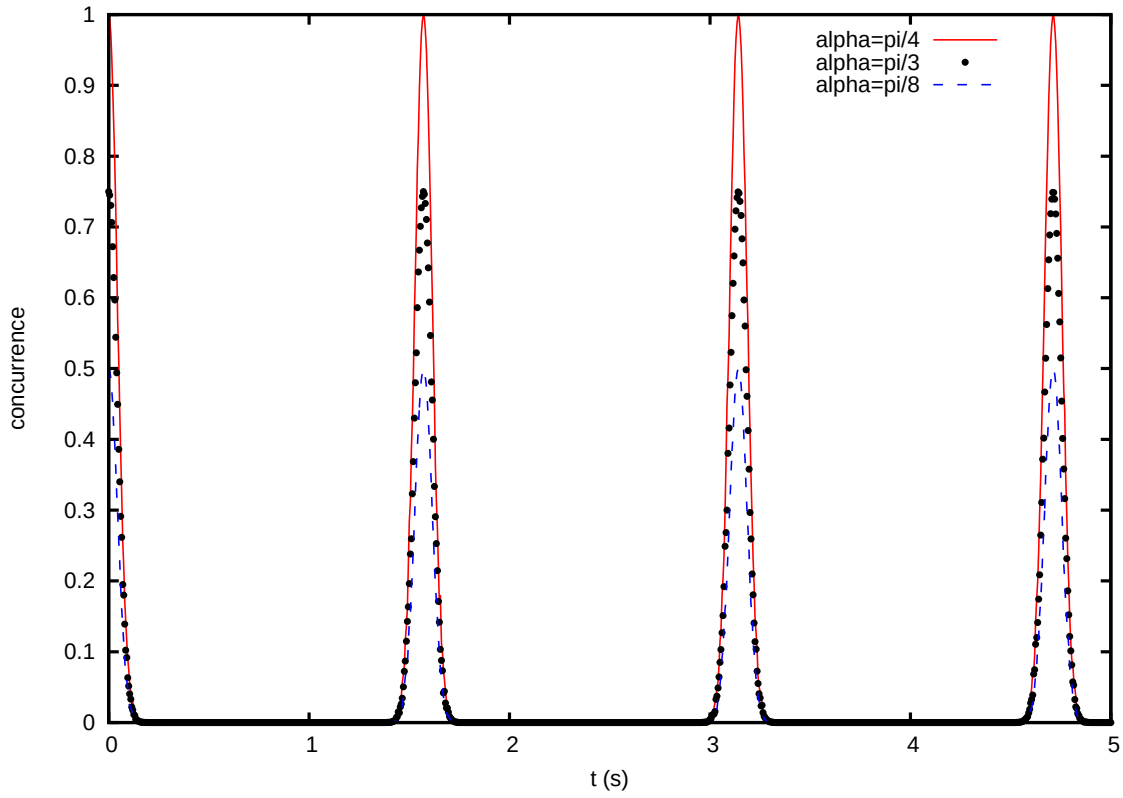

Figure 5.2: Concurrence as a function of the parametrized initial state of the system for a homogeneous environment. We used $g = 0.5\ s^{-1}$and $N = 200$.

In Fig. 5.2 we show the effect, on concurrence, of letting the initial state of the

system have different amounts of entanglement. This is translated in nothing but a different upper bound for the concurrence. This can be easily seen directly from the mathematical expression of concurrence as $|r_2(t)|$ is bounded between zero and one and therefore, $0 \leq C(\rho) \leq \sin^2(2\alpha)$. For the quantum discord the effect is the same as can be seen in Fig. 5.3. It can be noted looking at these two figures that

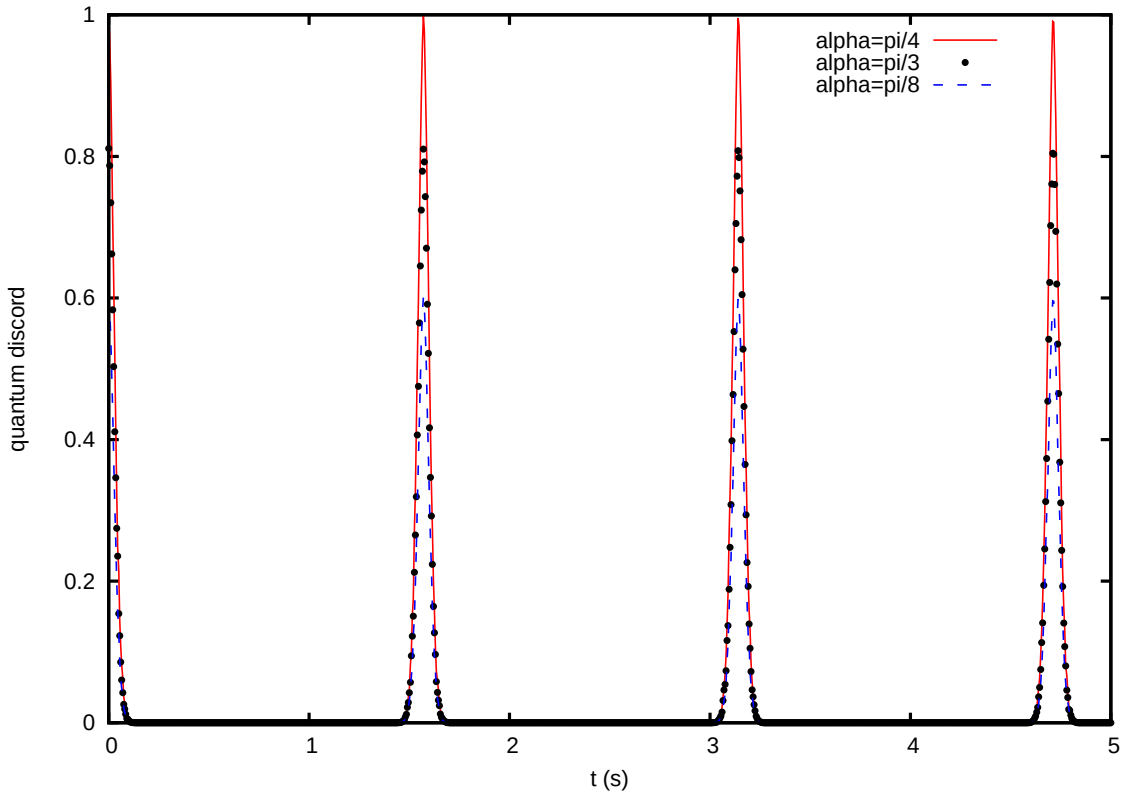

Figure 5.3: Effect of the initial amount of entanglement on the quantum discord in a homogeneous environment. We used $g = 0.5\ s^{-1}$ and $N = 200$.

the bound for the quantum discord is greater than for the concurrence. This occurs because of the following. The maximum amount of quantum discord occurs when the function $|r_2(t)|^2$ takes the value one. The upper bound for quantum discord is therefore (cf. 5.54) $-\cos^2(\alpha)\log[\cos^2(\alpha)] - \sin^2(\alpha)\log\left[\sin^2(\alpha)\right]$. Fig. 5.4 shows that for this state and this interaction the upper bound for quantum discord is greater than or equal to that of concurrence.

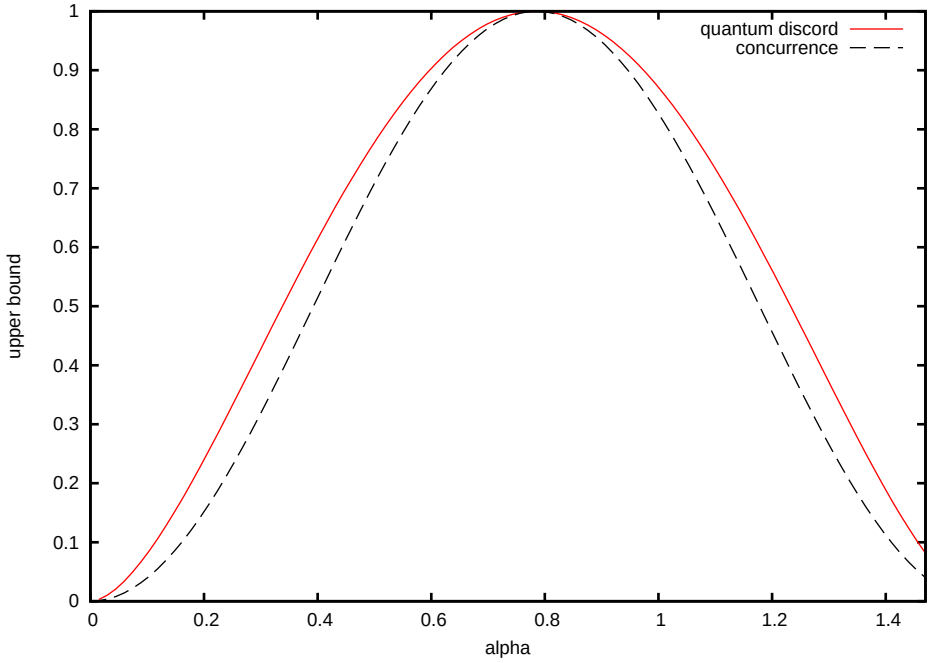

Figure 5.4: Upper bounds of quantum discord and concurrence.

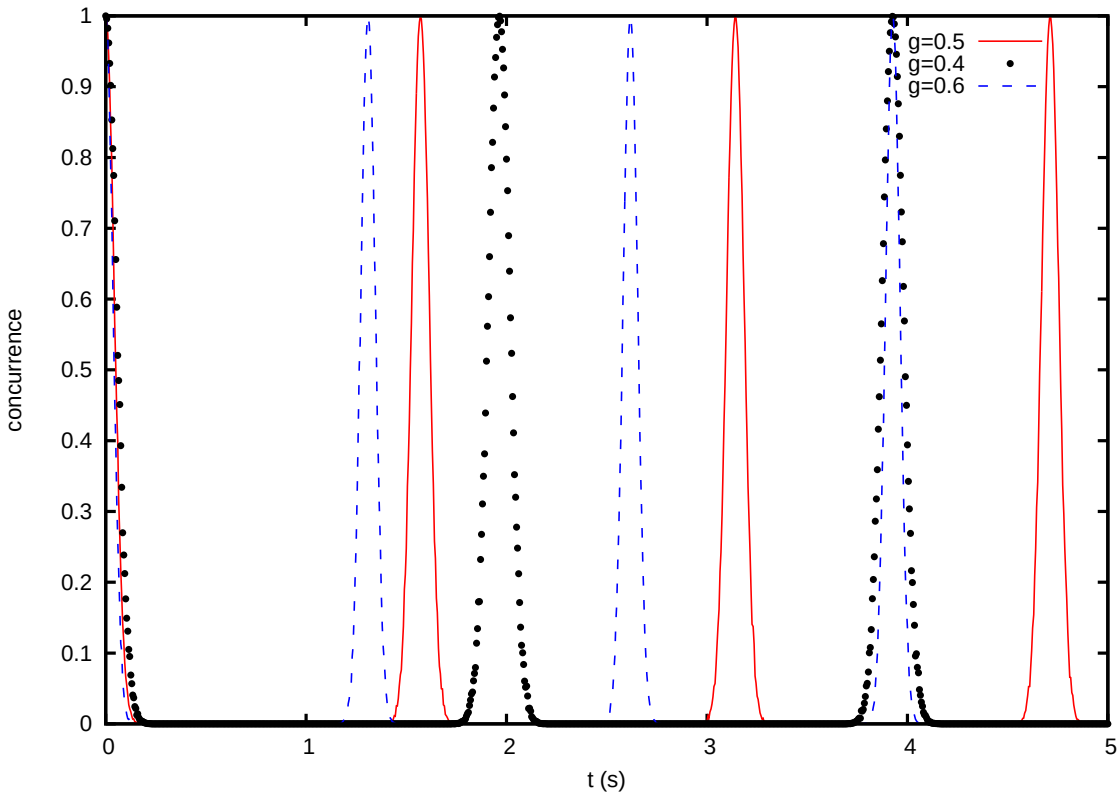

Figure 5.5: Concurrence for different couplings with the homogeneous environment. We used $N = 200$ and $\alpha = \pi/4$.

In Fig. 5.5 we show that changing the coupling constant changes the frequency of the concurrence with time. This is due to the periodicity of $|r_2(t)|$ and therefore it is reflected also in mutual information and quantum discord. This effect is going to be crucial when we consider the random case.

Now we are going to consider that situation where the coupling constants are selected from a uniform distribution in order to analyze the effects of randomizing the couplings. The distribution is uniform in the interval $[g - dg, g + dg]$, where we consider $g = 0.5$ and different amplitudes are given by dg. In Fig. 5.6 we show the time evolution of concurrence for three different intervals. As expected, the smaller

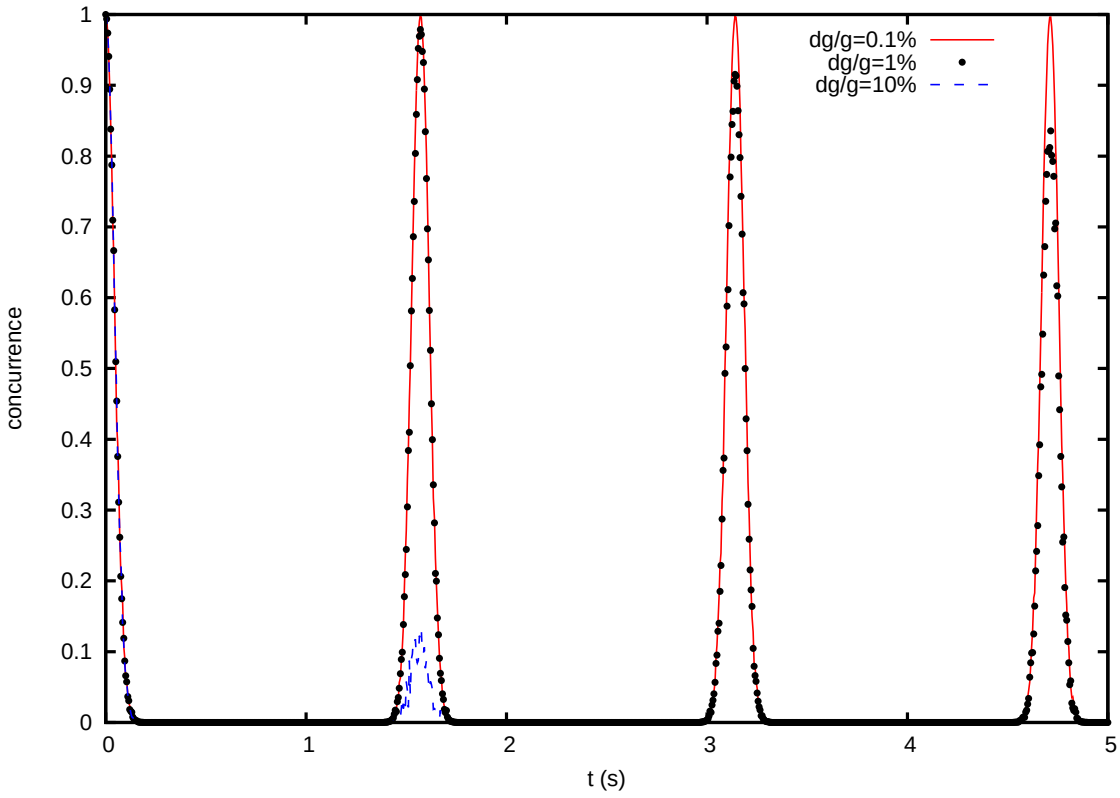

Figure 5.6: Effect of the rate dg/g of the uniform distribution on the dynamics of concurrence. We considered $g = 0.5\ s^{-1}$, $N = 200$ and $\alpha = \pi/4$ for all situations.

the interval is the more it will approximate the previous situation. For 1% relative size ($dg/g = 1\%$) it is already possible to note the decreasing of entanglement at each peak. For 10% relative size the entanglement is lost after the first peak. The same effects occur also with quantum mutual information and quantum discord.

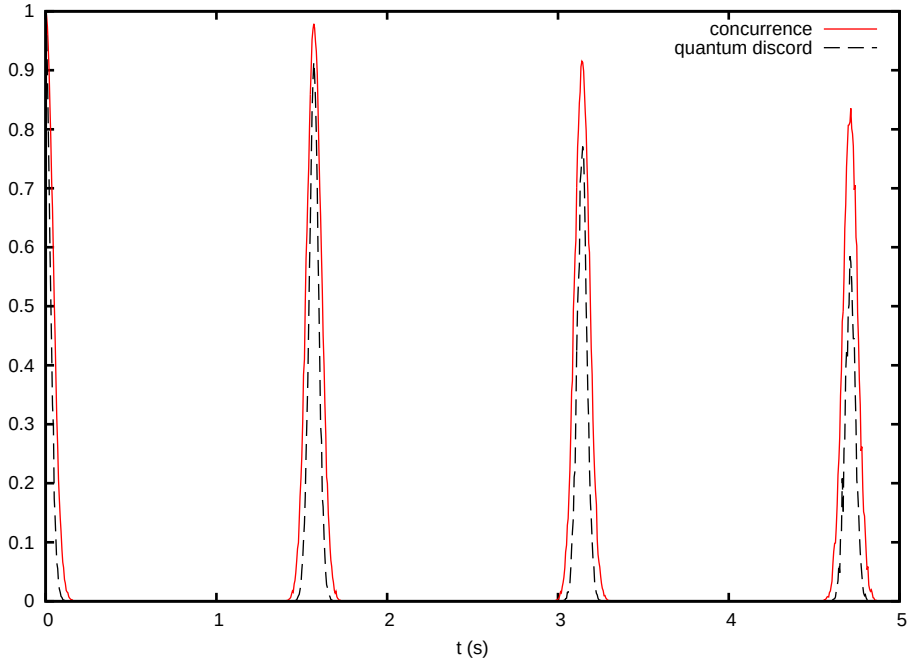

Figure 5.7: Behavior of the measures when $dg/g = 1\%$. We considered $g = 0.5\ s^{-1}$, $N = 200$ and $\alpha = \pi/4$.

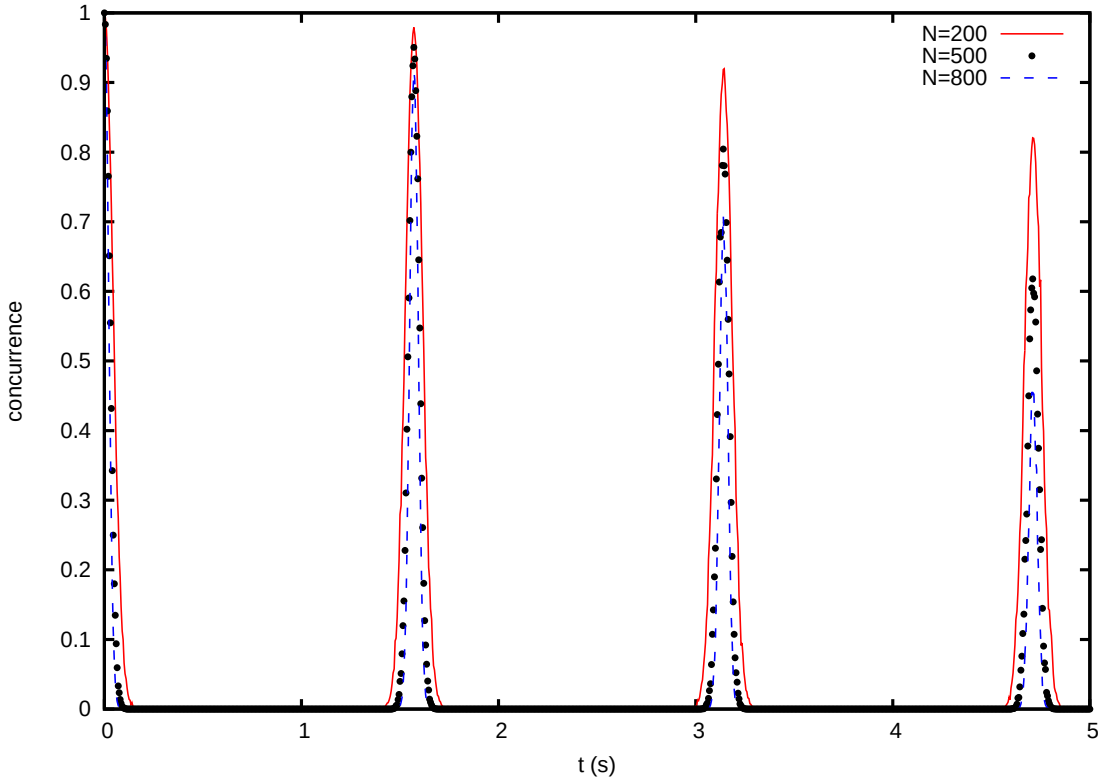

Figure 5.8: Effect of different environmental sizes on the concurrence. We considered $g = 0.5\ s^{-1}$, $dg/g = 1\%$ and $\alpha = \pi/4$.

In Fig. 5.7 we compare the behavior of the measures under the uniform coupling distribution with a 1% relative interval size. We can see that quantum discord decay and grow more rapidly than concurrence in the vicinity of each peak.

The next situation explores the change in the size of the environment. Figs. 5.8 and 5.9 show that both concurrence and quantum discord decrease more when the number of qubits in the environment increases. This can be understood, again, by the fact that the more qubits in the environment the more coherence the environment removes from the system.

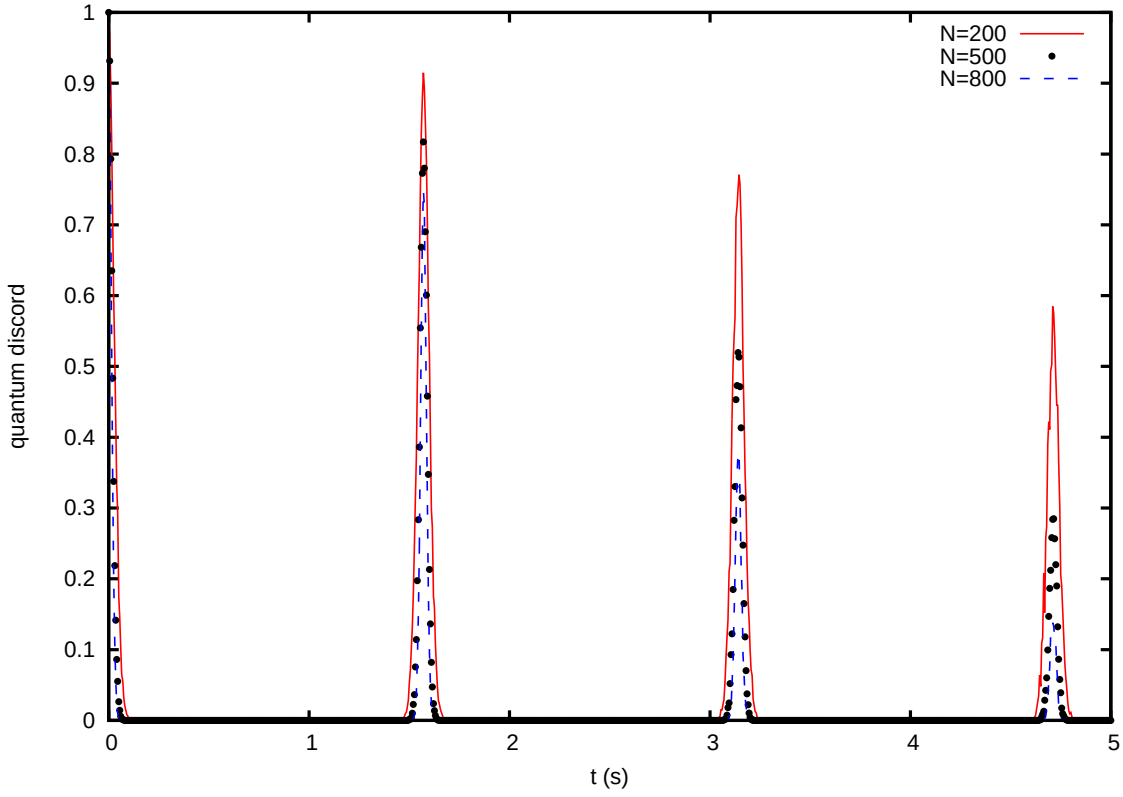

Figure 5.9: Effect of different environmental sizes on the quantum discord. We considered $g = 0.5\ s^{-1}$ and $dg/g = 1\%$.

The last distribution we are going to consider for the coupling constants is the normal distribution. Calling g the mean value of the normal distribution and dg its standard deviation, Fig. 5.10 shows that the larger the standard deviation the more severe the decoherence effect is. This effect is also present in quantum discord.

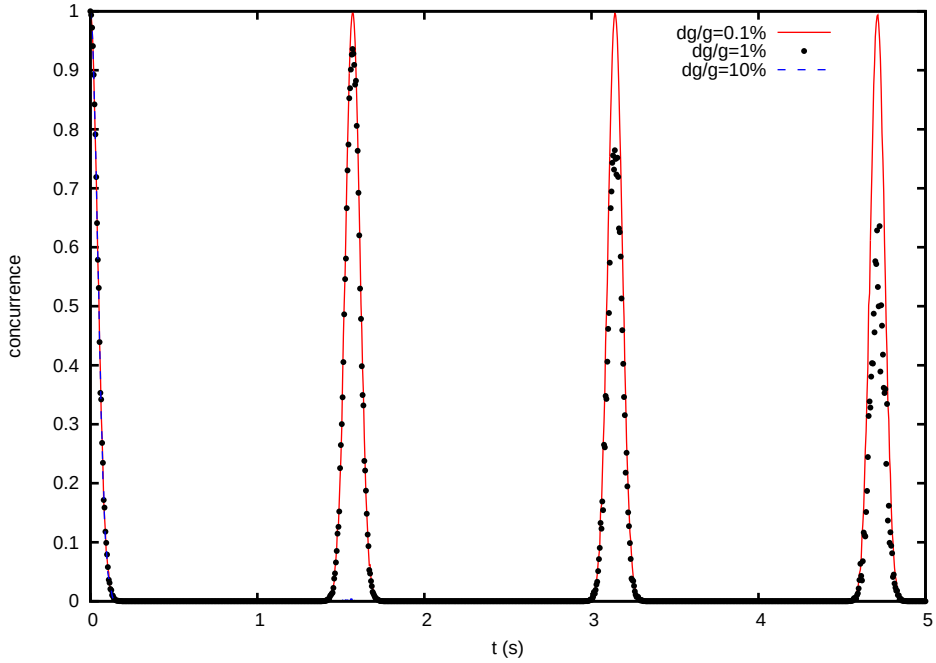

Figure 5.10: Effect of changing the standard deviation by mean value ratio on the concurrence. We considered the mean value $g = 0.5\ s^{-1}$, $\alpha = \pi/4$ and $N = 200$.

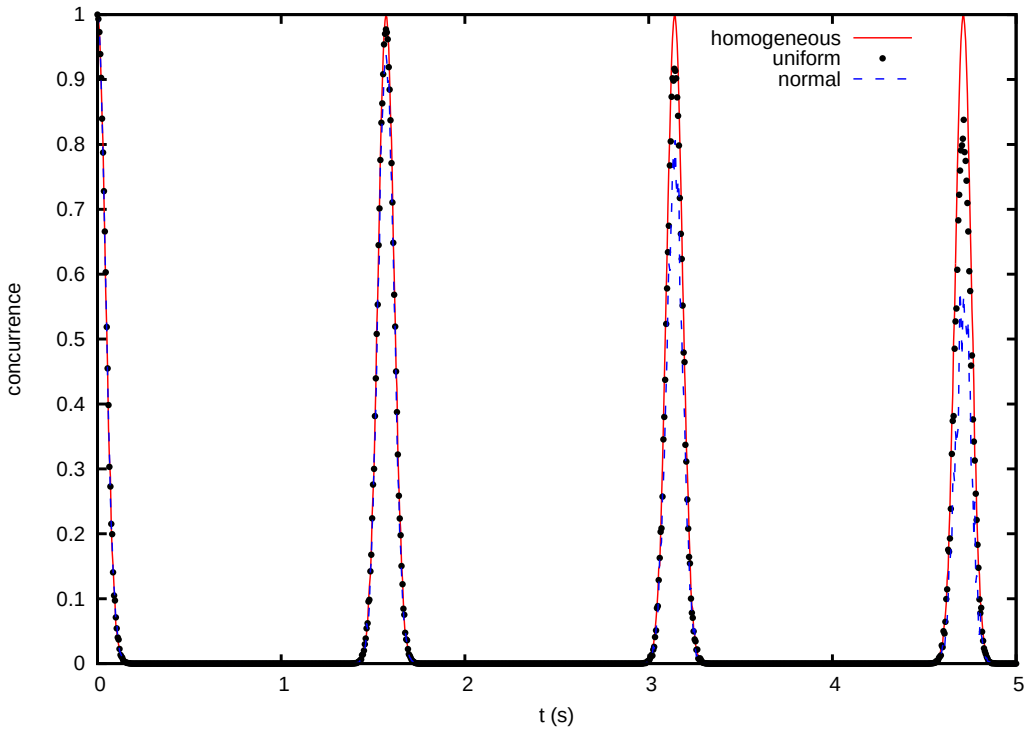

Figure 5.11: The dynamics of concurrence for the three distributions studied. For all of them we considered $\alpha = \pi/4$, $g = 0.5\ s^{-1}$ and $N = 200$. Both for the uniform and normal distributions we considered $dg/g = 1\%$.

The size of the environment again decreases the coherence of the systems. Figs. 5.11 and 5.12 show that the normal distribution imposes the most severe decoherence effect for both concurrence and quantum discord. Comparing these two figures it is easily seen that the quantum discord is more affected by the environment than the concurrence.

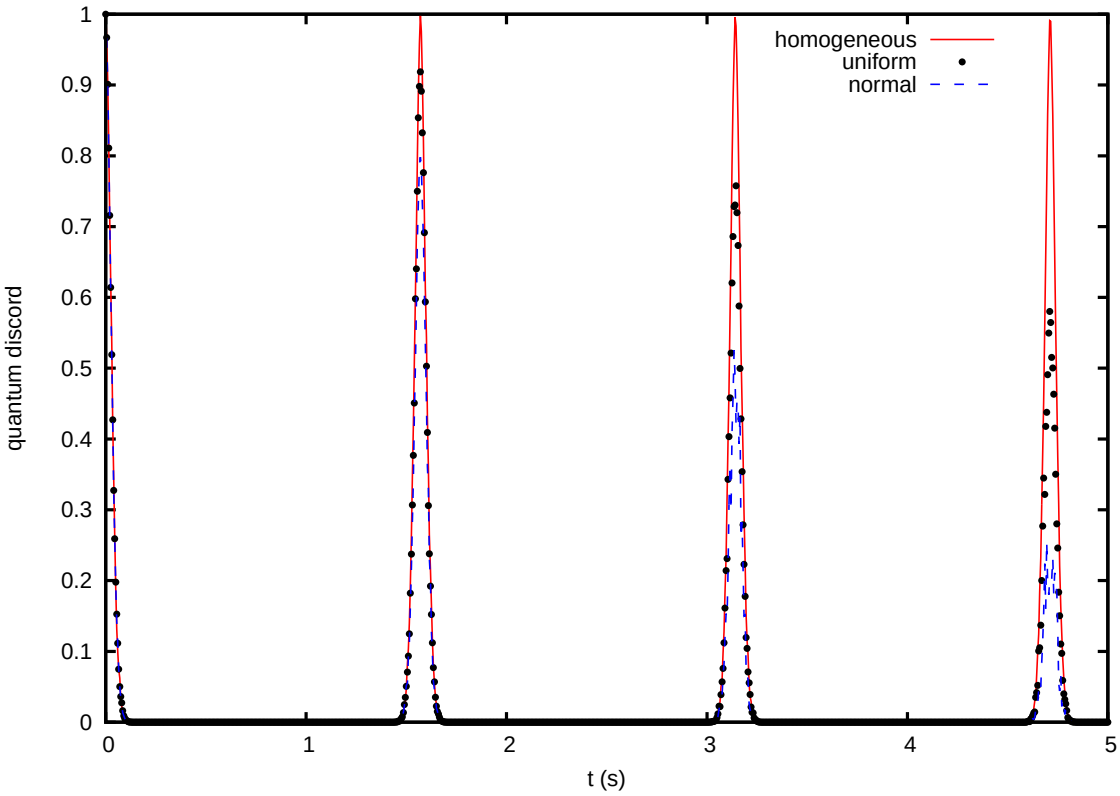

Figure 5.12: The time evolution of quantum discord for the three distributions studied. For all of them we considered $\alpha = \pi/4$, $g = 0.5\ s^{-1}$ and $N = 200$. Both for the uniform and normal distributions we considered $dg/g = 1\%$..

The main conclusion is that the effect of randomizing the coupling constants induces loss of quantum correlations. This occurs because the period of the function $|r_2(t)|$ depends on the period of each of the products for each environmental qubit. If we allow the environmental qubits to possess different coupling constants the periods of these functions will be different and therefore the period of this function, which is given by the least common multiple, will increase tremendously. In practice, the system never recovers the initial amount of quantum correlations. When the decreasing peaks occur, some of the products in $|r_2(t)|$ are closing their periods and

others are not.

Chapter 6

Conclusions

One of the purposes of this work was to provide an introduction to the field of quantum information theory. We began studying the classical theory of information in Ch. 2. We have introduced the Shannon entropy and seen two different interpretations for it. From one perspective, the Shannon entropy measures the amount of information (uncertainty) of a message (random variable) and, from the other, it tells us the minimum amount of binary digits needed to encode faithfully a given message. Then, we discussed two-messages entropies, the mutual information, which measures the correlations between two messages or the information they share in common; the conditional entropy, which gives the amount of information we have about some message when we have a previous knowledge of another; the joint entropy, which measures the non-redundant information of two messages; and the relative entropy, which measures the loss of information when we use an approximate message instead of the original. Then, in Ch. 3 we reviewed some basic concepts in quantum mechanics as well as introduced some more advanced ones such as the dynamics of open quantum systems and generalized measurement dynamics. In Ch. 4 we begin our exploration of quantum information. We began considering the von Neumann entropy and saw that it is the analogous of Shannon entropy in the quantum realm. In other words, it gives the amount of qubits necessary to transmit (transpose) faithfully a quantum signal (density operator). Then, we studied the concept of correlations in quantum theory as well as developed some measures of quantum correlations. For pure bipartite states, we can measure the entanglement with the entropy of entanglement, the purity or the Schmidt rank. For mixed states, we measure the amount of correlations with the entanglement of formation, the concurrence, the quantum mutual information and the quantum

discord. Concurrence measures the amount of entanglement in a quantum states whereas quantum discord measures the amount of quantum correlations. We have seen that quantum discord provides a notion of classicality given by the fact that the mutual information in the quantum realm can written in two ways. Finally, we have shown several applications of quantum information theory through some quantum algorithms. We presented the no-cloning theorem, quantum teleportation, entanglement swapping and deterministic quantum computation with one qubit. The last of which explores the quantum discord as a possible resource instead of entanglement, as in all previous algorithms.

The second purpose was to study the dynamics of quantum correlations when we consider the coupling between system and environment as a random distribution. Our model consisted of two central qubits interacting with a qubit environment.

For a homogeneous environment, where the distribution is constant, the correlations, as measured by concurrence, quantum mutual information and quantum discord, have simple periodic structure where the frequency is proportional to the coupling. The correlations are, therefore, never lost because they come and go with time.

For a uniform environment, i.e., when the couplings are selected uniformly from a interval $[g - dg, g + dg]$, the correlations are gradually lost. There are two factors that accelerates the lost of correlations, the size of the environment and the size of the interval for the coupling. The effect of the size of the environment is due to the fact that the larger the environment, more qubits available, the larger the capacity to steal correlation from the system. With respect to the size of the interval, the larger it is the more probable it is to the period become practically infinity, or even become nonperiodic, due to the structure of the measures.

For a normal environment, i.e., when the couplings are selected from a normal distribution with average value g and standard deviation dg, the correlation are, again, gradually lost. The reasons are the same as before. The relevant difference is that the normal environment imposes a more severe lost.

Our conclusion is that a random distribution for the coupling constants imposes permanent loss of quantum correlations. The lost can occur either progressively or almost instantaneously, depending on the size of the environment and the size of possible values the couplings may take.

Appendix A

Singular Value Decomposition

In this appendix we will demonstrate a theorem from linear algebra called singular value decomposition [65] that leads directly to the Schmidt decomposition, just by re-interpreting the terms within the context of quantum mechanics.

Let $A_{m\times n}$ be a complex matrix. Then $A^\dagger A$ is self-adjoint and therefore can be orthogonally diagonalized*. Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis formed by the eigenvectors of $A^T A$, and let $\lambda_1, \ldots, \lambda_n$ be the associated eigenvalues. Then, for eah $i \in \{1, \ldots, , n\}$

$$||Av_i||^2 = (Av_i)^\dagger Av_i = v_i^\dagger \left(A^\dagger A v_i\right) = v_i^T \lambda_i v_i = \lambda_i \geq 0. \tag{A.1}$$

So, the eigenvalues of $A^T A$ are all nonnegative. By renumering, if necessary, we may assume that the eigenvalues are arranged so that $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$.

The singular values of A are defined as the square roots of the eigenvalues of $A^T A$ and are denoted by $\sigma_1, \ldots, \sigma_n$. They are also arranged in decreasing order, $\sigma_i = \sqrt{\lambda_i}$. We will need some more definition before proceed. The column space of a matrix is denoted by Col A and it is composed of vectors b that are mapped from some x by A, i.e., $b = Ax$. The rank of a matrix is the dimension of its column space.

Theorem. If a $A_{m\times n}$ matrix has r nonzero singular values, $\sigma_1 \geq \cdots \geq \sigma_r$ and $\sigma_{r+1} = \cdots = \sigma_n = 0$, then rank of A is r.

Proof. Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis generated by the eigenvectors of $A^\dagger A$, ordered so that the corresponding eigenvalues of $A^\dagger A$ satisfy $\lambda_1 \geq \cdots \geq \lambda_n$. Then, for $i \neq j$

$$(Av_i)^\dagger (Av_i) = v_i^\dagger A^\dagger A v_j = v_i^\dagger (\lambda_j v_j) = 0 \tag{A.2}$$

*A $n \times n$ matrix is diagonalizable if and only if it has n linear independent eigenvectors.

since v_i and $\lambda_j v_i$ are orthogonal. Thus, $\{Av_1, \ldots, Av_n\}$ is an orthogonal set. Let r be the number of nonzero singular values of A. As $||Av_i|| = \sigma_i$, if λ_i is a zero eigenvalue then $Av_i = 0$. Then $\{Av_1, \ldots, Av_r\}$ is linearly independent and clearly is in Col A. Furthermore, for any y in Col A we may write $x = c_1 v_1 + \cdots + c_n v_n$, and

$$y = Ax = cAv_1 + \cdots + c_r Av_r + c_{r+1} Av_{r+1} + \cdots + c_n Av_n \tag{A.3}$$

$$= cAv_1 + \cdots + c_r Av_r + 0 + \cdots + 0. \tag{A.4}$$

Thus, y is in Span $\{Av_1, \ldots Av_r\}$, which shows that $\{Av_1, \ldots, Av_r\}$ is an orthogonal basis for Col A. Hence rank $A = r$.

Theorem. The Singular Value Decomposition (SVD). Let $A_{m\times n}$ be a matrix with rank r. Consider, now, the following matrix $\Sigma_{m\times n}$,

$$\Sigma = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}, \tag{A.5}$$

where $D_{r\times r}$ is a diagonal matrix with its entries being the first r singular values of A, $\sigma_1 \geq \cdots \geq \sigma_r \geq 0$. The theorem states that there exist an $U_{m\times m}$ unitary matrix and an $V_{n\times n}$ unitary matrix such that $A = U\Sigma V^\dagger$.

Proof. Using the results of the previous theorem, we define

$$u_i = \frac{1}{||Av_i||} Av_i = \frac{1}{\sigma_i} Av_i \tag{A.6}$$

so that

$$Av_i = \sigma_i u_i \ \ (1 \leq i \leq r)\,. \tag{A.7}$$

Then $\{u_1, \ldots, u_r\}$ is an orthonormal basis of Col A. Extend this set to an orthonormal basis $\{u_1, \ldots, u_m\}$ and let

$$U = [u_1 u_2 \ldots u_m] \ \text{ and } V = [v_1 v_2 \ldots v_n]\,. \tag{A.8}$$

Then U and V are unitary matrices (because their column vectors form an orthonormal set). In this way we have

$$AV = [Av_1 \ldots Av_r 0 \ldots 0] = [\sigma_1 u_1 \ldots \sigma_r u_r 0 \ldots 0]\,. \tag{A.9}$$

Now, consider the following product

$$U\Sigma = [u_1 u_2 \dots u_m] \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & \cdots & \cdots & \sigma_r & 0 \\ 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & \vdots \end{bmatrix}_{m\times n} = [\sigma_1 u_1 \cdots \sigma_r u_r 0 \cdots 0] \tag{A.10}$$

and thus

$$A = U\Sigma V. \tag{A.11}$$

With this result in hand we can easily prove the Schmidt decomposition and observe the recipe to find it. Suppose a general state of two particles

$$|\psi\rangle = \sum_{jk} a_{jk} |j\rangle |k\rangle . \tag{A.12}$$

Now, the SVD tells us that the decomposition of the matrix a_{jk} is given by $a_{jk} = \sum_i u_{ji} d_{ii} v_{ik}$. Substituting this in the above expression we obtain

$$|\psi\rangle = \sum_{ijk} u_{ji} d_{ii} v_{ik} |j\rangle |k\rangle . \tag{A.13}$$

Now, defining $\left|i^A\right\rangle = \sum_j u_{ji} |j\rangle$, $\left|i^B\right\rangle = \sum_k v_{ik} |k\rangle$ and $\sqrt{\lambda_i} = d_{ii}$ we obtain the known expression for the Schmidt decomposition

$$|\psi\rangle = \sum_i \sqrt{\lambda_i} \left|i^A\right\rangle \left|i^B\right\rangle . \tag{A.14}$$

Let us write some properties of a state written in its Schmidt decomposition. From the SVD these λ_i's coefficients are all nonnegative real numbers. These two sets $\left\{\left|i^A\right\rangle\right\}$ and $\left\{\left|i^B\right\rangle\right\}$ are orthonormal basis in the Hilbert space associated with the system A and B, respectively. Writing the density operator for such a state we have

$$\rho = \sum_{ij} \sqrt{\lambda_i \lambda_j} \left|i^A\right\rangle \left|i^B\right\rangle \left\langle i^A\right| \left\langle i^B\right| . \tag{A.15}$$

Evaluating the reduced density operator for both system we obtain

$$\rho^A = \sum_i \lambda_i \left|i^A\right\rangle \left\langle i^A\right| \tag{A.16}$$

$$\rho^B = \sum_i \lambda_i \left|i^B\right\rangle \left\langle i^B\right| , \tag{A.17}$$

so, the Schmidt basis is the basis that diagonalized both the reduced density operators and, even more, the eigenvalues of ρ^A and $\rho^B$are the same. We can see by the expressions that the following property is true, $\sum_i \lambda_i = 1$ (by the properties of the density matrices).

Appendix B

Operator-Sum Decomposition

This decomposition is called operator-sum decomposition, Kraus decomposition or Kraus' theorem. It is important because it provides a description for general quantum operations (cf. Sec. (3.5)).

Theorem: $\mathcal{E} : S(\mathcal{H}_N) \to S(\mathcal{H}_M)$ is a quantum operation if and only if

$$\mathcal{E}(\rho) = \sum_i K_i \rho K_i^\dagger, \tag{B.1}$$

where K_i are linear operators satisfying $\sum_i K_i^\dagger K_i \leq I$.

Before beginning the actual proof we have to define some states. Consider a bipartite system S^{AB} with state space $\mathcal{H}^{AB} = \mathcal{H}^A \otimes \mathcal{H}^B$ and, for simplicity, they have the same dimension d. Let $\{|a_n^A\rangle\}$ and $\{|b_n^B\rangle\}$ be orthonormal bases in $\mathcal{H}^A$ and $\mathcal{H}^B$, respectively. We can always construct a non-normalized maximally entangled state

$$\left|\tilde{\psi}^{AB}\right\rangle = \sum_{n=1}^{d} |a_n^A, b_n^B\rangle . \tag{B.2}$$

Observe that $|a_n^A\rangle = \left\langle b_n^B \middle| \tilde{\psi}^{AB} \right\rangle$. Let $|\phi^A\rangle$ be an arbitrary state

$$|\phi^A\rangle = \sum_n c_n |a_n^A\rangle , \tag{B.3}$$

and define the state

$$|\phi^{*B}\rangle = \sum_n c_n^* |b_n^B\rangle . \tag{B.4}$$

With these states we see the relationship

$$|\phi^A\rangle = \left\langle \phi^{*B} \middle| \tilde{\psi}^{AB} \right\rangle , \tag{B.5}$$

and likewise

$$|\phi^{*B}\rangle = \left\langle \phi^A \middle| \tilde{\psi}^{AB} \right\rangle . \tag{B.6}$$

Thus $\left|\tilde{\psi}^{AB}\right\rangle$ establishes an unambiguous map $\left|\phi^{A}\right\rangle \leftrightarrow \left|\phi^{*B}\right\rangle$, which is conjugate linear, i.e., if $\left|\phi^{A}\right\rangle = \alpha_1 \left|\phi_1^{A}\right\rangle + \alpha_2 \left|\phi_2^{A}\right\rangle$ then

$$\left|\phi^{*B}\right\rangle = \left\langle \phi^{A} \middle| \tilde{\psi}^{AB} \right\rangle = \alpha_1^* \left\langle \phi_1^{A} \middle| \tilde{\psi}^{AB} \right\rangle + \alpha_2^* \left\langle \phi_2^{A} \middle| \tilde{\psi}^{AB} \right\rangle = \alpha_1^* \left|\phi_1^{*B}\right\rangle + \alpha_2^* \left|\phi_2^{*B}\right\rangle . \tag{B.7}$$

Beginning the actual proof.

Proof. Suppose

$$\mathcal{E}(\rho) = \sum_i K_i \rho K_i^{\dagger}, \tag{B.8}$$

with K_i being linear and satisfying $\sum_i K_i^{\dagger} K_i \leq I$ and we want to show that it is a quantum operation (cf. Sec.(3.5)). If K_i are linear by hypothesis then $\mathcal{E}(\cdot)$ is also linear. The trace is given by

$$\begin{aligned} \mathrm{Tr}\left[\mathcal{E}(\rho)\right] =& \mathrm{Tr}\left[\sum_i K_i \rho K_i^{\dagger}\right] = \sum_i \mathrm{Tr}\left[K_i \rho K_i^{\dagger}\right] \\ =& \sum_i \mathrm{Tr}\left[K_i^{\dagger} K_i \rho\right] = \mathrm{Tr}\left[\left(\sum_i K_i^{\dagger} K_i\right)\rho\right] \leq 1, \end{aligned} \tag{B.9}$$

with $\sum_i K_i^{\dagger} K_i \leq I$. It is left to show that $\mathcal{E}$ is a completely positive map (see Sec. (3.3) for a definition). Let $\left|\psi^{AB}\right\rangle$ be an arbitrary state in $\mathcal{H}^A \otimes \mathcal{H}^B$. Define

$$\left|\phi_i^{AB}\right\rangle := \left(K_i^{A\dagger} \otimes I^B\right)\left|\psi^{AB}\right\rangle . \tag{B.10}$$

For an arbitrary positive operator π^{AB} we calculate

$$\left\langle\psi^{AB}\right| \left(K_i^A \otimes I^B\right) \pi^{AB} \left(K_i^{A\dagger} \otimes I^B\right) \left|\psi^{AB}\right\rangle = \left\langle\phi_i^{AB}\right| \pi^{AB} \left|\phi_i^{AB}\right\rangle \geq 0 \quad \forall i. \tag{B.11}$$

Therefore the average value of $\left(\mathcal{E}^A \otimes I^B\right)\left(\pi^{AB}\right)$ is

$$\begin{aligned} \left\langle\psi^{AB}\right| \sum_i \left(K_i^A \otimes I^B\right) \pi^{AB} \left(K_i^A \otimes I^B\right) \left|\psi^{AB}\right\rangle =& \\ =& \sum_i \left\langle\psi^{AB}\right| \left(K_i^A \otimes I^B\right) \pi^{AB} \left(K_i^A \otimes I^B\right) \left|\psi^{AB}\right\rangle \\ =& \sum_i \left\langle\phi_i^{AB}\right| \pi^{AB} \left|\phi_i^{AB}\right\rangle \geq 0, \end{aligned} \tag{B.12}$$

and thus $\mathcal{E}^A \otimes I^B$ is a positive map and $\mathcal{E}^A$ a completely positive map. $\mathcal{E}$ being a linear, trace-preserving and completely positive map it is a quantum operation.

Now we prove the reverse, suppose $\mathcal{E}^A$ is a quantum operation. Since $\mathcal{E}^A$ is a quatum operation then

$$\left(\mathcal{E}^A \otimes I^B\right)\left(\left|\tilde{\psi}^{AB}\right\rangle \left\langle\tilde{\psi}^{AB}\right|\right) =: C^{AB} \tag{B.13}$$

is a positive operator, where $\left|\tilde{\psi}^{AB}\right\rangle$ is the state of Eq. (B.2). Explicitly

$$\begin{aligned} C^{AB} &= \left(\mathcal{E}^A \otimes I^B\right) \left(\sum_n \left|a_n^A, b_n^B\right\rangle\right) \left(\sum_n \left\langle a_{n'}^A, b_{n'}^B\right|\right) \\ &= \sum_{n,n'} \mathcal{E}^A\left(\left|a_n^A\right\rangle \left\langle a_{n'}^A\right|\right) \otimes \left|b_n^B\right\rangle \left\langle b_{n'}^B\right|. \end{aligned} \tag{B.14}$$

Using the state of Eq. (B.6) we take

$$\begin{aligned} \left\langle \phi^{*B}\right| C^{AB} \left|\phi^{*B}\right\rangle &= \sum_{n,n'} \mathcal{E}^A\left(\left|a_n^A\right\rangle \left\langle a_{n'}^A\right|\right) \left\langle \phi^{*B}\right| b_n^B\rangle \left\langle b_{n'}^B\right| \phi^{*B}\rangle \\ &= \sum_{n,n'} c_n c_{n'}^* \mathcal{E}^A\left(\left|a_n^A\right\rangle \left\langle a_{n'}^A\right|\right) \\ &= \mathcal{E}^A\left(\left(\sum_n c_n \left|a_n^A\right\rangle\right) \left(\sum_{n'} c_{n'}^* \left\langle a_{n'}^A\right|\right)\right) = \mathcal{E}^A\left(\left|\phi^A\right\rangle \left\langle \phi^A\right|\right). \end{aligned} \tag{B.15}$$

As C^{AB} is a positive operator we can write its ensemble decomposition

$$C^{AB} = \sum_i \left|\tilde{c}_i^{AB}\right\rangle \left\langle \tilde{c}_i^{AB}\right|. \tag{B.16}$$

Hence

$$\mathcal{E}^A\left(\left|\phi^A\right\rangle \left\langle \phi^A\right|\right) = \left\langle \phi^{*B}\right| C^{AB} \left|\phi^{*B}\right\rangle = \sum_i \left\langle \phi^{*B}\right| \tilde{c}_i^{AB}\rangle \left\langle \tilde{c}_i^{AB}\right| \phi^{*B}\rangle. \tag{B.17}$$

Define $K_i^A \left|\phi^A\right\rangle := \left\langle \phi^{*B}\right| \tilde{c}_i^{AB}\rangle$ and we obtain

$$\mathcal{E}^A\left(\left|\phi^A\right\rangle \left\langle \phi^A\right|\right) = \sum_i K_i^A \left|\phi^A\right\rangle \left\langle \phi^A\right| K_i^{A\dagger}. \tag{B.18}$$

Due to linearity then it holds to any density operator

$$\mathcal{E}^A\left(\rho^A\right) = \sum_i K_i^A \rho^A K_i^{A\dagger}. \tag{B.19}$$

One last remark is that the decomposition is not unique. This makes sense since we are describing, in general, a nonunitary dynamics, i.e., non-reversible, of quantum systems and therefore we are not able to know from which state the present state came from.

Appendix C

Information Theoretic Inequality

In this appendix we will demonstrate a general inequality for the von Neumann entropy which, under some restricted conditions, turns out to result in some inequalities used in this work. Suppose $\rho = \sum_{k=1}^{N} p_k \rho_k$ is a general density operator, i.e., the operators ρ_k may not be pure states but instead mixed states. The goal of this appendix is to show that

$$S(\rho) \leq \sum_{k=1}^{N} p_k S(\rho_k) + H(\{p_k\}). \tag{C.1}$$

This inequality means that the uncertainty of ρ cannot exceed the uncertainty of the probability distribution alone plus the average uncertainty of the states composing the ensemble.

As for each k ρ_k is a density operator, then suppose that they have the following form

$$\rho_k = \sum_{j_k=1}^{d_k} \lambda_k^{j_k} \left|\lambda_k^{j_k}\right\rangle \left\langle\lambda_k^{j_k}\right|, \tag{C.2}$$

where j_k labels the states that compose the kth ensemble ρ_k and there are d_k of such states. The states $\left|\lambda_k^{j_k}\right\rangle \left\langle\lambda_k^{j_k}\right|$ are pure but not necessarily neither distinct nor orthogonal for all possible values of k and j_k. For each k the numbers $\lambda_k^{j_k}$ provide a normalized probaility distribution, i.e., $\lambda_k^{j_k} > 0$ and $\sum_{j_k=1}^{d_k} \lambda_k^{j_k} = 1$. We can rewrite the original ensemble as

$$\rho = \sum_{k=1}^{N} \sum_{j_k=1}^{d_k} p_k \lambda_k^{j_k} \left|\lambda_k^{j_k}\right\rangle \left\langle\lambda_k^{j_k}\right|. \tag{C.3}$$

If there are states $\left|\lambda_k^{j_k}\right\rangle$ which are equal we put them together and re-arrange all the terms in the above expression, leading us to the following decomposition

$$\rho = \sum_{\{k\}=1}^{n} q_{\{k\}} \left|\xi_{\{k\}}\right\rangle \left\langle\xi_{\{k\}}\right|. \tag{C.4}$$

The set of indices $\{k\}$ is finite and the brackets just emphasize the fact that we have changed the probability distribution $p_k\lambda_k^{j_k}$ to a new distribution $q_{\{k\}}$, but the sum over $\{k\}$ can be at any moment replaced by the original sum over k and j_k rearraging back the terms. The states $\left|\xi_{\{k\}}\right\rangle$ are distinct pure states but not necessarily orthogonal. All we have done is to state what we already knew, i.e., that any density operator can be seen as a convex combination of pure states. The reason of why we did not began with such a state is that the inequality we intend to prove depends explicitly on the original probability distribution $\{p_k\}$ and not on $\{q_{\{k\}}\}$. The next step is to see the density operator ρ as a reduced state operator of a pure state in a larger Hilbert space. Find such a pure state that gives the density operator ρ by tracing out the auxiliary degrees of freedom we are going to introduce by hand is called a purification of ρ.

Consider an auxiliary Hilbert space $\mathcal{H}^B$ with dimension $n_B = n$, i.e., the dimension is equal to the number of pure states in Eq. (C.4). Thus a purification of ρ^A is a state $\left|\psi^{AB}\right\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B$ such that $\rho^A = \mathrm{Tr}_B\left[\left|\psi^{AB}\right\rangle\left\langle\psi^{AB}\right|\right]$. The following state can be easily seen to be a purification of ρ^A,

$$\left|\psi^{AB}\right\rangle = \sum_{\{k\}=1}^{n} \sqrt{q_{\{k\}}}\left|\xi^A_{\{k\}}\right\rangle \otimes \left|\varphi^B_{\{k\}}\right\rangle, \tag{C.5}$$

where $\left\{\left|\varphi^B_{\{k\}}\right\rangle\right\}$ is an orthonormal basis in $\mathcal{H}^B$. Observe that this state is not the Schmidt decomposition of $\left|\psi^{AB}\right\rangle$ because if it were the sum should go from 1 to $\min(n_A, n_B)$, n_A being the dimension of $\mathcal{H}^A$, and $\left|\xi^A_{\{k\}}\right\rangle$ should be an orthonormal basis in $\mathcal{H}^A$. Tracing out B we obtain

$$\begin{aligned}\rho^A =& \mathrm{Tr}_B\left[\left|\psi^{AB}\right\rangle\left\langle\psi^{AB}\right|\right] = \sum_{\{k\},\{k\}'=1}^{n} \sqrt{q_{\{k\}}q_{\{k\}'}}\left|\xi^A_{\{k\}}\right\rangle\left\langle\xi^A_{\{k\}'}\right| \mathrm{Tr}_B\left[\left|\varphi^B_{\{k\}}\right\rangle\left\langle\varphi^B_{\{k\}'}\right|\right] \\ =& \sum_{\{k\}=1}^{n} q_{\{k\}}\left|\xi^A_{\{k\}}\right\rangle\left\langle\xi^A_{\{k\}}\right|, \end{aligned} \tag{C.6}$$

which shows that $\left|\psi^{AB}\right\rangle$ is indeed a purification of ρ^A. Suppose $\left\{\left|\phi^A_k\right\rangle\right\}$ is an orthonormal basis in $\mathcal{H}^A$, tracing out A gives us

$$\begin{aligned}\rho^B =& \mathrm{Tr}_A\left(\left|\psi^{AB}\right\rangle\left\langle\psi^{AB}\right|\right) = \sum_{\{k\},\{k\}'=1}^{n} \sqrt{q_{\{k\}}q_{\{k\}'}}\mathrm{Tr}_A\left[\left|\xi^A_{\{k\}}\right\rangle\left\langle\xi^A_{\{k\}'}\right|\right] \left|\varphi^B_{\{k\}}\right\rangle\left\langle\varphi^B_{\{k\}'}\right| \\ =& \sum_{\{k\},\{k\}'=1}^{n} \sqrt{q_{\{k\}}q_{\{k\}'}}\left\langle\xi^A_{\{k\}'}\middle|\,\xi^A_{\{k\}}\right\rangle\left|\varphi^B_{\{k\}}\right\rangle\left\langle\varphi^B_{\{k\}'}\right|. \end{aligned} \tag{C.7}$$

Moreover, the Schmidt decomposition guarantees that $S\left(\rho^A\right) = S\left(\rho^B\right)$.

Consider the following state

$$\rho'^B = \sum_{\{k\}=1}^{n} q_{\{k\}} \left|\varphi^B_{\{k\}}\right\rangle \left\langle\varphi^B_{\{k\}}\right| . \tag{C.8}$$

The relative entropy between ρ^B and ρ'^B is given by

$$S\left(\rho^B||\rho'^B\right) = -S\left(\rho^B\right) - \mathrm{Tr}_B\left[\rho^B \log \rho'^B\right] \geq 0. \tag{C.9}$$

Or equivalently

$$S\left(\rho^A\right) = S\left(\rho^B\right) \leq -\mathrm{Tr}_B\left[\rho^B \log \rho'^B\right] . \tag{C.10}$$

Let us handle the right-hand side. By the definition of ρ'^B we have that $\log \rho'^B = \sum_{\{k\}=1}^{n} \log\left(q_{\{k\}}\right) \left|\varphi^B_{\{k\}}\right\rangle \left\langle\varphi^B_{\{k\}}\right|$. Hence

$$\begin{aligned} -\mathrm{Tr}_B\left[\rho^B \log \rho'^B\right] &= -\sum_{\{k\}=1}^{n} \log q_{\{k\}} \mathrm{Tr}_B\left[\rho^B \left|\varphi^B_{\{k\}}\right\rangle \left\langle\varphi^B_{\{k\}}\right|\right] \\ &= -\sum_{\{k\}=1}^{n} \left\langle\varphi^B_{\{k\}}\right| \rho^B \left|\varphi^B_{\{k\}}\right\rangle \log q_{\{k\}}. \end{aligned} \tag{C.11}$$

Calculating the above expectation value,

$$\begin{aligned} \left\langle\varphi^B_{\{k\}}\right| \rho^B \left|\varphi^B_{\{k\}}\right\rangle &= \sum_{\{l\},\{l\}'=1}^{n} \sqrt{q_{\{l\}} q_{\{l\}'}} \left\langle \xi^A_{\{l\}'} \middle| \xi^A_{\{l\}} \right\rangle \left\langle \varphi^B_{\{k\}} \middle| \varphi^B_{\{l\}} \right\rangle \left\langle \varphi^B_{\{l\}'} \middle| \varphi^B_{\{k\}} \right\rangle \\ &= q_{\{k\}}. \end{aligned} \tag{C.12}$$

We have, thus, the following result

$$S\left(\rho^A\right) \leq -\sum_{\{k\}} q_{\{k\}} \log q_{\{k\}} = H\left(\left\{q_{\{k\}}\right\}\right) . \tag{C.13}$$

But, as we have said, we want to get back to the original probability distribution of the ensemble ρ, namely $\{p_k\}$. What we have to do now is just re-arrange the sum over $\{k\}$ to k and j_k again. This gives us

$$\begin{aligned} H\left(\left\{q_{\{k\}}\right\}\right) &= -\sum_{k=1}^{N} \sum_{j_k=1}^{d_k} p_k \lambda_k^{j_k} \log\left(p_k \lambda_k^{j_k}\right) \\ &= \sum_{k=1}^{N} p_k \left(-\sum_{j_k=1}^{d_k} \lambda_k^{j_k} \log \lambda_k^{j_k}\right) - \sum_{k=1}^{N} p_k \log p_k \left(\sum_{j_k=1}^{d_k} \lambda_k^{j_k}\right) \\ &= \sum_{k=1}^{N} p_k S\left(\rho_k\right) + H\left(\{p_k\}\right) . \end{aligned} \tag{C.14}$$

Substituting in Eq. (C.13) we obtain what we wanted

$$S(\rho) \leq \sum_k p_k S(\rho_k) + H(\{p_k\}). \tag{C.15}$$

Two further inequalities arise from this one. If the states ρ_k's of the ensemble ρ are all pure states, then $S(\rho_k)$ is zero for every k. Hence it follows that

$$S(\rho) \leq H(\{p_k\}). \tag{C.16}$$

The Shannon entropy of the probability distribution is an upper bound for ensembles of pure states. In this way we achieve $S(\rho) = H(\{p_k\})$. The next inequality that we can obtain from Eq. (C.15) is actually an equality. Considering that the set of operators ρ_k's has support on orthogonal subspaces*. This means that the image of the original ensemble can be decomposed into the direct sum of the images of each ρ_k and, moreover, the image sets are mutually orthogonal. Mathematically, the set ρ_k has support on orthogonal subspaces if and only if

$$\mathrm{im}(\rho) = \mathrm{im}(\rho_1) \oplus \cdots \oplus \mathrm{im}(\rho_N) \text{ and } \mathrm{im}(\rho_i) \perp \mathrm{im}(\rho_j)\, \forall i, j. \tag{C.17}$$

This is also known as orthogonal direct sum decomposition [66]. If this happens, then, there is no need to arrange the terms between Eqs. C.3 and C.4 or, in other words, $\langle \xi_{\{k\}} \,|\, \xi_{\{k\}'} \rangle = 0$. This in turn implies that ρ^B and $\rho^{B'}$ are the same and the equality of the relative entropy holds leading to an equality in Eq. C.15.

*I searched out in several linear algebra's textbooks and I did not find any such property. By the context and the result we intend to reach this definition must be equivalent to the one we are about to give.

BIBLIOGRAPHY

[1] A. Einstein, B. Podolsky, and N. Rosen, *Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?*, *Phys. Rev.* **47**, 777-780 (1935).

[2] J. S. Bell, *On the Einstein Podolsky Rosen Paradox*, *Physics* **1**, 195-200 (1964).

[3] W. Wootters, W. Zurek, *A Single Quantum Cannot Be Cloned*, *Nature* **299**, 802-803 (1982).

[4] D. Dieks, *Communication by EPR Devices*, *Phys. Lett. A* **92**, 271-272 (1982).

[5] H. Bennett ang G. Brassard, *Quantum cryptography: Public key distribution and coin tossing*, *In Proceedings of IEEE International Conference on Computers, Systems and Signal Processing* **175**, 175-179 (1984).

[6] Wikipedia, Search: Information Theory.

[7] E. Desurvire, *Classical and Quantum Information Theory, An Introduction for the Telecom Scientist*, Cambridge University Press, New York (2009).

[8] C. E. Shannon, *A Mathematical Theory of Communication*, *The Bell System Technical Journal* **27**, 379 (1948).

[9] S. Pinker, *Language Instinct: How the Mind Creates Language*, HarperCollins Publishers, New York (2007).

[10] J. Audretsch, *Entangled System: New Directions in Quantum Physics* (2007).

[11] Wikipedia, Search: Kullback-Leibler divergence.

[12] L. E. Ballentine, *The Statistical Interpretation of Quantum Mechanics*, *Rev. Mod. Phys.* **42**, 358-381 (1970).

[13] M. Schlosshauer, J. Kofler, A. Zeilinger, *A Snapshot of Foundational Attitudes Toward Quantum Mechanics*, *Stud. in Hist. and Philos. of Mod. Phys.* **44**, 222-230 (2013).

[14] H. Breuer, *Foundations and measures of quantum non-Markovianity*, *J. Phys. B: At. Mol. Opt. Phys.* **45**, 154001 (2012).

[15] K. Kraus, *States, Effects and Operations; Fundamental Notions of Quantum Theory*, Springer-Verlag, Berlin Heidelberg (1983).

[16] A. M. Gleason, *Measures on the Closed Subspaces of a Hilbert Space*, *J. Math. and Mech.* **6**, 885-893 (1957).

[17] C. C. Gerry and P. L. Knight, *Introductory Quantum Optics*

[18] V. Gorini, A. Kossakowski, E. C. G. Sudarshan, *Completely Positive Dynamical Semigroups of N-level Systems*, *J. of Math. Phys.* **17**, 821-825 (1976).

[19] G. Lindblad, *On the Generators of Quantum Dynamical Semigroups*, *Commun. Math. Phys.* **48**, 119-130 (1976).

[20] W. H. Zurek, *Pointer Basis of Quantum Apparatus: Into What Mixture Does The Wave Packet Collapse?*, *Phys. Rev. D* **24**, 1516 (1981).

[21] W. H. Zurek, *Environment-Induced Superselection Rules*, *Phys. Rev. D* **26**, 1862 (1982).

[22] W. H. Zurek, *Decoherence, Einselection, and the Quantum Origins of the Classical*, *Rev. Mod. Phys.* **75**, 715-775 (2003).

[23] M. Schlosshauer, *Decoherence, the Measurement Problem, and Interpretations of Quantum Mechanics*, *Rev. Mod. Phys.* **76**, 1267-1305 (2005).

[24] J. Baggott, *The Quantum Story: A History in 40 Moments* (2011).

[25] G. Barontini, R. Labouvie, F. Stubenrauch, A. Vogler, V. Guarrera, H. Ott, *Controlling the Dynamics of an Open Many-Body Quantum System with Localized Dissipation*, *Phys. Rev. Lett.* **110**, 035302 (2013).

[26] M. Schlosshauer, *Decoherence and the Quantum-to-Classical Transition*, Springer-Verlag, Berlin Heidelberg (2007).

[27] Wikipedia, Search: Transistor Count.

[28] R. Horodecki, P. Horodecki, M. Horodecki, K. Horodecki, *Quantum Entanglement*, *Rev. Mod. Phys.* **81**, 865-942 (2009).

[29] B. Schumacher, *Quantum Coding*, *Phys. Rev. A* **51**, 2738-2747 (1995).

[30] Macmillan Dictionary

[31] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, S. Wehner, *Bell Nonlocality* arXiv:1303.2849v1 (2013).

[32] A. Peres, *Separability Criterion for Density Matrices*, *Phys. Rev. Lett.* **77**, 1413-1415 (1996).

[33] M. Horodecki, P. Horodecki, R. Horodecki, *Separability of Mixed States: Necessary and Sufficient Conditions*, *Phys. Lett. A* **223**, 1-8 (1996).

[34] D. Bruss, *Characterizing Entanglement* arXiv:quant-ph/0110078v1 (2008).

[35] S. P. Walborn, P. H. Souto Ribeiro, L. Davidovich, F. Mintert and A. Buchleitner *Experimental Determination of Entanglement with a Single Measurement*, *Nature* **440**, 1022 (2006).

[36] M. P. Almeida, F. de Melo, M. Hor-Meyll, A. Salles, S. P. Walborn, P. H. Souto Ribeiro and L. Davidovich, *Environment-Induced Sudden Death of Entanglement*, *Science* **316**, 579 (2007).

[37] O. Jiménez Farías, C. Lombard Latune, S. P. Walborn, L. Davidovich and P. H. Souto Ribeiro, *Determining the Dynamics of Entanglement*, *Science* **324**, 1414 (2009).

[38] A. S. Coelho, F. A. S. Barbosa, K. N. Cassemiro, A. S. Villar, M. Martinelli and P. Nussenzveig, *Three-Color Entanglement*, *Science* **326**, 823 (2009).

[39] F. A. S. Barbosa, A. S. Coelho, A. J. de Faria, K. N. Cassemiro, A. S. Villar, P. Nussenzveig and M. Martinelli, *Robustness of Bipartite Gaussian Entangled Beams Propagating in Lossy Channels*, *Nature Photonics* **4**, 858 (2010).

[40] M. F. Cornelio, O. Jímenez Farías, F. F. Fanchini, I. Frerot, G. H. Aguilar, M. O. Hor-Meyll, M. C. de Oliveira, S. P. Walborn, A. O. Caldeira and P. H. Souto Ribeiro, *Emergence of the Pointer Basis Through the Dynamics of Correlations*, *Phys. Rev. Lett.* **109**, 190402 (2012).

[41] G. H. Aguilar, O. Jiménez Farías, J. Maziero, R. M. Serra, P. H. Souto Ribeiro and S. P. Walborn, *Experimental Estimate of a Classicality Witness via a Single Measurement*, *Phys. Rev. Lett.* **108**, 063601 (2012).

[42] D. O. Soares-Pinto, L. C. Céleri, R. Auccaise, F. F. Fanchini, E. R. deAzevedo, J. Maziero, T. J. Bonogamba and R. M. Serra, *Nonclassical Correlation in NMR Quadrupolar Systems*, *Phys. Rev. A* **81**, 062118 (2010).

[43] R. Auccaise, J. Maziero, L. C. Céleri, D. O. Soares-Pinto, E. R. deAzevedo, T. J. Bonagamba, R. S. Sarthour, I. S. Oliveira and R. M. Serra, *Experimentally Witnessing the Quantumness of Correlations*, *Phys. Rev. Lett.* **107**, 070501 (2011).

[44] R. Auccaise, L. C. Céleri, D. O. Soares-Pinto, E. R. deAzevedo, J. Maziero, A. M. Souza, T. J. Bonagamba, R. S. Sarthour, I. S. Oliveira and R. M. Serra, *Environment-Induced Sudden Transition in Quantum Discord Dynamics*, *Phys. Rev. Lett.* **107**, 140403 (2011).

[45] F. M. Paula, I. A. Silva, J. D. Montealegre, A. M. Souza, E. R. deAzevedo, R. S. Sarthour, A. Saguia, I. S. Oliveira, D. O. Soares-Pinto, G. Adesso and M. S. Sarandy, *Observation of Environment-Induced Double Sudden Transitions in Geometric Quantum Correlations*, *Phys. Rev. Lett.* **111**, 250401 (2013).

[46] I. A. Silva, D. Girolami, R. Auccaise, R. S. Sarthour, I. S. Oliveira, T. J. Bonagamba, E. R. deAzevedo, D. O. Soares-Pinto and G. Adesso, *Measuring Bipartite Quantum Correlations of an Unknown State*, *Phys. Rev. Lett.* **110**, 250401 (2013).

[47] C. Brukner, M. Zukowski, J. Pan, A. Zeilinger, *Bell's Inequalities and Quantum Communication Complexity*, *Phys. Rev. Lett.* **92**, 127901 (2004).

[48] W. Wootters, *Entanglement of Formation of and Arbitrary State of Two Qubits*, *Phys. Rev. Lett* **80**, 2245-2248 (1998).

[49] H. Ollivier, W. Zurek, *Quantum Discord: A Measure of Quantumness of Correlations*, *Phys. Rev. Lett* **88**, 017901 (2001).

[50] L. Henderson and V. Vedral, *Classical, Quantum and Total Correlations*, *J. Phys. A: Math. Gen.* **34**, 6899-6905 (2001).

[51] J. Maziero, L. C. Céleri, R. M. Serra and V. Vedral, *Classical and Quantum Correlations under Decoherence*, *Phys. Rev. A* **80**, 044102 (2009).

[52] S. J. Devitt, W. J. Munro, K. Nemoto, *Quantum Error Correction for Beginners*, *arXiv:0905.2794v4* (2013).

[53] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, W. K. Wootters, *Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels*, *Phys. Rev. Lett.* **70**, 1895-1899 (1993).

[54] X. Ma, T. Herbst, T. Scheidl, D. Wang, S. Kropatschek, W. Naylor, B. Wittmann, A. Mech, J. Kofler, E. Anisimova, V. Makarov, T. Jennewein, R. Ursin, A. Zeilinger, *Quantum Teleportation over* 143 *Km Using Active Feed-Forward*, *Nature* **489**, 269-273 (2012).

[55] E. Megidish, A. Halevy, T. Shacham, T. Dvir, L. Dovrat, and H. S. Eisenberg, *Entanglement Swapping Between Photons That Have Never Coexisted*, *Phys. Rev. Lett.* **110**, 210403 (2013).

[56] C. H. Bennett, S. J. Wiesner, *Communication via One- and Two-Particle Operators on Einstein -Podolsky-Rosen States*, *Phys. Rev. Lett.* **69**, 2881-2884 (1992).

[57] K. Modi, A. Brodutch, H. Cable, T. Paterek and V. Vedral, *The Classical-Quantum Boundary for Correlations: Discord and Related Measures*, *Rev. Mod. Phys.* **84**, 1655-1707 (2012).

[58] Z. Merali, *Quantum Computing: The Power of Discord*, *Nature* **474**, 24-26 (2011).

[59] M. Castagnino, S. Fortin, O. Lombardi, *The Effect of Random Coupling Coefficients on Decoherence*, *Mod. Phys. Lett. A* **25**, 611-617 (2010).

[60] S. Luo, *Quantum Discord for Two-Qubit Systems*, *Phy. Rev. A* **77**, 042303 (2008).

[61] M. Ali, A. R. P. Rau, G. Alber, *Quantum Discord for Two-Qubit X States*, *Phys. Rev. A* **81**, 042105 (2010).

[62] X. Lu, J. Ma, Z. Xi, X. Wang, *Optimal Measurements to Access Classical Correlations of Two-Qubit States*, *Phys. Rev. A* **83**, 012327 (2011).

[63] Y. Huang, *Quantum Discord for Two-Qubit X States: Analytical Formula with Very Small Worst-Case Errpr*, *Phys. Rev. A* **88**, 014302 (2013).

[64] Q. Chen, C. Zhang, S. Yu, X. X. Yi, C. H. Oh, *Quantum Discord of Two-Qubit X States*, *Phys. Rev. A* **84**, 042313 (2011).

[65] D. C. Lay, *Linear Algebra and Its Applications* 4th edition (2012).

[66] S. Roman, *Advanced Linear Algebra* 3rd edition, Springer Science+Business Media, LCC, New York (2008).